

OCHRE

*An Online Cultural and Historical
Research Environment*

J. DAVID SCHLOEN
AND
SANDRA R. SCHLOEN

Winona Lake, Indiana
EISEN BRAUNS
2012

© Copyright 2012 J. David Schloen and Sandra R. Schloen

All rights reserved.
Printed in the United States of America.

www.eisenbrauns.com

For the latest version of this book, please go to
<http://ochre.uchicago.edu>

Library of Congress Cataloging-in-Publication Data

Schloen, J. David, 1962-

OCHRE : an online cultural and historical research environment / by
J. David Schloen and Sandra R. Schloen.

pages cm

Includes bibliographical references and index.

ISBN 978-1-57506-264-8 (paperback : alkaline paper)

1. OCHRE (Computer system) 2. Learning and scholarship—
United States—Data processing. 3. History—Research—United
States—Data processing. 4. Research—United States—Data
processing. I. Schloen, Sandra R. II. Title.

AZ188.U5S24 2012

001.20285—dc23

2012042008

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984. ∞^{TM}

Contents

BASIC CONCEPTS

1. Purpose and Use	1
OCHRE Data Repositories	1
A Framework for Integrating Scholarly Data	2
An Online Database for Active Research Projects	4
A Means of Archiving and Preserving Data	6
How to Start OCHRE and Log into a Project	7
2. Representing a Project’s Information as Individual Items	11
Items that Represent Projects and Users	11
<i>Projects</i>	12
<i>Users</i>	13
Items that Represent Units of Space, Time, and Agency	14
<i>Locations and objects</i>	14
<i>Periods</i>	15
<i>Persons and organizations</i>	17
Items that Represent Project-defined Concepts	19
<i>Concepts</i>	19
Items that Represent Images and Other External Resources	20
<i>Resources</i>	20
Items that Represent Complex Texts and Writing Systems	22
<i>Texts, Epigraphic units, and Discourse units</i>	22
<i>Writing systems and Script units</i>	23
Items that Represent Research Tools	25
<i>Bibliographies</i>	25
<i>Dictionaries</i>	26
Items that Represent the Properties of Other Items	28
<i>Property variables and values</i>	28
<i>Property predefinitions</i>	32
<i>Taxonomy</i>	32
Items that Organize, Select, and Display Other Items	35
<i>Hierarchies</i>	35
<i>Sets</i>	36
<i>Queries</i>	37
<i>Presentations</i>	39

BASIC OPERATIONS

3. Accessing a Project to View or Modify Its Information	41
Levels of Access	41
<i>Anonymous Users and Named Users</i>	41
<i>View-only Mode and Data-entry Mode</i>	42
<i>Locking Items While You Modify Them</i>	43
Window Panes in the User Interface	44
<i>The Navigation Pane</i>	44
<i>The Item Pane</i>	45
<i>The Reference Pane</i>	46
4. Creating and Organizing Items	49
Creating Items	49
<i>Inserting New Items</i>	49
<i>Moving Items</i>	51
<i>Replicating Items</i>	51
<i>Repeating Items and Hierarchical Branches</i>	52
<i>Deleting Items</i>	54
<i>Item Names and Aliases</i>	55
<i>Automatic Naming of Items Based on Their Properties</i>	56
<i>Using an “Inbox” to Hold New Items</i>	57
Hierarchies of Items	58
<i>Containment Hierarchies and Subordination Hierarchies</i>	58
<i>Display Preferences for Hierarchies and Categories</i>	59
Sets of Items	61
<i>Creating a Set of Items</i>	61
<i>Query Result Sets</i>	63
<i>Viewing a Set of Items</i>	64
5. Describing Items with Properties and Notes	69
Entering Item Properties	69
<i>Types of Variables and Values</i>	69
<i>Assigning a Property to an Item as a Variable-Value Pair</i>	70
<i>Date, Serial Number, and True/False Properties</i>	71
<i>Indicating Uncertainty about a Property</i>	71
<i>Entering a Comment about a Property</i>	72
<i>Predefined Properties</i>	72
<i>Nested Properties</i>	73
<i>Inherited Properties</i>	74
<i>Multiple Observations and Interpretations</i>	76

Entering Free-form Notes for an Item	78
<i>Creating One or More Notes</i>	78
<i>Saving Notes and Exporting Them to PDF Files</i>	79
Entering Textual Content in Notes and Elsewhere	80
<i>Using the Text-formatting Toolbar</i>	80
<i>Using the Keypad of Special Characters</i>	81
<i>Specifying the Language or Function of a Selected Portion of Text</i>	82
<i>Embedding Properties and Hyperlinks within Textual Content</i>	82
<i>Displaying Translations of Textual Content</i>	85
6. Relating Items to Other Items	87
Simple Links	87
<i>General Links, Named Links, Hyperlinks, and Hotspot Links</i>	87
<i>Adding and Removing Links</i>	89
<i>Period Links</i>	91
<i>Viewing Reverse Links</i>	92
<i>Linking to Items in Other Projects</i>	93
Relational Properties	94
7. Recording Events that Affect an Item	97
Entering Events	97
<i>The "Events" Variable and Its Project-defined Values</i>	97
<i>Adding and Removing Events for an Item</i>	99
Using Events to Track Objects and Generate an Inventory	101
8. Describing and Organizing Units of Space, Time, and Agency	105
Describing and Organizing Locations and Objects	105
<i>Entering the Geographical Coordinates of a Location or Object</i>	105
<i>Entering the Local Coordinates of a Location or Object</i>	107
<i>Entering a Unique Code for an Object</i>	108
<i>Using Hierarchies to Represent and Query Spatial Relationships</i>	109
<i>Inheriting Properties in Spatial Hierarchies</i>	111
<i>Display Preferences for Spatial Hierarchies</i>	112
Describing and Organizing Chronological Periods	113
<i>Entering the Absolute Start, End, and Duration of a Period</i>	115
<i>Linking Other Items to Periods</i>	117
<i>Using Periods in Queries</i>	117
Describing and Organizing Persons and Organizations	118
<i>Entering Information about a Current Person</i>	119
<i>Entering Information about an Organization</i>	120
<i>Using Hierarchies to Organize Persons and Organizations</i>	121

RESOURCES and BIBLIOGRAPHIES

9. Managing Images and Other Resources	123
Internal Documents and External Resources	123
<i>Entering Information about a Resource</i>	124
<i>Entering Textual Content in an Internal Document</i>	126
<i>File Locations for External Resources</i>	128
<i>Image File Suffixes and File Formats</i>	130
<i>Image Watermarks and Preferred Audio and Video Formats</i>	131
<i>Importing External Documents into the Central Database</i>	132
External Resource Types and File Formats	134
<i>Images (BMP, GIF, JPEG, PNG, TIFF)</i>	134
<i>Documents (DOC, HTML, PDF, PPT, RTF, TXT, XLS, etc.)</i>	135
<i>Audio and Video Files (AIFF, AVI, MP3, MPEG-4, WAV)</i>	136
<i>2-D and 3-D Vector Drawings and Models (AI, SVG, U3D, X3D)</i>	136
<i>Geospatial Data (GeoTIFF, GRID, KML, KMZ, SHP)</i>	137
<i>Polynomial Texture Maps (PTM)</i>	138
Embedding Links to Other Items within an Image or Document	140
<i>Embedding Annotations and Hotspot Links within an Image</i>	140
<i>Linking to a Specific Page of a PDF Document</i>	145
Browsing Multiple Images	146
10. Creating and Using Bibliographic Items	151
Entering Bibliographic Data	151
<i>Entering Titles, Subtitles, and Abbreviations</i>	153
<i>Linking to Authors, Editors, and Publishers</i>	154
<i>Linking to Source Documents</i>	155
<i>Entering Publication Dates</i>	155
<i>Entering Volume Numbers, Issue Numbers, and Page Ranges</i>	156
<i>Entering Abstracts and Reviews</i>	156
<i>Entering Catalogue Codes and Linking to Zotero</i>	158
Types of Bibliographic Items	159

WRITING SYSTEMS, TEXTS, and DICTIONARIES

11. Working with Texts and Writing Systems	161
Representing a Writing System	161
<i>Creating a Hierarchy of Script Units</i>	163
<i>Entering the Allographs and Readings of a Script Unit</i>	165

Representing a Text as Physical Signs and as Meaningful Discourse	167
<i>Epigraphic Hierarchies and Discourse Hierarchies</i>	170
<i>Divergent Editions of the Same Text</i>	171
<i>Describing Epigraphic Units</i>	172
<i>Indicating Damage, Erasure, Gaps, and Emendation</i>	177
<i>Indicating Sign Placement and Other Characteristics</i>	179
<i>Describing Discourse Units</i>	180
<i>Grammatical Parsing of Discourse Units</i>	183
Displaying Texts	185
<i>Graphemic, Phonemic, and Hybrid Transcriptions</i>	185
<i>Viewing a Text</i>	188
<i>Standard Text View</i>	190
<i>Parallel Text View</i>	190
<i>Interlinear Text View</i>	191
<i>Restored Text Option</i>	192
<i>Highlighting Corresponding Units</i>	192
Associating a Text with Its Physical Medium	194
<i>The “Associated item” Relational Variable</i>	194
<i>Seals and Seal Impressions</i>	196
<i>Using Queries to Analyze Texts in Their Physical Contexts</i>	198
Importing Existing Digitized Texts	199
<i>Copying Textual Content from an External Source</i>	199
<i>Epigraphic and Discourse Processing Options</i>	201
<i>Previewing the Epigraphic and Discourse Hierarchies</i>	203
<i>Importing a Batch of Texts</i>	206
12. Constructing and Viewing Dictionary Entries	209
Entering Basic Information about a Lemma	210
<i>Entering the Citation Forms of a Lemma</i>	211
<i>Entering the Meanings of a Lemma</i>	214
<i>Linking a Lemma to a Primary Lemma or a Root Lemma</i>	215
<i>Entering Other Basic Information about a Lemma</i>	216
Entering Morphological Information about a Lemma	217
<i>Entering the Grammatical Forms of a Lemma and Their Parses</i>	217
<i>Entering the Textual Attestations of a Grammatical Form</i>	221
<i>Entering Textual References for an Attested Form</i>	223
Entering Comparative Information about a Lemma	228
Entering Semantic Information about a Lemma	229
<i>Entering Meanings, Submeanings, and Related Discussions</i>	230
<i>Entering Textual Citations for a Meaning or Submeaning</i>	232

Entering Bibliographic References and Dictionary Cross-references	235
Constructing Simplified Glossary Entries	237
Displaying Dictionary Entries	238
Searching Dictionary Entries	242
TAXONOMY, THESAURUS, and QUERIES	
13. Organizing Terms in a Taxonomy and Thesaurus 245	
Constructing a Project's Taxonomy	246
<i>Creating a New Variable</i>	247
<i>Prescribing the Values Allowed for a Nominal or Ordinal Variable</i>	252
<i>Representing Relationships between General and Specific Terms</i>	255
<i>Borrowing from another Project's Taxonomy</i>	257
Constructing a Project's Thesaurus	258
<i>Aliases versus Thesaurus Relationships</i>	259
<i>Entering Thesaurus Relationships for a Variable or Value</i>	261
<i>Chains of Thesaurus Relationships</i>	262
Working with Quantitative Properties	264
14. Using Queries to Retrieve Information 265	
Organizing and Describing Queries	266
Defining a Query's Scope	269
<i>Selecting the Projects to Include in a Query's Scope</i>	269
<i>Selecting the Category of Items to Include in a Query's Scope</i>	269
<i>Limiting a Query's Scope to Particular Periods</i>	270
Defining a Query's Search Criteria	270
<i>Searching Based on the Properties of Items</i>	271
<i>Searching for a Character String in the Textual Content of Items</i>	273
<i>Searching Based on Events that Affect Items</i>	275
<i>Searching Based on Other Criteria</i>	277
<i>Expansion of Search Criteria via Thesaurus Relationships</i>	279
Working with Query Results	280
<i>Viewing Items in a List of Query Results</i>	280
<i>Assigning Events to Items in a List of Query Results</i>	282
<i>Saving a List of Query Results as a Named Set of Items</i>	283
Using Queries in View-only Mode	284
<i>Displaying Queries for Use in View-only Mode</i>	285
<i>Performing Multiple Queries</i>	287
<i>Performing Contextual Queries</i>	290
Constructing a Generic Query to be Customized by Other Users	294

PRESENTING INFORMATION

15. Displaying Hierarchies, Tables, and Maps	297
Display Modes for Individual Items and Item Hierarchies	297
<i>Displaying Information about an Item or Hierarchy</i>	297
<i>Displaying Descendant Items Contained in an Item or Hierarchy</i>	298
<i>Displaying Items that are Linked to an Item or Hierarchy</i>	300
Displaying Tables of Data	301
<i>Customizing the Table View of a Set of Items</i>	301
<i>Printing and Exporting a Table of Data</i>	303
Displaying Geographical Maps	304
<i>Customizing the Map View of a Set or Hierarchy of Items</i>	305
Using Images and Maps to Browse a Project's Data	307
16. Public Presentations of a Project's Data	309
Presentation Styles and Components	310
<i>The "Vertical" Presentation Style</i>	313
<i>The "Image index" Presentation Style</i>	313
<i>The "Index" Presentation Style</i>	314
<i>The "Tabbed index" Presentation Style</i>	315
<i>The "Tabbed items" Presentation Style</i>	315
<i>The "Linear navigation" Presentation Style</i>	316
Web Links to OCHRE Database Items and Presentations	317

USERS, PREFERENCES, and EXTERNAL DATA FILES

17. User Accounts, User Preferences, and Project Preferences	319
Creating User Accounts and Controlling What Users Can See and Do	319
<i>Assigning Usernames and Passwords</i>	320
<i>Specifying User Privileges</i>	321
<i>Keeping Items Private and Making Them Public</i>	323
<i>Overriding Database Locks</i>	324
Individual User Preferences	324
<i>Password</i>	325
<i>Look-and-feel</i>	325
<i>Project to Open at Log-in</i>	325
<i>Confirm on close?</i>	326
<i>Check for locks on close?</i>	326
Project-wide Preferences	326
<i>Data-entry Preferences</i>	326

<i>View-only Preferences</i>	328
<i>Font Preferences</i>	331
<i>Resource Preferences</i>	332
<i>Epigraphic Sigla</i>	334
<i>Local Coordinates</i>	334
18. Importing and Exporting Data	337
Importing and Exporting Tables and Texts	338
<hr/>	
DESIGN PRINCIPLES and UNDERLYING STRUCTURES	
19. A Generic Ontology for Scholarly Research	339
An Item-based Approach	341
<i>Atomization and Dynamic Recombination</i>	344
<i>Hierarchies and the Power of Recursion</i>	346
<i>Beyond Gutenberg: Escaping the Digital Library Paradigm</i>	348
Ontology Alignment and Data Integration	351
<i>Integration without Forced Standardization</i>	351
<i>Ontological Heterogeneity and Semantic Authority</i>	353
<i>Interoperability with TEI, CIDOC CRM, and Other Standards</i>	356
<i>A Data Warehouse versus On-the-fly Mediation</i>	360
20. Implementing the OCHRE Ontology in XML	363
The Semistructured Data Model and XML Database Systems	364
<i>Advantages over the Strongly Structured Relational Data Model</i>	365
<i>Advantages over the Weakly Structured Web Data Model</i>	368
XML Document Schemas	372
<i>Ontological Concepts and XML Document Types</i>	373
<i>Normalized Database Form and Denormalized Archival Form</i>	375
Bibliography	377
Index	379

Acknowledgments

WE are grateful to our colleagues on the faculty and staff of the University of Chicago who have pioneered the use of the OCHRE system and made suggestions about how to improve it. They are (in alphabetical order): Philip Engblom, Harry Hoffner, Janet Johnson, Matthew Stolper, and Theo van den Hout. To this pioneering group in Chicago have gradually been added a number of scholars at other institutions, including Annalisa Azzoni of Vanderbilt University, Mark Garrison of Trinity University in San Antonio, Scott Lidgard of the Field Museum in Chicago, Sabrina Maras of the University of California at Berkeley, Lynn Nyhart of the University of Wisconsin at Madison, Martin Peilstöcker of the University of Mainz and the Israel Antiquities Authority, Lynn Swartz Dodd of the University of Southern California, Ron Tappy of Pittsburgh Theological Seminary, and Bruce Wells of St. Joseph's University. We greatly appreciate their comments on the practical and conceptual issues that emerged in the course of representing their information.

Several colleagues have used OCHRE very intensively, investing a lot of time to test new features and propose enhancements to the software. They are: Dennis Campbell and Miller Prosser of the University of Chicago, Aaron Burke of the University of California at Los Angeles, Daniel Master of Wheaton College, and Edward Stratford of Brigham Young University. It has been a pleasure to work closely with them as we worked through the complexities of philological and archaeological information.

In addition, many graduate and undergraduate students at the University of Chicago and elsewhere have used OCHRE over the years and have given us their feedback, for which we are grateful. They are too numerous to mention by name, so we thank them here collectively.

We owe a great deal to Lawrence Stager of Harvard University. Years ago, when we were still students, he encouraged our ambition to develop innovative software for use in archaeology. Moreover, he allowed us to test our ideas in practice in his own archaeological

project, the Leon Levy Expedition to Ashkelon. We thank him and the Ashkelon team for being willing to work with successive versions of the software and help us improve it.

Special thanks are due to Gene Gragg, a former director of the Oriental Institute of the University of Chicago, for his advice and encouragement. We are also indebted to Charles Blair, co-director of the Digital Library Development Center of the University of Chicago Library, and to his assistant Matthew Vincenz, for facilitating our use of the Library's XML database server and for providing prompt technical assistance.

Finally, we are grateful to Gil Stein, the current director of the Oriental Institute, and to Steve Camp, its executive director, for all they have done to enable the wider use of the OCHRE system in the University of Chicago and elsewhere.

J. David Schloen
Sandra R. Schloen

Chicago, Illinois
November 2012

BASIC CONCEPTS

Chapter 1

Purpose and Use

THIS book describes an “Online Cultural and Historical Research Environment” (OCHRE) in which scholars can record, integrate, analyze, publish, and preserve their data. OCHRE is a multi-project, multi-user database system that provides a comprehensive framework for diverse kinds of information at all stages of research. It can be used for initial data acquisition and storage; for data querying and analysis; for data presentation and publication; and for long-term archiving and curation of data. The OCHRE system was designed by the co-authors of this book, David Schloen and Sandra Schloen. The software for it was written by Sandra Schloen.

Although OCHRE was originally designed to facilitate cultural and historical research, it can be used equally well in other academic disciplines. It has a highly generic and flexible structure that can accommodate any observations and conceptual distinctions a scholar might make. Thus, it allows researchers to retain their own terminologies while integrating their data with data recorded using quite different taxonomies.

OCHRE Data Repositories

OCHRE is implemented in high-performance, professionally maintained databases that serve as repositories for scholarly information. An OCHRE database will typically be hosted by a university or other academic organization that has the mandate and resources to provide a scholarly data service with suitable technical support and to ensure that the data will be preserved and remain accessible.

A single OCHRE database can accommodate many different research projects and many simultaneous users, providing a collaborative environment for sharing information, either within one project or across multiple projects. Note that entering data into an OCHRE database does not entail loss of ownership or loss of control of the

data. It remains under the control of the project team who entered it and remains invisible to other people unless the project leader explicitly makes it public for all to see. Moreover, owners of data are free to remove their data from the database at any time.

Over the past several years, OCHRE has been thoroughly tested by a dozen different research projects based at the University of Chicago, Harvard University, the University of California at Los Angeles, the University of Southern California, the University of Wisconsin at Madison, the Field Museum of Chicago, Brigham Young University, and other institutions. The Oriental Institute of the University of Chicago—a world leader in archaeological and philological research since 1919—has now established a nonprofit “OCHRE Data Service” that is available to researchers everywhere. In conjunction with the University of Chicago Library, the Oriental Institute provides this service to assist scholars, both in the active phase of their research, when they are entering and analyzing data, and in the archival phase, when their data must be maintained indefinitely in a repository that conforms to modern library standards for access and preservation.

However, OCHRE is not restricted to the University of Chicago. Other academic organizations can set up their own OCHRE databases to serve other communities of scholars. This is in keeping with the flexible approach to scholarly data explained in the following chapters. OCHRE is not a monolithic database controlled by a single organization but rather an easily distributed technology for scholarly collaboration that can be implemented in multiple database servers in a variety of institutional and disciplinary settings. Having said this, it is important to emphasize that OCHRE databases hosted by different organizations are not isolated data silos. Their common database structure permits detailed linking of units of information from one database to another. It also enables complex queries that span many database servers in order to integrate data from diverse sources.

A Framework for Integrating Scholarly Data

OCHRE encompasses the varied evidence of past human activity, both written and unwritten, within a flexible yet coherent framework.

It has a simple basic structure within which it can integrate data of diverse origins. Archaeologists, philologists, historians, and social scientists routinely work with information from multiple sources that is recorded in diverse formats, including photographs, drawings, maps, structured tables, and unstructured or semistructured text. Entities of interest are described according to a wide variety of scholarly taxonomies that have been shaped by differing research traditions and research agendas. OCHRE is designed to organize this bewildering array of heterogeneous data within a coherent and easily searchable framework. It provides a way to integrate heterogeneous data, both within a single research project and across multiple projects.

The basic unit of organization in OCHRE is the “project.” OCHRE can support an unlimited number of research projects, enabling each project team to organize and share information while preserving their own terminology and restricting access to their data. A project team may be of any size, ranging from a single scholar pursuing individual research to a large group of collaborating scholars and students. Projects and their individual members are identified by name and are given credit for their data and interpretations.

Although OCHRE is an online system that employs a central database, control of the data is not centralized. Each project’s data remains under the control of the project leader, who determines who can see and modify the data. A project’s data is invisible to OCHRE users who have not been authorized to see it. The project leader also determines who among the project’s staff can modify particular kinds of data. Thus, OCHRE can be used as an internal data management system for a research team but it can also be used to share and publish data more widely, at the discretion of the project leader, who can make some or all of the project’s data public for anyone to see.

The University of Chicago’s OCHRE database is accessible on the Web at <http://ochre.uchicago.edu>. Users can click a link there to download a software application, free of charge, which will let them view and search information in the Chicago database that has been made public by participating projects. OCHRE databases at other institutions will provide access in a similar way.

To add or modify data in an OCHRE database, a research team (or individual researcher) will need their own OCHRE project. The project leader will be given a “project administrator” password to control user accounts and data entry for the project. A project administrator can create usernames and passwords for any number of project staff members, specifying the level of access each person possesses to view or modify the project’s data.

OCHRE users who do not have a user account assigned by a project’s administrator cannot log into that project to view or change its data. However, such users may view and query any data that the project administrator has made publicly available to “anonymous” users. Many projects will choose to keep their data invisible to anonymous users until they are ready to publish it.

To set up your own project, contact the OCHRE Data Service at the University of Chicago by sending e-mail to *ochre@uchicago.edu*.

An Online Database for Active Research Projects

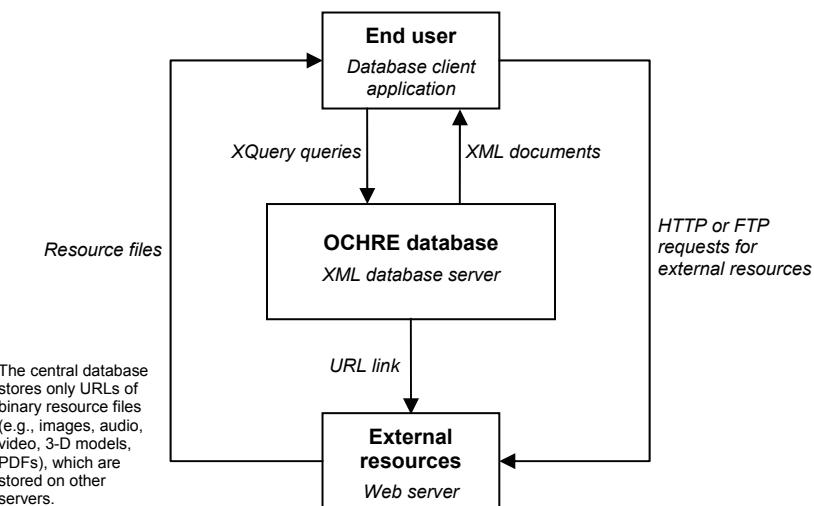
An OCHRE system consists of a central database server and user-interface software (a “database client application”) that runs on each user’s computer and enables him or her to enter, search, and display data (see the diagram on the facing page). The database structure reflects a generic ontology within which any project’s data can be represented as a collection of many interlinked items of information. Entities of interest and relationships among them are represented by structurally independent database items in a highly granular fashion.

This ontology is implemented in a set of eighteen document schemas that conform to the Extensible Markup Language (XML), a nonproprietary data format published by the World Wide Web Consortium. An OCHRE database consists of a large number of small XML data objects (“documents”) that conform to one or another of these eighteen XML schemas.

OCHRE’s item-based design is conducive to online citation and integration of scholarly data because of the high degree of granularity in the underlying digital representation of the information. Individual items of information can be easily retrieved from different projects

and combined in different ways because they are represented by individual XML documents. The highly atomized representation of information is made clear in the next chapter, which discusses the various categories into which items of information are grouped. OCHRE's underlying ontology and the XML document schemas used to express it are explained in Chapters 19 and 20 at the end of this book.

Note that OCHRE distinguishes between internal data stored in the XML text-based format within the central database and external digital resources such as images, spatial data, 3-D models, and video recordings (see the diagram below). Each project will normally provide its own Web server in which to store its images and other external resource files. These files are not stored in the central OCHRE database but are simply catalogued there, with URL links to the Internet locations where they can be found. See Chapter 9 for a detailed description of the types of resources and resource file formats that are supported by OCHRE.



OCHRE's user-interface software is written in the widely used Java programming language, which has the advantage of being “platform independent.” In other words, Java applications can run on a

wide range of computer operating systems, including Windows, Mac OS X, and Linux, provided that a piece of software called the “Java Runtime Environment” (JRE) is installed on the user’s computer. This software is preinstalled on most desktop and laptop computers. It can also be downloaded, free of charge, from the Java website at <http://java.com>. The use of Java enables the OCHRE user interface to be quite sophisticated with respect to its mechanisms for entering and viewing data while reducing the load on the central server by taking advantage of the computing power of each user’s computer.

A Means of Archiving and Preserving Data

Having a project in an OCHRE database (e.g., via the University of Chicago’s OCHRE Data Service) enables a research team to enter and analyze their data during the project’s active phase. However, the OCHRE system also allows long-term preservation of project data in a standards-compliant digital archive. Many funding agencies now require a credible plan for long-term data preservation. Accordingly, OCHRE is designed to make this as easy as possible.

Data is preserved in two different ways. In the first place, all data entered into an OCHRE database will remain permanently accessible via OCHRE’s own user interface, which allows researchers to search and display data in complex ways using a variety of advanced techniques. However, this kind of direct database access requires the use of specialized software, namely, the database client application that is described in the following chapters.

For this reason, OCHRE provides an alternative data-access and data-preservation mechanism that does not require specialized software. This entails exporting a stand-alone project archive from the central database (see Chapter 18). An exported archive consists of XML text files in which a project’s information is stored in a much less atomized fashion than in the central database. An exported archive is a snapshot of the dynamically changing database content at a particular time, so it will become out of date whenever a project’s data is modified in the central database, requiring the project administrator to regenerate the archive. However, such an archive has the

advantage of storing data in a simple form that does not require the use of specialized database software but can be retrieved by a search engine and displayed in an ordinary Web browser. Furthermore, an archive exported from OCHRE will contain embedded “metadata” (data about data) that conforms to professional library standards such as the Dublin Core. This permits the archive’s contents to be catalogued and found by any software that is based on these standards.

How to Start OCHRE and Log into a Project

At the present time, the OCHRE database client application requires the Java Runtime Environment (JRE), which comes preinstalled on most desktop and laptop computers. If you do not have the JRE, or you are not sure you have the latest version, visit the Java website at <http://java.com> to install it, free of charge.

The OCHRE database client application requires a high-speed Internet connection, at least one gigabyte of main memory, and a screen resolution of at least 1,024 by 768 pixels. The Java Web Start mechanism is used to launch the OCHRE client application from an ordinary Web link. To access the University of Chicago’s OCHRE database, follow the instructions at <http://ochre.uchicago.edu>.

The first time you start the OCHRE program you will be asked to accept software from the University of Chicago. Click the option to “always trust” this software publisher. The software will then be downloaded to your computer. You may also be asked to accept “unverified” software from the software companies Software AG and ESRI, Inc. This software is trustworthy and should be accepted.

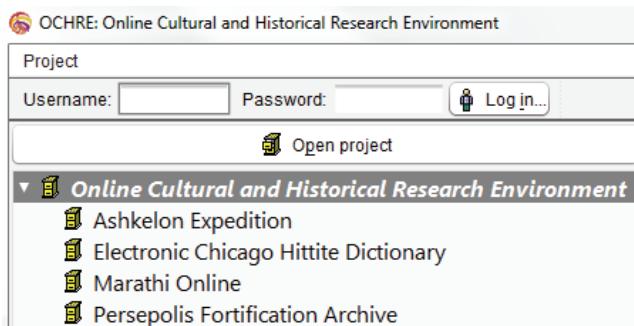
After OCHRE has been started once in this way, a desktop icon will be placed on your computer to make it easy to start it again without having to go back to the OCHRE website. Every time OCHRE is started it will make sure that the latest version of the software is installed on your computer. If not, an updated version will be downloaded automatically.

In rare cases, OCHRE may fail to start properly, requiring you to modify the Java settings on your computer by deleting all temporary files, thus removing the previous OCHRE installation. You can do

this in Windows by clicking the Java icon in the Control Panel and then clicking the “Settings...” button followed by the “Delete Files...” button.

After starting the software, you will see a list of the OCHRE projects that have made their data public for viewing by anonymous users. You can open a project to view its data by double-clicking on it or by clicking once to select it and then clicking the **Open project** button at the top of the project list. You will then see whatever information the project has decided to make visible to anonymous “view-only” users (see Chapter 16 for instructions on how to create public presentations of data for such users).

If you have a user account for one or more projects, you can log in by entering your username and password in the top left corner of the OCHRE window and then pressing the “Enter” key or clicking the **Log in** button. After you log in, you will see, in addition to the public projects, a list of all the projects for which you have been given access as a user or for which you are the designated project administrator. If you select one of those projects from the list and click the **Open project** button, you will enter data-entry mode and will see the database items that belong to that project, to the extent allowed by the privileges you have been assigned as a user.



To leave the currently open project and open a different project, click the **Close...** button in the top left corner of the OCHRE window. This will display again the list of projects to which you have been given access. Clicking the **Log out** button will close the current

project and log you out of the system, returning you to the original list of public projects.

Chapter 3 describes the window panes in the user interface, namely, the *navigation pane* on the left side of the OCHRE window, the *item pane* in the middle of the window, and the *reference pane* on the right side. It also discusses the distinction between view-only mode and data-entry mode, as well as the procedure for “locking” database items while you modify them. But before dealing with these and other topics that have to do with basic operations, it will be useful in the next chapter to define and describe the item categories that are listed alphabetically in the navigation pane of the user interface after you log into a project. These categories are:

Bibliographies	Presentations	Sets
Concepts	Property predefinitions	Taxonomy
Dictionaries	Property values	Texts
Locations & objects	Property variables	Thesaurus
Periods	Queries	Users
Persons & organizations	Resources	Writing systems

All of the many individual database items used to represent a project’s information will belong to one or another of these categories, with two exceptions. Items that represent hierarchies of other items are not shown in a separate “Hierarchies” category in the navigation pane; instead, they are used to organize other items within the listed categories, as described below in Chapter 2. And the item that represents the project as a whole does not belong to a category but is displayed as the parent of all the project’s item categories in the navigation pane in the top left corner of the OCHRE window. If you click on this project item to select it you will see information about the project as whole in the



item pane in the middle of the window, including the project's description and display preferences.

Note that your own user account will be the only one visible in the **Users** category unless you are logged in as the project administrator, in which case you will see all of the user accounts for the project. Other categories may also be invisible if they are not being used by the project and the project administrator has chosen not to display empty categories in the navigation pane.

Chapter 2

Representing a Project's Information as Individual Items

THE basic structural element in OCHRE is not a two-dimensional table that represents a class of similar items, as in many other database systems. The basic structural element is the individual item itself. Each item of interest is represented by a separate XML data object (“document”) that can be easily transmitted via the Internet. Database items are grouped into different semantic categories; but, regardless of their category, they are all basically similar. Every item is an individually manipulated unit of information that has its own name, properties, and links to other items.

OCHRE’s item-based design is structurally different from the class-based design of database systems whose basic structural elements are tables of data (see Chapter 19). In such systems, each item of interest must be assigned to a particular class that is predefined in the database structure in the form of a table. Items are represented as rows in a table and the items in a particular table are assumed to have a common set of properties, which are represented as the table’s columns.

The tabular, class-based design that is so widely used in the business world is not well suited for scholarly research in archaeology or textual studies. This will become clear in the following chapters, which demonstrate the benefits of an item-based representation of the entities and relationships typical of cultural and historical research.

Items that Represent Projects and Users

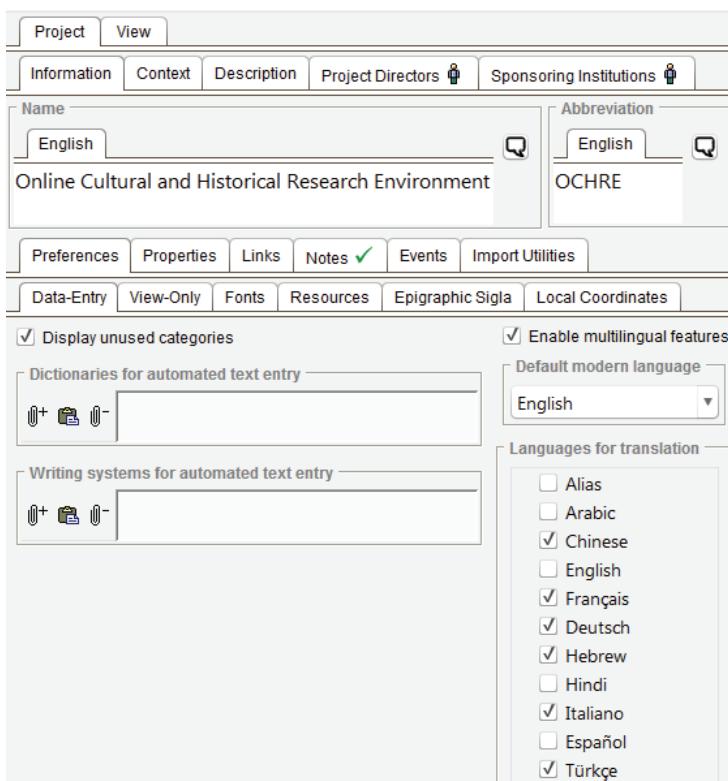
The OCHRE item categories are described below, one by one. Together, they are sufficient to represent any kind of archaeological, historical, or textual information that a scholar might want to record.

The explanations and screen images given below describe what you would see if you were logged into an OCHRE project as a named user. Anonymous users who are not logged in will see only what the project administrator has decided to present to such users (see Chapter 16 on creating public presentations accessible to anonymous

users). After surveying the OCHRE item categories in this chapter, we provide step-by-step instructions for logging in and other basic operations in Chapter 3.

Projects

Each research project in OCHRE is represented by a database item. After you have logged into a particular project, the item that represents the project is shown in the navigation pane as the hierarchical parent of all other categories of database items owned by that project. When this project item is selected, information about the project is displayed in the item pane in the middle of the OCHRE window. Like any other database item, a project item has a name and description. It may also have one or more properties, notes, and links to other items (see Chapters 5 and 6). Other fields include the project's preferences, director(s), and sponsoring institution(s).



All projects are themselves subprojects of an overarching project named “OCHRE,” from which they can borrow database items they may wish to use. This is especially useful when defining a taxonomy of descriptive properties, parts of which might be copied from the master taxonomy defined within the “OCHRE” project (see Chapter 13 for the procedure for borrowing items from another taxonomy).

Users

If you are logged in as a project administrator, you will see a **Users** category in the alphabetical list of item categories in the navigation pane. Each database item in this category represents a person who has been given a username and password by the project administrator that allows him or her to log into the project as a named user. Chapter 17 explains how to set up user accounts for a project.

If you are logged into a project but you are not the project administrator, you will see only one item in the **Users** category. This item represents your own user account. If you select it, you will be able to see your data-entry privileges and to modify your display preferences. The procedure for setting individual display preferences is discussed in Chapter 17.

Before a person can be designated as a user, he or she must first be represented as a database item in the **Persons & organizations** category (described below). Thus, to enter or view information about the user as a person, you must select the relevant item in the **Persons & organizations** category. Like any OCHRE item, a person item (and hence a user) may have a primary name, any number of aliases, a free-form description, structured properties, notes, events, and links to other items. However, when an item in the **Users** category is selected in the navigation pane, only the information relevant to that person as an OCHRE user is displayed in the item pane.

The user’s privileges are specified by a grid of checkboxes that determine his or her ability to insert, modify, or even view various categories of items. For example, a research assistant might be allowed to view and edit items in the **Locations & objects** category but not to modify, or even see, items in other categories.

The screenshot shows a user management interface with tabs for 'User' and 'View'. Under 'User', there is a form with fields for 'Name' (Schloen, Robbie) and 'Username' (xxxxxxx). Below the form are buttons for 'Activate account ...', 'Deactivate account ...', 'Reset password ...', and 'Remove user from project ...'. A navigation bar below these buttons includes 'Privileges', 'Preferences', 'Properties', 'Links', and 'Notes'. A section titled 'Update user privileges ...' contains a table with columns for 'Access' (Admin, Insert/Delete, Insert only, Edit content, View/Link, View only, No access) and rows for various categories: Bibliographies, Concepts, Dictionaries, Locations & objects, Periods, Persons & organizations, and Presentations. Each row has checkboxes corresponding to the access levels.

Access	Admin	Insert/Delete	Insert only	Edit content	View/Link	View only	No access
Bibliographies	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Concepts	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Dictionaries	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Locations & objects	<input checked="" type="checkbox"/>						
Periods	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Persons & organizations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Presentations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Items that Represent Units of Space, Time, and Agency

Locations and objects

Spatially situated units of observation are represented as items within the **Locations & objects** category. Items in this category can represent units of observation of any kind, at any scale of measurement. For example, an archaeological project might use such items to represent geographical regions, landscape features, settlement sites, buildings or monuments, excavated deposits, single artifacts, or components of artifacts. The nature and physical scale of the spatial units represented by items in this category are defined by each project according to its own recording system and research goals.

Item hierarchies in the **Locations & objects** category represent the contain-

The screenshot shows a hierarchical list of items under the 'Locations & objects' category. The tree structure starts with 'Locations & objects', which branches into 'Area 0', 'P10-20.17#56', 'P10-20.17#57', 'P10-20.17#61', and 'P10-20.17#62'. Each node is preceded by a red crossed-out icon.

- ▼ Locations & objects
 - ▼ Area 0
 - L10-0001
 - L10-0002
 - L10-0003
 - L10-0004
 - ▼ L10-0005
 - ▼ P10-20.17#56
 - C10-1134
 - C10-709
 - Unregistered sherd
 - Unregistered sherd
 - ▼ P10-20.17#57
 - R10-167
 - Soil sample
 - ▼ P10-20.17#61
 - C10-497
 - C10-498
 - Faunal remains
 - R10-211
 - R10-213
 - Unregistered sherd
 - Unregistered sherd
 - Unregistered sherd
 - P10-20.17#62
 - P10-20.17#56
 - P10-20.17#66
 - P10-20.17#67
 - P10-20.17#77

ment of smaller spatial units within larger ones. For example, a hierarchy might represent a large geographical region within which are subregions that contain individual sites, each of which in turn contains excavation areas that each contain particular artifacts. A project can create any number of containment hierarchies of this sort. Items may be nested within one another as deeply as is needed to represent the relevant containment relationships.

Furthermore, these hierarchies can overlap, in the sense that an individual spatial unit may appear in more than one hierarchy. This enables the use of multiple hierarchies to provide alternative views of the same units of observation and of the spatial relationships among them. For example, an artifact might appear in a hierarchy that represents its spatial context in the archaeological site in which it was discovered, while also appearing in a different hierarchy that represents its current museum location in a drawer in a storage cabinet.

In addition to allowing items in the **Locations & objects** category to appear in more than one hierarchy, OCHRE allows multiple observations of the same item, with each observation attributed to a particular observer. This is necessary because a unit of observation might have been observed by different people at different times, sometimes with contradictory results, and their observations must be kept distinct in the course of subsequent querying and analysis.

Periods

Items in the **Periods** category represent temporal units. These units may be of any type and duration, depending on the needs of the project. Periods are defined, named, and organized according to the chronological nomenclature used by the project. In archaeological and historical research it is often the case that the beginning, end, and duration of a chronological period are not known in terms of absolute calendar dates within a dating system such as our familiar calendar reckoned in years B.C. or A.D. However, scholars may well know the relative sequence of periods and their approximate durations in years, centuries, or millennia. Scholars usually rely on a widely understood relative periodization of archaeological or historical eras in their area

of study (e.g., the “Bronze Age” and its subdivisions) and they use this chronological scheme to synchronize entities and events assigned to the same period or subperiod.

In OCHRE, each period or subperiod of interest is represented as an item in the **Periods** category. Hierarchies of items are used to represent the relative sequence of the periods and the temporal containment of one period within another (e.g., the “Late Bronze Age” is the last part of the “Bronze Age”). The order in which periods are listed in a given branch of a hierarchy is understood to indicate their chronological order. Periods that are children of another period in a hierarchy are understood to be subdivisions of the parent period. Thus, hierarchies in the **Periods** category represent the strict containment of items within others of the same kind, just as in the **Locations & objects** category.

Note that the principle of strict containment of subperiods within larger periods in a hierarchy implies that there should be a separate hierarchy for each chronological scheme. You should not combine chronologies that are based on disparate criteria within a single hierarchy but should create a new hierarchy for each chronology. For example, a chronology based on archaeological criteria such as the changing styles of material remains in a particular geographical region should be in its own hierarchy distinct from chronologies based on political criteria such as the reigns of particular rulers and the succession of rulers in a dynasty, even though some archaeological periods may be conventionally named with respect to political periods, reflecting the impact of politics on style. Moreover, each political chronology should have its own hierarchy. For example, a chronology consisting of the successive dynasties that ruled ancient Egypt and the reigns of successive pharaohs within each dynasty would be represented by its own hierarchy of chronological periods separate from other hierarchies that represent concurrent or overlapping sequences of political regimes in Mesopotamia, Rome, and elsewhere.

If the absolute starting and ending dates of a period are known in terms of a standard calendar, these dates can be entered for the period at the appropriate level of precision (from seconds, minutes, and

hours on one end of the spectrum to days, months, years, centuries, or even millennia, as the case may be). Other OCHRE items that are linked to items in the **Periods** category can then be automatically arranged in chronological order in a timeline and placed in synchronic groups because of their association with temporal units whose relative sequence or absolute dates are known. Absolute dates also provide a means of synchronizing chronological schemes that are based on different criteria of periodization and are thus represented by different hierarchies of period items.

Unit	Start	End	Duration
Millennium			
Century			
Decade			
Year	-1173	-1142	
Month			

Persons and organizations

Unlike many database systems, OCHRE represents units of agency—creators, observers, authors, editors, publishers, institutions, historical and fictional persons, etc.—as database entities in their own right. They are represented as items in the **Persons & organizations** category. This allows the information about a particular unit of agency to be entered and stored in one place, avoiding the error-prone duplication of information.

Elsewhere in OCHRE, references to a unit of agency are made by linking to the appropriate item in the **Persons & organizations** category. For example, OCHRE users who have usernames and passwords must first be represented as items in the **Persons & organizations** category of a particular project, after which the project administrator can give them data-entry privileges via the **Users** category.

Item hierarchies in the **Persons & organizations** category are not containment hierarchies, in contrast to hierarchies in the **Locations & objects** and **Periods** categories. A person cannot directly contain another person in the way that a spatial or temporal unit may contain a smaller unit of the same kind. Instead, hierarchies in the **Persons & organizations** category are constructed using subordinate headings that are not themselves database items but are inserted where needed to create logical groupings of persons and organizations. (See Chapter 4 on containment hierarchies versus subordination hierarchies.)

An item in the **Persons & organizations** category can be identified as a “current person,” in which case there will be additional data-entry fields for such details as the person’s address and photograph and a place for entering links to institutions with which the person is affiliated. Similarly, an item can be identified as an “organization,” with relevant data-entry fields. Otherwise, the item is simply a “person” who may be living or dead, or may even be fictional.

Display name	<input type="text" value="English"/>		Abbreviation	<input type="text" value="UC-OI"/>		Type	<input type="text" value="organization"/>
Description							
<input type="text" value="English"/>							
<p>The Oriental Institute is a research organization and museum devoted to the study of the ancient Near East. Founded in 1919 by James Henry Breasted, the Institute, a part of the University of</p>							
Address line							
<input type="text" value="1155 East 58th Street"/>							
City	<input type="text" value="Chicago"/>	State/Province	<input type="text" value="IL"/>	Country	<input type="text" value="USA"/>	Postal code	<input type="text" value="60637"/>
Affiliations				Pictures			
<input type="button" value="0+"/> <input type="button" value="0-"/>				<input type="text" value="Oriental Institute"/>			
<input type="button" value="Preferences"/> <input type="button" value="Properties"/> <input type="button" value="Links ✓"/> <input type="button" value="Notes ✓"/> <input type="button" value="Events"/>							
<input type="checkbox"/> Keep this item private							

Event	Location	Person	Other	Period	Start Date
Born	Helmstedt				1822-10-07
Studied at University	Medical Faculty of the				
Student of	University of Göttingen	Wagner, Rudolf			
Promoted (degree) / graduated	University of Göttingen				
Employee of	University of Göttingen	Wagner, Rudolf			
Received (award)	University of Göttingen				

Events in a person's life (e.g., birth, death, or any other event) can be represented by entries in the **Events** tab in the person's item pane. The procedure for entering events for a person or another kind of item is described below in Chapter 7.

By representing persons and organizations in this way, OCHRE makes it easy to attribute scholarly data and interpretations to particular observers, authors, and editors. This is important, not just for assigning credit for work that is done (and perhaps assigning blame for errors that are made), but also to allow later researchers to disentangle the contributions of scholars to whom particular ideas and interpretations are attributed, and to see the intellectual and social connections among networks of scholars.

Items that Represent Project-defined Concepts

Concepts

A number of concepts are predefined in OCHRE, starting with units of space, time, and agency represented as items in the **Locations & objects**, **Periods**, and **Persons & organizations** categories. In addition to creating items in one of the predefined OCHRE categories, researchers can express their concepts very flexibly by means of their project's taxonomy of descriptive terms, as described in Chapter 13. Indeed, the flexible use of project-defined terms to describe other entities is fundamental to OCHRE's design.

However, some projects may wish to define certain concepts as database items that do not fit within any of the predefined item categories and do not belong in the taxonomy. By representing them as items in the **Concepts** category, rather than as taxonomic properties,

they can more easily be treated as entities to be studied in their own right. For example, a project in the history of religion might wish to represent the concept of monotheism as a database item.

Concept items can have their own properties, events, and links to other items. Moreover, OCHRE allows multiple interpretations of the same concept item by different interpreters in the same way that it allows multiple observations of the same spatial item in the **Locations & objects** category.

Items in the **Concepts** category can be organized in “containment hierarchies” that represent the strict containment of one concept within another. This provides one way of representing the semantic relationships among concepts. Another way is to link concept items directly to other concept items using the linking methods described in Chapter 6. Any item, including a project-defined concept, can be linked to any other item by means of either a “simple link” or a relational property. In combination, such links serve to represent the network of semantic relationships that a project wishes to express.

Items that Represent Images and Other Resources

Resources

In OCHRE, two-dimensional image files in various formats (e.g., TIFF or JPEG) are regarded as “external resources.” Other external resources include three-dimensional models, audio and video clips, PDF files, geospatial data such as vector map shapes and raster files, and several other kinds of files (see Chapter 9). A project’s external resources are registered in an OCHRE database as items in the **Resources** category. A new item must be created for each resource.

The **Resources** category also contains “internal documents,” which are not external resources but rather internal resources stored within the central OCHRE database. Chapter 9 contains instructions for composing internal documents or importing them into the database from an external file.

In contrast to internal documents, external resources are not stored within the central OCHRE database but are managed locally

by each project. A project will follow its own conventions in naming and organizing such data sources and will make them available on its own Web server, from which data will be retrieved as needed for display in the OCHRE user interface. Only the Internet addresses (URLs) of external resources are stored internally in an OCHRE database.

The screenshot shows the OCHRE interface for managing resources. At the top, there are tabs for 'Resource' (selected), 'View', 'Creators', 'Bibliography', 'Copyright', and 'Context(s)'. Below the tabs is a URL bar containing '/Resources/Geophysical images/Satellite image of Zincirli' with 'Refresh' and 'Go to' buttons. A navigation bar below the URL bar includes 'Information', 'Preferences', 'Properties', 'Links', 'Notes', 'Events', and 'Image Tools'. The main content area is divided into sections: 'Name' (containing 'English' and 'Satellite image of Zincirli'), 'Abbreviation' (containing 'English' and a dropdown menu), 'Type' (containing 'GIS'), 'Description' (containing 'English' and 'Area surrounding the site of Zincirli, 5 km²'), 'External resource' (containing 'File URI' with 'Regional/Digital%20Globe%20satellite%20image%20WGS84%20UTM37N.jpg' and 'Open...' and 'Browse...' buttons), 'Low-resolution (preview) file URI' (containing 'file:/C:/Zincirli/GIS/Regional/Digital%20Globe%20satellite%20image%20WGS84%20UTM37N thumbnail.gif'), and 'High-resolution file URI' (containing 'file:/C:/Zincirli/GIS/Regional/Digital%20Globe%20satellite%20image%20WGS84%20UTM37N.jpg').

Items in the **Resources** category can be organized in “subordination hierarchies,” in which items are grouped under subordinate headings. Any item in this category can be linked directly to any other OCHRE item. However, it may also be possible, depending on the resource type, to insert a link to a specific location within the resource, as opposed to a general link that pertains to the resource as a whole. For example, an item elsewhere in the database can be linked to a “hotspot” within an image or to a specific page within a PDF document. This capability is useful for archaeologists who want to

annotate a photograph in which several different architectural features and soil layers are shown. The part of the photograph in which a particular feature is visible can be linked directly to the item in the **Locations & objects** category that represents that feature as a spatial unit. This provides a way subsequently to label images automatically based on their links to other items and to navigate those items visually by clicking on the image.

Items that Represent Complex Texts and Writing Systems

Texts, Epigraphic units, and Discourse units

Database items in the **Texts** category represent written texts which are themselves objects of study and analysis. For example, items in this category are used to represent ancient inscriptions that employ a complex nonalphabetic writing system like Egyptian hieroglyphs or Mesopotamian cuneiform. Ordinary documents (e.g., scholarly articles) are not normally stored in the **Texts** category but are instead represented as **Resource** items—perhaps as internal documents or, in many cases, as external resources in the PDF format or in a plain-text ASCII or Unicode format. However, any text that will be subjected to a detailed epigraphic, grammatical, or linguistic analysis should be entered (or automatically imported) into OCHRE as a database item in the **Texts** category, regardless of its writing system, medium, or period of composition. Scholars perform detailed analyses, not just on ancient tablets and medieval manuscripts, but on printed texts of the modern era. See Chapter 11 for more information about working with texts and writing systems.

An item in the **Texts** category consists of an epigraphic hierarchy and/or a discourse hierarchy. The epigraphic hierarchy represents the physical structure of the text in terms of its division into epigraphic units at various levels of detail—usually sections, columns, lines, and individual graphic signs within a line. The discourse hierarchy represents the structure of the text in terms of its division into meaningful units of discourse at various levels of analysis; for example, paragraphs, sentences, clauses, phrases, words, and morphemes within

words. Each epigraphic unit in a text’s epigraphic hierarchy and each discourse unit in its discourse hierarchy is itself an OCHRE database item that can be described separately and linked to other OCHRE items. A text’s epigraphic units are linked to its discourse units in a way that represents the relationship between physical graphic signs and their readings as meaningful linguistic expressions.

As an alternative to the manual construction of epigraphic and discourse hierarchies, which can be quite tedious, OCHRE provides tools to automatically build these hierarchies and the cross-cutting links between them in the process of importing a digitized text. Note that OCHRE does not prescribe a particular structure of epigraphic and discourse units that must be used by all projects. The nature and scope of epigraphic and discourse units and their hierarchical arrangement are defined by the project, depending on how a particular text is analyzed. Some texts will be broken down only as far as the line or sentence level; other texts will be analyzed into their constituent graphemes and morphemes. The same text may have multiple epigraphic and discourse hierarchies, reflecting different analyses of the text by different editors, even within the same project. The procedure for entering epigraphic and discourse hierarchies is described in detail in Chapter 11.

Although a text in the **Texts** category can be analyzed into many subunits in this way, it can also be treated as a single database item with its own properties and links to other items. Items in the **Texts** category can be organized in subordination hierarchies, in which items are grouped under subordinate headings.

Writing systems and Script units

An item hierarchy in the **Writing systems** category represents a writing system whereas an individual item within such a hierarchy represents a “script unit”—a distinct graphic sign within the writing system. Item hierarchies in this category are subordination hierarchies, in which items are grouped under subordinate headings. This allows the construction of a hierarchy of script units organized into named groups representing classes of signs in a given writing system.

A script unit is an ideal entity that can be manifested in more than one graphic form or “allograph” (e.g., A and a are variant forms of the letter “a”). Each allograph of a script unit is represented in OCHRE by a font glyph (indicated by a font name and character code) or by an image. Epigraphic units of texts in the **Texts** category can be linked to script units in the **Writing systems** category in a way that specifies which allograph of a script unit is manifested at a particular place in a text.

In a logographic or syllabic writing system, quite different phonetic readings may be possible for a given script unit. For this reason, OCHRE allows multiple readings to be associated with each script unit, just as it allows multiple allographs. Discourse units of texts in the **Texts** category can be linked to script units in the **Writing systems** category in a way that specifies which reading of a script unit is to be used at a particular place in a text.

Note that a script unit’s readings (phonetic variants) are quite independent of its allographs (graphic variants). In general, all graphic forms of a script unit can be read phonetically in all of the possible ways that the script unit can be read. These independent graphic and phonetic dimensions of what is still recognizably the same script unit justify OCHRE’s treatment of script units as ideal entities capable of diverse modes of graphic and phonetic expression.

In some projects, the complexities of multiple allographs and multiple readings will need to be represented, not for a complex logographic writing system, but for a much simpler alphabetic writing system. To take a familiar example, a character (script unit) in the Latin alphabet may be manifested in quite different graphic forms (allographs) depending on the inscriptive tradition and medium of inscription. In modern handwriting, the “printed” form and the cursive form of an alphabetic character may be quite different and may also vary from writer to writer. Moreover, the same Latin character can have different phonetic readings when used for different languages (e.g., Turkish “c” corresponds to English “j” as in “jar”). Indeed, in English orthography the phonetic reading often depends on where the character is used (e.g., “c” in “cat” versus “nice”).

For many texts written in relatively simple alphabetic writing systems it will not be necessary to represent the alphabetic characters as script units in OCHRE. Epigraphic units in the **Texts** category are not required to be linked to script units in the **Writing systems** category; this is optional. However, certain kinds of textual study call for the explicit modeling of the writing system as part of the digital representation of a corpus of texts. For example, a researcher might wish to search for a particular allograph in relation to other properties of the texts in which it appears in order to reconstruct the cultural and temporal connections among the scribes responsible for the texts. The complexities of writing systems, and the use of OCHRE to represent them, are dealt with at greater length below in Chapter 11.

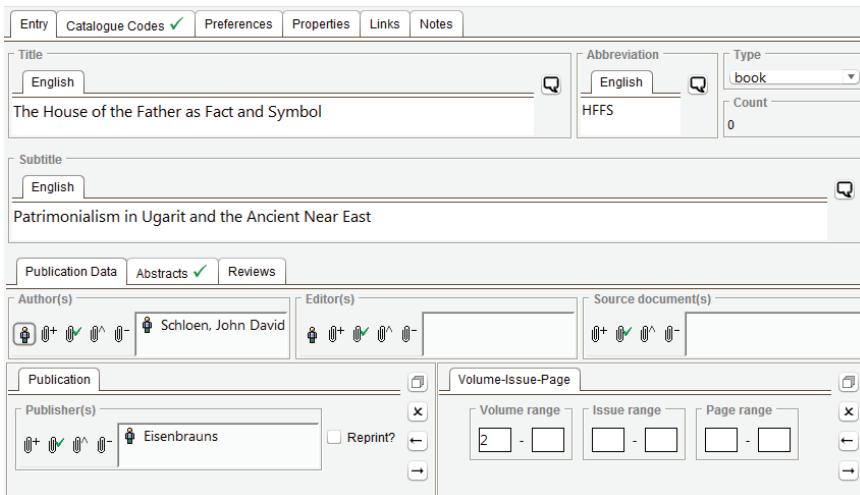
Items that Represent Research Tools

Bibliographies

An item in the **Bibliographies** category represents a bibliographic entry of some kind. A hierarchy in this category represents a group of related bibliographic entries. These are containment hierarchies, as in the **Locations & objects** category (see the section on “Containment Hierarchies and Subordination Hierarchies” in Chapter 4). They represent containment relationships of various kinds; for example, the containment of articles within journals, books within series, chapters within books, and so on. Using containment hierarchies in this way avoids the need to repeat bibliographic information about a series to which many books belong or a journal to which many articles belong, because this information is inherited from the parent in the hierarchy.

The **Bibliographies** category pertains to the published works, often in nondigital form, that are cited by the project. In contrast, the **Resources** category pertains to digital resources, often unpublished, to which a project has established links so they can be retrieved and viewed. Sometimes a digital resource will consist of a published work (e.g., a PDF version of a journal article) for which there is a corresponding bibliographic entry. In that case, the item in the **Resources** category can be linked to the relevant item in the **Bibliographies**

category. Keep in mind, however, that bibliography items do not necessarily have any connection to the digitized content of the works they reference; they simply contain bibliographic information about those works, which in many cases are nondigital printed works available only as physical objects in libraries.



The procedure for constructing and using bibliographic entries is described in Chapter 10. That chapter includes a discussion of how to link OCHRE bibliographies to the widely used Zotero online database, “a free, easy-to-use tool to help you collect, organize, cite, and share your research sources” (see <http://www.zotero.org>).

Dictionaries

The **Dictionaries** category provides a way to represent lexical information about one or more languages. This lexical information can be interlinked with other information in the database, especially with texts and their discourse units in the **Texts** category and with bibliographic data that has been entered in the **Bibliographies** category. You can define your own dictionary entries or you can copy entries from another project by repeating the relevant database items within your project’s **Dictionaries** category (see the section on “Repeating Items and Hierarchical Branches” in Chapter 4).

An individual item in the **Dictionaries** category represents a dictionary entry for a particular lemma (headword or lexeme) in the lexicon of a language. A hierarchy of items in this category represents an entire dictionary for a particular language. The order of the items in the hierarchy reflects the order of the entries within the dictionary.

Hierarchies in the **Dictionaries** category are subordination hierarchies (see the section on “Containment Hierarchies and Subordination Hierarchies” in Chapter 4). You can use headings within a hierarchy to organize dictionary entries into named groups. For example, the name of each volume in a multi-volume dictionary would be indicated by a heading within the hierarchy that represents the entire dictionary. Inserted under each heading would be items that represent the dictionary entries in a given volume.

In data-entry mode, many different data-entry tabs and fields are shown for a dictionary entry, reflecting the potential complexity of this kind of item, in which multiple forms, meanings, and citations may be listed for a given lemma. In view-only mode (and in the **View** tab of the data-entry item pane) these data-entry tabs and fields are not seen. Instead, their contents are combined and formatted to replicate the appearance of the dictionary entry as it would appear in a traditional printed dictionary.

When you select a hierarchy within the **Dictionaries** category the item pane will display data-entry fields for the hierarchy’s creators, date, name, abbreviation, description, properties, links, notes, and display preferences. These fields are the same as those used for hierarchies elsewhere (see the section on “Hierarchies of Items” in Chapter 4). However, there is an additional **Type of dictionary** field in which you can specify whether the dictionary entries contained in the hierarchy are “extended” entries or simpler “glossary” entries.

An extended entry can be quite complex, with many different meanings and submeanings for a given lemma and detailed morphological, comparative, and semantic information. This is in keeping with the complexity of comprehensive philological dictionaries, of which the *Oxford English Dictionary* is a well-known example. A glossary entry is much simpler; it contains a more restricted set of

information about the lemma. The procedure for constructing dictionary entries, both in the extended format and the glossary format, is explained below in Chapter 12.

Items that Represent the Properties of Other Items

Property variables and values

All OCHRE items have fundamentally the same structure, regardless of their category. Every item in the database has a primary name and may also have an abbreviated name, one or more aliases, a free-form description, one or more free-form notes, one or more links to other items, and one or more structured properties (these are often called “attributes” in other database systems). An item’s properties are themselves represented by means of database items that share this same basic structure and are grouped in two item categories: **Property variables** and **Property values**.

Two distinct property categories are needed because, in OCHRE, a property is understood to be a compound entity that consists of a property name such as “color” and a specific value such as “red.” In casual parlance, the word “property” is ambiguous because it can mean either the property name alone (“color”) or the entire name-value pair (“color = red”). To prevent confusion, OCHRE makes a clear distinction between the property name, which is called the variable, and the property as a whole, which consists of a variable-value pair.

OCHRE’s use of the term “variable” is borrowed from social research. In the social sciences, a variable, whether it be quantitative or qualitative, is a logical set of values that vary across a particular domain. For example, the variable “gender” has the values “male” and “female” (see *The Practice of Social Research* by Earl Babbie [12th ed.; Belmont, California: Wadsworth, 2010], pp. 14–15).

In social research, the values of a variable are often called “attributes” but this term is avoided in OCHRE because it has quite different meanings in different contexts. In social research, an attribute is a specific value of a variable; thus “gender” is a variable with the

two attributes “male” and “female.” In other settings, however, an attribute is a property name and so is equivalent to a social-scientific variable, or the term attribute may refer to an entire variable-value pair (e.g., “gender = male”).

Note that OCHRE does not prescribe the variables (i.e., property names) to be used when describing items. Unlike most other database systems, these names are not built into the software and imposed on users but are defined by the users themselves. A project will define its own variables as database items in the **Property variables** category or it will borrow such items by copying them from another OCHRE project (see Chapter 13).

The screenshot shows the OCHRE software interface for managing registered objects. The top navigation bar includes tabs for 'Information' (selected), 'Description' (with a green checkmark), 'Preferences', 'Events' (with a green checkmark), and 'Coordinates'. Below the tabs, there are sections for 'Name' (containing 'English' and 'R08-10'), 'Abbreviation' (containing 'English'), and 'Code'. A date field shows '07/20/08'. Under the 'Observers' section, there is a list of observers: 'Pace, Leann' with a timestamp of '07/20/08'. Below these sections are buttons for 'Properties' (with a green checkmark), 'Links' (with a green checkmark), and 'Notes'. A toolbar at the bottom includes icons for back, forward, search, and apply predefined properties. The main content area displays a table of variables and their values:

Variable	Value	?	Comment
Registered object no.	10	<input type="checkbox"/>	
Location or object type	Registered item	<input type="checkbox"/>	
Material	Metal	<input type="checkbox"/>	
Material	Gold	<input type="checkbox"/>	
Object type	Coin	<input type="checkbox"/>	
Object type	Stater	<input type="checkbox"/>	
Weight (g)	8.6	<input type="checkbox"/>	
Weight, postconservation (g)	8.5	<input type="checkbox"/>	
Diameter (cm)	1.8	<input type="checkbox"/>	
Thickness (cm)	0.4	<input type="checkbox"/>	
Elevation (m)	527.44	<input type="checkbox"/>	

At the bottom of the form are buttons for 'Save', 'Save and Close', and 'Cancel'.

The different types of variables are discussed below in Chapter 5. Some variables are quantitative and take numeric values in the form

of integers, decimal numbers, or dates. Other variables are qualitative and take character-string values; for example, “red,” “green,” and “blue” as values of the property “color.” In most cases, character-string values that are intended to be used with qualitative variables will be treated as database items in their own right and will be listed in the **Property values** category. A particular property will then be assigned to an OCHRE item by means of a link to the appropriate item in the **Property variables** category and, in the case of a qualitative property, by means of a second link to the appropriate item in the **Property values** category. This avoids the error-prone and space-consuming repetition of character strings throughout the database because a given character string, representing a particular property name or qualitative value, exists only once in the database and is referenced in a highly efficient manner by internal pointers.

Variables and their possible qualitative values are semantically related to one another by means of the **Taxonomy** category, which is discussed briefly below and is treated in greater detail in Chapter 13. Keep in mind that a value has no determinate meaning apart from the variable with which it is paired. For example, the value “light” means something quite different when it is paired with the variable “weight” (with values “light” and “heavy”) than when it is paired with the variable “color” (with values “light” and “dark”). The meaning of a value is determined contextually by the variable with which it is associated and differentially with respect to other possible values of the same variable, as prescribed in the taxonomy.

Apart from the project’s taxonomy, a variable can also be semantically related to another variable by means of a thesaurus relationship, which can be of one of two types: “synonym” or “related term” (i.e., semantically related in some way but not actually synonymous). A character-string value can be semantically related to another value in the same way. Thesaurus relationships for variables and values are entered in the **Thesaurus** data-entry tab of the currently selected variable or value item.

A taxonomic term (a variable or value) can be related in this way, not only to synonymous terms in the current project, but to synony-

mous terms in other OCHRE projects to which the user has access. This provides an important mechanism for integrating information across projects by making it easy to find synonymous but non-identical terms. A query can use thesaurus relationships to find, not just the item properties (variable-value pairs) that are explicitly specified in the query criteria, but also semantically related properties that may reflect a different nomenclature. Moreover, when performing a query, a user can choose to use the thesaurus relationships that have been defined by a different OCHRE project, in addition to, or instead of, the current project's thesaurus relationships.

The construction of thesaurus relationships is described in Chapter 13 and their use in queries is described in Chapter 14. You can display all existing thesaurus relationships by clicking the **Property variables** or **Property values** heading in the navigation pane to select the category and then viewing its **Thesaurus** tab in the item pane. This will show you an alphabetical list of the currently defined variables or values, as the case may be, together with their synonymous and related variables or values. Similarly, you can see a complete alphabetical list of both variables and values, intermingled in a single list, together with their semantically related terms, by clicking on the project name at the top of the navigation pane and then viewing the project's **Thesaurus** tab in the item pane.

Finally, variables and values can be semantically organized, not just by means of a taxonomic hierarchy and thesaurus relationships, as described above, but also via “subordination hierarchies” (see Chapter 4). A project will typically use hundreds of variables and thousands of qualitative values, so subordinate headings are a useful way to organize items in the **Property variables** and **Property values** categories into named groups and subgroups. But sometimes a user will want to flatten these hierarchies in order to see a complete alphabetical listing of the variables or values that have been defined by the project—for example, to determine whether a particular taxonomic term already exists or needs to be created. In that case, the user will select the **Property variables** or **Property values** category heading in the navigation pane in order to view the **Thesaurus** tab in the item

pane, as described above, selecting the option to **Hide linked terms**. (The navigation pane and item pane are described in Chapter 3.)

Property predefinitions

Items in the **Property predefinitions** category represent named groups of predefined properties (i.e., variable-value pairs) that have been saved so they can be assigned all at once to an item that is being described. Data entry can be done more quickly with the aid of predefined properties in cases where a series of items to be described have the same or similar properties (see Chapter 5).

Like any OCHRE database item, each item in the **Property predefinitions** category has properties. In this case, the item's properties are its primary content. They constitute the predefined group of properties that is available to be assigned to another item.

For some predefined properties, the value is not predefined but will be filled in during data entry; in other words, only the variable is predefined in the variable-value pair that makes up the property. In that case, a placeholder value called **<unassigned>** will be paired with the variable to indicate that its value is not yet defined.

Variable	Value
Location or object type	Registered item
▼ Material	Metal
Material	<unassigned>
Object type	<unassigned>
Diameter (cm)	0.0
Thickness (mm)	0.0
Degree of completeness	Partial

Taxonomy

Each project in OCHRE has a taxonomy that specifies the way in which property names (variables) and values may be used to describe other items. The taxonomy consists of a special hierarchy of database items drawn from the **Property variables** and **Property values** categories. This hierarchy specifies the logical relationships among variables and values. It defines the way in which particular properties (variable-value pairs) may be nested within other properties to repre-

sent explicitly the relationships between specific properties and more general properties (see Chapter 5 on assigning item properties and Chapter 13 on constructing a taxonomy). The taxonomic hierarchy is the only hierarchy allowed in the **Taxonomy** category.

At the top level of the taxonomic hierarchy is a list of variables. Quantitative variables that take numeric values cannot have child items in the hierarchy. However, underneath each qualitative variable are the possible character-string values of that variable, represented as its children in the hierarchy. For example, in a given project it may be that the variable “color” can have only the value “red,” “green,” or “blue.” This would be shown in the taxonomic hierarchy as follows, in the left-hand example:



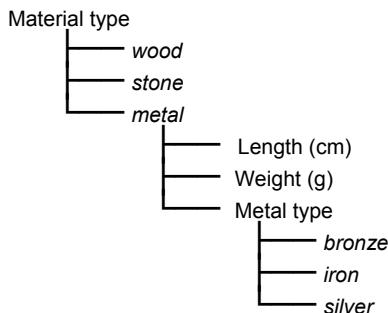
In the case of a qualitative variable of the **ordinal** type the order of the child values indicates their relative positions in an ordered series. For example, the ordinal variable “Size” might have the value “small”, “medium,” or “large,” as shown above in the right-hand example, and the fact that these values are listed in that order in the taxonomy indicates which is bigger and which is smaller. Thus, a query could use the taxonomy to find all items with “Size” greater than “small.”

At the next level of the hierarchy, under each of the qualitative values, there may be yet another list of qualitative and/or quantitative variables. Once again, the quantitative variables cannot have child items but each qualitative variable has character-string values as its children in the hierarchy. Each of those values may in turn have a child list of variables, creating a deeper level in the hierarchy. A taxonomic hierarchy can have as many levels as are needed to represent the logical relationships among variables and values.

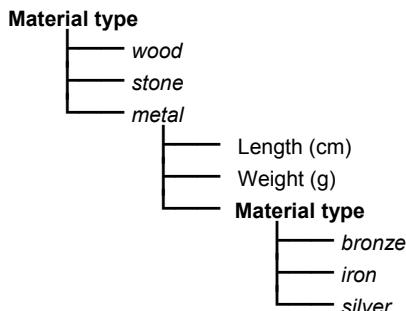
Only the variables at the top level of the taxonomic hierarchy will be available when one begins entering properties for an item (see Chapter 5). But if a top-level variable has other variables nested within it in the taxonomy (i.e., as its “grandchildren,” the children of

its child values) these will be available for data entry once the top-level variable has been assigned to the item.

For example, a user may wish to assign to an item in the **Locations & objects** category the top-level variable “Material type” with value “metal”; and nested within that property (variable-value pair) the user may wish to assign the more specific variable “Metal type” with value “bronze.” “Metal type” cannot be inserted at the top level because it is not allowed as a top-level variable in the taxonomic hierarchy but can only be a child of the value “metal,” which in turn is a child of the variable “Material type.” The project administrator has decided that “Metal type” is to be used only within the context of the more general property “Material type = metal.”



Alternatively, to avoid proliferating variables, the same variable can be nested recursively within itself. In that case, “Material type” would still contain “metal” as one of its values, but that value would in turn contain the same variable “Material type,” within which would be the more specific values “bronze,” “iron,” and “silver” at a lower level in the taxonomic hierarchy, as shown here:



In either case, the taxonomic hierarchy is an alternating hierarchy of variables that contain values, which in turn contain variables, and so on. For each qualitative variable you *must* enter one or more character-string values. For each of these values you *may* enter a subordinate variable, recognizing that some variables are quite specific and are needed only within the context of another, more general, variable; or perhaps only in the context of a particular value of the more general variable. The same variables and values can be reused at different locations in the taxonomic hierarchy, depending on what is needed in a particular context.

Organizing a project's property variables and values in a hierarchical taxonomy has two main benefits: it facilitates the entering of item properties and it represents explicitly the logical relationships between general and specific properties in a way that enables more powerful queries. The taxonomy constrains data entry, preventing errors and automatically generating context-specific pick-lists of permitted variables and values. The taxonomy also allows automatic query expansion in order to include more specific values of a recursively nested variable.

Using the example illustrated above, a database query whose search criteria included "Material type = metal" would automatically retrieve items with the more specific properties "Material type = bronze" or "Material type = iron" or "Material type = silver," as well as items with the general property "Material type = metal." The recursive taxonomic hierarchy formally represents the fact that bronze is a specific type of metal.

Items that Organize, Select, and Display Other Items

Hierarchies

Hierarchies are used extensively in OCHRE to organize database items. The **Bibliographies**, **Concepts**, **Locations & objects**, and **Periods** categories use "containment hierarchies" that represent the strict containment of entities within other entities. The **Dictionaries**, **Persons & organizations**, **Presentations**, **Property predefinitions**, **Property values**,

Property variables, Queries, Resources, Sets, and Users categories all use “subordination hierarchies,” in which items are grouped under headings in order to represent relationships of logical subordination rather than containment. The **Taxonomy, Texts, and Writing systems** categories make use of specialized hierarchies that have more complex structures. In all cases, however, the hierarchy is itself an independent database item that has its own name and (optionally) its own description, properties, notes, and links to other items. The procedure for creating and populating hierarchies is described in Chapter 4.

Note that a database item may appear in more than one hierarchy, or in more than one location within the same hierarchy, and yet still be the same item. By allowing the same item to appear in different hierarchical contexts, OCHRE can capture the multiple modes of classifying and organizing items that a project might need in order to represent its information.

Sets

Items in the **Sets** category represent named groups of items that exist in other categories but whose grouping into a set has been saved for some reason. For example, the list of database items found by a query can be saved as a named item in the **Sets** category. This set can then be used in various ways: its items can be displayed in a table, drawn on a map or timeline, or used to limit the scope of another query. In addition to creating a set automatically as the result of a query, a set can be constructed manually, by linking items to it one by one from other categories.

When the contents of a set are displayed as a table, the items in the set are shown as the rows of the table and the properties of the items are shown as columns (see Chapter 5). The leftmost column usually contains the names of the items. The user can choose to hide some of the columns (properties), if they are not needed in the table view, and can change the order of the columns.

The table view of a set of items can be printed out or exported to an external file in the Microsoft Excel format. Summary statistics can be calculated for table columns that contain the numeric values of

quantitative variables. The table view of a set can also be used for data entry; in other words, the contents of the table cells, which display the values of the properties of each item, can be directly edited by the user if he or she has data-entry privileges for the items in the set. In some cases, editing data is easier to do in table mode than in the usual hierarchical item-oriented data-entry mode (see Chapter 15).

Name	Abbreviation	Type
English	English	<Select type>
Current year's pottery pails (2010-07-13 08:06:38)		
Description		
English		
Pottery pails excavated to date in the 2010 season.		
Set Items ✓ Table Columns Preferences Utilities Properties Links Notes		
Associated query		Results date/time
Pottery buckets, 2010		07/23/11 20:13:19
Load set items Remove set items		Count
		1922
+/- ↑ ↓ 1922 items		
✖ P10-17.44#101 ✖ P10-17.44#102 ✖ P10-17.44#103 ✖ P10-17.44#104 ✖ P10-17.44#105		

Queries

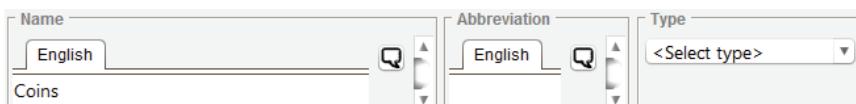
Each item in the **Queries** category represents a named query definition that contains search criteria and instructions about the scope of the search. (See Chapter 14 for detailed instructions about constructing and using queries.) Query definitions are saved as database items in their own right so that they can be used repeatedly and shared with other users.

A project administrator will normally construct a variety of pre-defined queries that are intended to be used frequently by members of the project team in data-entry mode and by view-only users who will study the project's data. The project administrator can choose predefined queries to display in a **Queries** tab in the navigation pane,

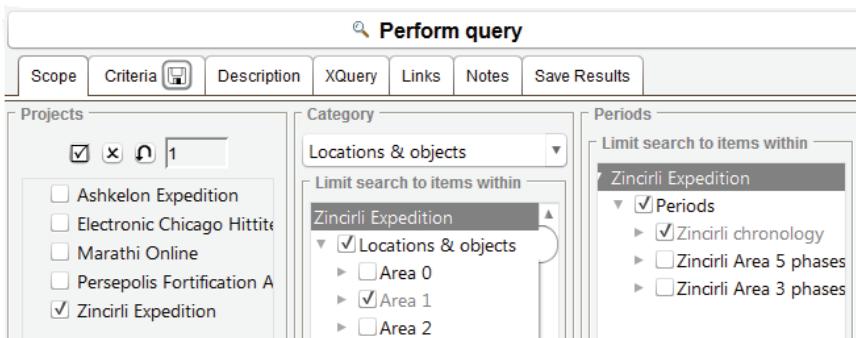
for the convenience of view-only users. This is done by means of the **View-only** subtab in the **Preferences** tab of the project's item pane (see Chapter 17).

The list of database items retrieved by a query can be inspected on the spot or saved for future use as a named set of items in the **Sets** category. The query item's **View** tab in the item pane shows a list of the items found by the most recent performance of the query. You can click on any individual item in the list to display its information or you can view all of the items together using various display formats.

The scope of a query is defined in terms of the category of items to be searched and the project (or projects) to be searched. A query can span multiple projects but it will retrieve only those items to which the user has access. Items marked "private" by a project will not be retrieved unless the user has been given a user account by that project and has access to the relevant item category. Items that are public and thus visible to anonymous users will be retrieved by a query regardless of the user's access privileges.



```
FIND ALL { items } FROM Zincirli
WITHIN { Locations & objects [Area 1] }
DURING { Periods [Zincirli chronology] }
WHERE
    Location or object type is Registered item
    Object type is Coin
```



A query can search for particular properties (variable-value pairs) and combinations of properties. Parentheses and logical operators can be used to construct Boolean algebraic expressions of considerable complexity which incorporate multiple properties. With respect to textual content, a query can search for a character string (a word or phrase) within the names, descriptions, and free-form notes of items.

The search criteria of a query make use of the intrinsic properties of database items, their textual content, and the events that affect them. However, queries can also be combined to create compound queries that search, not just the intrinsic characteristics of items, but also the extrinsic spatial and temporal containment relationships among items and the links between pairs of items.

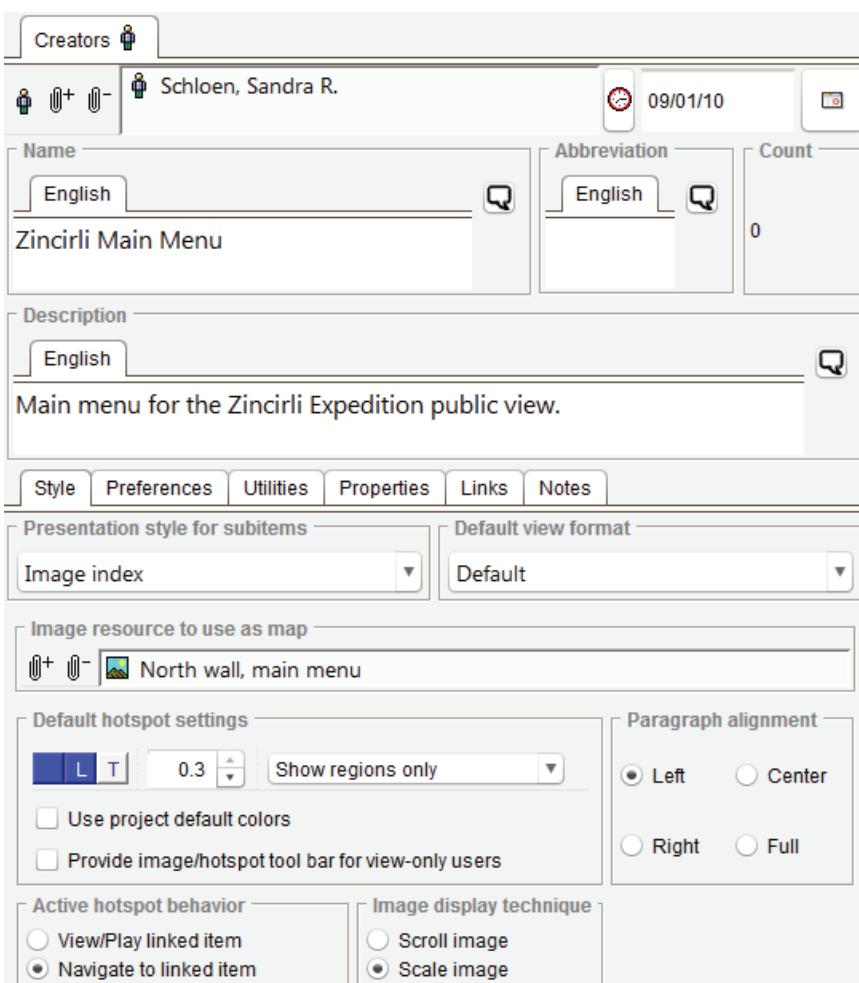
Presentations

Items in the **Presentations** category are used to specify how a project’s data should be presented to the public, that is, to anonymous users who do not have user accounts and are not logged into the project, but who wish to view and query its data. This category provides a way to customize a project’s appearance for public use in view-only mode without displaying the more complex data-entry mode.

Each item hierarchy in the **Presentations** category represents an entire presentation; the items within a hierarchy are components of that presentation. The sequence of components within the hierarchy and the nesting of components within the hierarchy’s headings and subheadings determine their order of presentation. The project administrator can specify in the project’s **Preferences** tab whether a presentation hierarchy is to be used automatically to display the project’s data to any anonymous user who opens the project (see the “View-only” subsection of “Project-wide Preferences” in Chapter 17). The project administrator can also specify which presentation to use, in case more than one presentation hierarchy has been created.

Presentations might be created for pedagogical purposes, to be used as scripted lessons, or as a form of scholarly publication for use by researchers and advanced students. An OCHRE presentation can

include database items from any category. A particularly useful presentation style is the “image index,” which uses a resource item that represents an annotated image (e.g., a photograph) in which “hotspot” links to other database items, including other image indexes, have been embedded. Users can navigate a project’s data visually by clicking on such images and drilling down to increasing levels of detail. See Chapter 16 for detailed instructions about creating and using presentations.



BASIC OPERATIONS

Chapter 3

Accessing a Project to View or Modify Its Information

AFTER starting OCHRE (see Chapter 1) you must choose a project to work with from the list of projects that will be displayed in the initial window. An OCHRE database can support a large number of research projects; how many will depend on the particular database and the organization that hosts it. The leaders of projects can organize and share information while preserving their own terminology and restricting access to their data. OCHRE allows multiple levels of access to a project, leaving it up to the project's leader to determine what can be viewed and modified.

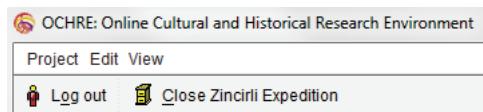
Levels of Access

Anonymous Users and Named Users

If you do not log in but remain anonymous, the list of projects displayed to you will consist only of those projects that have made their data public for all OCHRE users to view and query. However, if you log in with a username assigned to you by a particular project, then that project—and any other projects that have given you a user account with the same username—will be added to the list of available projects. You will have data-entry privileges for each project in which you have a user account, at a level determined by the project administrator (see Chapter 17).

To select a project from the list, click on it to highlight it and then click the **Open project** button at the top of the OCHRE window, or simply double-click the project name to open it. The item categories of the selected project will appear in the navigation pane on the left side of the window. You may then proceed to view and query this project's data. If you have logged in with a username for which the project has given you data-entry privileges, you will be able to add or modify data.

At any time, you may change the currently selected project by clicking the **Close project** button at the top of the OCHRE window. Doing so will clear the project information from the navigation pane and cause the list of available OCHRE projects to be displayed again. The same thing will happen if you log in with a different username by clicking on the **Log out** button and then clicking on the **Log in** button at the top of the window.



View-only Mode and Data-entry Mode

You will often work with projects for which you do not have a user account, either because you have not logged in or because you have logged in as a member of a project different from the one you are currently viewing. The information for such projects will be displayed to you in view-only mode. On the other hand, if you have opened a project for which you have a user account, the information will be displayed in data-entry mode. The main difference between these two modes is that data may be added or modified only in data-entry mode, thus the individual data fields, which can be quite complex, are usually not shown in view-only mode.

A project administrator can designate a particular way of navigating and displaying a project's information in view-only mode by selecting the appropriate options in the **View-only** subtab in the **Preferences** tab of the project's item pane (see Chapter 17). Thus the view-only mode can be customized in a manner that is suitable for the project. It is often desirable to provide a simplified user interface for anonymous users who wish only to view and query a project's data, in contrast to the more complex user interface that is needed in data-entry mode. For this purpose, a scripted presentation can be created in the **Presentations** category and then specified in the project's preferences as the vehicle for displaying the project's data to anonymous users who do not have access to the data-entry interface. (See Chapter 16 for more information about presentations and the various display modes available in them.)

Locking Items While You Modify Them

OCHRE is an online, multi-user database system. As a result, it may happen that more than one user will try to access the same database item at the same time. This is not a problem in view-only mode because the central database server can send a copy of the item's information to be viewed by each user. However, in data-entry mode the item must be "locked" by the person who is modifying its information in order to prevent anyone else from trying to modify it at the same time.

To lock the item that is currently selected in the navigation pane, you would click the  "lock" button in the main toolbar. As soon as you lock an item, any other users who select it will receive a message saying that it has been locked and cannot be modified by them. When you are finished with it, you must click the  "unlock" button to release it. The  "show all locks" button displays all of the items that you have locked and not yet unlocked. The  "whose lock?" button displays the name of the user who has locked the currently selected item.

Like any other database item, a hierarchy of items must be locked before it can be modified; for example, before items can be inserted, deleted, or moved within the hierarchy. It is possible to lock all of the item categories in a project by selecting the project item in the top left corner of the navigation pane and then clicking the  button, although this is not usually necessary and it may cause problems for other members of your project team, who will be prevented from entering data while your lock is in effect.

You may modify an individual item (i.e., an item that is not a hierarchy) without explicitly locking it, in which case an automatic lock will be placed on it until you leave the item. As soon as you leave the item by selecting a different item or by closing the project, any changes you have made will be automatically saved and the item will be unlocked and made available for another user to modify. Nonetheless, it is a good practice to explicitly lock any item that you will be modifying for a long time, especially if it is likely that another user will try to modify it at the same time.

Window Panes in the User Interface

The OCHRE user interface is divided into three resizable window panes: the navigation pane on the left, the item pane in the middle, and the reference pane on the right.

The Navigation Pane

The navigation pane contains an alphabetical list of item categories used by the project (see Chapter 2 for an explanation of these categories). Each category contains hierarchies of individual database items that can be expanded and collapsed by clicking on the ► and ▼ icons to the left of the hierarchy or item names. A hierarchy item or an item within a hierarchy is selected by clicking on its name.

As an alternative to the standard “tree” view, you can click the  toggle button in the main toolbar to display a “list” view of the navigation pane. This view uses drop-down pick-lists to show and select item categories, hierarchies within each category, and the individual items within the hierarchies.

In view-only mode, you will see only the item categories for which data has been entered and, within these categories, only the item hierarchies that have been made publicly available to view-only users. A project administrator may keep some of the project’s data invisible to people who do not have user accounts for the project while exposing other data for them to see. A project administrator can also customize the view-only mode for the project by replacing the usual list of item categories and their item hierarchies with a different kind of index (e.g., tabbed lists), or by hiding the navigation pane altogether. This is done by selecting the appropriate options in the **View-only** subtab of the **Preferences** tab of the project’s item pane (see Chapter 17).

In data-entry mode, however, the list of item categories will always appear in the navigation pane. In this mode, items can be inserted, moved, and deleted within each category. The project administrator may choose to hide item categories that are not used by the project. The project administrator assigns to each named user specific privileges with respect to each category. For a given category,

a user may be authorized to: (1) view items; (2) view and link items; (3) view and link items and edit item names, properties, and notes; (4) view, link, edit, and insert items; or (5) view, link, edit, insert, and delete items. Anonymous users are allowed only to view the items that a project has made public, and the project's data is shown to them in view-only mode.

The Item Pane

Information about the item currently selected in the navigation pane is shown in the item pane in the middle of the OCHRE window. Its contents will vary depending on the item category.

In data-entry mode, the item pane has two overlapping tabs: a data-entry tab and a **View** tab that displays a formatted view of the data stored for the current item. In view-only mode, the data-entry tab is normally not visible, although it may sometimes appear, albeit without the ability to modify the data, depending on the display format that the project administrator has chosen for view-only users (see Chapter 17).

The **View** tab contains a formatted, printable view of the information stored for the current item. When you are entering data in data-entry mode, you can look at the **View** tab at any time to see how the data will be formatted and displayed in view-only mode. The contents of the **View** tab can be copied into other software for use outside of OCHRE. At the bottom of the **View** tab is a **Citation URL** that uniquely identifies the database item whose information is being displayed; this URL (Uniform Resource Locator for the World Wide Web) can be copied and pasted as an external Web link to be used to start OCHRE and display a formatted view of the cited item within a stand-alone window (see Chapter 16).

The title of the data-entry tab in the item pane reflects the category of the item currently selected in the navigation pane; for example, it will be entitled **Resource** for an item in the **Resources** category. This tab contains data fields and subtabs in which you can add and change information about the current item. In all categories, it contains subtabs for **Properties**, **Links**, and **Notes**. In some, but not

all, categories it contains a subtab for **Events** that pertain to the item. Other subtabs are specific to individual item categories.

Any number of structured properties (variable-value pairs) can be assigned to an item via its **Properties** subtab. These properties may then be used as search criteria in queries. The **Notes** subtab contains one or more free-form notes associated with the item. The procedure for entering properties and notes is described in Chapter 5.

The **Links** subtab contains a list of the other database items that have been linked to the current item by means of “general links” (see Chapter 6). There are four kinds of links: *general links* that pertain to the item as a whole; *named links* that pertain to a particular data field for the item; *hyperlinks* that pertain to a selected portion of the item’s textual content; and *hotspot links* that pertain to a location within an image represented as a resource item. Hyperlinks can be associated with a selected portion of any textual content that belongs to the item, including its name, description, and notes (see Chapter 5).

In data-entry mode, general links are listed in the **Links** subtab of the item pane; named links are shown in various data-entry fields in the item pane, such as the **Creator** field; hyperlinks are shown visually as shaded portions of textual content and are listed in the **Links** tab of the reference pane (described below); and hotspot links are shown visually as shaded regions within an image in the **Image Tools** tab in the item pane of the resource item that represents the image (see Chapter 9). In view-only mode and in the data-entry **View** tab, general links and hotspot links are listed on the right; named links are shown in the item’s information as the contents of the data fields to which they are linked; and the presence of a hyperlink is indicated by a shaded region in the item’s textual content, which can be clicked to display the linked items.

The Reference Pane

On the right side of the OCHRE window is the reference pane. It contains two overlapping tabs entitled **Linked Items** and **Query Results**. These tabs are not always needed, so they may be turned off by de-selecting them in the drop-down **View** menu at the top of the

OCHRE window or by clicking the appropriate toggle buttons on the main toolbar ( for **Linked Items** and  for **Query Results**). If both tabs are turned off, the reference pane will disappear altogether, allowing the item pane to expand horizontally.

In data-entry mode, the **Linked Items** tab allows you to add and remove links between database items, including items in other OCHRE projects to which you have been given access. Within the **Linked Items** tab are two separate areas: the “link target” area at the top and the “textual properties and hyperlinks” area at the bottom. Either of these areas can be hidden from view, if it is not needed, by clicking the **Hide link target** button at the top of the tab or the **Hide textual properties and links** button at the bottom. When the area is hidden, the label on the button will change from **Hide...** to **Show...** and clicking the button again will display the area. Textual properties and hyperlinks are normally hidden by default whereas the link target area is shown by default.

The procedure for linking items is discussed in Chapter 6. Textual properties and hyperlinks pertain to specific portions of the textual content of an item, as opposed to general properties and links, which pertain to the item as a whole. General properties and links are shown in the item pane in the middle of the OCHRE window, in the **Properties** and **Links** subtabs. Textual properties and hyperlinks are shown in the reference pane in similar **Properties** and **Links** tabs; but these pertain, not to the whole item, but to selected portions of the item’s textual content (see Chapter 5).

Named links are a third type of link; they pertain to particular named data fields such as **Periods** and **Authors** and supply their content. The fourth type consists of hotspot links, which pertain to a particular location within an image.

The **Query Results** tab of the reference pane contains a list of the database items that match the search criteria of the most recently performed query. At the top of the tab are the date and time when the query was performed and the number of items found by the query. Any item in the list of query results can be viewed by clicking on its name. This will pop up a floating window that contains all the

information entered for that item. See Chapter 14 for detailed instructions on performing a query and working with query results.

There is a  “table view” button on the **Query Results** tab that enables you to produce a tabular listing of the items in the query results list. This can be useful when you have retrieved a group of similar items and want to display the data about them compactly in the form of table rows and columns, where each item is shown as a row and the item properties are shown as columns. Other buttons allow you to display all of the images, events, and map shapes associated with the items in the query results list.

In some cases, it is desirable to convert the reference pane into a floating window that you can move around on the screen. This creates more space for the item pane to expand horizontally. If you have two monitors, you might want to move the reference pane to the second monitor, freeing up space on your main monitor. To “undock” the reference pane and turn it into a floating window, click the  toggle button on the main toolbar. To re-dock it, simply click the  button again.

Chapter 4

Creating and Organizing Items

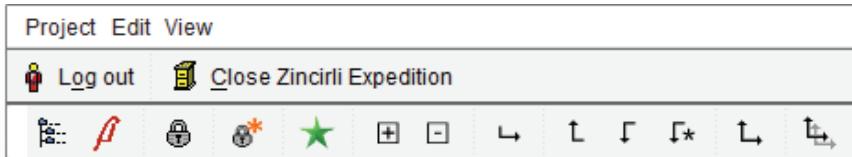
IN AN OCHRE database, a research project's information is represented in a highly atomized fashion as many individual items, each of which has its own name, description, properties, and links to other items. The various categories of items are described above in Chapter 2. This chapter describes the procedures for creating items and for organizing them by means of two special kinds of items: hierarchies and sets. The procedures are the same for every category.

Creating Items

Inserting New Items

Items are inserted in data-entry mode via the navigation pane on the left side of the OCHRE window. After you have logged in and opened a project for which you have data-entry privileges (see Chapter 1), you will see a list of item categories in the navigation pane. To insert an item or hierarchy at the top level within a category, click the category name to select it and then click the ↲ “insert below” button in the main toolbar. Not all categories allow an individual item to be inserted at the top level; some categories allow only hierarchies at this level. It is usually a good practice to avoid inserting individual items at the top level.

To insert an item within an existing hierarchy in a category, click the ► icon to the left of the category name to show its hierarchies, and then click on the desired hierarchy to select it before clicking the ↲ button to insert the item at the top level of that hierarchy. If you want to insert the item at a deeper level, first click the ► icon to the left of the hierarchy name to expand its contents and then select the item that will be the parent of the new item before clicking the ↲ button to insert the new item. You may be prompted to lock the hierarchy for your exclusive use before inserting new items into it. Click the  button in the main toolbar to lock the hierarchy. When you have finished inserting items, click  to unlock it.



If you want to insert a new item as the sibling of another item in the hierarchy (i.e., in the same branch), click the item's name to select it and then click the \leftarrow “insert before” or \downarrow “insert after” button on the toolbar to insert the new item before or after the currently selected item. In contrast, the \hookleftarrow “insert below” button inserts the new item, not as a sibling, but as a child of the currently selected item or hierarchy, nesting it more deeply in the hierarchy.

After you have clicked \leftarrow or \downarrow or \hookleftarrow , a new unlabeled item will appear in the navigation pane. You may then be given a choice in the item pane to make the new entry a hierarchy or an individual item of a particular type. In some contexts, no choice will be available and a hierarchy or an individual item labeled “New...” will be created immediately. The item pane will then show information about the hierarchy or individual item that you have created. You can enter a **Name** and **Abbreviation** in the item pane; the name will then be shown in the navigation pane. You can also enter a free-form **Description**. There will be data-entry tabs for the item’s **Properties**, **Links**, and **Notes**, and perhaps other specialized data-entry tabs, depending on the category.



In the case of a hierarchy, a tab will appear at the top of the item pane for entering the **Creators** of the hierarchy and the date when it was created. The button to the left of the date field lets you insert the current date. The button to the right of this field displays a calendar from which you can select a date. Below the date will be a **Count** of how many items are currently contained in the hierarchy.



The item pane of a hierarchy (as opposed to an individual item) will also have a **Preferences** data-entry tab that lets you choose to **Keep this hierarchy private** (i.e., make it invisible to anonymous users) and also lets you choose the sorting method for items within the hierarchy and determine what to include in the labels of items in the hierarchy (e.g., abbreviation, aliases, description, etc.).

The  “fully expand” and  “fully collapse” buttons in the main toolbar are used to expand or collapse an entire hierarchy, or a branch within a hierarchy, depending on what is currently selected. Be careful when fully expanding a hierarchy, in case it is very large.

Moving Items

At a given level of the hierarchy, you can move the selected item up or down within a list of sibling items by means of the  “move before” and  “move after” toolbar buttons. You can move an item one level higher in the hierarchy, making it a sibling of its parent, by means of the  “move up one level” button.

Items and entire hierarchical branches can also be moved by “dragging and dropping” with your mouse. This is done by clicking to select what you want to move and holding the mouse button down while moving the cursor to the desired location. An entire hierarchy can be grafted into another hierarchy by dragging and dropping. Conversely, a branch of a hierarchy (with or without headings) can be dragged onto the category name in the navigation pane and thereby converted into a separate hierarchy. In this way, complex hierarchies may be easily constructed and modified in order to represent relationships of logical subordination or physical containment among items.

Large hierarchies consisting of thousands of items may be displayed and updated slowly. If this becomes a problem, it is advisable to drag and drop hierarchical branches to the top level of the category in order to create a larger number of smaller hierarchies.

Replicating Items

Various buttons on the main toolbar are used to copy items. In most cases, copying does not actually create a new item but simply

displays the same item in a different hierarchical location. This is called “repeating” the item (discussed in the next section), as opposed to “replicating,” which does create a new item.

Replicating an item entails the insertion of one or more new database items that are duplicates of an existing item. This is done by selecting an item and then clicking on the  “replicate” toolbar button. (Note that you must lock the hierarchy that contains the item before replicating it.) A window will be displayed prompting you to enter the number of new items to insert. After entering this number and clicking the **OK** button, the replicated item(s) will be created as siblings of the currently selected item within the same hierarchical branch. They will be exact duplicates of that item.

Replicating an existing item is useful when you need to enter a number of similar items in the same hierarchical location, which is a common occurrence. For example, if several similar coins had been found in an archaeological layer, you could insert a database item for one of the coins as a child of the item that represents the layer and then automatically replicate it to create items for the other coins within the same layer. The properties you had entered for the first coin would be replicated for the others, saving data-entry time if only slight changes to these properties are needed to produce a correct description of each item.

Repeating Items and Hierarchical Branches

In contrast to replication, which creates new database items, you can copy an item by repeating the same item in a different hierarchical location. A given item may thus appear in more than one hierarchy or in more than one location within the same hierarchy. This is a useful feature because it allows the creation of overlapping hierarchies that present alternative views of the same information by organizing the same items in different configurations. Furthermore, you can borrow items from other OCHRE projects by repeating their items within your own project’s categories.

To repeat an existing item in a new location, you would use the **Linked Items** tab of the reference pane in conjunction with the 

“repeat link target” toolbar button. The procedure for selecting a link target is described in Chapter 6. After the link target has been selected in the reference pane, you would select a hierarchy or individual item in the navigation pane and click the button. This will insert a copy of the link target as a child of the selected item or hierarchy. Keep in mind, however, that *this copy is not a new item but just another manifestation of an existing item that was originally inserted elsewhere.* Any changes to the names, description, properties, links, or notes of the item, in any place where it appears, will be shown automatically in every other place where that item has been repeated.

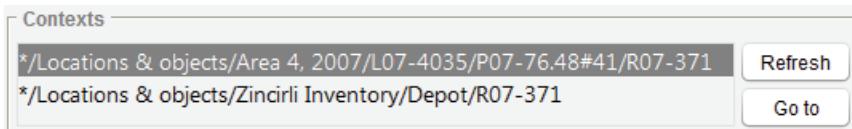
If the item being repeated has child items, then the entire branch of which it is a parent will be copied with it, and all of its descendants will also be repeated in the new location as children of their repeated parent. These child items may then be individually moved or deleted, like any other items, allowing you to modify the copied branch and customize its appearance. However, any modifications subsequently made to the structure of this branch in its original location will not be shown in the new location. The connection between the hierarchical structure in the original location and its structure in the new location will have been lost, even though each child item within the copied branch will still remain simply a repeated instance of the corresponding item in the original branch.

In some cases, you may want to repeat an item and the hierarchical branch descended from that item in such a way that a connection is maintained between the structure of the copied branch and the structure of the original branch, with the result that any changes in the original hierarchical structure due to the addition or removal of child items is automatically reflected in the new location. This is especially useful when you borrow a branch from another project’s taxonomy and wish to keep your version of the taxonomy synchronized with the original version (see Chapter 13).

In OCHRE this form of repetition is called a “bound branch.” To create a copy of an item together with a bound branch of all its hierarchical descendants, use the **Linked Items** tab of the reference pane to select a link target and then click the button “repeat link

target with bound branch” on the main toolbar. This will insert a copy of the link target and its descendants as the descendants of the item or hierarchy currently selected in the navigation pane. But in contrast to an ordinary item repetition, the internal organization of the copied branch cannot be changed. Its child items cannot be moved, deleted, or inserted because *the branch structure is still connected to the original branch in such a way that any changes made in the original branch are automatically shown in the copied branch*. Copied branches are displayed in a dimmed fashion to show that their branch structure cannot be modified.

To make it easier to keep track of the various hierarchical locations in which a given item is repeated, a list of **Contexts** is displayed for the currently selected item at the top of the item pane. Each context is displayed as a slash-separated “path” that shows the item’s location within a particular hierarchy. The first part of each path indicates the project in which the item is located, recognizing that items can be copied from other projects (the current project is indicated not by name but simply by an asterisk). Clicking the **Go to** button to the right of the **Contexts** area will scroll to the hierarchical location of the selected path in the navigation pane. Clicking the **Refresh** button will refresh the list of context paths in case the list is empty or out of date.



Deleting Items

To delete a hierarchy or an individual item, click on its name to select it in the navigation pane and then click on the **x** “delete” button in the main toolbar. If the item to be deleted is a repetition of an item that exists elsewhere, only the selected copy will be removed; other instances of the same item will not be deleted. If there are no other instances of the item, you will be prompted to confirm its permanent deletion.

Deleting items should be done carefully. Deletion will be prevented if links exist from other items to the item to be deleted. If an item has child items in a containment hierarchy, you will be prompted to delete them or move them to another parent before deleting the parent item. The project administrator determines which users have the power to delete items in a given category.

Item Names and Aliases

For both hierarchy items and individual items, the **Name** data-entry area in the item pane has a tab labeled with the project's default language (e.g., **English**, **Français**, **Deutsch**, **Italiano**, **Español**, etc.) in which the item's primary name is entered. For many items, this will be the only name. But some items may have alternative names in languages other than the default language, or they may have alternative names in the project's default language (these are called "aliases" in OCHRE).

For example, in a project whose default language is English, an item in the **Locations & objects** category might have the primary name "House" but it might be desirable also to enter the alias "Dwelling" and the French translation "Maison." A project's default language is set by the project administrator in the **Data-entry** subtab in the **Preferences** tab of the project's item pane, as described in Chapter 17. There is no requirement that the default language be English—or any other European language. OCHRE supports Arabic and Hebrew, which are written right-to-left, as well as East Asian languages and scripts such as Chinese, Japanese, and Korean.

To enter an alternative name, click the  "translation" button on the right-hand side of the **Name** data-entry area and choose either **Alias** or the desired language. The choice of languages available is specified in the project's preferences by the project



administrator, who must also click the checkbox to **Enable multilingual features** for the project. A new tab will appear beside the English tab (or the tab labeled Français, Deutsch, Italiano, Español, etc., depending on the project's default language) in which you can enter the alternative name. The new tab will either be labeled with **Alias**, if you have chosen this option in order to enter an alternative name in the project's default language, or it will be labeled with the name of the selected language.

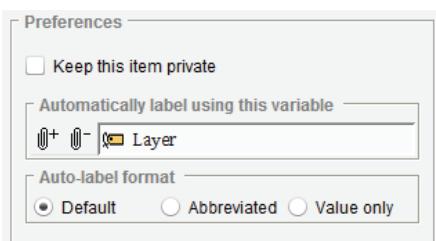
An item can have multiple aliases in the project's default language and it can have alternative names in multiple non-default languages. These names, like the primary name, can be used to find the item in a query. The item's aliases can be used in its label in the navigation pane in addition to (or instead of) its primary name, if this preference is chosen for the hierarchy in which it appears.

Moreover, the alternative-name mechanism can be used to develop a bilingual or even multilingual project database. If all, or even some, of the item names have been entered in an alternate language, the user can select this language to override the project's default language wherever the item name is displayed. This is done via the **Preferences** tab of the item pane for the particular item in the **Persons & organizations** category that represents the user. (Every user who has data-entry privileges for a project must be represented as an item in the project's **Persons & organizations** category.)

Automatic Naming of Items Based on Their Properties

In some cases, it is convenient to name items automatically based on their properties rather than to enter their names directly. For example, if a group of items in the **Locations & objects** category represents archaeological layers identified as "Layer 52," "Layer 28," "Layer 91," etc., each item's name might best be stored as a property, that is, a variable-value pair in which the variable is named "Layer"





and the value is the layer number. In that case, you would enter a link to the “Layer” variable in the **Automatically label using this variable** field in the **Preferences** tab of the relevant items. The following automatic labeling formats are available: **Default**, in which the item name consists of the variable’s primary name (in the project’s default language) followed by its value; **Abbreviated**, in which the item name consists of the variable’s abbreviation followed by its value; and **Value only**, in which the item name consists of only the value.

Using an “Inbox” to Hold New Items

There are six item categories in which a new item may be created automatically in the course of data entry for an item in a different category. In such a case, you would not be required to select the new item’s parent or sibling in the navigation pane or be required to click a toolbar button to perform the insertion. Instead, the new item would be put automatically into a special hierarchy called “Inbox.”

The six categories that have “Inbox” hierarchies are **Persons & organizations**, **Property variables**, **Property values**, **Resources**, **Sets**, and **Texts**. New person items may be created automatically while data is being entered in the **Bibliographies** category; new variable and value items may be created automatically while data is being entered in the **Taxonomy** category; new resource items while data is being entered in the **Locations & objects** category; new set items when query results are saved via the **Queries** category; and new text items when texts are linked to objects as “associated items” (see Chapter 11).

For example, when a photograph is to be associated with an item in the **Locations & objects** category, you can click the $\mathbb{0}^+$ “add link to new item” button to create a resource item for the image in the “Inbox” hierarchy of the **Resources** category and automatically add a link from the current spatial item to the new resource item (see Chapter 6).

It is a good practice to clean out the “Inbox” hierarchies regularly and move their items to other hierarchies within the same category in order to organize these automatically inserted items in a more meaningful fashion. You also have the option of hiding the “Inbox” hierarchy by de-selecting the **Show inbox even if empty** option in the **Preferences** data-entry area of the item pane for the parent category.

Hierarchies of Items

Containment Hierarchies and Subordination Hierarchies

In four of the OCHRE item categories, database items may be nested directly within other items to form a hierarchy that represents the containment of the child item within the parent item. This is the case in the **Locations & objects** category, in which hierarchies represent relationships of spatial containment among physical places and artifacts. In that category, a child item is understood to be physically contained within its parent. Likewise, in the **Periods** category, child items represent temporal subdivisions of their parent items; in the **Concepts** category, child items represent subconcepts that are semantically contained within their parent items; and in the **Bibliographies** category child items represent the containment of chapters within a book, of books within a series, of articles within a journal, etc.

In all the other categories, however, an item in a hierarchy is not understood to be directly contained within another item. Instead, hierarchies are built up using subordinate headings. They represent relationships of logical subordination rather than strict containment. In such “subordination hierarchies” an item cannot be directly inserted as the child of another item. Instead, a subordinate heading must be inserted as a child of the parent item, at which point a subordinate database item can be inserted as a child of the heading. The headings and subheadings within the hierarchy are not themselves items that represent external entities but are simply a means of grouping items in the desired fashion.

When you select a subordination hierarchy in the navigation pane and click the ↲ “insert below” button in the main toolbar, you will

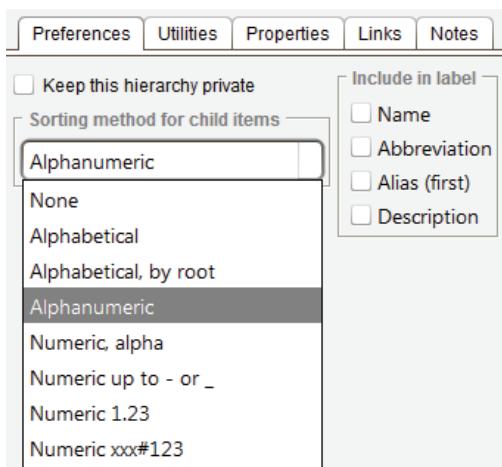
be given a choice to create a heading or an individual database item of the relevant category. A heading may in turn contain database items or additional headings, permitting the construction of a logical hierarchy. In contrast, when you select a containment hierarchy and click the ↳ button, there is no option to create a heading; a child item is created immediately.

Spatial and temporal containment hierarchies in the **Locations & objects** and **Periods** categories, respectively, are used in queries to determine the spatial and temporal relationships among items. A query can use such hierarchies to search for items based on specific spatial and temporal relationships or, more generally, as a way to define the spatial and temporal scope within which to constrain the search (see Chapter 14).

Keep in mind that any hierarchy, whether it is a containment hierarchy or a subordination hierarchy, is represented as a database item in its own right. Like any item it can have a name, aliases, description, properties, links, and notes.

Display Preferences for Hierarchies and Categories

The project administrator can set the display preferences for a hierarchy or category by clicking on its name in the navigation pane in order to select it and then changing the display options in the **Preferences** tab of the item pane. A hierarchy can be made invisible to anonymous users by selecting **Keep this hierarchy private**. Likewise, the option to **Keep this category private** can be selected for an entire category. The content of the item labels within a hierarchy is determined by the checkboxes in the **Include in label** area. An item's label can consist of one or more of the



item's primary name, abbreviation, aliases, or description, in any combination. The primary name is used by default.

You can also specify the sorting method used to order the items within each branch of a hierarchy, or to order the top-level items or hierarchies within a category. Items may be sorted in any of the following ways:

- **Alphabetical**—using strict alphabetical sorting that ignores the numeric value of sequences of numerals (e.g., “11” is before “2”);
- **Alphanumeric**—using the numeric value of sequences of numerals (e.g., “2” is before “11”);
- **Numeric (alpha)**—using the numeric value of sequences of numerals but only up to the first non-numeric letter, after which sorting is strictly alphabetical (e.g., “2A11” is before “11A2” but “11A2” is after “11A11”);
- **Numeric (up to - or _)**—using the numeric value of sequences of numerals but only up to the first hyphen or underscore character, after which sorting is strictly alphabetical (e.g., “2-11” is before “11-2” but “11-2” is after “11-11”);
- **Numeric (1.23)**—with sorting of numeric item names in which a period is used as a separator and not as a decimal point (e.g., “1.2” is before “1.11” which is before “1.20”);
- **Numeric (xxx#123)**—with a customized sorting method that has been devised for the project.

Other options for controlling how a hierarchy will be displayed are found on the hierarchy's **View** tab. A drop-down list at the top of this tab lets you determine the display mode. The **Information** and **Images** display modes show only data that pertains to the hierarchy itself and not any of the data that pertains to items contained in the hierarchy. In contrast, the various **Descendants...** display modes show all of the items contained in the hierarchy in different formats and levels of detail. The **Linked items** mode shows not just the items in the hierarchy but other database items that have been linked to them. See Chapter 15 for a detailed discussion of these display modes.

Sets of Items

Creating a Set of Items

Each item in the **Sets** category represents a named group of items that exist in one or more other categories. Among other uses, the items in a set can be viewed and edited in a table, displayed as a map layer, or used to limit the scope of a query. Like other categories of items, sets can themselves be organized into subordination hierarchies with headings. The items grouped within a set may include other set items; indeed, a set can include itself, though there are few, if any, situations in which this would be necessary.

Unlike a hierarchy, the items grouped within a set cannot have child items of their own. Note also that a set is not restricted to a single item category but may contain items of different kinds drawn from any of the categories. Thus, a hierarchy belongs to a particular category and is used to organize items within that category whereas a set may span categories, providing a different way to organize items that have been previously inserted into hierarchies.

The screenshot shows the 'Set Items' dialog box. At the top, there are three main input fields: 'Name' (containing 'English'), 'Abbreviation' (containing 'English'), and 'Type' (a dropdown menu). Below these are two text input fields: 'Description' (containing 'Pottery pails excavated to date in the 2010 season.') and 'Associated query' (containing 'Pottery buckets, 2010'). A status bar at the bottom indicates '1923 items'. A scrollable list below shows five items, each preceded by a red 'X' icon:

- P10-17.44#101
- P10-17.44#102
- P10-17.44#103
- P10-17.44#104
- P10-17.44#105

A set can be populated automatically by saving a list of query results or it can be populated manually by linking items to it one by one. To manually link an item to a set, click on the set's name in the navigation pane in order to display the set in the item pane; then specify the item to be added to the set by selecting that item as the link target in the **Linked Items** tab of the reference pane (see Chapter 6); and then click the \oplus “add link” button in the **Set Items** data-entry tab of the item pane.

To remove an individual item from a set, select it from the list in the **Set Items** tab and click the \ominus “remove link” button. You can remove all of the items in a set at once, making it empty, by clicking the **Remove set items** button in the **Set Items** tab. This operation is not reversible and should be done with care.

You can change the order of items in the list by means of the \uparrow “move up” and \downarrow “move down” buttons. The total number of items currently in the list is shown to the right of these buttons.

Note that when you first display a set in the item pane, the list of items in the **Set Items** tab may be empty even though the set contains items. The fact that the set is not truly empty is indicated by a green checkmark beside the **Set Items** label in the tab heading. In order to display the list of items contained in the set, you must click on the **Load set items** button. At that point, the list will appear and you can add and remove individual items or move them up and down. Requiring you to load the set's items explicitly in this way spares you a long wait when the number of items in the set is very large and you want only to inspect the set's name, description, properties, and other information without actually loading all of its items.

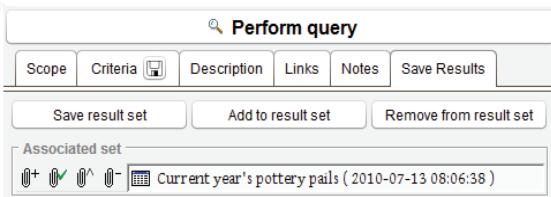
You can also add items to a set by clicking the **Add by code...** button or the **Add by name...** button (these are not available for sets created by saving query results). When you add by code a window will appear in which you can enter a list of codes of the objects to be added based on the contents of the **Code** fields of the relevant items in the **Locations & objects** category. When you add by name you can enter a list of names of items to be added based on the contents of their **Name** fields. In either case, you can type the list directly or paste

it in from another source (e.g., a word-processing document). When you click the **OK** button the database items whose codes or names match those in the list will all be added to the set.

Query Result Sets

A set can be created automatically via the **Save Results** data-entry tab in the item pane for a query item (see Chapter 14). After you have performed a query, open this tab and click the **Save result set** button. This will create a new item in the “Inbox” hierarchy of the **Sets** category. The name of the new set item will consist of the query’s name followed by the date and time when the query was performed. The items in the set will consist of the query results. You can then remove items from this set or add items to it, as usual, but keep in mind that this will cause the set to differ from the query results.

If you perform the same query again and open its **Save Results** tab, you will see that the name of the previously created set of query results is shown in the **Set of query results** area. If you want to create an entirely new result set for the query, you should use the **Ø-** button to remove the current set (which will still exist in the **Sets** category) and then click the **Save result set** button to save a new set whose name reflects the date and time of the most recent performance of the query. In other words, if the **Set of query results** area is empty, clicking the **Save result set** button will create a new result set whose name incorporates the most recent querying date and time.



However, you may wish to keep the existing set and simply update it with the most recent query results, which may have changed if the query’s criteria were modified or the data being queried was changed. In that case, do not remove the set shown in the **Set of query results** area but simply click the **Save result set** button to update this set and replace its contents. Alternatively, you can click the **Add to result set** button to add any new items in the current

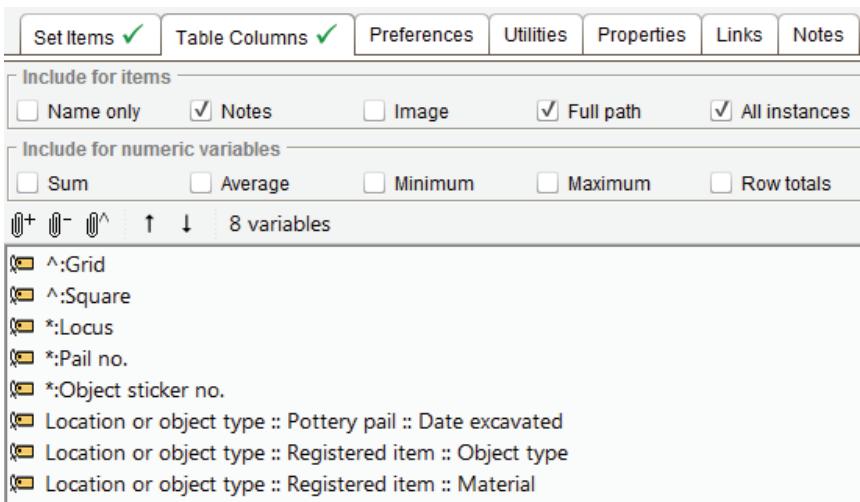
query results to the existing set, or you can click the **Remove from result set** button to remove from the existing set any items found in the current query results—this can be used to “subtract” from an existing set the results of a different query.

Note that the name of the set shown in the **Set of query results** area will not be changed automatically when you update its contents. This might be misleading if the name contains the date and time of a previous performance of the query and thus does not reflect the most recent query performance. In that case, you may wish to edit the set’s name, which you can easily do on the spot by double-clicking on the name in the **Set of query results** area in order to display the set’s item pane in a floating window. But note that the corresponding item label in the **Sets** category of the navigation pane will not be updated to reflect the change until you collapse and re-expand that category or until you click the “refresh” button in the main toolbar (usually found at the top of the OCHRE window).

Viewing a Set of Items

The **View** tab of a set’s item pane allows you to display the contents of the set in various ways. You can select the display mode by clicking on the “list view,” “table view,” “form view,” “image view,” or “map view” button at the top of the **View** tab. The “list view” will be displayed by default when you first open the tab. It consists of a simple list of the items in the set. Clicking on an item in the list will display its information within the **View** tab (i.e., the contents of its item pane). The number of items is shown in the **Count** field at the top of the **View** tab. To the right of that field, you can use the **Sort by** drop-down field to determine how the listed items will be sorted (alphabetical, alphanumeric, etc.). Farther to the right, the **Include in label** checkboxes allow you to determine what is included in the item labels (primary name, abbreviation, aliases, description, etc., or any combination of these).





The table view displays the items in the set in a tabular form in which each item is shown as a row in the table and the items' properties are shown as columns. The table view can be customized via the **Table Columns** data-entry tab for the set item. This tab contains links to variables in the **Property variables** category. The variables entered here determine which properties are shown as table columns in the table view of the set. The order of the variables in the tab determines the left-to-right order of the table columns. If no variables are entered in the **Table Columns** tab, the “table view” button in the **View** tab will display a table whose columns correspond to all of the variables currently used by the items in the set. Column variables can be added or removed individually by selecting them as link targets in the **Linked Items** tab of the reference pane and then clicking the “add link” or “remove link” buttons (see Chapter 6).

In addition to the list of variables that define the table columns, other options are available in the **Table Columns** tab to customize the table view. In the **Include for items** area you can specify that any or all of the following be included for each item as additional columns in the table: **notes** (i.e., the item's free-form notes as opposed to its properties), **image** (i.e., the first image resource that is linked to the item), **full path** (i.e., the full hierarchical context of the item), and **all instances** (i.e., all observations of locations and objects and all

interpretations of concepts). In the **Include for numeric variables** area you can specify that any or all of the following summary statistics be included as additional rows with values for each column that represents a numeric property: **sum**, **average**, **minimum**, **maximum**, and **row totals**.

The “form view” of a set displays the items in the set one by one, with a navigation bar under the buttons at the top of the **View** tab that lets you browse sequentially through the items. Clicking the ► button advances you to the next item; the ◀ button takes you back to the previous item; and the |◀ and ►| buttons take you to the first and last items, respectively. You can also go directly to a chosen item via the drop-down list to the right of the navigation buttons or by typing in a sequence number.

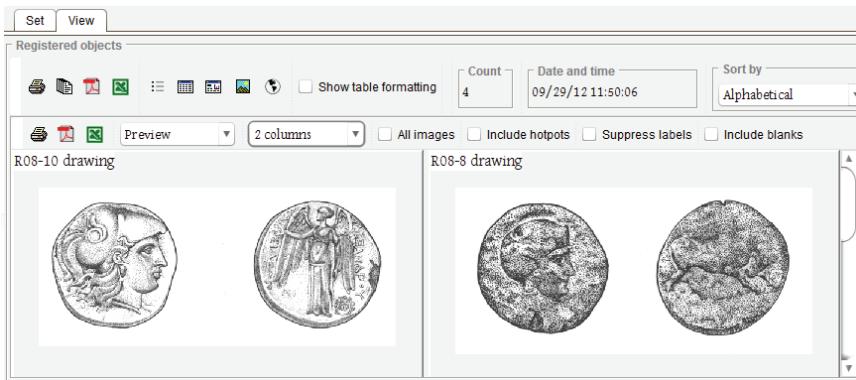
The screenshot shows the 'Registered objects' form view for object R08-10. The top header bar includes icons for file operations (New, Open, Save, Print, etc.), a date/time selector, and a 'Show table formatting' checkbox. To the right are 'Count' (4) and 'Date and time' (10/07/12 09:37:01).

The main content area has tabs for 'Context(s)', 'Code', 'Name', 'Description', 'Observations', 'Observers', and 'Properties'. The 'Name' tab shows 'R08-10' in English. The 'Description' tab shows 'Alexander the Great gold stater.' in English, Chinese, Italiano, and Türkçe. The 'Observations' tab shows the date '07/20/08'. The 'Observers' tab lists 'Pace, Leann' with a timestamp of '07/20/08'. The 'Properties' tab contains a table with the following data:

Variable	Value	Comment
Registered object no.	10	
Location or object type	Registered item	
Material	Metal	
Material	Gold	
Object type	Coin	
Object type	Stater	

At the bottom are 'Save' and 'Cancel' buttons.

The “image view” of a set displays all of the image resources that are linked to items in the set. The display options for viewing images are discussed below in Chapters 9 and 15.



The “map view” of a set displays the geospatial data associated with items in the set. If none of them has such data, clicking on the “map view” button will produce a message informing you that the view cannot be generated. Otherwise, a map will be displayed in the View tab with a special pane to the left of the map that lists each item as a separate map layer.

Any database item can be linked to any number of geospatial resource items. A geospatial resource will contain either a group of vector map shapes or a raster image (see Chapter 9). The vector shapes that are linked to an item collectively constitute a map layer that can be uniquely colored or shaded and independently hidden or shown. Each raster image constitutes its own map layer.

The View tab also allows you to print the information about a set of items or export it to an external file, as either a Microsoft Excel spreadsheet or a PDF document. You can select the export format by clicking on the “export as Excel” button or the “export as PDF” button. The contents of the printed report or exported file will depend on the current view mode. For example, clicking the button while in “table view” will create an Excel spreadsheet with rows and columns that duplicate the current view, whereas clicking this button while in “list view” will create an Excel spreadsheet that

contains only one column with the item names. Two types of reports can be printed. Clicking the  “print summary” button will print a more condensed version of the current view, whereas clicking the  “print details” button will print a full report that contains all available information. Printing is disabled for the map view.

Chapter 5

Describing Items with Properties and Notes

EVERY hierarchy and individual item in an OCHRE database has a data-entry tab entitled **Properties**. The procedure for entering properties (variable-value pairs) in this tab is the same in every case. Before properties can be entered, the project administrator must create a taxonomic hierarchy of variables and values in the project's **Taxonomy** category (see Chapter 13).

Entering Item Properties

Types of Variables and Values

A variable will belong to one of nine different types. These types are shown in the item pane when a variable item is selected in either the **Property variables** or the **Taxonomy** category (see Chapter 13). The type determines the kind of value with which the variable can be paired to form a property. The nine variable types are:

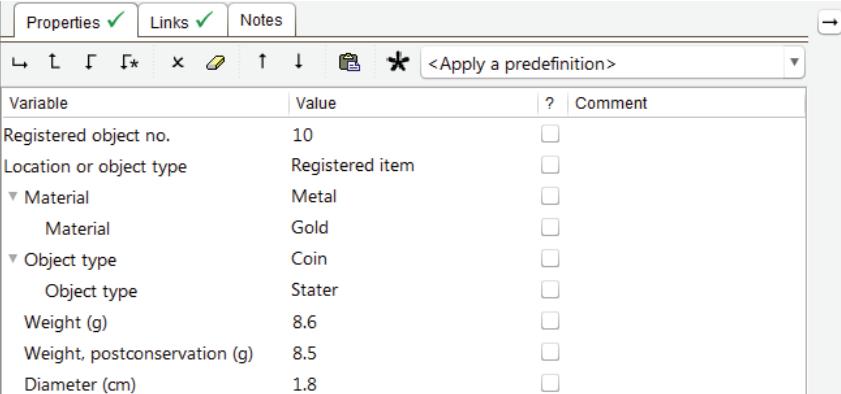
- **nominal**, which means that the variable must be paired with a character-string value that is represented as a database item in the **Property values** category, and the variable's values consist of names or classifications that have no inherent order;
- **ordinal**, which means that the variable must be paired with a character-string value that is represented as a database item in the **Property values** category, and the variable's values consist of names that do have an inherent order;
- **true/false**, which means that the variable must be paired with a Boolean value representing “true” or “false”; this choice between two and only two values is represented in the interface by means of a checkbox, that is, a check entered in the checkbox represents the value “true”;
- **integer**, which means that the variable must be paired with a numeric integer value;
- **decimal**, which means that the variable must be paired with a numeric decimal value;

- **date**, which means that the variable must be paired with a value that represents a calendar date;
- **string**, which means that the variable must be paired with a character-string value that may contain letters or numbers, or both, but does not exist as a database item in the **Property values** category;
- **serial no.**, which means that the variable must be paired with an integer value that is automatically incremented and is guaranteed to be unique for that variable;
- **relational**, which means that the variable must be paired with a value that is a link to another database item in the current project or in a different project (see Chapter 6).

Assigning a Property to an Item as a Variable-Value Pair

The toolbar on the **Properties** tab contains buttons to insert, delete, and reorder properties. To insert a new property, click on the ↴ “insert after” button. This will insert a blank property immediately after the property that is currently selected. Double-clicking in the **Variable** column will provide a drop-down pick-list of the variables that are allowed at that level of the taxonomy. If a **nominal** or **ordinal** variable is entered, then double-clicking in the **Value** column will provide a drop-down pick-list of the available character-string values for the variable, as specified in the taxonomy. Otherwise, double-clicking in the **Value** column does not provide a pick-list but allows the value to be typed directly into the field. Values must be directly entered for **integer**, **decimal**, **string**, **date**, **true/false**, **serial no.**, and **relational** variables.

Any number of properties can be inserted in this way. To delete a property, select it by clicking on it and then click the ✘ “delete” button in the **Properties** toolbar. To delete all of the properties at once, click the ⌂ “clear all” button. To move a property up or down in the list of properties, click the ↑ “move up” or ↓ “move down” button. Note that all changes must be saved by clicking the **Save** button at the bottom of the tab. The existence of properties is indicated by a green checkmark in the **Properties** tab heading.



The screenshot shows the 'Properties' tab of a software application. At the top, there are three tabs: 'Properties' (highlighted), 'Links', and 'Notes'. Below the tabs is a toolbar with icons for back, forward, search, and other functions. A dropdown menu says '<Apply a predefinition>'. The main area is a table with columns 'Variable', 'Value', '?', and 'Comment'. The data entries are:

Variable	Value	?	Comment
Registered object no.	10	<input type="checkbox"/>	
Location or object type	Registered item	<input type="checkbox"/>	
Material	Metal	<input type="checkbox"/>	
Material	Gold	<input type="checkbox"/>	
Object type	Coin	<input type="checkbox"/>	
Object type	Stater	<input type="checkbox"/>	
Weight (g)	8.6	<input type="checkbox"/>	
Weight, postconservation (g)	8.5	<input type="checkbox"/>	
Diameter (cm)	1.8	<input type="checkbox"/>	

Date, Serial Number, and True/False Properties

The value of a **date** property is entered by typing it in or by clicking the “calendar” button that will appear to the right of the value after a variable of this type has been inserted and you have clicked on the empty value field. This will display a calendar in which you can select the year, month, and day.

A variable of the **serial no.** type is used in properties whose values are integers that must be guaranteed to be unique within the project or hierarchy (i.e., unique for the particular variable used in the property). When a **serial no.** variable is inserted in an item’s **Properties** tab, its integer value will be set automatically to one more than the most recent value used with that variable. You can type in a different integer value, but if it has already been used with the variable, it will not be accepted. If the integer you enter is a higher number than has previously been used with that variable, it will be accepted and the serial number counter will be reset to the new value.

The value of a **true/false** property is entered by clicking on the checkbox that will appear in the value field after a variable of this type has been inserted and you have clicked on the empty value field. A checkmark indicates a value of “true.”

Indicating Uncertainty about a Property

In some cases, there is uncertainty about a property. For example, you may believe that an iron artifact is an arrowhead but you are not

sure. To indicate that a property is uncertain, click the checkbox to the right of the property's value in the column headed by a question mark. This allows you to enter the property, capturing the information it represents, while reserving the option of excluding it from later queries or displays in which you want to ignore uncertain properties.

Entering a Comment about a Property

If you wish to record a free-form comment about a property, you may do so by clicking on the **Comment** field to the right side of property's value. This will cause a floating window to pop up, in which you can enter the comment using various text-formatting tools (see the instructions on entering textual content below). Comments that are specific to individual properties are often a useful alternative to free-form notes entered via the **Notes** tab because such notes apply to the entire item, not just to one property.

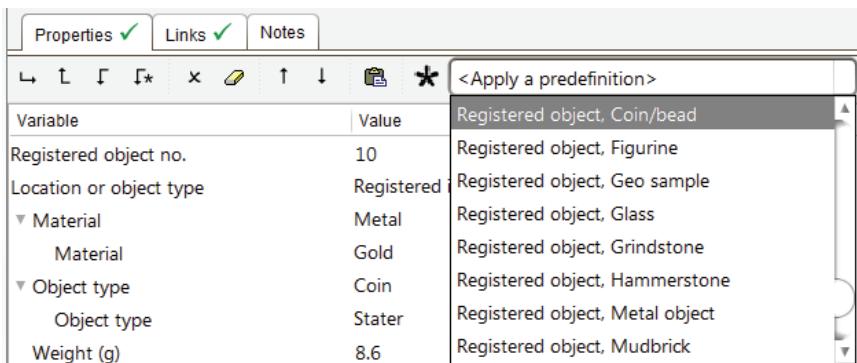
Predefined Properties

It will often be necessary to describe many different items using the same group of variables and/or variable-value pairs. For this reason, predefined properties can be stored and named as a group in the **Property predefinitions** category. To insert a predefined group of properties, select the appropriate item from the **Predefinitions** drop-down list at the top of the **Properties** tab. Any properties currently in the tab will be replaced by the predefined properties, which can then be modified or deleted individually, if necessary. To avoid having to re-select the same predefinition repeatedly from the drop-down list, you can click the * “repeat predefinition” button on the toolbar to apply the most recently used predefinition.

A variable may be specified in a predefined property without specifying its value. This is done by using a special **<any value>** placeholder in the course of defining the properties of a predefinition item in the **Property predefinitions** category. When the predefined properties are subsequently assigned to an item, the **<any value>** placeholder will be converted to a blank value, making it easy at that point to enter the correct value in its place. Similarly, during data entry, a

special <unassigned> placeholder value can be used for variables whose value is not known at the time but will be filled in at a later date.

In some cases, all you want to do is copy the properties you have just saved for an item to the next item you are describing. To do this, click the  “copy last saved” toolbar button in the **Properties** tab in order to assign the most recently saved group of properties to the current item. Or, you may just want to duplicate a property that has already been entered. To do this, select the property to be duplicated and click the  “replicate” toolbar button in the **Properties** tab. An identical property will be inserted after the selected property.



The screenshot shows the OCHRE software interface with the 'Properties' tab selected. The main pane displays a table of properties:

Variable	Value
Registered object no.	10
Location or object type	Registered
Material	Metal
Material	Gold
Object type	Coin
Object type	Stater
Weight (g)	8.6

A dropdown menu titled '<Apply a predefinition>' is open on the right side, listing various object types:

- Registered object, Coin/bead
- Registered object, Figurine
- Registered object, Geo sample
- Registered object, Glass
- Registered object, Grindstone
- Registered object, Hammerstone
- Registered object, Metal object
- Registered object, Mudbrick

Nested Properties

OCHRE allows “nested properties.” A nested property is logically subordinate to another property of the same item, in keeping with the semantic relationships among variables and values prescribed in the project’s taxonomic hierarchy (see Chapter 13). This is a powerful mechanism for describing an item in a flexible manner by relating more specific properties to more general properties. For example, the property “Material = metal” could have the subordinate property “Material = gold” nested within it in the description of a particular item, as shown above.

To insert a nested property, click the  “insert below” toolbar button at the top of the item’s **Properties** tab instead of the  “insert after” button. This will insert a child property that is hierarchically

nested within the currently selected property instead of inserting a sibling property at the same level of the hierarchy. The variable and value of a nested property are entered in the usual way.

The ability to insert a nested property in a particular location depends on the project's taxonomy. Only those variable-value combinations that have a variable nested within them in the taxonomic hierarchy are permitted to have a property nested within them when entering data for an item. Moreover, the character-string value to be paired with a **nominal** or **ordinal** variable to form a nested property must be one of the values allowed in that particular context in the taxonomy; there may be different values allowed for the same variable in a different location within the taxonomic hierarchy, but these would be irrelevant.

Inherited Properties

An inherited property is a property that, having been assigned to a particular item, can be viewed among the properties of all other items contained within that item in a containment hierarchy. For example, a property that specifies the “Date of excavation” of an archaeological layer could be inherited by all the items that are hierarchically contained within the layer’s database item in the **Locations & objects** category. In this example, the descendant items would represent artifacts found in the layer. By definition, they would have the same excavation date as the layer. The use of inherited properties therefore avoids the tedious repetition of the same property within a hierarchy in the course of data entry.

To designate an inherited property, select its variable item in the **Property variables** category or in the **Taxonomy** category and click the **Inherited by contained items** checkbox in the **Preferences** tab of the item pane. Any property that uses this variable can now appear automatically among the properties of every item that is hierarchically contained within the item to which it was originally assigned. Inherited properties will be shown in a different color. They are hidden from view by default and can be made visible by checking on the “show inherited properties” option in the **View** menu.

R08-10

*/Locations & objects/Area 3, 2008/L08-3002/P08-55.48#7/R08-10

*/Locations & objects/Zincirli Inventory/Gaziantep Museum/R08-10

Description: Alexander the Great gold stater.

Observation date: 07/20/08 Pace, Leann

Grid: 55

Locus: L08-3002

Square: 48

Pail no.: 7

Location or object type: Pottery pail

Date excavated: 07/20/08

Registered object no.: 10

Location or object type: Registered item

Material: Metal

Material: Gold

Object type: Coin

Object type: Stater

Weight (g): 8.6

Weight, postconservation (g): 8.5

Diameter (cm): 1.8

Thickness (cm): 0.4

Elevation (m): 527.44

Inherited properties should not be confused with nested properties. A nested property is related to a parent property of the *same* item, in accordance with the project's taxonomy, whereas an inherited property duplicates a property of a *different* item that contains the inherited property's item as a descendant within a containment hierarchy of items.

Furthermore, an inherited property is not actually assigned to the descendant items. It can be viewed within a descendant item, but only as a reflection of a property assigned to an ancestor item. It will not be found by a query that searches the properties of the descendant items, although it will be found in the ancestor item to which it was assigned, where it functions as an ordinary property. The querying of inherited properties in descendant items is instead accomplished via the ability of a compound query to consider the extrinsic containment relationships of the items it seeks (see Chapter 14).

Multiple Observations and Interpretations

In the **Locations & objects** and **Concepts** item categories, but not in other categories, you can assign multiple groups of properties to a single item. Each group of properties represents a distinct observation of an item (in the case of a spatial location or object) or a distinct interpretation of an item (in the case of a concept).

OCHRE permits multiple observations of the same physical location or object because, in many cases—for example, an archaeological discovery—the object will have been observed by different observers at different times, sometimes with contradictory results, and these observations need to be kept distinct in the course of subsequent querying and analysis. Similarly, OCHRE permits multiple (and perhaps conflicting) interpretations of the same abstract concept, each authored by a different interpreter.

The item pane for an item in the **Locations & objects** category contains a data-entry tab entitled **Obs. 1** (“Observation 1”). At the top of this tab is an **Observers** field, in which you can enter links to one or more person items. Such links are made in the usual way by selecting the link target in the **Linked Items** tab of the reference pane and clicking the \oplus “add link” button to the left of the field (see Chapter 6). If there are multiple links, they can be reordered by means of the \uparrow “move up” and \downarrow “move down” buttons to the left of the field. A link can be removed by selecting it and clicking the \ominus “remove link” button. The observer is optional and need not be entered.

To the right of the **Observers** field is a date field into which a date can be typed directly or entered by clicking the calendar button to the right of the field. Alternatively, the button to the left of the date field lets you insert the current date as the date of the observation.

To save time in cases where you (as the current user) are the observer, you can simply click the  “current user” button to the left of the **Observers** field. This will automatically enter a link to the person item that represents you and will automatically enter the current date in the date field. No matter how a date is entered, it will replace “Obs. 1” as the tab heading for the observation; this allows

observations to be conveniently labeled and displayed in order by date. The date is optional, however, and need not be entered.

The screenshot shows the Observation tab interface. At the top, there are three main sections: 'Name' (containing 'English' and 'L08-5001'), 'Abbreviation' (containing 'English'), and 'Count' (containing '467' and 'Code'). Below these are two rows of dates. The first row contains: 07/26/09, 07/14/09, 07/14/09, 06/20/10, 06/20/10, 06/20/10, 07/04/10, and 07/20/10. The second row contains: 07/15/08, 07/15/08, 07/15/08, 07/15/08, 07/15/08, 07/14/09, 07/14/09, 07/14/09, and 07/17/09. Underneath these is the 'Observers' section, which lists 'Cannon, Joshua' with a date of '07/15/08'. Below the observers is a toolbar with buttons for Properties (green checkmark), Links (green checkmark), and Notes. A dropdown menu shows '<Apply a predefined>' and other options like 'Grid', 'Square', 'Locus', 'Location or object type', 'Locus date begun', and 'Locus date completed'. The 'Value' column for these properties shows: Grid (17), Square (55), Locus (L08-5001), Location or object type (Stratigraphic unit), Locus date begun (07/15/08), and Locus date completed (07/21/08). There are also checkboxes for 'Comment' next to each value.

Each observation tab contains subtabs for properties, links, and notes. These function in the normal manner. They constitute the data recorded by a particular observer (or observers) on a particular date with respect to the unit of observation represented by the item. A second observation can be created by clicking the “create new” button to the right of the date field. This will create a new tab entitled **Obs. 2** (“Observation 2”), into which you can enter a different observer and observation date, together with the properties, links, and notes that constitute this observation. You can create as many observation tabs as are needed.

An observation tab can be removed by clicking the “delete” button to the right of the observations data-entry area. The tabs can be reordered horizontally by clicking the “move left” and “move right” buttons. The tabs will be automatically renumbered as they are moved left and right. Multiple interpretations of concepts are entered in exactly the same way, except that the **Obs.** tabs are labeled **Interp.** and the **Observers** field is labeled **Interpreters**.

Entering Free-form Notes for an Item

Creating One or More Notes

In data-entry mode, every database item has a **Notes** tab in its item pane. This can be used to create one or more free-form notes that pertain to the current item. The notes are displayed in numbered subtabs arranged vertically along the left side of the **Notes** tab.

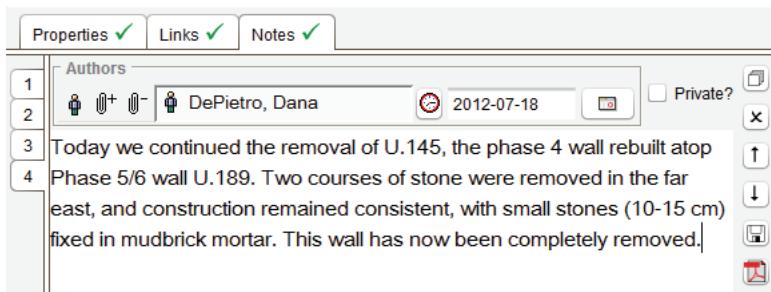
Clicking the  “create new” button on the right side of the **Notes** tab will create a new note. The  “delete” button deletes the selected note. The  “move up” and  “move down” buttons move the selected note up or down in the sequence of notes. Clicking the **Private?** checkbox will make the note invisible to anonymous users. If the project administrator has selected the **Enable multilingual features** option in the project’s preferences, the  “translation” button will be available on the right side of the **Notes** tab to enable you to insert a translation of the note in a different language (translations are discussed in more detail below).

The author and date of a note are entered in the same way as the observer and date of an observation. Both the author and the date are optional. At the top of the **Notes** tab is an **Authors** field, in which you can enter links to one or more person items. Such links are made in the usual way by selecting the link target in the **Linked Items** tab of the reference pane and clicking the  “add link” button to the left of the field (see Chapter 6). If there are multiple links, they can be reordered by means of the  “move up” and  “move down” buttons to the left of the field. A link can be removed by selecting it and clicking the  “remove link” button.

To the right of the **Authors** field is a date field into which a date can be typed directly or entered by clicking the calendar button to the right of the field. Alternatively, the button to the left of the date field lets you insert the current date as the date of the note.

To save time in cases where you (as the current user) are the author, you can simply click the  “current user” button to the left of the **Authors** field; this will automatically enter a link to the person item that represents you and will automatically enter the current date.

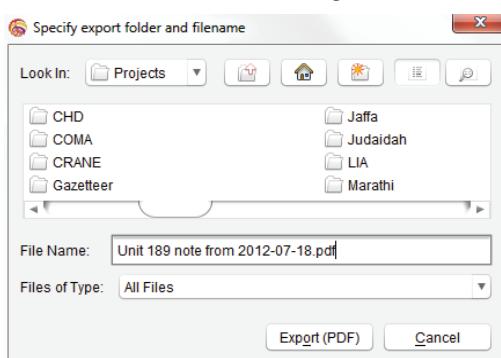
For items in the **Locations & objects** category or the **Concepts** category, you also have the option of specifying that the observer(s) or interpreter(s) are the author(s) of the note by clicking the larger  button on the far left of the tab.



Saving Notes and Exporting Them to PDF Files

Any textual content you enter will be saved automatically to the central database as soon as you leave the data-entry field, whether by clicking on another field, selecting a different item, closing the project, or logging out of OCHRE altogether. If the textual content is lengthy you might not leave the field for a long time, so to avoid losing what you have entered due to an unexpected power failure that shuts down your computer you should periodically save what you have typed by clicking the  “save” button on the right side of the **Notes** tab.

To export the contents of a note to an external file in the PDF format, click the  “export as PDF” button on the right side of the tab. This will display a window that allows you to specify the name and location of the exported file. If you hold down the shift key while clicking this button, the file will be named automatically using the item’s primary name and will be saved to your computer’s desktop.



Entering Textual Content in Notes and Elsewhere

A note is one of several places in which you can enter free-form textual content. Other places for textual content include item names, aliases, and descriptions; and, most notably, resource items that are “internal documents” (see Chapter 9). The procedure for entering and editing textual content is the same in all cases. (Textual content cannot be entered or edited in view-only mode.)

Using the Text-formatting Toolbar

In data-entry mode, a toolbar for formatting textual content can be displayed by clicking on the **F** “text-formatting tools” button in the main toolbar or by selecting the **Text-formatting tools** option in the drop-down **View** menu. On the left end of the text-formatting toolbar are buttons that change the style of the currently selected portion of text.

To select a portion of text, hold down the mouse button while dragging the mouse to highlight the desired characters and then click the appropriate button to format them. Click **T** for plain text (remove all formatting), **B** for bold, **I** for italicized, **U** for underlined, **T^s** for superscripted, or **T_s** for subscripted. To the right of the style buttons is the “centered text” toggle button. Clicking it will center a left-justified line of text or will left-justify a centered line, depending on the current alignment of the selected line.

By default, the font used for entering textual content is the project’s “document” font, which is specified by the project administrator in the **Fonts** subtab in the **Preferences** tab of the project’s item pane (see Chapter 17). However, there are buttons in the text-formatting toolbar to apply a special font to the currently selected portion of text. Special fonts include the project’s fonts for phonemic transcription (normally rendered using Latin characters with diacritics or using the International Phonetic Alphabet) and for graphemic transcription (a sign-by-sign rendering of the graphic signs that make up a text). These fonts, too, are specified by the project administrator in the **Fonts** subtab. (See the

section on “Graphemic, Phonemic, and Hybrid Transcriptions” below in Chapter 11.)

Click the  button in the text-formatting toolbar to apply the phonemic transcription font and click the  button to apply the graphemic transcription font. In addition, there is a built-in Egyptian hieroglyphic font that can be applied to the selected portion of text by clicking the  button. All fonts in OCHRE are Unicode fonts because all textual content is encoded using the Unicode character-encoding standard (see <http://www.unicode.org>).

A project's graphemic font will normally contain non-Latin characters, either from another alphabetic script (e.g., Arabic, Cyrillic, Devanagari, Greek, Hebrew) or from a nonalphabetic script (e.g., Chinese or Egyptian hieroglyphs). Its phonemic font will normally contain Latin characters with diacritics or IPA characters suitable for representing the types of texts dealt with by the project (note that the latest versions of the "Times New Roman" and "Arial" fonts contain almost all the diacritics one might need). A project's transcription fonts are used by default in the relevant data-entry fields of database items in the **Texts** category but they can also be applied to selected portions of notes and other textual content.

Using the Keypad of Special Characters

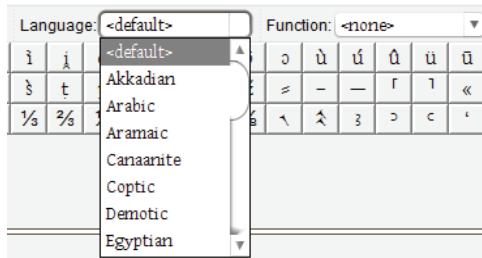
Click the  button in the text-formatting toolbar to display a “keypad” that facilitates entry of characters not found on the user’s keyboard. The contents of the keypad are determined by the **Keypad** drop-down list to the right of the  button. By default, the keypad includes a selection of Latin characters with diacritics and some special symbols. If the current project uses the **Writing systems** item category to represent a complex writing system (e.g., Egyptian hieroglyphs or Sumero-Akkadian cuneiform), then the user can use the **Keypad** list to choose a writing system from which to generate the keypad. Click the  button again in order to hide the keypad.



The keypad is useful as a means of entering characters that are not found in one of the standard keyboard layouts available on the user's computer. For many modern languages, however, the user can simply choose the appropriate keyboard layout on his or her computer and type the characters directly.

Specifying the Language or Function of a Selected Portion of Text

The text-formatting tool bar has two drop-down lists for entering **Language** and **Function**. It is often useful to search textual content based on the language or function of particular parts of the text. The **Language** drop-down list allows the user to indicate that the currently selected portion of text is written in a language other than the default language for the current project. The **Function** drop-down list allows the user to indicate the "function" (however that be defined) of the currently selected portion of text.



For example, a sign in a cuneiform text may function as a "determinative" that provides information about what follows or precedes it in the text but is not itself intended to be pronounced. The functions available in the drop-down list will vary from project to project. The project's default language and list of possible functions are specified by the project administrator in the **Data-entry** subtab in the **Preferences** tab of the project's item pane (see Chapter 17).

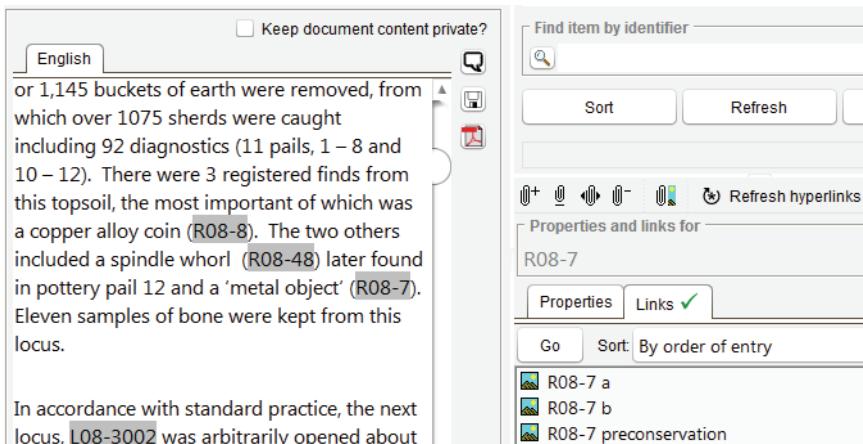
Embedding Properties and Hyperlinks within Textual Content

In addition to language and function, other properties can be associated with particular words or phrases. The item properties mechanism described earlier in this chapter may be used for this purpose. Project-defined properties that conform to the project's taxonomy can be assigned to the currently selected portion of text. To do this, click the **Show properties and links** button at the bottom of the **Linked Items** tab of the reference pane (if this area is not already

displayed). Then click the  “assign properties to selected text” button above the **Properties and links for...** field in order to fill this field with the selected text portion. At that point, you can enter properties via the **Properties** tab in the bottom of the reference pane, according to the procedure describe above. As usual, the properties must be explicitly saved to the database by clicking the **Save** button at the bottom of the tab. Properties inserted and saved in this way will be assigned to the portion of text that is currently selected in the item pane, in contrast to general properties that pertain to the item as a whole. The latter are inserted in the item pane, not in the reference pane.

If you select a different note (or, in general, a different textual content field), the **Properties** tab in the reference pane will be cleared. To view the properties again, you must move the mouse cursor over the relevant text portion; the cursor will turn into a pointing hand when it is positioned over any portion of text for which properties or hyperlinks have been saved. Click on the desired text portion to display a floating window that shows its properties and/or hyperlinks; this will simultaneously load them into the **Properties** and **Links** tabs of the reference pane, where they can be modified, if necessary.

The procedure for embedding hyperlinks in textual content is very similar. After a text portion has been selected in the item pane and the properties and hyperlinks area is displayed in the reference pane, select a link target (see Chapter 6) and click the  “add link” button above the **Properties and links for...** field. This will add a link from the currently selected portion of text to the currently selected link target. The hyperlink will appear in the **Links** tab of the reference pane. The  “remove link” button can be used to remove existing hyperlinks. If you want the name of the link target to be inserted automatically at the current cursor location within the textual content in the item pane, use the  “add link with name” button instead of the  button. Item names that are automatically inserted into textual content in this way will become out of date if the linked item’s name is modified, so a **Refresh hyperlinks** button is provided that lets you regenerate the names and update the textual content.



To insert a linked image into textual content at the current cursor location, select an image resource item as the link target and click the “add image link” button. The image will then appear in the textual content in the item pane, embedded at the chosen cursor location.

To relocate an existing hyperlink to a different portion of the textual content, select the hyperlink that is to be moved by clicking on it in the **Links** tab of the reference pane, select the text portion to which it is to be moved, and click the “relocate link” button. This will save you the trouble of first having to remove the hyperlink in order to embed it in a new location.

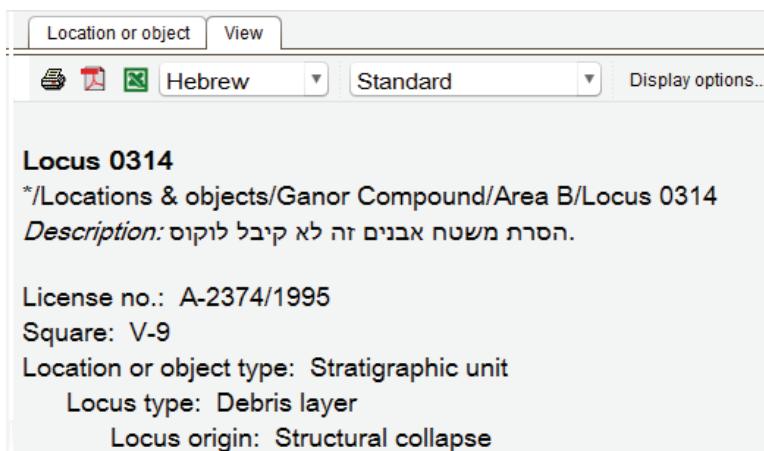
Finally, to view the hyperlinks for a portion of text after you have left the text-entry area and returned to it, you must move the mouse cursor over the relevant text portion, which will be shaded to indicate that it has a hyperlink. The cursor will turn into a pointing hand when it is positioned over any portion of text for which properties or hyperlinks have been saved. Click on the desired text portion to display a floating window that shows its properties and/or hyperlinks; this will simultaneously load them into the **Properties** and **Links** tabs of the reference pane, where they can be modified, if necessary.

Displaying Translations of Textual Content

Unless otherwise indicated, all textual content is assumed to be written in the project’s default language, which is specified by the

project administrator in the **Data-entry** subtab in the **Preferences** tab of the project's item pane. As was mentioned above, selected portions of textual content can be identified as having been written in a different language by means of the **Language** drop-down list on the text-formatting toolbar. This provides a way to identify foreign words and phrases that occur within a document.

Alternatively, you may wish to store a translation of an entire body of textual content, such as an item name or description, an entire note, or an entire “internal document” resource. To do this, click the  “translation” button adjacent to the name, description, note, or document. (Note that this button will be visible only if the checkbox to **Enable multilingual features** is checked in the **Data-entry** subtab in the **Preferences** tab of the project's item pane.) You will be prompted to select the language of the translation, after which a new tab will appear for the name, description, note, or document, into which you can enter the translation. The new tab will be labeled with the name of the language in which the translation is written.



In view-only mode, only one of the available languages will be shown in the formatted, printable view of an item's information that is displayed in its **View** tab in the item pane. This tab normally uses the project's default language; however, you can choose a different language to display via the drop-down list of languages at the top of

the **View** tab. Translations in the chosen language, wherever they exist, will then be substituted for textual content written in the default language. If no translation in the chosen language exists for a given body of textual content, the project's default language will be used instead.

In addition to textual content from item names, descriptions, notes, and internal documents, the **View** tab displays item properties. Property names (variables) and values can be translated into other languages, just like any textual content. Here, as elsewhere, OCHRE's translation features enable researchers to store information in more than one language and to switch easily from the project's default language to a different language.

Chapter 6

Relating Items to Other Items

THREE are two ways to relate items to other items: simple links and relational properties. A simple link indicates that two items are related but says little about the nature of the relationship. A relational property makes use of the item-properties mechanism (see Chapter 5) to establish a relationship between two items in such a way that the relationship itself is named with a project-defined name and so can be described and queried as an entity in its own right. Furthermore, a relational property explicitly indicates the direction of the relationship and allows you to give a different name to the inverse relationship.

For example, instead of using a simple link to relate a person item to another person item in an ambiguous fashion in which the nature of the relationship and its direction are unclear, you could use a relational variable named “Teacher of” to represent the fact that one person, represented by the item to which the property is being assigned, is the teacher of another person, who is represented by the value of the relational variable. The name of the inverse relationship would be “Student of.” After this property (a variable-value pair) has been assigned to the person item that represents the teacher, the nature and direction of the relationship between the two persons is clearly represented in the database. A query could then subsequently find all persons who are students of a particular teacher or all teachers of a particular student.

Simple Links

General Links, Named Links, Hyperlinks, and Hotspot Links

There are four kinds of simple links: general links, named links, hyperlinks, and hotspot links. A *general link* pertains to the item as a whole rather than to a particular component of the item’s information. It represents the fact that an item is related in some way to another database item. The relationship is bidirectional and has no

name or description. A query can find all of the items related to a given item by general links, but it cannot search more precisely based on the type of relationship.

In contrast, a *named link* pertains to a specific data field for the item such as the **Periods** field or the **Authors** field in a note. The link represents the fact that the data field is related to another database item, which provides the content of the field. Unlike general links, named links have directionality and are by definition named with the name of the field. Thus, they are much less ambiguous because the nature of the relationship and the category of the linked item are prescribed. For example, the **Authors** field of an item note can be linked only to an item in the **Persons & organizations** category and the simple link embedded in this data field represents the fact that the person referenced is the author of the note. A query can easily find all of the notes authored by a particular person.

Hyperlinks pertain to a selected portion of the item's textual content rather than to a specific data field (see Chapter 5). They can be embedded in any textual content that belongs to the item such as its names, aliases, description, and notes. An "internal document" resource item consists mainly of textual content (see Chapter 9).

A hyperlink is similar to a general link insofar as it is bidirectional and is not named. For example, the phrase "Battle of Marathon" in the description of an item can be linked to a person but there is no way to know the nature of the relationship. You might be able to guess, based on the wording of the phrase and the category of the item it is linked to, but there would still be some ambiguity.

Finally, a *hotspot link* pertains to a selected location within an image represented as a resource item (see Chapter 9). Hotspot links are inserted via the **Image Tools** tab of the image's item pane. Image hotspot links are like textual hyperlinks, in the sense that they are bidirectional and are not named, so there may be some ambiguity concerning the nature of the relationship between the image and the item to which it is linked.

In data-entry mode, general links are listed in the **Links** subtab of the item pane; named links are shown in various data-entry fields in

the item pane, such as the **Creator** field; hyperlinks are shown visually as shaded portions of textual content and are listed in the **Links** tab of the reference pane (described below); and hotspot links are shown visually as shaded regions within an image in the **Image Tools** tab in the item pane of the resource item that represents the image (see Chapter 9). In view-only mode and in the data-entry **View** tab, general links and hotspot links are listed on the right; named links are shown in the item's information as the contents of the data fields to which they are linked; and the presence of a hyperlink is indicated by a shaded region in the item's textual content, which can be clicked to display the linked items.

Adding and Removing Links

The **Linked Items** tab of the reference pane is used to select the “target” item for a simple link. The “link target” area at the top of this tab can be hidden when it is not in use by clicking the button labeled **Hide link target** and can later be restored by clicking **Show link target**. If the **Linked Items** tab itself is not visible, select it in the drop-down **View** menu at the top of the OCHRE window or click the  toggle button in the main toolbar.

To choose a target item from a list of potential targets, you must first specify the target’s project and category. Choose the project from the drop-down list entitled **Project of link target** and the category from the drop-down list entitled **Category of link target**. All of the items in the chosen category of the chosen project will be displayed.

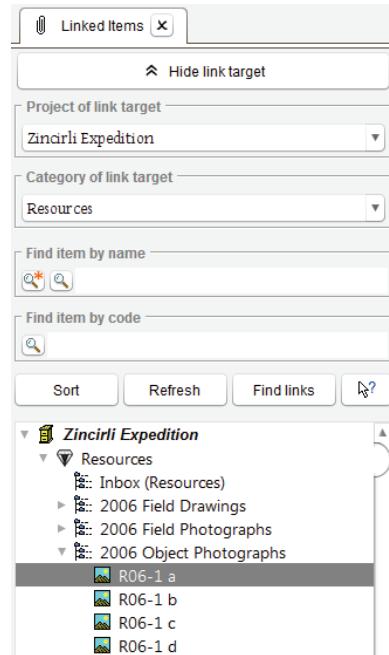
You can then scroll through the hierarchies in the chosen category and click on the name of the desired item to select the link target. Alternatively, there are look-up fields that let you **Find item by name** or **Find item by code** (but note that only items in the **Locations & objects** category have **Code** fields; see Chapter 8). Enter a name (or part of a name) or a code, as the case may be, and click the  “search” button to the left of the field. When finding a target by name, this will search only within the currently selected category. Click the  “search all” button to search within every category of the currently selected project.

The items available for linking can be sorted alphabetically by clicking the **Sort** button. This will make it easier to find a particular link target within a long list. Note that the list of potential link targets may become out of date if the **Linked Items** tab of the reference pane remains open while items are inserted, moved, or deleted in the navigation pane. In that case, the list can be updated by clicking on the **Refresh** button.

In the course of creating links, you may want to know which items, if any, are already linked to a potential link target. Clicking the **Find links** button will perform a query to find all the items linked to the item that is currently selected in the list of potential targets. The results of this query will be displayed in the **Query Results** tab of the reference pane.

The **⊕** “add link” button is used to add a link to the currently selected link target. This button appears in various locations in both the item pane and the reference pane, depending on where links are to be inserted. As noted above in Chapter 5, it is used in the **Links** subtab in the item pane to add general links that pertain to the item as a whole. To save time when working with query results, the **⊕✓** “add link from checklist” button is used to add a link to the item that is currently selected in the **Query Results** list (see Chapter 14). Both the **⊕** “add link” and **⊕✓** “add link from checklist” buttons allow multiple items to be linked at once.

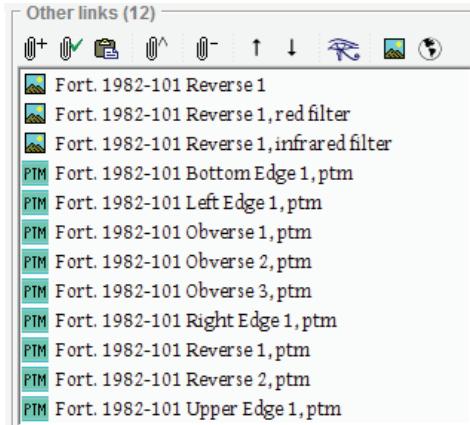
In the reference pane, however, the **⊕** button is used to add a hyperlink from a selected portion of textual content to another item. If the name of the target item is to be inserted automatically into the textual content, the **⊕ “add link with name”** button is used instead of the **⊕** button.



The \ominus “remove link” button always appears with the \oplus “add link” button. Clicking on it will remove the link currently selected in the adjacent **Links** tab or data field. You can change the order of the links by means of the \uparrow “move up” and \downarrow “move down” buttons.

Double-clicking any link displays the link’s target item in a pop-up window. It will be displayed either in data-entry mode or in view-only mode, depending on the current setting of the edit/view toggle button in the main toolbar. The  “view” button opens it in view-only mode, even if the edit/view toggle is set to “edit.”

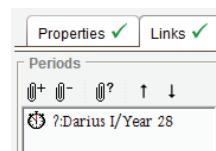
In the **Links** subtab in the item pane, where you enter general links from the current item to other items, there are two additional buttons. Clicking on the  “show linked images” button will open a window that displays all of the images to which the current item is linked. Similarly, clicking the  “show linked geospatial data” button will open a window that displays all of the geospatial data resources to which the current item is linked.



For items in the **Locations & objects** and **Concepts** categories, you will also see the \oplus^+ “add link to new item” button. Clicking it will open a window that lets you name and save a new resource item in the “Inbox” hierarchy of the **Resources** category and, at the same time, add a link to this image item from the current item. This is a useful shortcut during data entry because it avoids the necessity of creating the resource item separately before linking another item to it.

Period Links

In most item categories there will be a **Periods** data field in the item pane. This field can store one or more named links to link targets selected



from the **Periods** category. The organization and use of chronological periods are discussed in detail below in Chapter 8.

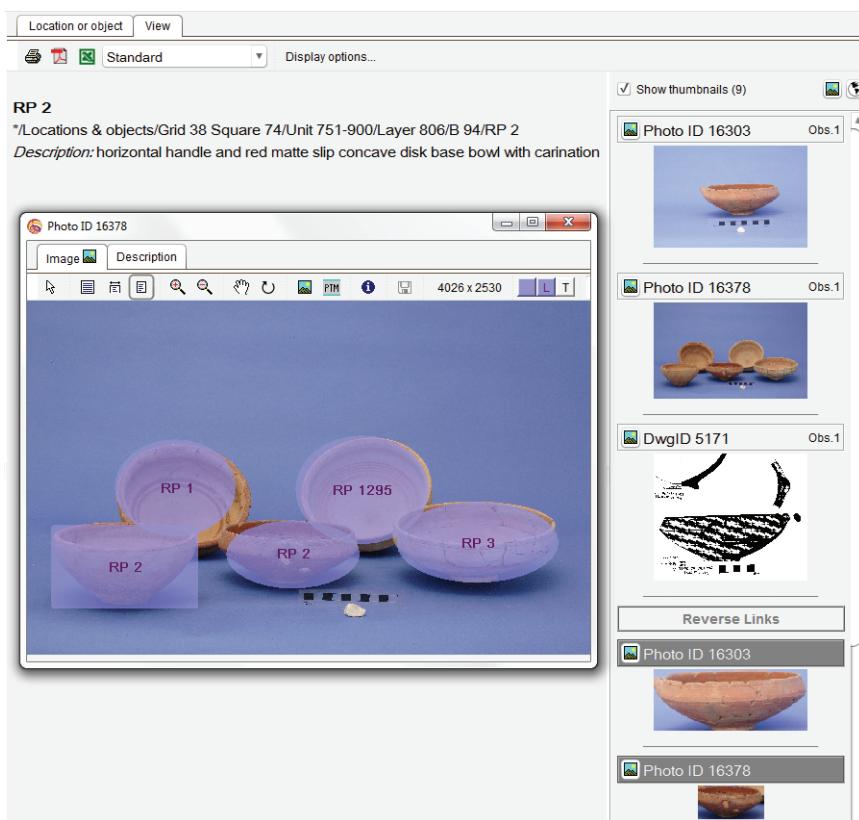
Each link entered in the **Periods** field represents the fact that the current item belongs to the designated chronological period and thus can be found by a query that seeks items of that period or can be displayed on a timeline as an entity associated with that period. If you are not sure about the attribution to a particular period but still want to indicate that the item probably or possibly belongs to it, click the $\emptyset?$ “add uncertain period” button instead of the $\emptyset+$ “add link” button. As with uncertain properties (discussed in Chapter 5), this allows you to reserve the option of excluding the item from later queries or information displays that encompass this period, enabling you to use only those period attributions of which you are sure, or to use all period attributions, both certain and uncertain.

For items in the **Locations & objects** and **Concepts** categories, you will see two additional buttons next to the **Periods** data field: the $\emptyset!$ “add predominant period” button and $\emptyset\sim$ “add preliminary period” button. In cases where an item belongs to more than one period, the $\emptyset!$ button provides a way to represent the fact that the item is mainly associated with one period. For example, an item in the **Locations & objects** category that represents a basketful of pottery sherds excavated from an archaeological site could be linked to a “predominant” period, indicating that most of the pottery belongs to that period even though pottery of other periods is contained in the basket. The $\emptyset\sim$ button provides a way to represent the fact that the assignment of the item to a particular period is “preliminary” and might be changed after further study.

Viewing Reverse Links

In the **Links** subtab of the data-entry tab in the item pane, you can see all of the links *from* the current item to other items. But it is often useful to see a list of all the items that have been linked *to* that item. For example, a photograph of an archaeological site may show a number of architectural features and artifacts, each of which is represented by a different item in the **Locations & objects** category. For

each of those spatial items there would be a general link to the item in the **Resources** category that represents the photograph. You can see those spatial items in the **View** tab of the resource item, which shows not only the “forward” links *from* the current item to other items but also the “reverse” links *to* the current item from other items. Both forward and reverse links (including general links and image hotspot links) are listed on the right side of the **View** tab. Clicking on any item in the list of linked items will display that item’s information in a window.

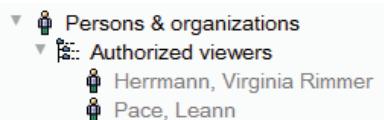


Linking to Items in Other Projects

The ability to link to items in other projects is a powerful means of integrating information derived from multiple sources. For example, a group of image resources in one project may be used by

several other projects. Linking to another project's data is quite easy. It is simply a matter of selecting the desired project in the **Project of link target** drop-down list of the **Linked Items** tab in the reference pane. This list contains the names of all the projects to which you have access, either because they have made their data public for all users to see, or because they have given you a user account that makes use of the username and password by means of which you are currently logged into OCHRE.

A link to an item in a project other than your current project will be displayed in some contexts with the other project's name (or an abbreviated form of the name) prefixed to the linked item's name. The link will be dimmed to indicate that the linked item belongs to a different project and its information cannot be edited within your current project. This is illustrated above with links to persons.



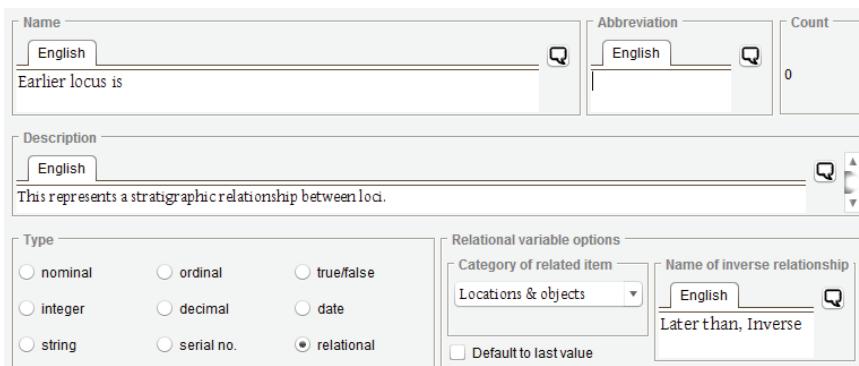
Relational Properties

In contrast to simple links, relational properties permit each project to define its own types of named relationships, in addition to the categories and named links that are built into OCHRE. These relationships can then be described and queried as entities in their own right. Like any other property, a relational property consists of a variable-value pair. Variables of the relational type are defined in the **Property variables** category and are constrained within the project's taxonomy in the usual way (see Chapter 13).

However, there are two additional data-entry fields in the item pane for this type of variable. You must enter the **Category of related item** by selecting the appropriate category in a drop-down list. Secondly, you have the option of entering **Name of inverse relationship**. If you do not enter a name, the inverse relationship will have the same name as the relational variable.

For example, an archaeologist could create a relational variable named "Earlier locus is," specifying that the category of the related item is **Locations & objects** and the name of the inverse relationship is

“Later than.” When assigning properties to an item, the relational variable would be inserted in the **Variable** column of the **Properties** tab, at which point the corresponding **Value** field would require a link to a target item in the appropriate category—in this case, the **Locations & objects** category.



To enter the value, the user would double-click the **Value** field to display **⊕+** “add link” and **⊖-** “remove link” buttons to the left of the field. After selecting the link target in the usual way via the **Linked Items** tab of the reference pane, the user would click the **⊕+** button to insert a link to the target item as the value of the relational variable. In this example, it would be a link to an item that represents an archaeological locus of excavation that was built or deposited earlier than the locus to which the relational property is being assigned.

Properties ✓		Links ✓	Notes ✓	<Apply a predefinition>	
↶	↶	↶*	✖	↶	↷
Variable	Value				
Color	Brown				
▼ Locus type	Debris layer				
Particle size	Medium				
Locus origin	Fill				
Locus component	Mudbrick detritus				
Later locus is	L08-5005				
Later locus is	L08-5013				
Earlier locus is	L08-5007				
Equals or corresponds to	L08-5085				
Equals or corresponds to	⊕+ ⊖- 🔗 L08-5102				

Relational properties provide a simple yet powerful way to represent relationships between items. An item in the **Property variables** category can have its own description and notes, so a great deal of

information can be stored about a relational variable. In queries, relational properties can be used to find database items in two different ways: first, by finding all items that have the same relational property in which the same related item is the “value” of the same relational variable (e.g., find all teachers of the same student); and secondly, by finding all items that are themselves the target “values” of a relational variable (e.g., find all students of the same teacher).

More importantly, the “value” of a relational variable—in this case, a student—is not just a name but a full database item for which a great deal of other information is available and can be incorporated into complex queries that consider the intrinsic properties of related items while at the same time considering the type of relationships among those items. A relational property stores a link to an item, not just its name; thus if any information about the item is changed, including its name, this will be automatically reflected throughout the database wherever the property is used.

Chapter 7

Recording Events that Affect an Item

THE various events that affect a person, object, or other entity can be recorded in the **Events** tab of the item pane for the relevant database item. For example, after an artifact has been unearthed in an archaeological site, its subsequent movements and handling should be recorded. The archaeologists will want to keep track of the steps that have been taken to clean, photograph, draw, measure, analyze, and store the artifact; and they will want to record when, where, and by whom these tasks were performed. In the case of a historical person, a researcher may want to record his or her birth, death, and other life events or achievements, and when and where these took place.

Events are most often recorded for items that belong to the **Locations & objects** and **Persons & organizations** categories but they can also be recorded for **Concepts**, **Resources**, and **Texts**. The event names available within a project for labeling different kinds of events are represented as the values of a special “Events” variable in the project’s **Taxonomy** category.

Entering Events

The “Events” Variable and Its Project-defined Values

A special nominal variable called “Events” is predefined within the parent “OCHRE” project, from which this variable can be borrowed for use by other projects (see Chapter 13 on borrowing items from a different project). To borrow a variable from another project—in this case, the “OCHRE” project—you would select it as the link target, either from the lending project’s **Property variables** category or its **Taxonomy** category, and then click the  “repeat link target” or  “repeat link target with bound branch” button on the main toolbar (see Chapter 4 on the difference between these two forms of copying). If you copy the variable from the **Taxonomy** category, you will also copy its descendants in the taxonomic hierarchy, which are its possible values.

Before anyone can enter events for a project's items, the project administrator must first have copied the "Events" variable from the "OCHRE" project into the top level of the current project's taxonomic hierarchy and must have specified this variable's values as child items in the taxonomy. These values are the names of the different types of events defined by the project. They will appear in the drop-down pick-list in the data-entry tab used for entering events.

When copying the "Events" variable, the project administrator can copy just the variable, without any of the predefined values it has in the "OCHRE" taxonomy. This is done by selecting the "Events" variable as a link target from the "OCHRE" project's **Property variables** category, rather than from its **Taxonomy** category, and then clicking the  "repeat link target" button. The values (event names) to be used by the current project must then be inserted or borrowed individually by the project administrator as child items of the "Events" variable in the project's taxonomic hierarchy.

Alternatively, the project administrator can copy the "Events" variable together with its values from the "OCHRE" project's taxonomy, in such a way that individual values can later be added or removed within the current project's taxonomy. This is done by selecting the "Events" variable as a link target from the "OCHRE" project's **Taxonomy** category and clicking the  "repeat link target" button. This method of borrowing allows the borrowed values to be customized to suit the needs of the current project. The borrowed values (and the "Events" variable itself) will be shown in a different color to indicate that they belong to a different project.

A third option is to copy the "Events" variable and its values in such a way that the values cannot be changed but will always reflect the values (event names) specified in the "OCHRE" parent project. This is done by selecting the "Events" variable as a link target from the "OCHRE" project's **Taxon-**

- ▼  OCHRE Taxonomy
- ▼  Events
 - ◊ Analyzed
 - ◊ Authorized
 - ◊ Checked
 - ◊ Cleaned
 - ◊ Collated
 - ◊ Conserved
 - ◊ Copied
 - ◊ Created
 - ◊ Discarded
 - ◊ Disregarded
 - ◊ Done
 - ◊ Drawn
 - ◊ Edited

omy category and clicking the “repeat link target with bound branch” button. The borrowed values and the “Events” variable will be dimmed to indicate that the hierarchical branch remains tied to the branch from which it was copied in a different project (see Chapter 4 on “bound branch” copying).

A value of the “Events” variable can be designated as the “event fulfillment” for another value of that variable. To do so, select the “unfulfilled” value from the list of children of the “Events” variable in the **Taxonomy** category and then, in that value’s item pane, add a link to another child of the “Events” variable in the **Event fulfillment value** field. For example, the value “To photograph” represents an unfulfilled event whose counterpart is the value “Photographed.” The latter value would be linked as the event fulfillment value for the former. By distinguishing the unfulfilled and fulfilled states of an event in this way you can keep track of the fact that something is waiting to happen to, or needs to be done to, an entity (e.g., an artifact needs to be photographed) and you can issue a query to find out whether the fulfillment event has occurred (see Chapter 14).

The screenshot shows the item pane for an event named "English". The "Name" section contains the name "English" and a search icon. Below it is a "Description" section with the text "To photograph". At the bottom of the pane are tabs for "Preferences", "Thesaurus", and "Links". Underneath these tabs are several checkboxes and dropdowns:

- Keep this item private
- Sorting method for child items: None
- Is core value?
- Event fulfillment value: Photographed

Adding and Removing Events for an Item

To record events for an item, select the item in the navigation pane and open its **Events** tab in the item pane. To insert a new event, click the “insert after” button in the toolbar of this tab and then assign a name for the event by double-clicking in the **Event** column to display a drop-down list of available event names (i.e., the values of the “Events” variable). When you have finished entering a group of events, you must click the **Save** button at the bottom of the **Events** tab in order to save them to the database.

You can enter additional information about an event in the six optional fields entitled **Location**, **Person**, **Period**, **Other**, **Date**, and **Comment**. To enter the location of an event, double-click in the **Location** column to display “add link” and “remove link” buttons that let you add or remove a link to a target item in the **Locations & objects** category, which you must select in the usual way via the **Linked Items** tab of the reference pane (see Chapter 6). In the same fashion, in the **Person** column you can enter a person who is associated with an event via a link to an item in the **Persons & organizations** category; in the **Period** column you can enter the period of an event via a link to an item in the **Periods** category; and in the **Other** column you can enter another entity relevant to an event via a link to an item in some other category. In the **Date** column you have the option of entering the calendar date of the event (the format in which you must enter the date depends on the project preference chosen by the project administrator in the **Data-entry** subtab of the **Preferences** tab in the project’s item pane; see Chapter 17). Finally, in the **Comment** column you can enter a free-form comment about the event (click the button to open a window with various text-formatting options).

In some cases, you may wish to enter both a starting date and an ending date for an event. To do so, click the “show all dates” button in the toolbar of the **Events** tab. This will display additional columns for the event’s **End Date**, **Start Year**, **Start Month**, **End Year**, and **End Month**. If you do not know the full date, you may type in just the year and/or the month when the event began and ended.

To save time when entering a series of similar events, you can insert a duplicate of an event (which you would subsequently modify) by selecting an existing event and clicking the “replicate” button. To insert a duplicate event in which the starting date of the new event is the same as the ending date of the preceding event, click the “replicate as subsequent” button. Another time-saving method is to insert the most recently saved group of events, which you had presumably saved for a different item, by clicking the “copy last saved” button in the toolbar of the **Events** tab.

To delete an event, select it in the list of events and click the **x** “delete” button. To delete all of the events at once, click the **✖** “clear all” button. To move an event up or down in the list of events, select it and click the **↑** “move up” or **↓** “move down” button.

Be careful when using the **Person** column to indicate a person associated with an event assigned to an item that itself belongs to the **Persons & organizations** category. The person entered in the **Person** column of the **Events** tab should not be confused with the person for whom the event was created. For example, there might be an item in the **Persons & organizations** category that represents a student and another item that represents a teacher of that student. An event could be inserted for the student using the **Event** value “Studied with” and the teacher, as the person associated with that event, could be indicated by a link in the **Person** column.

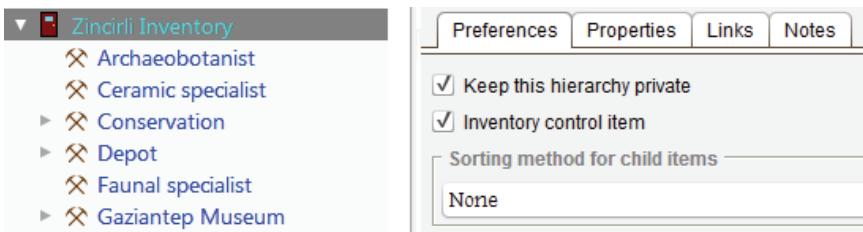
Events					
Event	Location	Person	Other	Period	Start Date
Born	Helmstedt				1822-10-07
Studied at University	Medical Faculty of the				
Student of	University of Göttingen	Wagner, Rudolf			
Promoted (degree) / graduated	University of Göttingen				
Employee of	University of Göttingen	Wagner, Rudolf			
Received (award)	University of Göttingen				

Using Events to Track Objects and Generate an Inventory

Events can be used to automatically generate an inventory of objects in which the objects are listed according to their current locations. This entails creating a hierarchy in the **Locations & objects** category to represent the inventory and checking the **Inventory control item** checkbox in the **Preferences** tab of this hierarchy’s item pane. The items inserted within the hierarchy will represent locations at which objects are stored. The entire hierarchy will be shown in blue to mark it out as an inventory hierarchy.

To indicate that a particular object was moved to one of the locations represented in an inventory hierarchy, select the item that represents that object in the **Locations & objects** category in the

navigation pane and then enter an event for that item in the **Events** tab using the “Moved to” value in the **Event** column. The “Moved to” value is a special event name that is predefined in the “OCHRE” parent project. In order to be available in the current project, it must first be copied by the project administrator from the “OCHRE” taxonomy to the current project’s taxonomy, either individually or together with all of the predefined values of the “Events” variable, as described above. The “Moved to” value is shown in blue to indicate that it pertains to inventory generation.



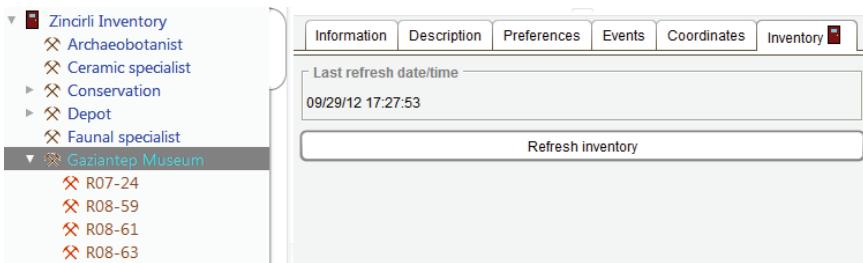
In the **Location** column of the “Moved to” event, you would insert a link to the item within the inventory hierarchy that represents the object’s current location. This event represents the fact that the object was moved to the location specified. When a listing of objects is generated for a location within the inventory hierarchy, the most recent “Moved to” event for each object (i.e., the event with the latest date) will be used to determine where the objects are currently located.

The screenshot shows the Events tab of the item pane. It lists three events: 'Photographed' (Location: Struble, Eudora), 'Moved to Depot' (Location: Struble, Eudora), and 'Moved to Gaziantep Museum' (Location: Gaziantep Museum). The 'Moved to' events are highlighted in blue, indicating they are inventory control items.

Event	Location	Person	Other
Photographed		Struble, Eudora	
Moved to	Depot	Struble, Eudora	
Moved to	Gaziantep Museum		

Finally, to generate an inventory of objects, select a location within the inventory hierarchy by clicking on the relevant item in the navigation pane and open its **Inventory** tab in the item pane. Then click the **Refresh inventory** button in this tab. All of the objects that are currently in that location, as determined by their most recent

“Moved to” event, will be displayed in the navigation pane as child items of the selected location. They will remain in the inventory hierarchy until the parent location’s inventory is refreshed with updated information based on subsequent “Moved to” events.



An inventory can be printed out or exported to an Excel spreadsheet or a PDF document by selecting the inventory hierarchy in the navigation pane and opening its **View** tab. In this respect, an inventory hierarchy functions like any other hierarchy whose contents can be viewed in various formats, depending on the display mode selected (see Chapter 15).

Another way to find objects based on their events is to use a query (this is discussed in detail in Chapter 14). In the **Criteria** tab of a query’s item pane is an **Events** subtab in which you can specify search criteria based on particular events associated with the items to be found by the query, taking into account the location, date, and other information about each event. You can also restrict the search to “fulfilled events” or “unfulfilled events” (these are discussed above).

Chapter 8

Describing and Organizing Units of Space, Time, and Agency

THREE of the OCHRE item categories represent the fundamental concepts of space, time, and agency. These categories are briefly described in Chapter 2. They will be used very frequently in most OCHRE projects so it is worth devoting the present chapter to an in-depth description of these categories and how best to use them.

Items in the **Locations & objects** category represent spatially situated units of observation; hierarchies of such items represent relationships of spatial containment. Items in the **Periods** category represent temporal units; hierarchies of such items represent temporal sequences of chronological periods and the containment of sub-periods within periods. Items in the **Persons & organizations** category represent units of agency; hierarchies of such items represent named groups and subgroups of individual or collective agents. For each of these three categories we describe below the data fields unique to items in the category and explain the use of item hierarchies and item-to-item links to represent relationships among units of space, time, and agency.

Describing and Organizing Locations and Objects

Entering the Geographical Coordinates of a Location or Object

Like all OCHRE items, each item in the **Locations & objects** category has data fields for the item's primary name, abbreviated name, aliases in the project's default language, and names in other languages. These are contained in an **Information** tab, which contains subtabs for the item's **Properties**, **Links**, **Notes**, and **Events** (see Chapters 5, 6, and 7). The latter are organized according to "observation number" because an item in this category can have multiple observations, each attributed to a different observer at a different time. The item's description is contained in a separate **Description** tab, in which free-form textual content can be entered (see Chapter 5). There is another tab for **Preferences**, which are discussed below.

What is unique to items in this category is the **Coordinates** tab, which has subtabs for entering **Geographical** and **Local** coordinates. The **Geographical** subtab contains data fields for the geographical coordinates of a spatial location or object. These fields are optional because the geographical coordinates of many locations and objects—their absolute spatial positions on the surface of the earth—are either unknown or irrelevant. In other cases, you will want to record geographical coordinates so that a location or object's position can be compared to those of other spatial entities and so that it can be plotted on a map. (Note that geospatial data about an item can also be recorded by linking the item to a resource item that represents a spatial data file, such as a “shapefile” that contains the coordinates of vector map shapes; see Chapter 9.)

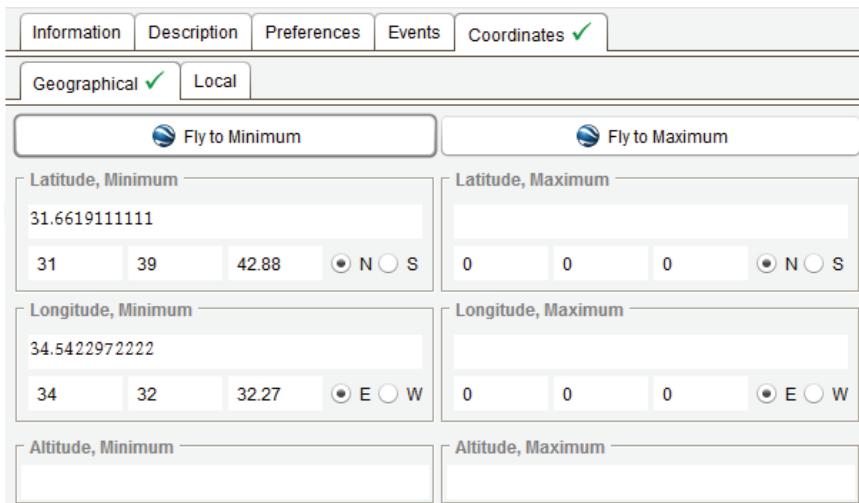
Geographical coordinates are entered as degrees of latitude and longitude, either in decimal degrees or in degrees ('), minutes (''), and seconds (''''') of arc. When you open the **Geographical** subtab of the **Coordinates** tab, you will see fields for entering the **Latitude** and **Longitude** in decimal degrees with up to eight decimal places, ranging from -90 to +90 for latitude and from -180 to +180 for longitude. There is also a field for entering the **Altitude** in meters above sea level, although this may be unknown or irrelevant and can be left blank.

To switch from decimal degrees to degrees, minutes, and seconds of arc, click the button labeled **Deg-min-sec**. This will replace the decimal-degrees field with three separate fields showing the degrees as an integer value from 0 to 90 (for latitude) or from 0 to 180 (for longitude); the arc-minutes as an integer value from 0 to 60; and the arc-seconds as a decimal value from 0.0000 to 60.0000. There will also be a fourth field for the compass direction: either N(orth) or S(outh) of the equator for the latitude, or E(ast) or W(est) of the prime meridian for the longitude. Decimal degrees will be automatically converted to degrees, minutes, and seconds of arc when you click the **Deg-min-sec** button. The button label will then change to **Decimal**. Clicking it again will switch back to decimal degrees.

In the **Geographical** subtab is a button labeled **Region...**. Clicking this button will display data-entry fields for two sets of latitude and

longitude coordinates, allowing you to specify not just a point but a rectangular region on the earth's surface. The existing latitude and longitude (if any) will become the **Minimum (SW) latitude** and **Minimum (SW) longitude**, which identify the southwestern corner of the region. The **Maximum (NE) latitude** and **Maximum (NE) longitude** identify the northeastern corner of the region. Similarly, two altitude fields will be shown instead of just one, allowing you to specify the **Minimum altitude** and **Maximum altitude**. The altitude fields may not be relevant and can be left blank. When the minimum and maximum latitudes, longitudes, and altitudes are displayed, the label of the **Region...** button will be changed to **Point...**. Clicking it again will switch back to a single latitude, longitude, and altitude.

Finally, above each set of latitude-longitude coordinates in the **Geographical** subtab is a  “Google Earth” button. Clicking it will launch the Google Earth application, if it is installed on your computer, and “fly to” the specified geographical location.



Entering the Local Coordinates of a Location or Object

The **Coordinates** tab also has a **Local** subtab for entering the local coordinates of a location or object. Local coordinates are Cartesian coordinates that specify a position in two-dimensional space (x, y) or in three-dimensional space (x, y, z). Many archaeological projects

record the positions of architectural features, debris layers, and individual artifacts in terms of local coordinates rather than geographical coordinates. The **Z** coordinate value is often recorded as the elevation above mean sea level, although it is sometimes the vertical distance from a local datum.

There can be a convergence between local and geographical coordinates when a project uses as its local coordinate space a planar projection of geographical coordinates such as the Universal Transverse Mercator (UTM) projection. In that case, the **X** value may be referred to as the “easting” (i.e., the distance east of the central meridian of the projected planar zone, which in the UTM projection is arbitrarily assigned an easting value of 500,000 meters) and the **Y** value may be referred to as the “northing” (i.e., the distance north of the equator, expressed in meters in the UTM system).

When you open the **Local** subtab you will see fields for the **X (easting)**, **Y (northing)**, and **Z (elevation)** coordinate values. These are decimal values whose units (e.g., meters or feet) are specified by the project administrator in the **Local Coordinates** subtab in the **Preferences** tab of the project’s item pane (see Chapter 17).

As in the **Geographical** subtab, there is a button in the **Local** subtab labeled **Region...**. Clicking this button will display data-entry fields for two sets of coordinates, allowing you to specify not just a point but a rectangular region. The existing coordinates will become the **Minimum X (easting)** and **Minimum Y (northing)**, which identify the southwestern corner of the region. The **Maximum X (easting)** and **Maximum Y (northing)** identify the northeastern corner of the region. Similarly, you can specify the **Minimum Z (elevation)** and **Maximum Z (elevation)**. When minimum and maximum coordinates are displayed, the label of the **Region...** button is changed to **Point...**. Clicking it again will switch back to a single set of coordinates.

Entering a Unique Code for an Object

In the **Information** tab for each item in the **Locations & objects** category is a special **Code** field not found in other categories. This allows you to enter a unique code for an object in addition to the

display name entered in the **Name** field. For example, objects (or the boxes in which they are contained) may be labeled with barcodes so their movements can be tracked with the aid of a barcode scanner. The barcode number can be quickly entered into the **Code** field by scanning the barcode rather than by typing it in, if you use a barcode scanner as an input device.

Objects can subsequently be selected by their codes rather than by their names or properties. This can be done when choosing a link target in the **Linked Items** tab of the reference pane, where you can enter an object's code (by typing it in or scanning a barcode) in the **Find item by code** field (see Chapter 6). Similarly, you can automatically add a group of coded objects to a set of items by clicking the **Add by code...** button in the **Set Items** tab of the **Sets** category. This will display a window in which you can type or paste in a list of codes of the objects to be added based on the contents of the **Code** fields of the relevant items (see Chapter 4).

Name	Abbreviation	Code
MC 66222		ASH0014341

Using Hierarchies to Represent and Query Spatial Relationships

Item hierarchies in the **Locations & objects** category are containment hierarchies (see Chapter 4). Each item in such a hierarchy can directly contain one or more other items that represent units of observation which are understood to be spatially contained within the parent unit. There is no limit on the depth of a spatial hierarchy. Researchers are free to set up hierarchies with as many levels as they need to represent the observed entities, scales of observation, and analytical distinctions they deal with in their project.

Each item in a hierarchy represents a spatially situated unit of observation. The spatial scale of the unit can be of any size, depending on the needs of the project. For an archaeological project there might be large geographical regions at the highest level of a spatial hierarchy. At the next level would be subregions within each region; then sites within each subregion; then stratigraphic units such as architectural features or debris layers within each site; then artifacts and other portable finds within each stratigraphic unit; and finally

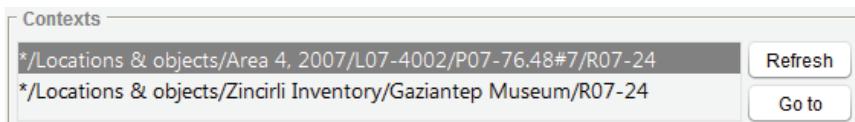
components or analyzed aspects within each artifact. In other domains of research, hierarchies might be quite different in terms of content and spatial scale of observation.

It is often the case that a single hierarchy does not capture all of the spatial relationships among items that need to be represented. Thus, OCHRE permits a given item to appear in any number of different hierarchies, or to appear in different branches of the same hierarchy, while remaining in essence the same item. This is done by repeating an item in more than one hierarchical location via the  “repeat link target” button or the  “repeat link target with bound branch” button in the main toolbar (see Chapter 4).

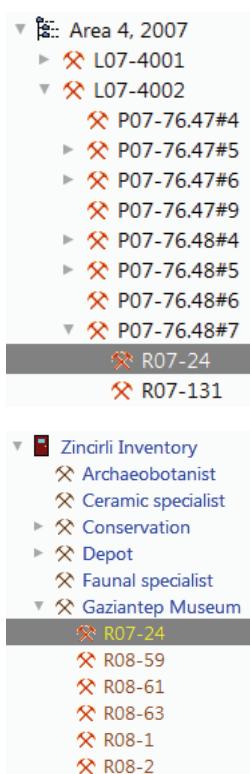
For example, the spatial hierarchy of geographical regions, sites, and excavated artifacts described above might be complemented by a quite different hierarchy that represents the physical organization of a large museum. At the top level of the hierarchy would be the main wings of the museum. At the next level would be the exhibit galleries and storage areas within each wing; then the rooms within each gallery or storage area; then the display cases or storage cabinets within each room; then the shelves within each case or cabinet; and finally the artifacts on each shelf. As should be obvious by now, an artifact on display or in storage in a museum can be represented in OCHRE by the database item that represents the same artifact in its original geographical context at a particular archaeological site. The two different spatial hierarchies, in which the same item appears in relation to a different group of locations and objects, represent two different contextualizations of the same entity at different times.

In most projects there will be analogous situations that call for the creation of multiple overlapping hierarchies to represent the different spatial contexts in which locations and objects are situated. All of the contexts in which a given item is situated are displayed in the **Contexts** field at the top of the item pane, which shows the hierarchical “paths” to the item currently selected in the navigation pane. The first part of each path indicates the project in which the item is located, recognizing that items can be copied from other projects (the current project is indicated not by name but simply by an asterisk).

Clicking the **Go to** button to the right of the **Contexts** area will scroll to the hierarchical location of the selected path in the navigation pane. Clicking the **Refresh** button will refresh the list of context paths in case the list is empty or out of date.



Spatial hierarchies are used in queries in two different ways: to limit the spatial scope of a query or to find items that contain, or are contained in, another item. To limit a query's spatial scope, select the relevant item in the **Queries** category in the navigation pane, open the **Scope** tab in the item pane, and choose the **Locations & objects** category in the **Category** drop-down list. You can then **Limit search to items within** a location selected from this category. To find items that are in a particular spatial relation to other items, open the **Locations** subtab in the **Criteria** tab in the query item pane. There you can enter links to one or more items in the **Locations & objects** category and indicate whether you want to find items that contain, or are contained within, the specified locations. The use of spatial relationships in queries is discussed in more detail in Chapter 14.



Inheriting Properties in Spatial Hierarchies

In addition to representing the spatial context of each item, containment hierarchies in the **Locations & objects** category allow the properties of descendant items to be “inherited” from their ancestors in the hierarchy. Inherited properties are discussed above in Chapter

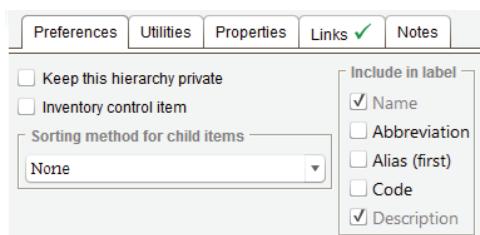
5. An inherited property is a property that, having been assigned to a particular item, can be viewed among the properties of the items contained within that item in a containment hierarchy.

For example, a property that specifies the “Date of excavation” of an archaeological layer could be inherited by all the items that are hierarchically contained within the layer’s database item. In this example, the descendant items would represent artifacts found in the layer. By definition, they would have the same excavation date as the layer. The use of inherited properties therefore avoids the tedious repetition of the same property within a hierarchy in the course of data entry. Note that an inherited property is not actually assigned to the descendant items. It can be viewed within a descendant item, but only as a reflection of a property assigned to an ancestor item.

To designate an inherited property, select its variable item in the **Property variables** or **Taxonomy** category and click the **Inherited by contained items** checkbox in the **Preferences** tab of the item pane. Any property (variable-value pair) that uses this variable can now appear automatically among the properties of every item that is hierarchically contained within the item to which it was originally assigned. Inherited properties will appear in a different color. Note, however, that they are hidden from view by default. They can be shown by checking on the “show inherited properties” option in the **View** menu.

Display Preferences for Spatial Hierarchies

As with hierarchies everywhere in OCHRE, you can indicate your preferences for a hierarchy in the **Locations & objects** category by selecting the hierarchy in the navigation pane and then opening the **Preferences** tab in the item pane. You can choose to **Keep this hierarchy private**, making it invisible to all “anonymous” users. You can also indicate that a spatial hierarchy is an “inventory” hierarchy whose items represent

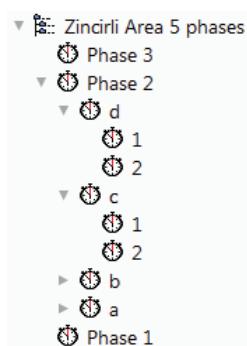


locations at which objects are stored by checking the **Inventory hierarchy** checkbox (see Chapter 7 on tracking objects via events and generating inventories). In the **Preferences** tab you can also choose the sorting method for items within the hierarchy (see the discussion of the various sorting methods in the section on “Display Preferences for Hierarchies and Categories” in Chapter 4); and you can determine what to include in the label of each item in the hierarchy in addition to, or instead of, the item’s primary name (i.e., its abbreviated name, aliases, unique code, and/or description).

You can specify similar preferences, not just for the hierarchy as a whole, but for each item within it. To do so, click on the item to select it in the navigation pane and then open the **Preferences** tab in the item pane. You will see a checkbox to **Keep this item private** (i.e., invisible to anonymous users). You will also be able to choose the sorting method for the child items of the current item—this will override the sorting method for the whole hierarchy. Finally, you can specify a property variable to be used to automatically label the item, choosing one of the following automatic labeling formats: **Default**, in which the item name consists of the variable’s primary name (in the project’s default language) followed by its value; **Abbreviated**, in which the item name consists of the variable’s abbreviation followed by its value; and **Value only**, in which the item name consists of only the value (see the section on “Automatic Naming of Items Based on Their Properties” in Chapter 4).

Describing and Organizing Chronological Periods

Items in the **Periods** category represent temporal units. Items in this category can represent spans of time of any duration, from nanoseconds (or less) to millennia (or more), depending on the needs of the project. Periods are defined, named, and organized according to the chronological schemes and nomenclature used by the project. A temporal sequence of periods is represented by their sequence as



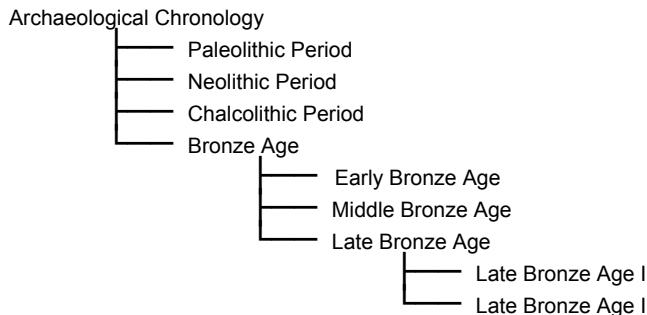
sibling items in a given branch of a hierarchy. Periods that are the children of another period in the hierarchy are understood to be subdivisions of the parent period.

Other database items that are linked to items in the **Periods** category can be automatically arranged in chronological order and placed in synchronic groups by virtue of their period association. OCHRE's explicit representation of units of time and their inter-relationships as database entities in their own right allows queries to determine automatically whether a given item existed before, after, or at the same time as other items.

Note that item hierarchies in the **Periods** category are “containment” hierarchies that represent the strict containment of sub-periods within larger periods (see Chapter 4). This means that there should be a separate hierarchy for each chronological scheme. You should not combine chronologies that are based on disparate criteria within a single hierarchy but should create a new hierarchy for each chronology. For example, a chronology based on archaeological criteria such as the changing styles of material remains in a particular geographical region should be in its own hierarchy distinct from chronologies based on political criteria such as the reigns of particular rulers and the succession of rulers in a dynasty even though some archaeological periods may be conventionally named with respect to political periods, reflecting the impact of politics on style. Moreover, each political chronology should have its own hierarchy. For example, a chronology consisting of the successive dynasties that ruled ancient Egypt and the reigns of successive pharaohs within each dynasty would be represented by its own hierarchy of chronological periods separate from other hierarchies that represent concurrent or overlapping sequences of political regimes in other regions.

In archaeological and historical research, the date and duration of a period are often known, not absolutely, but only relative to earlier and later periods. These would appear before or after it in a hierarchy. Each project can define its own period hierarchies, in keeping with the chronological schemes used in the project's area of research. Hierarchies can also be copied from other projects (see Chapter 4).

For example, an archaeological project might have the following (partial) hierarchy of periods:



Entering the Absolute Start, End, and Duration of a Period

Like all OCHRE items, items in the **Periods** category have data fields for the item's primary name, abbreviated name, aliases in the project's default language, names in other languages, and free-form description. As usual, there are also data-entry tabs for the item's **Properties**, **Links**, and **Notes** (see Chapters 5 and 6).

Unit	Start	End	Duration
Millennium			
Century			
Decade			
Year	-1173	-1142	
Month			

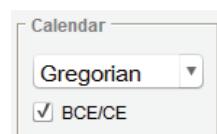
Unique to period items are fields for entering the **Absolute Dates** of the period (when it began, when it ended, and its duration) and the **Calendar** by which the absolute dates are reckoned. These fields are optional because it is often the case that the beginning, end, and duration of a chronological period are not known in terms of absolute calendar dates within a dating system such as the familiar Gregorian calendar reckoned in years B.C. or A.D. However, scholars may well

know the relative sequence of periods and their approximate durations in years, centuries, or (in the case of prehistoric and paleontological research) millennia or millions of years.

If the absolute starting and ending dates of a period are known in terms of a standard calendar, these can be entered at the appropriate level of precision. In the **Calendar** drop-down pick-list, the Gregorian calendar is used by default. This is the well known calendrical system, reckoned in years B.C. and A.D., which is now widely used around the world. For some projects, it will be helpful to define the starting and ending dates of a chronological period in terms of the Julian calendar that was used in Europe before the calendrical reform promulgated by Pope Gregory XIII in 1582 and that continued to be used in Protestant countries, Russia, and elsewhere for centuries thereafter.

Instead of the Gregorian or Julian calendar, you can choose the Islamic calendar, the Jewish calendar, or a calendar reckoned in “Years BP” (Years Before Present). The latter is commonly used by researchers who study prehistoric periods for which the main dating method is radiocarbon dating or some other isotopic or radiometric dating technique (note that, by convention, the “present” year for dates expressed in years B.P. is assumed to be A.D. 1950). OCHRE can convert dates automatically from one calendar to another.

Under the **Calendar** drop-down field is a checkbox labeled **BCE/CE**. If it is checked, then all dates reckoned according to the Gregorian and Julian calendars will be displayed using the suffix “B.C.E.” (Before the Common Era) instead of “B.C.” (Before Christ) and using “C.E.” (Common Era) instead of “A.D.” (Anno Domini). Dates reckoned according to the Islamic calendar will be suffixed with “A.H.” (After Hegira) and dates reckoned according to the Years Before Present calendar will be suffixed with “B.P.”



The starting and ending dates of the period can be entered with the desired precision in the **Start** and **End** columns of the **Absolute Dates** area using some combination of the following units: millennium, century, year, month, day, hour, minute, and second. If you have specified the Gregorian or Julian calendar, you will have a

choice between B.C. and A.D. (or B.C.E. and C.E.) when entering the millennium, century, or year. Only the relevant fields need be filled in. A typical historical date expressed in years would simply be entered in the **Year** row and the other fields would be left blank.

In some cases, you do not know the starting or ending date of a period but you do know its duration to some level of approximation. You can enter this in the **Duration** column.

Linking Other Items to Periods

In most of the item categories there will be a **Periods** data field in the item pane for the currently selected item. This field can store one or more links to items in the **Periods** category (see Chapter 6 on entering period links). Such a link represents the fact that the current item belongs to the designated period. Thus, the item can be found by a query or displayed on a timeline as an entity associated with that period.

An item should be linked to the most specific period possible. In terms of the archaeological example shown above, an artifact that has been ascribed to the “Late Bronze Age II” but has not been dated more precisely should be linked to that period and not to a narrower or broader period. It would not be wrong to link it to the “Late Bronze Age” or even to the “Bronze Age” but those links would be redundant. The fact that it belongs to the Late Bronze Age II automatically implies that it belongs to the Late Bronze Age and to the Bronze Age because these are its direct ancestors in the period hierarchy. A query that searched for all Bronze Age artifacts would find all artifacts been linked to the Late Bronze Age II or to any other subperiod of the Bronze Age.

Using Periods in Queries

Periods are used in queries to limit the temporal scope of a query to a particular period (see Chapter 14 for a detailed discussion of this). To limit a query’s temporal scope, select the relevant query item in the **Queries** category in the navigation pane and then open the **Scope** tab in the item pane. You can then choose to **Limit search to**

items within a period selected from the **Periods** category of the current project.

For example, the phases and subphases of architectural construction at an archaeological site could be represented as a hierarchy of **Periods** items, which could in turn be linked to particular architectural features, debris layers, and small finds represented as **Locations & objects** items. A subsequent query might search for all artifacts that belong to (i.e., are linked to) a constructional phase later than another specified phase. A more complex (but quite typical) archaeological query that makes use of both temporal and spatial relationships might search for all artifacts of a certain kind (i.e., with specified intrinsic properties) found in debris layers of a certain kind that are stratigraphically contemporary with other debris layers that contained another kind of artifact. For example, an archaeologist might want to find all the pottery of a given type that was found in excavated contexts contemporary with (or earlier than, or later than) other contexts in which were found coins of a given date. See Chapter 14 for a more extensive discussion of how to construct queries that exploit both the intrinsic properties of items and their extrinsic temporal and spatial containment relationships.

Describing and Organizing Persons and Organizations

Units of agency are represented as database items in the **Persons & organizations** category. Such items represent OCHRE users as well as observers, creators, authors, editors, publishers, institutions, and any other kind of person or organization, past or present. Items in this category may represent not just current or historical agents but also fictional persons and organizations, if these are regarded as agents or objects of study by a given project.

In many places in OCHRE, a reference is made to a particular agent who plays a specific role as an author, observer, creator, etc. This is done by means of a “named link” from a data field to the appropriate item in the **Persons & organizations** category. For example, the author of a note or the creator of a resource can be identified in this way (see Chapter 6).

Entering Information about a Current Person

An item in the **Persons & organizations** category can be identified as a “current person,” in which case there will be additional data-entry fields for such details as the person’s address and photograph and a place for entering links to institutions with which the person is affiliated. Similarly, an item can be identified as an “organization,” with relevant data-entry fields. Otherwise, the item is simply a “person” who may be living or dead and who may be historical or fictional.

To specify the type of person or organization, select the relevant item in the navigation pane and open its **Information** tab in the item pane. To the right of the usual **Name** and **Abbreviation** fields there will be a drop-down data-entry field labeled **Type** in which you can choose the type. If you choose “current person” or “organization,” click the  “refresh” button in the main toolbar in order to display the additional data fields associated with the selected type.

For a current person there are data-entry fields for the person’s display name in the database and title/position, first name, middle name, last name, initials, address (e.g., street address or post office box), city, state/province, country, postal code, and e-mail address. There are also “named link” fields for the person’s institutional affiliation and picture (e.g., a photograph of the person or some other image associated with him or her). An affiliation can be entered by selecting the relevant organization item as a link target in the **Linked Items** area of the reference pane and then clicking the  “add link” button to the left of the **Affiliations** field. A picture can be entered by selecting the relevant resource item as a link target and clicking the  button to the left of the **Pictures** field.

In the **Affiliations** field you can enter one or more links to institutions represented as items of the “organization” type in the **Persons & organizations** category. Likewise, in the **Pictures** field you can enter one or more links to items in the **Resources** category. These will often be image resources but they could be resources of any type, such as video or audio files (see Chapter 9).

There is an **Authorized user** checkbox below the **Type** field. If the type is set to “current person” and if you are logged in as the project administrator, you can check this box to create a user account for the person that will appear in the **Users** category in the navigation pane. This category is visible only to the project administrator and is the place in which each person’s username, password, and level of access to the project’s data are specified (see Chapter 17).

As in other OCHRE item categories, there are tabs in the item pane for **Properties**, **Links**, and **Notes** (see Chapters 5 and 6). Events in a person’s life—birth, death, or any other event—are represented by entries in the **Events** tab (see Chapter 7). In the **Preferences** tab is a checkbox that provides the option to **Keep this item private**, that is, make it invisible to anonymous users.

Display name	English	Alias	Q	Title/position	English	Q	Type	current person
Schloen, John David				Associate Professor of Archaeology			<input checked="" type="checkbox"/> Authorized user	
First name	J.	Middle name	David	Last name	Schloen	Initials	JDS	
Address line								
1155 E. 58th St.								
City	Chicago	State/Province	IL	Country	USA	Postal code	60430	
Email address	dschloen@uchicago.edu							
Affiliations	<input type="button" value="+"/> <input type="button" value="-"/> <ul style="list-style-type: none"> University of Chicago, Oriental Institute University of Chicago, Divinity School 							
Pictures								

Entering Information about an Organization

If you specify “organization” in the drop-down **Type** field of an item in the **Persons & organizations** category and click the “refresh” button in the main toolbar you will see the data fields associated with this type. These include the organization’s display name, abbreviated name, and free-form description, as well as an address consisting of the street address or post office box, city, state/province, country, and postal code. There are also fields for the

organization's affiliations to other institutions and for pictures associated with it. The data-entry fields are used in the same was as those for a "current person," described above. The main difference between an organization and a current person is that a field for a free-form description replaces the fields for first name, middle name, last name, and initials.

Using Hierarchies to Organize Persons and Organizations

Item hierarchies in the **Persons & organizations** category are "subordination hierarchies." They are not "containment hierarchies," in contrast to hierarchies in the **Locations & objects** category and the **Periods** category. A person cannot directly contain another person in the way that a spatial or temporal unit may contain a smaller unit of the same kind. Instead, hierarchies in this category are constructed using subordinate headings that are not themselves database items but are inserted where needed to create logical groupings of items (see Chapter 4 on subordination hierarchies).

Thus, the fact that a person belongs to a particular organization is not represented by a hierarchy in the **Persons & organizations** category but by a "general link" between two different items within that category (see Chapter 6). To associate a person with an organization you must select the relevant person item in the navigation pane, select the organization item as the link target in the reference pane, and add a link to that organization in the **Links** tab in the person's item pane. Bidirectional item-to-item links of this sort can be used to link any kind of item to a person or organization, depending on the needs of the project.

RESOURCES and BIBLIOGRAPHIES

Chapter 9

Managing Images and Other Resources

IN AN OCHRE database system, bitmapped images in various binary formats (e.g., TIFF and JPEG) are regarded as “external resources.” Other external resources include documents, video clips, two-dimensional vector drawings, three-dimensional models, geo-spatial data files, and so on. The types of resources that OCHRE can manage and display are discussed in the present chapter.

External resource files are not stored within the central OCHRE database but are managed locally by each project administrator. A project will follow its own conventions in naming and organizing such files. It will make them available on its own Web server, from which the files will be retrieved as needed for display in the OCHRE user interface. Only the Internet address of an image or other external resource (its “Uniform Resource Locator” or URL) is stored in the central database. By special arrangement, the University of Chicago’s OCHRE Data Service may be able to store a project’s external resources if it does not have its own Web server (for more information, send e-mail to *ochre@uchicago.edu*).

Internal Documents and External Resources

A project’s external resources are registered within the central database as items in the **Resources** category. A new item must be created for each external resource file. However, there is another kind of resource called an “internal document”; such documents are not stored in an external Web server but are stored within the central database. When you create a new resource item you will specify whether it is an external resource or an internal document.

To insert a new resource item in an existing resource hierarchy, select the hierarchy in the navigation pane and click the ↗ “insert below” button in the main toolbar (see Chapter 4). You will be given the choice to insert a heading, an external resource, or an internal

document. If you have selected the **Resources** category rather than a hierarchy, you can insert a hierarchy, an external resource, or an internal document at the top level of the category.

Item hierarchies in the **Resources** category are “subordination hierarchies.” They are not “containment hierarchies,” in contrast to hierarchies in the **Locations & objects** category and the **Periods** category. A resource does not directly contain another resource in the way that a spatial or temporal unit may contain a smaller unit of the same kind. Instead, hierarchies in this category are constructed using subordinate headings that are not themselves database items but are inserted where needed to create logical groupings of items (see Chapter 4 on subordination hierarchies).

Entering Information about a Resource

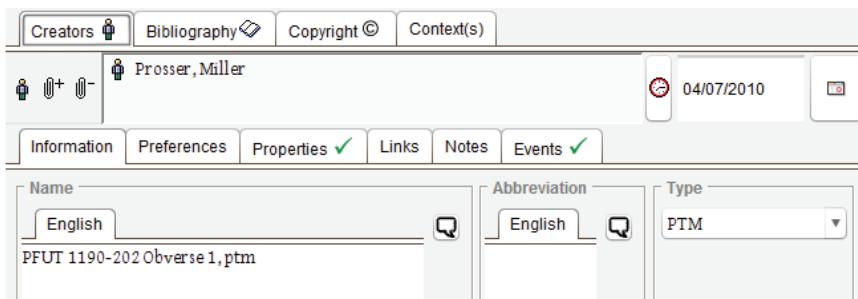
If an item in the **Resources** category is selected in the navigation pane, there will be data-entry tabs at the top of the item pane entitled **Creators**, **Bibliography**, **Copyright**, and **Contexts**. In the **Creators** tab is a field in which you can enter one or more links to items in the **Persons & organizations** category; these represent the creators or authors of the resource. To the right of this field is a data-entry field for entering the creation date (click the button to the left of this field to insert the current date; the button to the right displays a calendar from which you can select a date).

In the **Bibliography** tab you can enter one or more links to items in the **Bibliographies** category; these represent bibliographic references that pertain to the resource (see Chapter 10). In the **Copyright** tab is a field in which you can enter a copyright notice for the resource. A copyright notice can be entered by the project administrator for all the resources in the project via the **Copyright notice** field in the **View-only** subtab of the project’s **Preferences** tab (to see this tab select the parent project’s item in the top left corner of the navigation pane); however, anything entered in the **Copyright** field of an individual resource item will override the project-wide copyright notice.

The **Contexts** tab works in the usual way to show all of the hierarchical contexts in which the resource item appears (see Chapter 4).

Each context is displayed as a slash-separated “path” that shows the item’s location within a particular hierarchy. The first part of each path indicates the project in which the item is located, recognizing that items can be copied from other projects (the current project is indicated not by name but simply by an asterisk). Clicking the **Go to** button to the right of the **Contexts** area will scroll to the hierarchical location of the selected path in the navigation pane. Clicking the **Refresh** button will refresh the list of context paths in case the list is empty or out of date.

Below these four tabs is an **Information** tab that contains fields for entering the resource’s **Name** (click the “translation” button to enter an alias in the project’s default language or a name in another language), **Abbreviation**, and free-form **Description**. A resource item can be named automatically based on one of its properties (see Chapter 4). This is done by entering a link to a property variable in the **Automatically label using this variable** field in the resource item’s **Preferences** tab. The following automatic labeling formats are available: **Default**, in which the item name consists of the variable’s primary name followed by its value; **Abbreviated**, in which the item name consists of the variable’s abbreviation followed by its value; and **Value only**, in which the item name consists of only the value.



To the right of the **Information** tab are tabs for the resource item’s **Preferences** (see below) and for its **Properties**, **Links**, **Notes**, and **Events** (see Chapters 5, 6, and 7). If it is an external resource, there will be a drop-down list to indicate the resource **Type**, which can be one of the following types: **image**, **document**, **audio**, **video**, **drawing**

(i.e., a 2-D vector drawing), **model** (i.e., a 3-D model), **geospatial**, or **PTM** (polynomial texture map). These external resource types and their associated file formats are discussed below.

Other fields in the **Information** tab specify the resource's external file location. You can either type in the **File URL** (see the explanation of this field below) or click the **Browse...** button to select it from a folder on your computer. After the file location has been specified, click the **Open...** button to display the resource (or, if it is an audio or video resource, to play it).

For an external resource whose type is "image," an **Image Tools** tab will appear in the item pane. The image will be displayed in this tab together with a toolbar for embedding links to other database items at designated "hotspot" locations within the image. In this way, an image can be annotated or turned into an "image index" to be used as a means of presenting a project's information (see Chapter 16). The **Image Tools** tab is described in more detail below in the section on "Annotating Resources with Links to Other Items."

Entering Textual Content in an Internal Document

If the resource is an internal document, there will be a large area in the **Information** tab in which to enter the textual content of the resource. The procedure for entering textual content is described above in Chapter 5. Different styles can be applied and special characters inserted with the aid of the text-formatting toolbar and the keypad of special characters. Properties and links to other database items can be embedded in an internal document by attaching them to a particular word, phrase, or other portion of the textual content.

It is often better to use an internal document to store textual content instead of an external resource of the **document** type. An external resource can be displayed in a window but it cannot be as fully integrated into the database as an internal document.

For example, you might have a document that is contained in an external file. In order to integrate the textual content of such a file into the OCHRE database, it should be copied and pasted into an internal document. Selected portions of the document can then be

annotated and hyperlinked with reference to other database items via embedded properties and links. Moreover, translations of the document into other languages can be entered by clicking the “translation” button on the right side of the field. (This button will not be visible unless the project administrator has checked the checkbox to **Enable multilingual features** in the project’s preferences; see Chapter 17.) In some cases, however, the appearance of an external document cannot be replicated within an internal OCHRE document or it consists of scanned page images rather than ASCII or Unicode characters, so it must remain an external resource.

The screenshot shows the OCHRE interface for entering document metadata and content. The fields are as follows:

- Name**: Contains "English" in a dropdown menu. Below it is the identifier "LIA ID 111".
- Abbreviation**: Contains "English" in a dropdown menu.
- Description**: Contains "English" in a dropdown menu. Below it is the text: "William SHARPEY to Richard OWEN, London (?), 12 February (?) 1855".
- Internal document**: Contains "English" in a dropdown menu. Below it is the text: "35 Gloucester Crescent
12 Feby[?] 1855
My dear Owen:
Vogt's memoir which you showed me was what Huxley referred to. On reperusal he at once most frankly admitted his mistake. It was long since he had read the memoir, + the figures had obviously left an impression on his mind not warranted by the text. Before sending his communication to the R.S. he had wished to publish his "Zoological Society" T B I U T^s T_s ☰ t ? Keypad: <default>".
- Text Area**: Contains the message: "I will take care that Vogt's Memoir + your No. of the Proceedings shall be duly returned.
Yours very faithfully
W. Sharpey."

Any textual content you enter for an internal document will be saved automatically to the central database as soon as you leave that field, whether by clicking on another field, selecting a different item, closing the project, or logging out of OCHRE altogether. If the textual content is lengthy you might not leave the field for a long

time, so to avoid losing what you have entered due to an unexpected power failure that shuts down your computer, you should periodically save what you have typed by clicking the  “save” button on the right side of the **Internal document** field.

To export the contents of an internal document to an external file in the Adobe PDF format, click the  “export as PDF” button on the right side of the tab. This will display a window that lets you specify the name and location of the exported file. If you hold down the shift key while clicking this button, the file will be named automatically using the resource item’s primary name and will be saved to your computer’s desktop.

File Locations for External Resources

The project administrator can set preferences for all of the resources in the project via the **Resources** subtab in the **Preferences** tab of the project’s item pane (see Chapter 17). In the field labeled **Resource root folder (high-resolution)** the project administrator will specify the Internet location of the full-size versions of the external resource files. In the case of images, in particular, there is often a distinction between the uncompressed, high-resolution version of an image and a much smaller “preview” or “thumbnail” version of the same image. The high-resolution version contains more detail but may require a long time to be transmitted to the user’s computer.

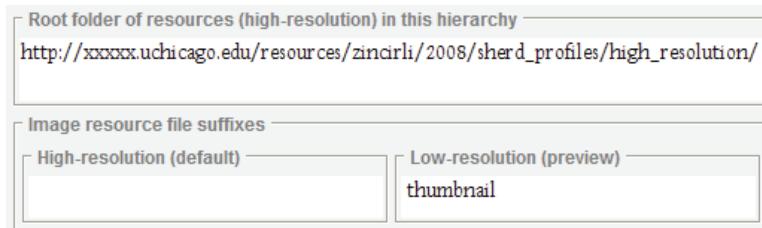
For this reason, there is a second field in the resource preferences subtab entitled **Preview root folder (low-resolution)**. In this field the project administrator will specify the Internet location of the smaller, compressed versions of the project’s external resource files. When viewing an image in OCHRE the low-resolution version will be downloaded first, allowing the user to view it before deciding whether to download the high-resolution version.

In either case, the “root folder” is a character string that specifies only the first part of the URL (Uniform Resource Locator) for a given resource. Each resource item has a data-entry field in its item pane in which to enter the second part of the URL, including the actual file name. This is the “relative path” of the external file that contains the

resource. It is automatically concatenated to the root folder specified in the project's preferences in order to form the complete URL. For example, the relative path *images/image_1.tif* would be the second part of a URL that would be concatenated to the root folder with the name *http://myinstitution.edu/myproject/resources/* in order to form a complete URL which specifies the file's Internet location, namely, *http://myinstitution.edu/myproject/resources/images/image_1.tif*.

The resource root folder is distinguished from each resource's relative path in this way in order to allow the project's external resources to be moved en masse to a different server location (i.e., to a different root folder) without disrupting the hierarchical organization of the files into folders and subfolders. For faster retrieval of resource files, a project may wish to copy them to each user's computer, in which case the resource root folder would be a local folder on the user's disk drive such as *file:/c:/myproject/* (on a Windows computer) rather than a remote folder on a Web server. Of course, this is more feasible when the project's information has not been made public and its users are all members of the project staff.

In some cases, a project may wish to store different sets of resources at different Internet locations. For example, image files might be stored on a different server than PDF documents. For this reason, a root folder can be specified, not just at the project level in the **Resources** subtab of the project's **Preferences** tab, but for each resource hierarchy—assuming, in this example, that image files and PDF files are represented by two different hierarchies. The hierarchy preferences will override the project preferences.



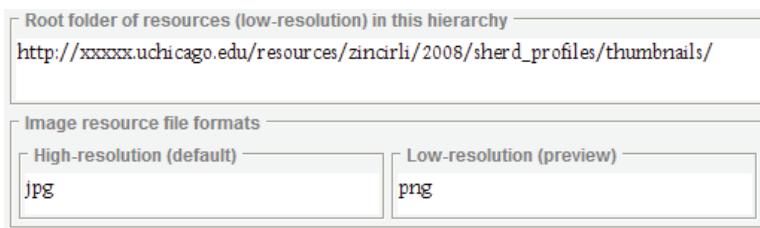
To override the project's resource root folder for a particular hierarchy, select the hierarchy in the navigation pane and then, in the

Preferences tab in the item pane, enter the first part of the URL in the field labeled **Root folder of resources (high-resolution) in this hierarchy**. Here, too, you can specify a different root folder for compressed, low-resolution versions of the same resources via the **Root folder of resources (low-resolution) in this hierarchy** field.

Image File Suffixes and File Formats

For images, in particular, you can distinguish between the high-resolution and low-resolution versions of the same image resource by means of file suffixes. For example, the file that contains the high-resolution version of an image might have appended to its name the suffix “*_big*”, as in *image_1_big.tif*, while the low-resolution version has the suffix “*_small*”, as in *image_1_small.tif*. More likely, the high-resolution version will have no suffix, since it is the primary resource, while the low-resolution version will have the same file name with the addition of some kind of suffix. If a suffix is specified in the **Image resource file suffixes** area of the **Resources** subtab in the project’s **Preferences** tab, it will be used to find the appropriate file, depending on which version is to be displayed. This allows you to mix high-resolution and low-resolution versions of images in the same folder, if you wish to do so, although it is often a good practice to keep them in separate folders.

High-resolution and low-resolution versions of the same image resource can also be distinguished by means of their file formats. For example, you might store the high-resolution versions of images using the TIFF format, as in *image_1.tif*, and the low-resolution versions using the JPEG format, as in *image_1.jpg*, avoiding the need to use different file suffixes or different root folders. The file extensions used to indicate the differing formats of your high-resolution and low-resolution files can be entered in the **Image resource file formats** area of the **Resources** subtab in the project’s **Preferences** tab. Do not include the dot that separates the file name from the file extension; enter only the file extension (e.g., “*tif*” for TIFF and “*jpg*” for JPEG). If a file format is entered for either high-resolution or low-resolution images, it will be used to find the appropriate file.



As with the resource root folder, the image file suffixes and file formats specified for the whole project in the **Resources** subtab in the project’s **Preferences** tab can be overridden for a particular hierarchy of image resources. To do so, select the hierarchy in the navigation pane and enter the desired suffixes or formats in the hierarchy’s **Preferences** tab in the item pane.

Image Watermarks and Preferred Audio and Video Formats

In the **View-only** subtab in the project’s **Preferences** tab there is a data-entry area labeled **Watermark on images**. In the **Watermark** text field you can enter textual content that will appear as a semi-transparent “watermark” overlaid on all images displayed in view-only mode. This watermark will not appear in data-entry mode. It is intended to prevent an anonymous user who does not have a user account for the project from creating unmarked copies of the project’s images to be used without attribution to the project. To the left of the watermark text field is a field for entering a **Copyright notice**. Check the **Use copyright notice as watermark** checkbox to use it as the watermark for images.

You can override the watermark specified in the project’s preferences by opening the **Preferences** tab of a resource hierarchy and entering a different watermark in the area labeled **Watermark on images in this hierarchy**. A watermark specified for the hierarchy will apply to all image resources in the hierarchy. Likewise, you can override a project-level or hierarchy-level watermark for an individual image by entering a watermark in the **Preferences** tab of the relevant resource item, in the area labeled **Watermark on this image**. At each level—the project, hierarchy, or individual item—there is a radio-

button option in the watermark preferences area with the following choices: **Lowest**, meaning that the watermark displayed on an image will be the one entered at the lowest level (i.e., an item-level watermark will override a hierarchy-level watermark, which will override the project-level watermark); **Current**, meaning the currently specified watermark will override watermarks entered at a lower level (i.e., the project-level watermark will override a hierarchy-level watermark, which will override an item-level watermark); and **Suppress**, meaning that no watermark will be displayed on the relevant image or images regardless of the settings at other levels (i.e., no watermark will be shown on the image represented by the current item, or on images represented by items in the current hierarchy, or on images in the entire project, depending on the level at which this option is specified).

In the **Resources** subtab in the **Preferences** tab of the project's item pane, the project administrator can also specify a **Preferred audio format** for audio resources used by the project and a **Preferred video format** for video resources. This is useful in cases where an audio or video clip is stored in two different versions in two files that use different formats (e.g., Microsoft's AVI format and Apple's MOV QuickTime format). Both versions of the resource can be stored in same folder because only the preferred format will be used if there is a choice of formats.

Importing External Documents into the Central Database

External resources of the **document** type comprise a wide variety of digital documents. The following types of documents can be displayed in the OCHRE user interface: unformatted plain text (TXT); word-processing documents in the Microsoft Word format (DOC or DOCX) or the Rich Text Format (RTF); files in Adobe's Portable Document Format (PDF); presentations in the Microsoft PowerPoint format (PPT or PPTX); spreadsheets in the Microsoft Excel format (XLS or XLSX); and Web pages (HTML). Each kind of document is displayed on the user's computer by means of the default software installed for viewing that file type (e.g., Adobe Reader for

PDF files; Microsoft Office for DOC, PPT, and XLS files; and the default Web browser for HTML files).

In some cases, however, it is useful to import or enter the content of such documents into the central OCHRE database. As was noted above, textual content from external documents can be copied and pasted into an “internal document” resource item where it can be annotated, translated, and linked to other database items.

In addition, OCHRE has special features for bibliographies and dictionaries, as well as for texts that are objects of study in their own right. Bibliographies, dictionaries, and texts often have complex internal structures that ought to be represented within the central OCHRE database to permit more powerful linking and querying. These kinds of documents can be represented as items in three special categories discussed in separate chapters below: **Bibliographies** (Chapter 10), **Texts** (Chapter 11), and **Dictionaries** (Chapter 12). Of course, such documents could also be represented and viewed as external resources; but entering them into the central database allows their components to be treated individually as distinct pieces of information that can be related to other information in a highly atomized fashion while still permitting the entire document to be reconstituted so it can be viewed as a whole and exported to an external file.

For example, the epigraphic (physical) structure and the discourse (grammatical) structure of an ancient inscription can be represented in great detail as a hierarchy of items within the **Texts** category, down to the level of individual graphemes and morphemes, or even as sub-elements of these. Special tools are available for importing transliterations of texts encoded as strings of ASCII or Unicode characters. The importing process automatically creates cross-linked hierarchies of database items that represent the epigraphic units and discourse units that make up each text.

Likewise, structured tables of data can either be represented as external resource items in the Microsoft Excel (XLSX) format or, more usefully, they can be imported into the central database, where the rows and columns of each data table are represented as items (one item per row) and as properties of those items (one property per

column). Importing a data table entails automatically decomposing its rows and columns into items with properties; the items are inserted in whichever categories are relevant, depending on the content of the table. Conversely, any set of items in the database (e.g., the result set of query) can be displayed as a data table and, if necessary, exported in turn to an external Excel file.

External Resource Types and File Formats

The types of external resources that can be displayed in the OCHRE user interface, and their associated file formats, are listed below. The data-entry fields common to all external resources (e.g., **Creators**, **Bibliography**, **Copyright**, **Name**, **Abbreviation**, **Description**, **Type**, and **File URL**) are described above. Information and options specific to each resource type are discussed here.

Images (BMP, GIF, JPEG, PNG, TIFF)

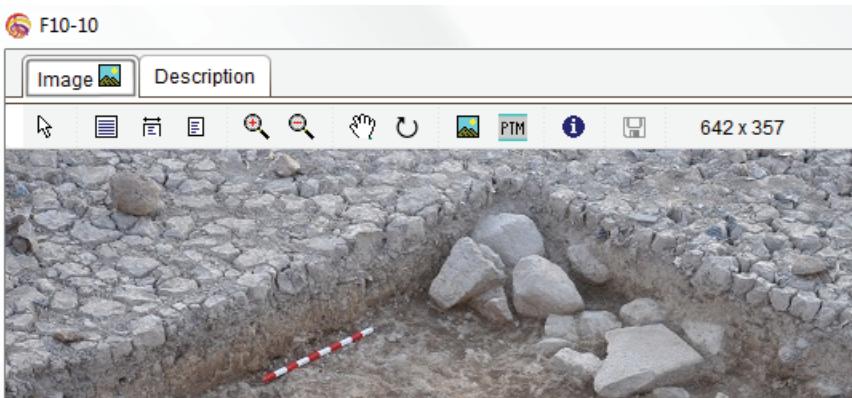
Two-dimensional bitmapped images (e.g., photographs) are represented by resources of the **image** type. OCHRE can display an image stored in any of the following file formats: BMP, GIF, JPEG, PNG, or TIFF. When an image resource is selected in the navigation pane, an **Image Tools** tab will appear in the item pane with toolbar buttons that let you annotate the image and insert “hotspot” links in it. The procedure for doing this is discussed below in the section on “Embedding Annotations and Hotspot Links within an Image.”

If you select an image resource in the navigation pane and click the **Open...** button or open the **View** tab in the item pane, or if you click on an image link in the list of links for another item, you will see subtabs entitled **Image** and **Description**. The **Description** tab shows a formatted, printable view of the names, description, properties, notes, and links stored for the image resource. The **Image** tab shows the image itself. It has toolbar buttons for panning, zooming, resizing, and rotating the image and for displaying a high-resolution version of the image, if one exists.

The toolbar buttons perform the following functions:  “full size” shows the image in its full size (i.e., scaled to 100% of its

original size); “fit width” resizes the image to display the full width of the image; “fit to window” resizes the image to display all of it; “zoom in” enlarges the image; “zoom out” makes the image smaller; “pan” lets you move the image horizontally or vertically by holding down the mouse button and moving the mouse; and “rotate” causes the image to be rotated clockwise by 90 degrees.

By default, the low-resolution version of an image is displayed in the **Image** tab. To display the high-resolution version (if there is one) click the “high-resolution” button in the toolbar. The current resolution is shown at the right edge of the toolbar, expressed as the number of pixels the image contains horizontally and vertically. To display metadata about the image (e.g., the camera model and exposure settings, for a photograph) click the “show metadata” button.



Documents (DOC, HTML, PDF, PPT, RTF, TXT, XLS, etc.)

External resources of the **document** type comprise a wide range of digital documents stored in various file formats. An external document is displayed by the user’s computer using the default software installed for viewing that document’s file format. Widely used digital documents include: plain text files (TXT); word-processing files in the Microsoft Word format (DOC or DOCX) or the Rich Text Format (RTF); files in Adobe’s Portable Document Format (PDF); presentations in the Microsoft PowerPoint format (PPT or PPTX); spreadsheets in the Microsoft Excel format (XLS or XLSX); and Web

pages (HTML). Most users will already have the appropriate software, or can easily obtain it, to view a project's external document resources (e.g., Adobe Reader for PDF files; Microsoft Office for DOC, PPT, and XLS files; and a Web browser for HTML files).

If you select an external document resource in the navigation pane and click the **Open...** button in the item pane, or if you click on an external document link in the list of links for another item, a window will appear that displays the document using the default software installed on your computer for that document's file format.

Audio and Video Files (AIFF, AVI, MP3, MPEG-4, WAV)

An external resource item of the **audio** type can be in the AIFF, AVI, MP3, MPEG-4, or WAV file format. If you select an audio resource in the navigation pane and click the **Open...** button in the item pane, or if you click on an audio resource in the list of links for another item, the sound file will be played through your computer's speakers.

An external resource item of the **video** type can be in the AVI or MPEG-4 (MP4) file format. If you select a video resource in the navigation pane and click the **Open...** button in the item pane, or if you click on a video link in the list of links for another item, the video file will be played in a window.

2-D and 3-D Vector Drawings and Models (AI, SVG, U3D, X3D)

An external resource item of the **drawing** type can be in the AI (Adobe Illustrator Artwork) file format or the SVG (Scalable Vector Graphics) file format. The OCHRE user interface can display SVG files natively. An external resource in the AI format must be saved from Adobe Illustrator with the "PDF compatibility" option in order to be viewed using Adobe Reader, a free program that is already installed on most computers and can be easily downloaded, if necessary. Otherwise, the Adobe Illustrator software itself must be installed to display AI files that are saved without PDF compatibility.

If you select a drawing resource in the navigation pane and click the **Open...** button in the item pane, or if you click on a drawing

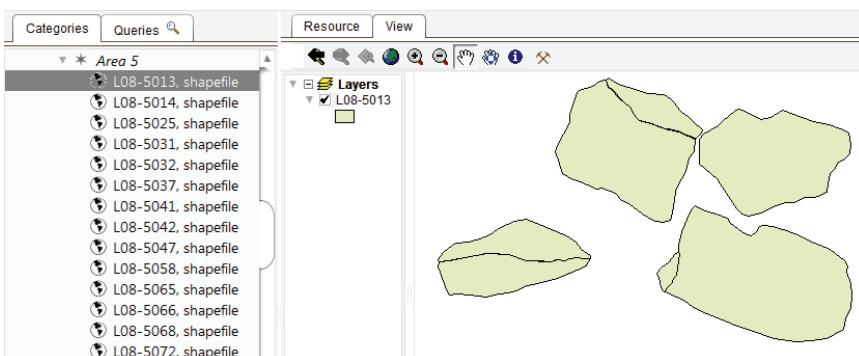
resource link in the list of links for another item, the drawing will be displayed in a window.

An external resource item of the **model** type can be in the U3D (Universal 3D) file format or the X3D (Extensible 3D) file format. A U3D model will be displayed in Adobe Reader. If you select a 3-D model resource in the navigation pane and click the **Open...** button in the item pane, or if you click on a model resource link in the list of links for another item, the 3-D model will be displayed in a window.

Geospatial Data (GeoTIFF, GRID, KML, KMZ, SHP)

External resources consisting of geospatial data are used to display maps (see Chapter 15). An individual database item in any category can be linked to any number of geospatial resources. These resources can then be used to depict that item in a map or plan.

A geospatial resource item can represent a vector shapefile consisting of subfiles with SHP, SHX, DBF, PRJ, and SBN extensions, following the shapefile standard established by ESRI, Inc. A shapefile contains the coordinates for one or more points, polylines, or polygons, depending on the shapefile's type.



Alternatively, a geospatial resource item can represent a raster file: either a georeferenced raster in the GeoTIFF or GRID format or an ordinary raster image in TIFF or JPEG format paired with a “world” projection file that allows the image to be displayed within a planar coordinate system (e.g., *raster_1.tif* paired with a *raster_1.tfw* world file or *raster_1.jpg* paired with *raster_1.jpw*).

When you display a map view of a group of items, all of the shapefiles that have been linked to a particular item will be combined to form a single vector map layer that corresponds to that item. This vector map layer will consist of all of the item's points, polylines, and polygons. However, each raster resource that has been linked to the item will be treated as a separate raster map layer.

Another kind of vector geospatial resource is a KML (Keyhole Markup Language) file, which contains georeferenced “placemarks” and perhaps also ground overlays, paths, and polygons. The compressed form of a KML file has the KMZ extension. If a KML or KMZ file is linked as a resource to another database item it will not be included in the map generated for a set in which that item is found. Instead, the KML data can be viewed by selecting the relevant geospatial resource in the navigation pane and clicking the **Open...** button in the item pane or by clicking on the relevant link in the list of links for another item. Regardless of how you invoke the KML resource, a window will appear—either a Google Earth window, if that software is installed on your computer, or a Google Maps browser window—in which the KML data will be shown.

Polynomial Texture Maps (PTM)

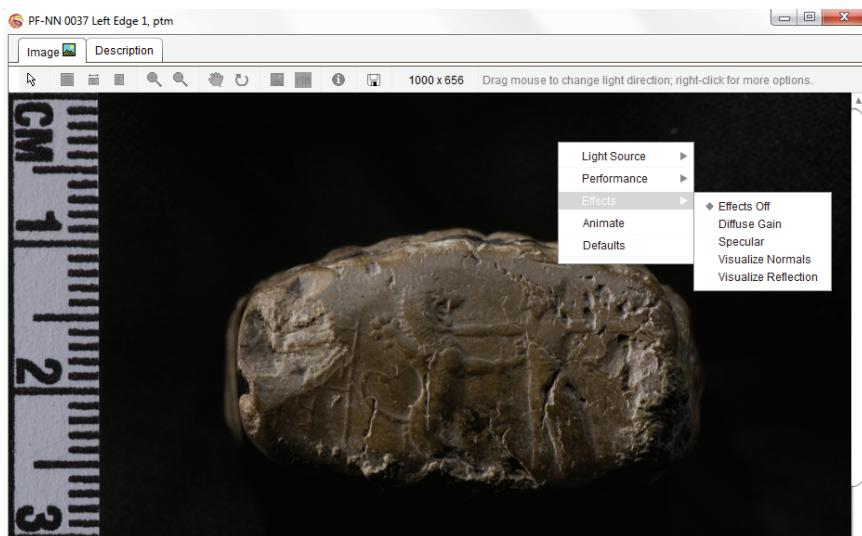
An external resource item of the **PTM** type represents a “polynomial texture map” created by a scanning device that takes a large number of digital photographs of an object, lighting it from a different angle for each photograph. The resulting images are analyzed and combined into a single PTM file. Specialized display software can then dynamically adjust the lighting of the object in response to the user's mouse movements.

To display a polynomial texture map, select the desired resource item in the navigation pane and then click the **Open...** button in the item pane or else open the **PTM** subtab in the **View** tab. (As usual, the **Description** subtab in the **View** tab shows a printable view of the resource's names, description, properties, notes, and links.) Or you can click on a PTM resource link in the list of links for another item or in a query results list. Regardless of how the PTM resource is

invoked, you will initially see a low-resolution bitmapped image extracted from the PTM. This is shown for identification purposes so you can confirm that you wish to load the full PTM file, which may take some time. Click the  “load PTM” button in the image-viewing toolbar in order to load the PTM file and display it with a toolbar that allows you to adjust the display settings.

The PTM display mechanism allows you to adjust the angle and direction of the apparent light source by holding down the mouse button and dragging the mouse. This is very useful when examining objects that have subtle surface details that you wish to highlight. Right-clicking and choosing the “specular” option will make the object being viewed appear to be highly reflective (mirror-like); clicking this button again will turn off the specular mode. Right-clicking and choosing the “animate” option will animate the display, automatically changing the angle and direction of the light source in a repetitive loop. Click this option again to turn off the animation.

To display metadata about the PTM click the  “show metadata” button. You can capture the current view of a PTM (e.g., a particular lighting angle) and save it in a local image file on your computer by clicking the  “export PTM view” button.



Embedding Links to Other Items within an Image or Document

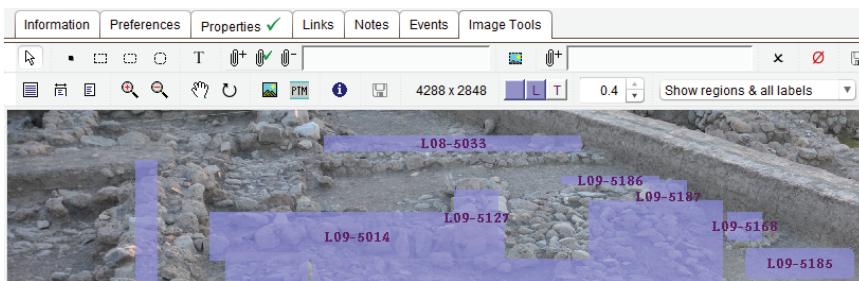
A database item in the **Resources** category can be linked to another item using a bidirectional “general link” that pertains to the resource item as a whole. This is described above in Chapter 6. For example, a photograph of an artifact might be represented as a resource item that is linked to the database item in the **Locations & objects** category that represents the artifact.

It is also possible to embed a link to another item at a particular location within an image or document (i.e., an internal document or an external PDF document). For example, in a photograph of an archaeological site several different architectural features and artifacts might be visible in different areas of the photographic image. A demarcated region within the image can be linked as a “hotspot” to the particular item in the **Locations & objects** category that represents the relevant feature or artifact. Moreover, if an image’s hotspots are linked to other image resources, which in turn have hotspot links to other images, a powerful presentation of the project’s data can be constructed in which users navigate visually by clicking on images and can drill down to increasing levels of detail (see Chapter 16).

The procedure for embedding links at a specified location within a resource depends on the type of resource. For image resources, you would use the **Image Tools** tab, which is discussed in the next section. For internal documents, you would use the procedure described above in Chapter 5 in the section on “Embedding Properties and Hyperlinks within Textual Content.” For external documents in the Adobe PDF format, the procedure is described below in the section on “Linking to a Specific Page of a PDF Document.”

Embedding Annotations and Hotspot Links within an Image

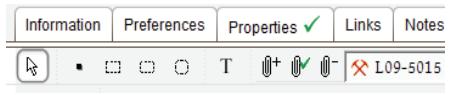
When an image resource is selected in the navigation pane, a tab entitled **Image Tools** will appear in the item pane. When this tab is opened, the image data will be retrieved from the external file location and the image will be displayed in the tab. At the top of the tab are toolbar buttons for annotating the image and for embedding hotspot links at specified locations within it.



By default, the low-resolution version of an image is displayed in the **Image Tools** tab (see the discussion of low-resolution and high-resolution versions of images above in the section on “File Locations for External Resources”). To display the high-resolution version (if there is one) click the “high-resolution” button in the lower toolbar. The current resolution is shown to the right of this button, expressed as the number of pixels the image contains horizontally and vertically. The pixel coordinates of the current cursor location are shown on the right edge of the lower toolbar. Note that the (0,0) origin of the coordinate space is the top left corner of the image. To display metadata about the image, click the “show metadata” button beside the “high-resolution” button in the lower toolbar.

The “pointer” button on the left end of the upper toolbar is used to select the default cursor. Other cursors are activated by means of the buttons to the right of this button; these allow the insertion of a hotspot link or text annotation embedded at a specified point or region within the image.

To insert a hotspot link, select the link target in the **Linked Items** tab of the reference pane (see Chapter 6) and then click the appropriate hotspot-cursor button. Clicking the “hotspot point” button will change the cursor to a “crosshairs” cursor; the next mouse click will insert a link to the target item at the current cursor location. Clicking the “hotspot rectangle” button will also change the cursor to crosshairs; subsequent clicking and dragging of the mouse will demarcate a rectangular region for the hotspot link. Similarly, the “hotspot rounded rectangle” button and “hotspot ellipse” buttons allow



the insertion of hotspot links in demarcated regions that have those shapes. The link field associated with the hotspot drawing tools allows you to change the link target of the hotspot.

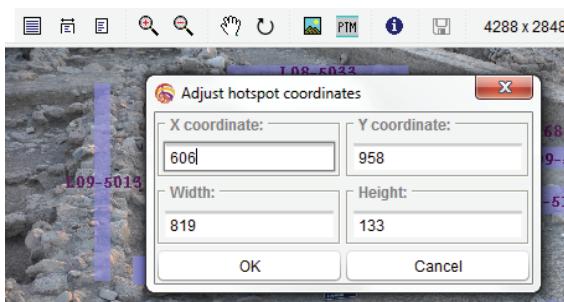
Clicking the **T** “text annotation” button will change the cursor to a text-insertion I-shaped cursor. The next mouse click will open a window in which you can type text that will be inserted at the current cursor location. In this case, the link target is irrelevant; you do not need to have a link target selected in the reference pane in order to insert a text annotation. An annotation is not a hotspot; it does not link the image to another item but simply overlays the annotation text on the image.

Finally, the  “image hotspot” button and its associated link field in the upper toolbar allow you to create a special hotspot link that is labeled with another image. For example, you may wish to label a hotspot with a colorful bitmapped logo that has a transparent background and is stored as an image file in the PNG format. To do this: (1) select the image you want to use as a hotspot label by choosing the appropriate image resource item as the link target in the **Linked Items** tab of the reference pane; (2) click the  “add link” button to the right of the  “image hotspot” button to load this image into the adjacent field in the toolbar; (3) select the hotspot’s own link target in the **Linked Items** tab; and (4) click on the  “image hotspot” button to change the cursor to crosshairs and then move the crosshairs and click once more to specify the top left corner of the embedded image. The image specified in the image-hotspot link field in the toolbar will appear over the main image. It will function as a hotspot region that, when clicked in view-only mode, will call up the item to which the hotspot is linked.

On the right edge of the **Image Tools** tab are buttons for sequentially selecting the hotspots and text annotations in the current image. Clicking the  “previous” button takes you back to the previously selected hotspot or text annotation; clicking the  “next” button selects the next one.

The label of the currently selected hotspot or annotation is shown to the left of the  and  buttons. The  “delete” button in the

upper toolbar deletes the currently selected hotspot or text annotation. The “save” button saves all of the current hotspots and text annotations. The “cancel” button cancels all changes you have made since the last time you saved.



Note that if you wish to adjust the size or location of a hotspot region after it has been created, you do not have to delete and re-enter the hotspot. You can adjust it by

double-clicking on the hotspot in the **Image Tools** tab to display a window in which you may change its pixel coordinates. Hotspots can be inserted in either the low-resolution or the high-resolution version of an image. A hotspot will appear at the same place in the image, regardless of which version is used, because its pixel coordinates will be scaled to match the current resolution of the image.

The lower toolbar in the **Image Tools** tab contains buttons for panning, zooming, resizing, and rotating the image. Adjusting the image display in these ways can make it easier to insert a hotspot precisely where it needs to be within the image. The “full size” button shows the image in its full size (i.e., scaled to 100% of its original size). The “fit width” button resizes the image to display the full width of the image. The “fit to window” button resizes the image to display all of it. The “zoom in” button enlarges the image. The “zoom out” button makes the image smaller. The “pan” button lets you move the image horizontally or vertically by holding down the mouse button and moving the mouse. The “rotate” button rotates the image clockwise by 90 degrees.

On the right side of the lower toolbar are three buttons that let you change the color of hotspot regions, hotspot labels, and text annotations. Clicking any of these color buttons displays a window in which you can specify the color. The “spinner” control to the right of the color buttons lets you adjust the transparency of hotspot regions.

The lower toolbar also contains a drop-down field in which you can specify whether to display hotspot labels and regions and in what combination. The choices are:

- **No labels or regions** (i.e., make the hotspot regions and labels totally invisible, even though clicking anywhere within a hotspot region while in view-only mode will call up the item to which the hotspot is linked);
- **Show labels on rollover** (i.e., do not show the hotspot regions but show a hotspot's label when the cursor passes over its region, in the currently specified color for hotspot labels);
- **Show labels only** (i.e., show all hotspot labels, each centered within its region, but do not show the hotspot regions);
- **Show regions only** (i.e., show all hotspot regions, using the currently specified color and transparency, but do not show the hotspot labels);
- **Show regions & point labels** (i.e., show all hotspot regions and show the labels of “point” hotspots, which are otherwise hard to see, but do not show the labels of rectangular, elliptical, or rounded-rectangle hotspots);
- **Show regions & all labels** (i.e., show all hotspot regions and labels using the currently specified colors and transparency);
- **Show text annotation points** (i.e., do not show any hotspot regions or labels but do show the insertion points for text annotations—note that text annotations will remain visible (in the specified color) after they have been inserted, regardless of the display settings for hotspots).

Thus, in addition to linking an image to other database items, hotspots can be used to annotate the image. The names of linked items can be automatically overlaid on the image, indicating the areas to which embedded links were added. For example, a photograph of an archaeological site could be automatically labeled with the names of the architectural features and artifacts visible in different parts of the photograph. But these labels would also be “live,” in the sense that they are really the names of linked database items and clicking

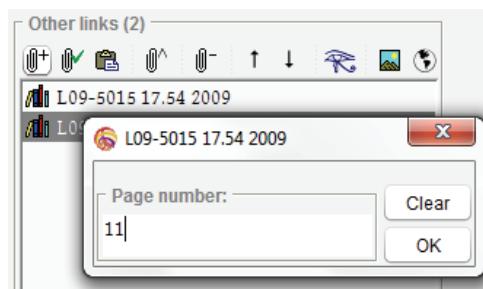
on one of them would call up information about a particular feature or artifact. See Chapter 16 for a more detailed discussion of the use of hotspots to construct image-based presentations of linked data.

Linking to a Specific Page of a PDF Document

When you link a database item to an external resource item of the **document** type that represents an Adobe PDF file, you will have the option of entering a page number that indicates which page of the PDF should be displayed whenever the resource is opened (i.e., it will be automatically scrolled to that page). If no page number is specified, the first page of the document will be displayed.

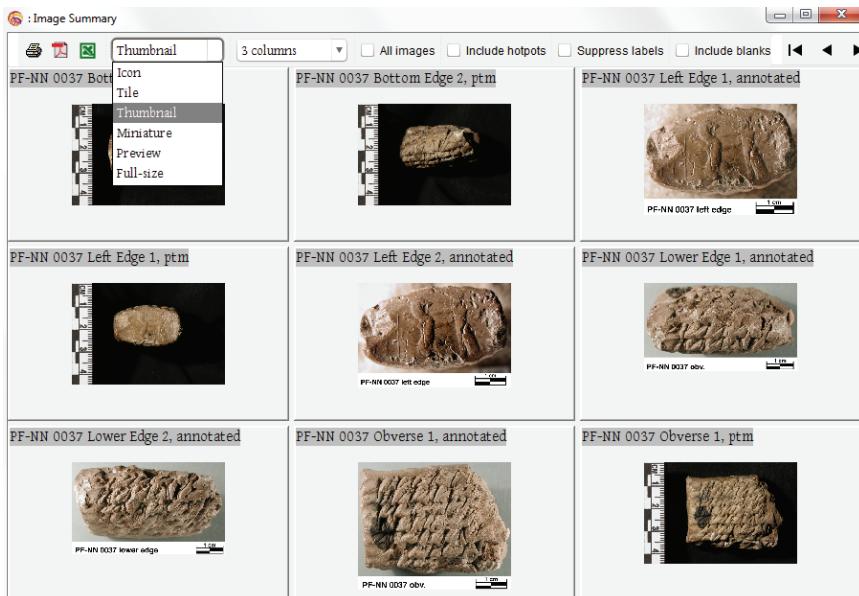
For example, an item in the **Locations & objects** category that represents an archaeological layer might be linked to a document item in the **Resources** category that represents the handwritten field notebook of the excavator, which has been scanned and stored as a PDF file. Such a notebook may contain many pages, so it is desirable to link the database item for a particular layer to the particular page in the PDF that describes that layer.

This kind of page-specific linking can be done by selecting the relevant item in the navigation pane, opening its **Links** tab in the item pane, selecting a PDF document as the link target in the reference pane, and clicking the **+ “add link”** button in the **Links** tab. A window will pop up in which the page number can be entered. The page number field may be left empty but if a number is entered, it will be stored with the link and it will appear in square brackets after the name of the linked PDF resource wherever the item’s links are listed. Whenever a page-specific PDF link is selected (e.g., by double-clicking in the list of links shown in the **View** tab of an item), the PDF document will be shown, having been automatically scrolled to the specified page.



Browsing Multiple Images

All of the image resources that have been linked to an individual item (in any category) can be displayed as a group. To do so, select an item in the navigation pane and open its **View** tab in the item pane. (This tab will be displayed automatically if you are in view-only mode.) Then select the **Images** display mode in the drop-down field shown at the top of the **View** tab. This will display all the images in a scrolling region in the tab, together with various options and buttons that let you resize them and view them in ways that are described below.



Note that this applies to hierarchies as well as items within a hierarchy. Fundamentally, a hierarchy is itself a database item that can be linked to other items including image-resource items (see Chapter 4). In many cases, it will be appropriate to attach an image to a hierarchy instead of attaching it to an individual item within the hierarchy. Thus, when a hierarchy is selected in the navigation pane, selecting the **Images** display mode will display the images linked to the hierarchy itself but it will not display the images linked to items contained within the hierarchy (see Chapter 15).

To display all of the images linked to the items contained within a hierarchy you must select the **Descendants (images)** display mode. If you select a lower-level item that has children of its own within a containment hierarchy (e.g., in the **Locations & objects** category) or if you select a heading in a subordination hierarchy, the **Descendants (images)** display mode will be available in the **View** tab and selecting this mode will display the images of all items descended from the currently selected item or heading.

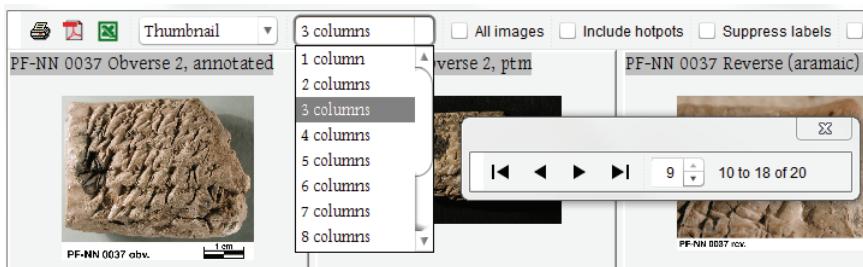
Another way to see all of the images linked to a particular group of items is by means of a list of query results shown in the reference pane. Click the  “image view” button in the **Query Results** tab after performing a query in order to display the images in a window.

Alternatively, you can display all the images linked to the items contained in a named set in the **Sets** category. A set can be created by saving a query result set or by manually adding items. Select the set in the navigation pane, open its **View** tab in the item pane, and then click the  “image view” button at the top of the tab.

Note that the number of images might be very large if a hierarchy or a list of query results or a set contains a large number of items, each of which is linked to image resources. For this reason, the project administrator can place a limit on how many images will be displayed. The maximum number of images that can be displayed at one time is entered in the **View-only** subtab in the project’s **Preferences** tab (see Chapter 17).

Regardless of whether you selected the **Images** display mode from a drop-down list or clicked an  “image view” button, and regardless of whether the images are shown in a floating window or within the **View** tab, you will see the same toolbar with buttons and settings that pertain to the image resources being displayed. If you click the  “print report” button in this toolbar, a report will be printed that contains a summary of the information stored about the images. If you click the  “export as PDF” button or the  “export as Excel” button, information about the images will be exported to an external file in either the Adobe PDF or Microsoft Excel format, depending on which button you have pressed.

To the right of these reporting buttons in the image-viewing toolbar is a drop-down field in which you can select the size at which to display the images. The options are, in order of increasing size: **Icon**, **Tile**, **Thumbnail**, **Miniature**, **Preview**, or **Full-size**. The image-viewing toolbar also contains a drop-down field in which you can select the number of columns to be used to arrange the images horizontally, ranging from one to twelve columns, as well as a “spinner” data-entry field that lets you specify the number of images to be displayed at a time. The allowable number of images in the spinner field will be a multiple of the number of columns. For example, if you have specified four columns, you can choose to display four images at a time, or eight images, or twelve, or sixteen, etc.



The image-viewing toolbar contains buttons that let you scroll sequentially through the images. The **◀ first** button scrolls back to the first image. The **◀ previous** button scrolls the display backward by the number of images specified in the spinner field. The **▶ next** button scrolls the display forward by the number of images specified. The **▶ last** button scrolls forward to the last image.

In many cases, an item will have more than one image linked to it; for example, there may be several photographs of an artifact that are linked to it as image resources. But often you will want to display only the first image for each item, especially if you are viewing the images associated with a large number of items in a hierarchy or set and not just the images for one item. By default, only the first image will be shown for each item; to display all of the images you must check the **All images** checkbox in the image-viewing toolbar.

The demarcation of “hotspot” regions within images is discussed above in the section on “Embedding Annotations and Hotspot Links within an Image.” If you check the **Include hotspots** checkbox the hotspot regions (if any) that are embedded within each image currently being displayed will also be displayed as additional, separate images. This is useful, for example, when you have a photograph of a complex artifact and have marked out several areas of the photograph as hotspots linked to other database items via the **Image Tools** tab (see below). You may wish to display these hotspot regions together with the original photograph as close-ups of different parts of the artifact.

Two other checkboxes in the image-viewing toolbar allow you to customize the image display even further. Clicking the **SUPPRESS labels** checkbox will hide the names of the image resources. Clicking the **Show blanks** checkbox will display a blank image for each resource whose external image file is unavailable; otherwise, these image resources will not be shown at all.

Chapter 10

Creating and Using Bibliographic Items

ITEMS in the **Bibliographies** category represent published works that are cited somewhere within a project's data. These cited works often will not exist in digital form but only as printed books and articles, for which the OCHRE database stores the bibliographic reference data. Storing a bibliographic reference as a database item in its own right allows the published work to be cited in many different places via links from other database items to the relevant bibliography item. This prevents the redundant and error-prone duplication of such information; an update or correction to bibliographic data will appear instantly everywhere it is cited.

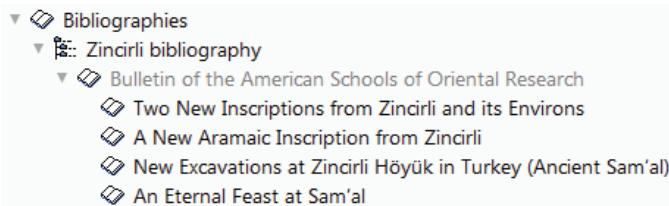
As was discussed above in Chapter 9, there are many kinds of digital resources that a project may wish to represent as items in the **Resources** category so they can be retrieved from external sources and linked to other database items. In many cases these digital resources will be unpublished, consisting of images, documents, drawings, and other materials that belong to the project and have not been formally published. Thus, there will be no need to store bibliographic data about them.

However, in some cases a digital resource will consist of a published work for which there is a corresponding bibliographic entry. For example, a PDF version of a published book or journal article could be represented as an item in the **Resources** category. In that case, the resource item would be linked to an item in the **Bibliographies** category that stores the reference data about the book or article. In general, though, an OCHRE bibliography item need not have any link to the content of the work it references; it simply contains bibliographic data about the work, which in many cases will be a nondigital printed work available only as a physical object in a library.

Entering Bibliographic Data

An item in the **Bibliographies** category represents a bibliographic entry of some kind. An item hierarchy in this category represents a

group of related bibliographic entries. These are “containment hierarchies,” as in the **Locations & objects** category (see Chapter 4). They represent containment relationships of various kinds, such as the containment of articles within journals, chapters within books, books within series, and so on. Using containment hierarchies in this way avoids the need to repeat bibliographic data about a series to which many books belong or a journal to which many articles belong, because this data is associated with the parent in the hierarchy and can be inherited from it.

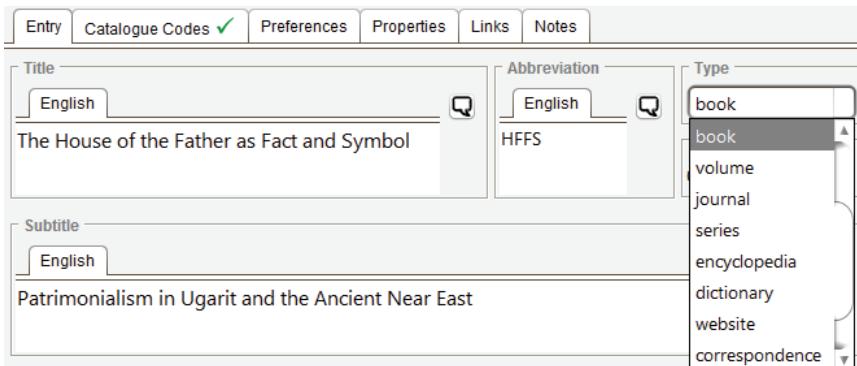


After you select a hierarchy or individual item in the **Bibliographies** category and click the ↪ “insert below” button in the main toolbar, a new bibliographic item will be inserted as a child item contained within the currently selected hierarchy or item. Information about the new item will appear in the item pane. As in other item categories, there will be a **Contexts** area at the top of the pane showing the paths to all the hierarchical locations at which the item has been repeated in the database, and there will be data-entry tabs for the item’s **Preferences**, **Properties**, **Links**, and **Notes** (see Chapters 4, 5, and 6).

Data fields specific to bibliographic items are found in the tabs entitled **Entry** and **Catalogue Codes**. The **Entry** tab has fields for the bibliographic item’s **Title**, **Subtitle**, **Abbreviation**, and **Type**. Another field displays a **Count** of the number of descendant items that are hierarchically contained with the current item (e.g., the number of books contained in a series). Below these fields are subtabs for entering **Publication Data**, **Abstracts**, and **Reviews**.

The following types can be specified via the **Type** drop-down field: **archive**, **article**, **chapter**, **book**, **volume**, **journal**, **series**, **encyclopedia**, **dictionary**, **website**, **correspondence**, **zotero**, **other**. This list is

not meant to be exhaustive; the last option, **other**, can be used if none of the more specific types suits the current entry. The **zotero** type pertains to the Zotero online database, which is discussed below. The use of different types is discussed in the next section. Regardless of the type, the data-entry fields and tabs are the same for every bibliographic item.



Entering Titles, Subtitles, and Abbreviations

The **Title**, **Subtitle**, and **Abbreviation** fields of the **Entry** tab are self-explanatory. Note that you can use the text-formatting toolbar and the keypad of special characters when entering textual content in any of these fields (see Chapter 5). These tools may be necessary if there are italics, diacritics, or special characters to be entered.

In the case of a website, its URL should be entered into the **Title** field and its name, if any, should be entered in the **Subtitle** field. For example, for a bibliographic reference to the online newspaper “The Huffington Post” you would enter <http://www.huffingtonpost.com> in the **Title** field and The Huffington Post in the **Subtitle** field.

If the project administrator has checked the **Enable multilingual features** checkbox in the **Data-entry** subtab of the **Preferences** tab for the overarching project item, a “translation” button will be available on the right side of each of these fields (see Chapter 4). To enter an alternative title, subtitle, or abbreviation, click this button and choose either **Alias** or a language different from the project’s default language. A new tab will appear, entitled either **Alias** or with the name of the chosen language. An alternative or translated title,

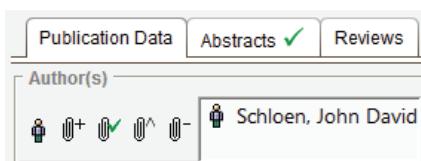
subtitle, or abbreviation can then be entered into the new tab. As in other item categories, you can enter any number of alternative titles in the project's default language (these are called "aliases") and any number of alternative titles in non-default languages.

Note, however, that a bibliographic reference to a translation of a published work should usually be represented, not by means of a translated title within a single bibliographic item, but by means of another bibliographic item. This is because the publication data of the translation will differ in some respects from that of the original (i.e., there will be a different date of publication and sometimes a different publisher).

Linking to Authors, Editors, and Publishers

The **Publication Data** subtab of the **Entry** tab contains fields for authors, editors, source documents, publishers, publication dates, volume numbers, issue numbers, and page ranges. The fields for **Authors**, **Editors**, and **Publishers** store links to database items in the **Persons & organizations** category. Each of these fields is optional and may be empty, in case there is no author or editor or publisher.

In each field, one or more links can be entered by using the $\mathbb{0}+$ "add link" button to the left of the field in conjunction with a link target selected in the **Linked Items** tab of the reference pane (see Chapter 6). If the relevant item does not yet exist in the **Persons & organizations** category, click the $\mathbb{0}^+$ "add link to new item" button instead of the $\mathbb{0}+$ "add link"



This allows you to insert a new person or organization item in the "Inbox" hierarchy of the **Persons & organizations** category, linking it automatically to the current data field, without having to leave the bibliographic item for which you are entering data.

Automatically inserted items can accumulate rapidly, so it is a good practice to clean out the "Inbox" hierarchy regularly, moving its items to other hierarchies in the category in order to organize them in a more meaningful fashion.

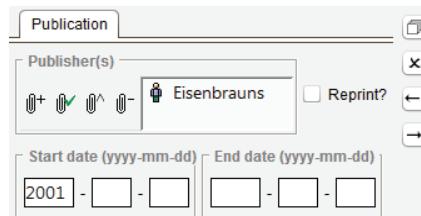
Linking to Source Documents

The **Source documents** data-entry field stores links to items in the **Resources** category. This field is optional but is useful in cases where a digital file exists (or perhaps a group of files) in which is contained the actual work to which the bibliographic item refers. For example, a journal article or book may be available as a PDF file. In that case, OCHRE users will be able to go from the bibliographic citation to the work itself by means of a document resource link in this field. Links to more than one resource item can be entered, if necessary.

The **Source documents** link field functions in the same way as the **Authors**, **Editors**, and **Publishers** fields. It, too, gives you the option to use the “add link to new item” button instead of the “add link” button, in cases where the relevant item does not yet exist. In this case, however, new items will be automatically inserted in the “Inbox” hierarchy of the **Resources** category, not the **Persons & organizations** category.

Entering Publication Dates

The publication date is entered in the **Start date (yyyy-mm-dd)** and **End date (yyyy-mm-dd)** fields below the **Publishers** link field. Often you will enter only the year of publication, which you would type as a four-digit number into the leftmost subfield of the **Start date** field, leaving all the other date fields blank. In some cases, there will be a month and perhaps a day of publication (entered as two-digit numbers in the middle and rightmost subfields, respectively), in addition to the year. There may even be a date range rather than a single date of publication, requiring an **End date** as well as a **Start date**.



The fields for publishers and publication dates are contained in a separate **Publication** tab within the larger **Publication Data** subtab. If there is a reprint of a book or article produced by a different publisher or at a later date, you may wish to record this publisher and publication date within the bibliographic entry for the original work

instead of creating an entirely new bibliographic item for the reprint. To do so, you would create a new **Publication** tab that overlaps the existing one by clicking the  “create new” button at the right edge of the tab and then checking the **Reprint?** checkbox in the new tab to indicate that it pertains to a reprint.

You can create as many tabs as you need to store multiple publishers and publication dates, not just in the case of reprints but for whatever reason this may be necessary. Click the  “delete” button to delete the currently selected **Publication** tab. Click the  “move left” button or the  “move right” button to move the selected tab left or right in the sequence of overlapping tabs.

Entering Volume Numbers, Issue Numbers, and Page Numbers

Volume numbers, issue numbers, and page numbers are entered in the **Volume range**, **Issue range**, and **Page range** fields in the **Volume-Issue-Page** tab within the larger **Publication Data** subtab. Each of these fields is optional. In each case, there are two subfields to allow for a range of numbers. For example, a journal article will usually have a starting page number and an ending page number. If there is only one volume number, issue number, or page number—as will often be the case—you would enter it into the leftmost subfield of the relevant range field.

There may be multiple volume ranges, issue ranges, or page ranges for a particular bibliographic reference. In that case, you would create a new **Volume-Issue-Page** tab that overlaps the existing one by clicking the  “create new” button at the right edge of the tab. You can create as many tabs as you need to store multiple ranges. Click the  “delete” button to delete the currently selected tab. Click the  “move left” button or the  “move right” button to move the selected tab left or right in the sequence of overlapping tabs.

Entering Abstracts and Reviews

Overlapping the **Publication Data** subtab in the **Entry** tab are two other subtabs for entering **Abstracts** and **Reviews** of the work referred to by the current bibliographic item. The **Abstracts** tab functions in

exactly the same way as a **Notes** tab (see the section on “Entering Free-form Notes for an Item” in Chapter 5). The **Reviews** subtab is also intended to work in the same way; however, this feature is not yet implemented.

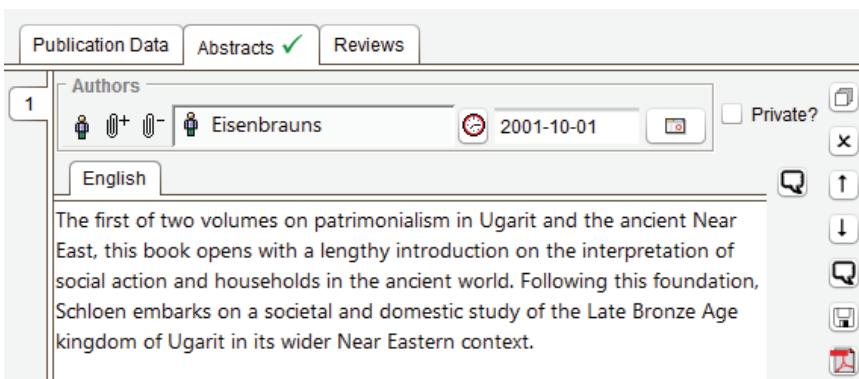
To enter an abstract, open the **Abstracts** tab and create a new abstract by clicking the “create new” button at the right edge of the tab. A numbered subtab will appear in which you can type or paste textual content. Repeatedly clicking the “create new” button will create multiple abstracts displayed in numbered subtabs arranged vertically along the left side of the **Abstracts** tab. The “delete” button deletes the currently selected abstract. The “move up” and “move down” buttons move the selected abstract up or down in the sequence of abstracts. Clicking the **Private?** checkbox will make the abstract invisible to anonymous users. If the project administrator has selected the **Enable multilingual features** option in the project’s preferences, the “translation” button will be available on the right side of the **Abstracts** tab to enable you to insert a translation of the abstract in a different language.

At the top of each abstract is an **Authors** field in which you can enter one or more links to items in the **Persons & organizations** category. Such links are made in the usual way by selecting the link target in the **Linked Items** tab of the reference pane and clicking the “add link” button to the left of the field (see Chapter 6). If there are multiple links, they can be reordered by means of the “move up” and “move down” buttons to the left of the field. A link can be removed by selecting it and clicking the “remove link” button.

To the right of the **Authors** field is a date field into which a date can be typed directly or entered by clicking the calendar button to the right of the field. Alternatively, the button to the left of the date field lets you insert the current date as the date of the abstract. Both the author and the date are optional. To save time in cases where you (as the current user) are the author of the abstract, you can simply click the “current user” button to the left of the **Authors** field. This will automatically enter a link to the person item that represents you and will automatically enter the current date.

Any textual content you enter will be saved automatically to the central database as soon as you leave the data-entry field, whether by clicking on another field, selecting a different item, closing the project, or logging out of OCHRE altogether. If the textual content is lengthy you might not leave the field for a long time, so to avoid losing what you have entered due to an unexpected power failure that shuts down your computer, you should periodically save what you have typed by clicking the “save” button on the right side of the **Abstracts** tab.

To export the contents of an abstract to an external file in the PDF format, click the “export as PDF” button on the right side of the tab. This will display a window that lets you specify the name and location of the exported file. If you hold down the shift key while clicking this button, the file will be named automatically using the bibliographic item’s title and saved to your computer’s desktop.



Entering Catalogue Codes and Linking to Zotero

At the top of the item pane beside the **Entry** tab is a data-entry tab for **Catalogue Codes**. This tab contains optional fields for entering the **Library of Congress Control Number (LCCN)** and/or the **International Standard Book Number (ISBN)** of the work referred to by the current bibliographic item. In addition, there is an optional field entitled **Online Computer Library Center (WorldCat)** for entering a reference to information stored in WorldCat, which is “a global catalog of library collections” (see <http://www.worldcat.org>).

To the right of each of these fields is a **Find** button. Clicking it will perform an online search to find the LCCN, ISBN, or WorldCat reference for the current bibliographic item and will enter it automatically into the relevant field. The bibliographic data already in the **Entry** tab will be used to guide the search. (Note that this automatic searching capability is not yet fully implemented.)

OCHRE bibliographies can also be integrated with the Zotero online bibliographic database (<http://www.zotero.org>). If you have a Zotero user account and wish to link a bibliographic item in OCHRE to information stored in the Zotero database, you must first set the item's **Type** field to the **zotero** option in the **Entry** tab. Moreover, the item must be contained in an item hierarchy for which Zotero access keys have been specified. Item hierarchies in the **Bibliographies** category have a special tab in the item pane entitled **Zotero Access**. In this tab are fields in which you can enter the **User key**, **Group key**, and **Authentication key** which are required for access to the Zotero database. This makes it easy to link to bibliographic information in Zotero and display it in the OCHRE user interface.

Types of Bibliographic Items

A bibliographic item can belong to one of the following types, as specified in the **Type** drop-down field in the **Entry** tab: **archive**, **article**, **chapter**, **book**, **volume**, **journal**, **series**, **encyclopedia**, **dictionary**, **website**, **correspondence**, **zotero**, **other**. This list is not meant to be exhaustive; the last option, **other**, can be used if none of the more specific types suits the current item. However, you should assign a specific type wherever possible in order to facilitate subsequent searching and managing of bibliographic references. You can use a query to search for bibliographic items of a particular type, perhaps in combination with other search criteria, and then generate lists of the selected references based on the query results (see Chapter 14). For example, you might want to list only the books cited in a project's notes and descriptions, ignoring the citations to articles.

Managing a large number of bibliographic references is made easier by organizing bibliographic items into containment hierarchies

in which items of different types are inserted at different levels. For example, the bibliographic data about a monograph series would be represented by a bibliographic item of the **series** type; hierarchically contained within this item as child items would be items of the **book** type that represent books in the series; moreover, a separately cited chapter within a book in the series could be represented by an item of the **chapter** type inserted as a child of the relevant **book** item.

The hierarchical relationships among bibliography items are used to generate full bibliographic references that include the information entered at each level. For example, the **article** item that represents a journal article would contain in its data fields the article's title and author, its volume and issue numbers, and its page range; but it would not contain the title of the journal itself, which would be stored in a parent **journal** item. When a bibliographic reference is generated from a bibliography item in order to be displayed elsewhere in OCHRE, relevant pieces of information are collected from each of its hierarchical ancestors in order to construct the full reference. The full reference for a bibliographic item can be seen by selecting it in the navigation pane and opening its **View** tab in the item pane.

Note that the place of publication is not entered into any of a bibliographic item's data fields. The place of publication is derived from the **City** field (and perhaps also the **State/Province** or **Country** fields) of the organization item in the **Persons & organizations** category that represents the publisher. A bibliographic item is linked to such an item via its **Publishers** field. The publisher's name and city are fetched from the organization item whenever the full bibliographic reference is displayed.

Links to bibliographic items can be made from any other database item. The links can be in the form of general item-to-item links, named links from **Bibliography** data fields, or hyperlinks embedded in textual content (see Chapter 6 on the different kinds of links and how to enter them). In each case, only a link is inserted; the bibliographic data is not duplicated. Changes or corrections entered in the data fields of a bibliographic item will be instantly propagated throughout the database and will appear wherever that item is referenced.

Chapter 11

Working with Texts and Writing Systems

ITEMS in the **Texts** category represent written texts which are themselves objects of study and analysis. For example, items in this category can represent ancient inscriptions that use complex non-alphabetic writing systems such as Egyptian hieroglyphs or Mesopotamian cuneiform. Modern documents such as scholarly articles are not normally stored in the **Texts** category but are instead represented as **Resource** items—perhaps as resources of the “internal document” type or, in many cases, as external document resources in Adobe PDF format or some other format (see Chapter 9). However, any text that will be subjected to epigraphic, grammatical, literary, or linguistic analysis should be entered (or automatically imported) as a database item in the **Texts** category, regardless of its writing system, medium of inscription, or period of composition.

Before entering such a text, it is often desirable to have a representation of the text’s writing system already in the database. This enables more powerful querying and analysis of the epigraphic characteristics of texts that use the same writing system. It also makes the writing system an object of study in its own right. The study of writing systems is a major part of scholarly work on ancient non-alphabetic texts, in particular.

Representing a Writing System

An item hierarchy in the **Writing systems** category represents a writing system whereas an individual item within such a hierarchy represents a “script unit,” that is, a distinct graphic sign within the writing system. Item hierarchies in this category are “subordination hierarchies,” in which items are grouped under subordinate headings (see Chapter 4). This allows the construction of a hierarchy of script units organized into named groups representing classes of signs in a given writing system.

A script unit is an ideal entity that can be manifested in more than one graphic form or “allograph.” For example, A and a are variant forms of the letter “a.” Each allograph of a script unit is represented in OCHRE by a font glyph (indicated by a font name and character code) or by an image. Epigraphic units of texts in the **Texts** category can be linked to script units in the **Writing systems** category in a way that specifies which allograph of a script unit is manifested at a particular place in a text.

In a logographic or syllabic writing system, quite different phonetic readings may be possible for a given script unit. For this reason, OCHRE allows multiple readings to be associated with each script unit, just as it allows multiple allographs. Discourse units of texts in the **Texts** category can be linked to script units in the **Writing systems** category in a way that specifies which reading of a script unit is to be used at a particular place in a text.

Note, however, that a script unit’s readings (phonetic variants) are independent of its allographs (graphic variants). In general, all graphic forms of a script unit can be read phonetically in all of the possible ways that the script unit can be read. These independent graphic and phonetic dimensions of what is still recognizably the same script unit justify OCHRE’s treatment of script units as ideal entities capable of diverse modes of graphic and phonetic expression.

In some projects, the complexities of multiple allographs and multiple readings will need to be represented, not for a complex logographic writing system, but for a much simpler alphabetic writing system, depending on the research goals of the project. To take a familiar example, a character (script unit) in the Latin alphabet may be manifested in quite different graphic forms (allographs) depending on the inscriptive tradition and the medium of inscription. In modern handwriting, the “printed” form and the cursive form of an alphabetic character may be quite different and may also vary from writer to writer. Moreover, the same Latin character can have quite different phonetic readings when used for different languages (e.g., Turkish “c” corresponds to English “j” as in “jar”). Indeed, in English orthography the phonetic reading often depends on where the

character is used (e.g., “c” in “cat” versus “nice”). For many texts written in relatively simple alphabetic writing systems it will not be necessary to represent the alphabetic characters as script units in OCHRE. Epigraphic units in the **Texts** category are not required to be linked to script units in the **Writing systems** category; this is optional. However, certain kinds of textual study call for the explicit modeling of the writing system as part of the digital representation of a corpus of texts. For example, a researcher might wish to search for a particular allograph in relation to other properties of the texts in which it appears in order to reconstruct the cultural and temporal connections among the scribes responsible for the texts.

A few writing systems (e.g., Mesopotamian cuneiform and Middle Egyptian hieroglyphs) are predefined within the “OCHRE” project, which is the parent of all other projects. Other projects are free to link their texts to the script units predefined in the **Writing systems** category of the “OCHRE” project by selecting this category in the **Linked Items** tab of the reference pane when manually linking items (see Chapter 6) or by selecting this category when automatically importing texts (see the section in this chapter entitled “Importing Existing Digitized Texts”).

Creating a Hierarchy of Script Units

When you select the **Writing systems** category in the navigation pane and then click the ↲ “insert below” button in the main toolbar you will be given a choice to create a hierarchy or an individual item. When you select a hierarchy within the **Writing systems** category and click the ↲ “insert below” button you will be given a choice to create a heading or an individual item that represents a script unit. A heading may contain script-unit items or subordinate headings. This permits the construction of a hierarchy of script units organized into named groups.

For example, in Sir Alan Gardiner’s *Egyptian Grammar* the hundreds of Middle Egyptian hieroglyphs are organized in a sign-list with sections entitled “Man and his Occupations,” “Woman and her Occupations,” “Anthropomorphic Deities,” “Parts of the Human

Body,” “Mammals,” “Parts of Mammals,” and so on, for a total of twenty-five sign categories, plus an extra category of unclassified signs.

Sometimes a grapheme in a text will contain more than one graphic sign, each of which might, in another context, constitute an entire grapheme by itself. For example, in the ancient Mesopotamian writing system, the cuneiform sign , which when written by itself represents the Sumerian word for “god,” can be inscribed inside the sign , which means “house,” in order to produce a new compound sign that means “mother.”

Note that the combination of graphic signs into a compound sign is not represented by means of a hierarchy of script units in the **Writing systems** category because such hierarchies are not strict containment hierarchies. Instead, there will normally be a separate script unit for the compound sign. In the example given above, the Sumerian sign for “mother” would have its own script unit, distinct from the signs for “god” and “house.” In cases where it is desirable to represent the compounding of signs more explicitly, this can be done by means of a containment hierarchy of epigraphic units in the **Texts** category, in which the compound sign is represented by an epigraphic unit that has child units, each of which is linked to a script unit that represents one of the graphic signs that make up the compound sign (see the section below on “Describing Epigraphic Units”).

If a script-unit item is selected in the navigation pane, the item pane will display data-entry fields for its **Name** and **Description**, as well as tabs for its **Properties**, **Links**, and **Notes**. These pertain to the script unit as an ideal entity and not to a particular allograph or reading of the script unit. The **Name** field should contain the name or code normally used by scholars to refer to the script unit; for example, the Middle Egyptian hieroglyph  has the name “C 3,” which is the code assigned to it in the sign-list published by Sir Alan Gardiner in his *Egyptian Grammar* (it is the third sign in Section C of his sign-list, which contains signs for anthropomorphic deities such as the god with the head of an ibis, Thoth, who is depicted in sign C 3). The naming of script units is usually straightforward for alphabetic scripts; for example, the script unit for the Greek letter Σ would have the name “Sigma.” Confusion between the script unit’s name and its allographs is possible in the case of the Latin alphabet (i.e., in projects that use a Western language as their default language). For example, the script unit for the letter A in the Latin alphabet would have the name “A” (or perhaps “Letter A”) but it should be understood that the allographs of this script unit include the graphic forms A and a (among others).

As elsewhere in OCHRE, the **Name** may contain alternative names (aliases) in the project’s default language, and both the **Name** and **Description** fields may contain translations into languages other than the project’s default language (see the section on “Item Names and Aliases” in Chapter 4).

Entering the Allographs and Readings of a Script Unit

The various graphic forms taken by the script unit are entered in the **Allographs** tab that appears in the item pane when a script-unit item is selected in the navigation pane. To enter an allograph, click the  “create new” button on the right edge of the tab. This creates a numbered tab with data fields and subtabs for **Properties**, **Links**, and **Notes** that are specific to the allograph. An allograph tab can be removed by clicking the  “delete” button. If there are multiple allograph tabs, they can be reordered by clicking the  “move left”

and → “move right” buttons. The tabs will be automatically renumbered as they are moved left and right. Allograph “1” is assumed to be the most common or standard graphic form of the script unit.

Each allograph tab contains data-entry fields for entering a **Character code** and **Font**. Together, these identify a particular font glyph that depicts the allograph. This glyph is displayed in the upper right-hand corner of the allograph tab. Unicode character codes should be used wherever possible and the font should be a Unicode font, although a non-Unicode character code and font may be used if necessary. The character code can be entered either as an integer or in hexadecimal form (for Unicode hexadecimal character codes, see the charts published at <http://www.unicode.org/charts>).

The screenshot shows the 'Allographs' tab selected (indicated by a green checkmark). The tab bar also includes 'Readings', 'Properties', 'Links', and 'Notes' (also with a green checkmark). Below the tabs, the number '1' is displayed. Under the heading 'Font glyph', there are two input fields: 'Character code, integer' containing '100' and 'Character code, hexadecimal' containing '64'. To the right of these fields is a small square preview window showing a stylized hand-like font character. Below the preview window is another input field labeled 'Font' containing 'CleoAfc'.

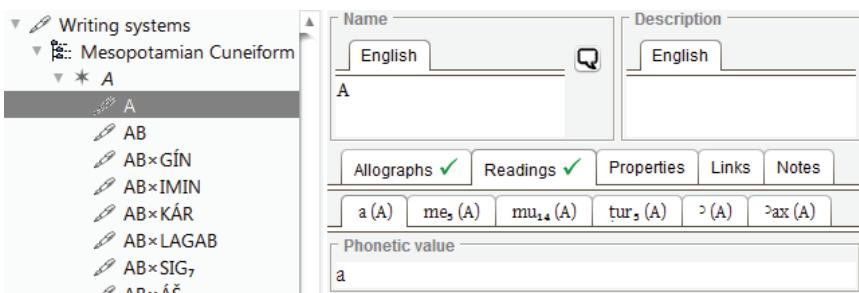
Epigraphic-unit items contain links to script unit allographs (see the section below on “Describing Epigraphic Units”). Thus, the font glyphs of the allographs will be used to display script units wherever they appear in graphemic transcriptions (sign-by-sign transliterations) of epigraphic units.

In addition to a font glyph, or instead of it, an image that depicts the allograph (e.g., a drawing or photograph) can be entered in the **Allograph image** link field. The image must first be represented by an image resource item and then selected as the link target in the **Linked Items** tab of the reference pane, after which you can click the \oplus “add link” button to the left of the field in order to enter the image link (see Chapter 6).

The various possible phonetic readings of a script unit are entered in the **Readings** tab. A script unit's readings (phonetic variants) are independent of its allographs (graphic variants), as explained above. Different readings of the same graphic sign occur frequently in logo-graphic, syllabic, and logosyllabic writing systems. Even in simple alphabetic writing systems an alphabetic character can have more than one phonetic reading (e.g., “c” in English “cat” and “nice”).

To enter a script unit reading, click the  “create new” button on the right edge of the tab. This creates a numbered tab with data fields and subtabs for **Properties**, **Links**, and **Notes** that are specific to the reading. A reading tab can be removed by clicking the  “delete” button. If there are multiple reading tabs, they can be reordered by clicking the  “move left” and  “move right” buttons. The tabs will be automatically renumbered as they are moved left and right. Reading “1” is assumed to be the most common phonetic reading.

Each reading tab contains a data-entry field entitled **Phonetic value** in which a phonetic transcription of one possible reading of the script unit is entered, usually in the form of Latin characters with diacritics or characters from the International Phonetic Alphabet (the project’s phonetic transcription font is used to display the phonetic value). The phonetic transcription of the script unit reading appears in transcriptions of texts in which the text’s discourse units are linked to epigraphic units that are linked in turn to script-unit items.



Representing a Text as Physical Signs and as Meaningful Discourse

When you select the **Texts** category in the navigation pane and then click the  “insert below” button in the main toolbar, you will be

given a choice to create a catalogue of texts or an individual text item. Unlike other item categories, there is no option to create a named hierarchy of text items, but you can create a named catalogue which is a “flat” list of texts rather than a hierarchy. There are containment hierarchies *within* each text item that represent the epigraphic and grammatical structure of the text, as explained below. To create an individual text item within a catalogue, select the catalogue in the **Texts** category and click the ↗ “insert below” button.

If a text item is selected in the navigation pane there will be five data-entry tabs at the top of the item pane entitled **Contexts**, **Creators**, **Editions**, **Bibliography**, and **Copyright**. The **Contexts** tab works in the usual way to show all of the hierarchical contexts in which the text item appears (see Chapter 4). Each context is displayed as a slash-separated “path” that shows the item’s location within a particular hierarchy. The first part of each path indicates the project in which the item is located, recognizing that items can be copied from other projects (the current project is indicated not by name but simply by an asterisk). Clicking the **Go to** button to the right of the **Contexts** area will scroll to the hierarchical location of the selected path in the navigation pane. Clicking the **Refresh** button will refresh the list of context paths in case the list is empty or out of date.

In the **Creators** tab is a field in which you can enter one or more links to items in the **Persons & organizations** category. Each such link indicates a person or group who constructed the text in OCHRE on the basis of previously published editions of the text or by making a new edition. To the right of this field is a data-entry field for entering the creation date (click the button to the left of this field to insert the current date; the button to the right displays a calendar from which you can select a date).

In the **Editions** tab is a field in which you can enter one or more links to items in the **Persons & organizations** category. Each such link indicates an editor or group of editors responsible for a particular edition of the text. If there are multiple editions of the text, each edition will be represented by a separate link in the **Editions** tab. In the **Bibliography** tab you can enter one or more links to items in the

Bibliographies category. These represent bibliographic references that pertain to the text represented by the currently selected text item (see Chapter 10). In the **Copyright** tab you can enter a copyright notice for the text. This notice is optional. It pertains to the creative work done by the scholars who have created the text in OCHRE. It does not pertain to previously published editions of the text.

Below these five tabs are tabs entitled **Information**, **Dictionary References**, and **Import**. The **Import** tab is discussed below in the section on “Importing Existing Digitized Texts.” In the **Dictionary References** tab you can generate a list of all the references to the current text within dictionary entries that have been entered via the **Dictionaries** category (see Chapter 12).

The screenshot shows the OCHRE software interface with the 'Information' tab selected. At the top, there are tabs for 'Text' and 'View', followed by 'Context(s)', 'Creators', 'Editions', 'Bibliography', and 'Copyright'. Below these are buttons for adding ('+') and removing ('-') items from a list, which currently contains 'Pardee, Dennis'. The 'Information' tab has several input fields: 'Name' (English, 'Inscription of KTMW'), 'Abbreviation' (English, 'KTMW'), 'Type' (alphabetic), 'Language' (Aramaic). There are also sections for 'Preferences' (checkbox for 'Keep this item private') and 'Description' (English, 'Aramaic inscription on well-preserved basalt stele'). At the bottom, there are tabs for 'Properties', 'Links' (with a green checkmark), 'Notes', and 'Events'. The 'Links' tab is active. On the right, there is a sidebar titled 'Other links (4)' containing links to 'BASOR 356 2009; Pardee', 'R08-13', 'R08-13 drawing', and 'KTMW Stele, handcopy'. Navigation icons for periods and other tabs are also visible.

The **Information** tab contains fields for entering the text’s **Name** (click the “translation” button to enter an alias in the project’s

default language or a name in another language) and its **Abbreviation**, **Type**, **Language**, and **Description**. A **Preferences** area contains a checkbox to **Keep this item private**, that is, to keep it invisible to anonymous users who are not logged into the project. At the bottom of the **Information** tab are subtabs for entering the text item's **Properties**, **Links**, **Notes**, and **Events** (see Chapters 5, 6, and 7).

Epigraphic Hierarchies and Discourse Hierarchies

When you select a text item in the navigation pane and then click the ↪ “insert below” button in the main toolbar you will be given a choice to create an epigraphic hierarchy or a discourse hierarchy. An epigraphic hierarchy represents the physical structure of a text in terms of its division into epigraphic units at various levels of analysis. For example, the editor of a cuneiform text inscribed on a clay tablet or of a medieval text written on parchment pages in a codex may want to distinguish, in descending order of analysis, the sides or pages of the text, columns within each side, lines within each column, and individual signs (graphemes) within each line. In contrast, a discourse hierarchy represents the meaning of a text in terms of its division into discourse units at various levels of analysis. For example, an editor might distinguish sentences within the text, clauses within each sentence, phrases within each clause, words within each phrase, and morphemes within each word. The epigraphic structure of a text is conceptually distinct from its discourse structure.

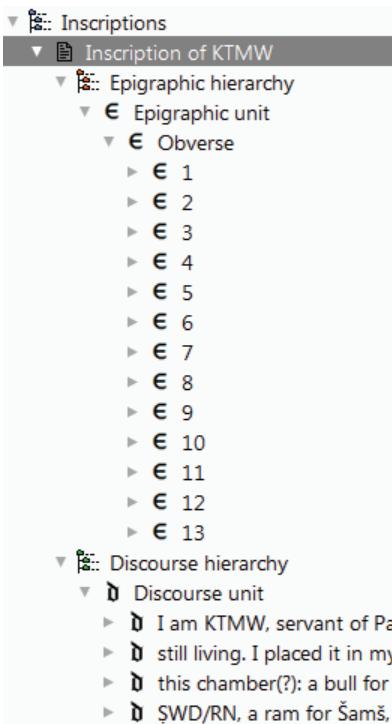
Note, however, that OCHRE does not prescribe the number of levels of analysis or the nature and scope of the epigraphic and discourse units at any level in the hierarchy. These are determined by the text's editor. Some texts will be broken down only as far as the line or sentence level; others will be broken down to their constituent graphemes and morphemes.

Each epigraphic unit and each discourse unit is represented by a distinct database item that has its own name, description, properties, links, and notes. Links are used to connect epigraphic units to the corresponding discourse units that represent the meaning of the physical signs represented by the epigraphic units. Links can also be

made between an epigraphic unit and script units in the **Writing systems** category, if the writing system used in the text is represented explicitly in the database.

The manual construction of epigraphic and discourse hierarchies can be quite tedious so there are tools for the automated construction of these hierarchies and of the links between them. Automated entry of digitized texts pasted in from other sources makes use of the white space and punctuation in the textual content to parse out the epigraphic units and discourse units and link them to one another (see the section below on “Importing Existing Digitized Texts”).

Thus, a “text” in OCHRE typically consists of an epigraphic hierarchy, a discourse hierarchy, and cross-cutting links between the constituent items in these two hierarchies. The sign-by-sign graphemic transcriptions of the text pertain to its epigraphic units. Word-by-word phonemic transcriptions of the text and translations of the text into other languages pertain to its discourse units. In the case of an undeciphered text, there is no discourse hierarchy because the meaning of the signs is not known.



Divergent Editions of the Same Text

A given text may be understood differently by different scholars. There may be disagreements about a text’s epigraphic units or its discourse units, or both. For example, scholars may disagree about which sign was intended by the scribe in cases where the sign is damaged or its shape is ambiguous. Likewise, scholars may disagree about the meaning of a sign or of a sequence of signs, even if they

agree on which signs were intended. For this reason, you can specify the editor (or editors) of any epigraphic unit or discourse unit within a text and you can do so at any level within an epigraphic or discourse hierarchy. This is done via the **Editors** field in the **Information** tab of an epigraphic-unit item or discourse-unit item (see below). Subordinate units in an epigraphic or discourse hierarchy are assumed to belong to the same edition as the closest parent unit in the hierarchy for which an editor has been specified. If no editor has been specified for any higher-level unit in the hierarchy, all of the text's epigraphic and discourse units will belong to all of the editions specified in the **Editions** tab of the text item from the top of the epigraphic or discourse hierarchy down to the first unit in the hierarchy for which an editor has been specified.

Thus, wherever an editor diverges from other editions in his or her analysis of a text you can indicate the divergent interpretation simply by inserting an additional epigraphic or discourse unit at the appropriate level and specifying the relevant editor in that unit's **Editors** field. Users will be able to choose whose edition to view whenever the text is displayed. Only epigraphic and discourse units associated with a given editor, either directly or by inheritance from a higher level, will be displayed in that editor's edition of the text.

Column 1	Column 1
(1) [sp] ^r l . bnš . ml ^k l	(1) siprü . bunuši . malki
(2) [dt/.] b ^l . riš	(2) dūti/dū bi . ra ^š i
(3) [y/m] ⁿ ḥm	(3) yanhamu
(4) ilb ^l	(4) ʔiliba ^l lu
(5) c ^b dyr ^r ḥl	(5) cabdiyarihu
(6) ttph	(6) ttph

Describing Epigraphic Units

After you select an epigraphic hierarchy in the **Texts** category and click the ↗ “insert below” button in the main toolbar, a new epigraphic-unit item will be created. When an epigraphic-unit item is

selected in the navigation pane, the item pane will contain three overlapping subtabs named **Information**, **Contexts**, and **Import**.

The **Import** tab is discussed below in the section on “Importing Existing Digitized Texts.” The **Information** tab contains optional data-entry fields for the epigraphic unit’s **Editors**, **Name**, and **Abbreviation**. There is also a drop-down **Type** field to indicate the type of epigraphic unit, namely, **side**, **region**, **column**, **line**, **sign**, or **separator**.

The **Contexts** tab works in the usual way to show all of the hierarchical contexts in which the epigraphic-unit item appears (see Chapter 4). Each context is displayed as a slash-separated “path” that shows the item’s location within a particular hierarchy. The first part of each path indicates the project in which the item is located, recognizing that items can be copied from other projects. Clicking the **Go to** button to the right of the **Contexts** area will scroll to the hierarchical location of the selected path in the navigation pane. Clicking the **Refresh** button will refresh the list of context paths.

Below these tabs are four overlapping tabs entitled **Contents**, **Properties**, **Links**, and **Notes**. The latter three function in the usual fashion, allowing you to enter properties, links, and notes for the epigraphic unit (see Chapters 5 and 6). The **Contents** tab has data-entry areas where you can specify in great detail the signs that make up the epigraphic unit and their physical condition and placement:

- The **Signs** area is for entering the graphic signs that make up the epigraphic unit (there may be only one sign if the unit lies at the bottom of the text’s epigraphic hierarchy).
- The **Damage, Erasure, and Gaps** area indicates whether the signs have suffered accidental damage or intentional erasure.
- In the **Partial Damage Details** area you can describe more precisely the damage suffered by the epigraphic unit, specifying which part of the epigraphic unit has been damaged.
- In the **Emendation** area you can indicate whether the epigraphic unit contains a scribal error, omission, or duplication, and if so, whether there was a scribal correction or editorial emendation.

- In the **Original Signs** area you can indicate the original signs inscribed before the epigraphic unit was changed by a scribal erasure or correction, or by an editorial emendation.
- In the **Sign Placement** area you can indicate the physical placement of the sign or signs that make up the epigraphic unit (i.e., above the line, below the line, etc.).
- In the **Separator** area you can specify the length and style of the line to be used to display an epigraphic unit whose **Type** has been defined as **separator**.
- In the **Other** area you can enter other information about the epigraphic unit (i.e., that it is unusual, that its identification is uncertain, that it has been collated by the editor via examination of the original document, etc.).

The **Signs** data-entry area contains a **Script units** link field and a display field that shows the actual signs. In the **Script units** field you can enter one or more links to script-unit items in the **Writing systems** category (a cuneiform example is shown below). Script-unit links should be entered only at the lowest level of the epigraphic hierarchy. If the epigraphic unit represents a single sign at the bottom of a detailed hierarchy it will normally have only one script-unit link.



To add a script-unit link, select the desired script-unit item as the link target in the **Linked Items** tab of the reference pane (see Chapter 6) and click the **0+** “add link” button to the left of the field. Alternatively, if you are viewing the hierarchy of script units in the **Writing systems** category in the navigation pane, you can select the required script unit in the hierarchy, use the **Edit** option on the main menu to **Copy Item**, and then use the **Clip** “paste” button in the **Script units** link field to insert the copied script unit into the list of links.

To remove a script-unit link, click on it to select it in the list of links and then click the  “remove link” button.

Instead of entering script-unit links manually you can enter them automatically on the basis of a character string that has been typed or pasted into the field located to the right of the **Script units** link field. In other words, you can leave the **Script units** link field empty and directly enter a string of characters, perhaps with the aid of the keypad of special characters (see Chapter 5). These characters will be displayed using the **Graphemic font** specified by the project administrator in the **F**onts subtab of the project’s **P**references tab. After entering the characters, you can click the  “find script units” button in the **Script units** link field in order to find all of the script-unit items in the project’s **Writing systems** category that match the characters displayed in the field and automatically create links to these items.

For example, a cuneiform text can be typed in (or copied-and-pasted) in the form of a Latin-alphabet transliteration of the signs. The Latin characters could then be automatically linked to cuneiform script units in OCHRE by clicking the  “find script units” button. But note that the character string does not have to be alphabetic; for example, a text in a nonalphabetic cuneiform font could be used and linked automatically to OCHRE script units. The use of specialized nonalphabetic graphemic fonts is made easier by the fact that Egyptian hieroglyphs, Mesopotamian cuneiform, and other unusual writing systems are now part of the Unicode character-encoding standard. (See <http://www.unicode.org/charts/PDF/U12000.pdf> for cuneiform signs and <http://www.unicode.org/charts/PDF/U1300.pdf> for Egyptian hieroglyphs.)

The **I**mport tab provides another way to enter character-string transcriptions in the **C**ontents tabs of a text’s epigraphic units. All the epigraphic units and their character strings will be inserted automatically for an entire imported text (see the section below on “Importing Existing Digitized Texts”).

Regardless of how character strings are entered, you can click the  “find script units” button in the **Script units** link field to find all the script-unit items in the project’s **Writing systems** category that

match the character strings and to create links to these items. This provides a way to build script-unit links quickly from the character strings, with the advantage that the epigraphic unit becomes firmly integrated with an explicit representation of the writing system. This enables more flexible and powerful database queries; for example, to study the pattern of usage of allographs in a text corpus.

Note that manually entered script-unit links will be overridden by any character string entered in the transcription field. However, if no character string has been entered, but only script-unit links, the transcription field beside the **Script units** link field will display the font glyphs for the script-unit allographs.

Also keep in mind that sign-by-sign transcriptions (or links to the OCHRE script units from which such transcriptions are generated) are normally entered only at the lowest level of the epigraphic hierarchy. This is because an epigraphic hierarchy is a “containment hierarchy” in which each item can contain child items that are understood to be subdivisions of the parent item (see Chapter 4).

For example, at the top level of an epigraphic hierarchy there might be sibling items that represent the sides or pages of the text; each side would in turn contain child items that represent the columns of text in that side; each column would contain child items that represent the lines in the column; and each line would contain child items that represent the words in the line; and each word would contain child items that represent the signs used to write the word. The order in which epigraphic units are listed in a given branch of the epigraphic hierarchy indicates their sequential order in the text; for example, the physical order of the lines of a text is represented by the order of sibling epigraphic units of the **line** type. Whenever a portion of the text, or the text as a whole, is displayed in the OCHRE interface, it is constructed from its components by “bubbling up” from the lowest level in the epigraphic hierarchy to a higher level, concatenating the signs of the sibling items at each level and then rising to the next level to concatenate the lower-level components. If a transcription is entered at a higher level, it will override the transcription produced by concatenating epigraphic units at lower levels.

Indicating Damage, Erasure, Gaps, and Emendation

An epigraphic unit may be affected by accidental damage or intentional erasure. In some cases, an epigraphic unit will be created by an editor's emendation of a text, when a scribal error is corrected or a scribal omission is filled in. Brackets and other sigla are used in transcriptions to indicate damage, erasure, and emendation.

Note that these sigla are not part of the text; they are part of the scholarly apparatus that provides information about the physical condition of the text and the editorial improvements that have been made to it. Thus, unlike most other software for studying texts, OCHRE makes a clear distinction between the text itself and sigla for indicating damage, erasure, gaps, and emendation. These sigla are generated on-the-fly whenever a transcription of the epigraphic unit is displayed; they are not stored as part of the transcription.

In the remainder of this section, reference is made to the default epigraphic sigla used to indicate various kinds of damage, erasure, and emendation (e.g., half brackets, square brackets, angle brackets, curly braces, exclamation points, etc.). However, a project administrator can specify different font characters in the **Epigraphic Sigla** subtab of the **Preferences** tab in the project's item pane in order to customize the sigla for his or her project.

In the **Contents** tab of an epigraphic unit's item pane there is a data-entry area for **Damage, Erasure, and Gaps**. The options for indicating accidental damage are **none**, **partial**, and **illegible**. An epigraphic unit that is partially damaged but still legible will be surrounded by half-brackets whenever its transcription is displayed, whereas an illegible unit will be surrounded by full square brackets.

The options for intentional erasure are **none**, **partial**, and **full**. An epigraphic unit that is partially erased will be surrounded by curly braces, whereas an epigraphic unit that has been fully erased will be surrounded by angle brackets. You can specify an estimated number of missing signs and also the **Missing sign character**, that is, the font character to be used to represent missing signs (e.g., "x"). Similarly, you can specify an estimated number of missing lines and also the **Missing line character** to be used to represent missing lines.

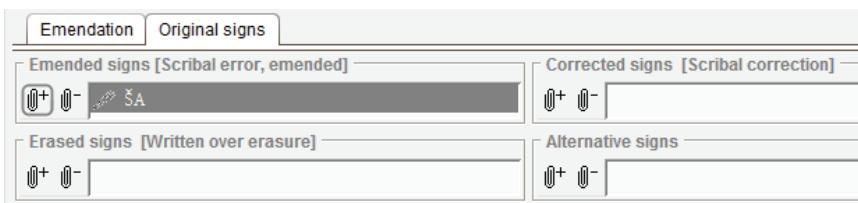
In the case of partial damage or erasure, you can enter **Partial Damage Details** in a separate area in order to describe more precisely the damage suffered by the epigraphic unit. There are options to specify whether the leading or trailing part of the epigraphic unit is partially or wholly illegible and a “slider” to indicate the position within the epigraphic unit at which to stop applying the “leading” damage indicator and begin applying the “trailing” damage indicator.

The data-entry area for **Emendation** should be used if the epigraphic unit contains a scribal error, omission, or correction. Several options are available:

- **Scribal error, emended**, i.e., the epigraphic unit contains the editor’s correction of a scribal error; this is indicated by adding the siglum (!) after the epigraphic unit’s transcription.
- **Scribal error, unemended**, i.e., the epigraphic unit is a scribal error that has not been corrected by the editor; its transcription will be surrounded by angle brackets.
- **Scribal omission**, i.e., the epigraphic unit is supplied by the editor to compensate for an accidental scribal omission; its transcription will be surrounded by double angle brackets.
- **Scribal duplication**, i.e., the epigraphic unit is an accidental scribal duplication; its transcription will be surrounded by ordinary parentheses.

- **Scribal correction**, i.e., the epigraphic unit is a scribal correction of the scribe's own error; its transcription will be surrounded by percentage signs.
- **Restored from duplicate**, i.e., the epigraphic unit contains the editor's restoration of missing signs based on a better-preserved duplicate of the current text; its transcription will be surrounded by double parentheses.

The data-entry area for **Original Signs** contains script-unit link fields in which you can indicate, in the case of scribal erasure or correction, the original sign or signs (if they can be reconstructed), or, in the case of a scribal error, the original sign or signs that have been replaced by an editorial emendation. There are three fields in which you can enter links to script-unit items in the **Writing systems** category to indicate **Emended signs**, **Corrected signs**, and **Erased signs**.



To add a script-unit link in any of these fields, select the desired script-unit item as the link target in the **Linked Items** tab of the reference pane (see Chapter 6) and click the $\Omega+$ “add link” button to the left of the field. To remove a link, click on it to select it in the list of links and then click the $\Omega-$ “remove link” button.

Indicating Sign Placement and Other Characteristics

The data-entry area for **Sign Placement** at the bottom of the **Contents** tab pertains primarily to epigraphic units that represent individual signs. This area contains the following sign-placement options: **Normal**, **Above line**, **Below line**, **Inside previous sign**, **Ligature with previous sign**, and **Over erasure**. “Normal” placement is assumed to be in line with neighboring signs. The sigla that will be used to indicate sign placement in transcriptions of the epigraphic unit can be

customized in the **Epigraphic Sigla** subtab of the **Preferences** tab in the project's item pane. The default sigla are as follows: ^ for placement above the line, || for placement below the line, × for placement inside the previous sign, + for a ligature with the previous sign, and ° for placement over an erasure.



The **Other** data-entry area contains checkboxes for indicating that the epigraphic unit has an **Uncertain identification** (default siglum ?), has an **Unusual form** (siglum %), is the result of a **Seal impression** (→...←), is a **Collated sign** (*), and/or is an **Uninterpreted sign** (/...*l*). If the identification of the epigraphic unit's signs is uncertain, you can use the **Alternative signs** field to enter links to script units that represent signs that are possible alternatives to the signs entered in the **Sigla** field at the top of the **Contents** tab.



Describing Discourse Units

When you select a discourse hierarchy in the **Texts** category and click the ↗ “insert below” button in the main toolbar, a new discourse-unit item will be created. When a discourse-unit item is selected in the navigation pane, the **Discourse Unit** tab in the item pane contains three overlapping subtabs at the top of the pane named **Information**, **Contexts**, and **Import**.

The **Import** tab is discussed below in the section on “Importing Existing Digitized Texts.” The **Information** tab contains optional data-entry fields for the discourse unit's **Editors**, **Name**, and **Abbreviation**. There is also a drop-down **Type** field to indicate the type of discourse unit, namely, **paragraph**, **sentence**, **clause**, **phrase**, **word**,

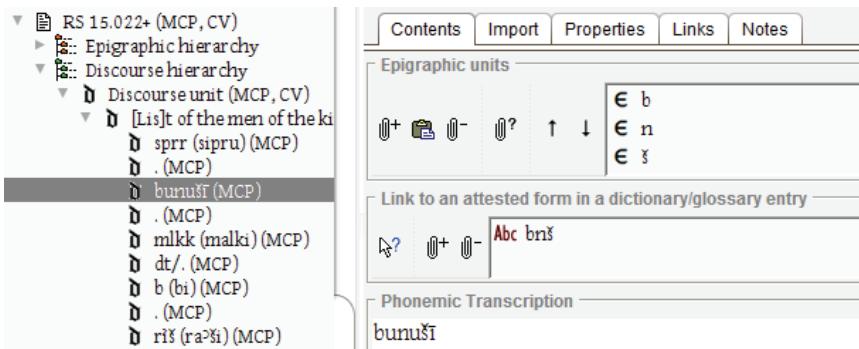
morpheme, number, logogram, determinative, or heading. The **Contexts** tab works in the usual way to show all of the hierarchical contexts in which the discourse-unit item appears (see Chapter 4). Each context is displayed as a slash-separated “path” that shows the item’s location within a particular hierarchy. The first part of each path indicates the project in which the item is located, recognizing that items can be copied from other projects. Clicking the **Go to** button to the right of the **Contexts** area will scroll to the hierarchical location of the selected path in the navigation pane. Clicking the **Refresh** button will refresh the list of context paths in case the list is empty or out of date.

Below these tabs are four overlapping tabs entitled **Contents**, **Properties**, **Links**, and **Notes**. The latter three function in the usual fashion, allowing you to enter properties, links, and notes for the discourse unit (see Chapters 5 and 6). Within the **Contents** tab is an **Epigraphic units** field for entering links to the epigraphic-unit items to which the discourse unit corresponds. For example, a word in a text can be represented by a discourse unit that is linked to several epigraphic units, each of which represents one of the signs that make up the word. To add an epigraphic-unit link, select the desired epigraphic-unit item as the link target in the **Linked Items** tab of the reference pane (see Chapter 6) and click the  “add link” button to the left of the field. Alternatively, if you are viewing the hierarchy of epigraphic units in the **Texts** category in the navigation pane, you can select the required epigraphic unit in the hierarchy, use the **Edit** option on the main menu to **Copy Item**, and then use the  “paste” button in the **Epigraphic units** link field to insert the copied epigraphic unit into the list of links. To remove an epigraphic-unit link, click on it to select it in the list of links and then click the  “remove link” button.

You can indicate that you are uncertain about an epigraphic-unit link by using the  “add uncertain link” button instead of the  “add link” button. You can change the order of the links in the list by means of the  “move up” and  “move down” buttons.

Each discourse unit that represents a word can be linked to the project’s dictionary as an attested form of the word. To do this, select

the relevant attested form of a dictionary entry as the link target in the **Linked Items** tab of the reference pane (see Chapter 6) and click the **⊕+** “add link” button to the left of the field entitled **Link to an attested form in a dictionary/glossary entry**. Alternatively, you can use the **↗?** “find dictionary form” button to search for the attested form in the dictionary and make the link automatically.



The phonemic transcription will be displayed using the **Phonemic font** specified by the project administrator in the **Fonts** subtab of the project’s **Preferences** tab. This font is usually a Latin-based or International Phonetic Alphabet (IPA) font that has suitable diacritics for transcribing a text’s spoken sound in accordance with the scholarly conventions current for the text’s language and genre.

Finally, the **Contents** tab of the discourse-unit item pane has a **Translation** field in which you can enter a translation of the discourse unit into a modern language. The translation is displayed in the “document font” specified by the project administrator in the **Fonts** subtab of the project’s **Preferences** tab. Like any textual content, a translation can be entered with the aid of the text-formatting toolbar and the keypad of special characters (see the section on “Entering Textual Content in Notes and Elsewhere” in Chapter 5). You can also embed hyperlinks and properties within the textual content, allowing you to annotate the translation with reference to other database items (see the subsection in Chapter 5 on “Embedding Properties and Hyperlinks within Textual Content”). Note that, in addition to properties and hyperlinks that are embedded in the

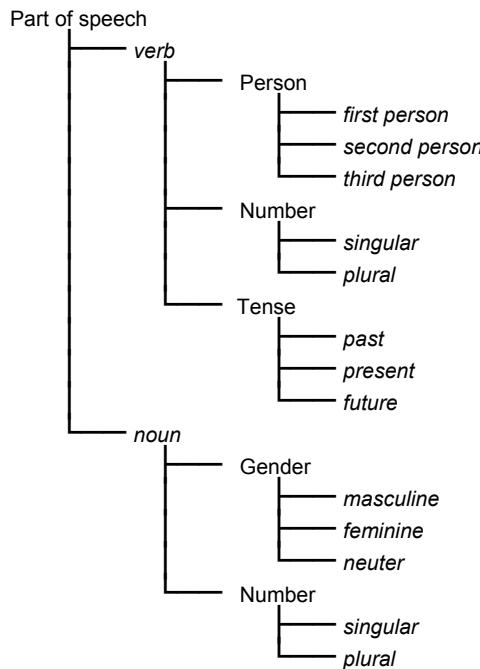
textual content of a translation, you can enter general links from the translation as a whole to other database items (e.g., to items in the **Bibliographies** category). Likewise, you can enter properties for the translation as a whole. These general links and properties are entered by means of the “add link,” “remove link,” and “assign properties” buttons on the right side of the **Translation** field, in conjunction with the **Properties** and **Links** subtabs located at the bottom of the **Linked Items** tab of the reference pane.

Translations will normally be entered at a relatively high level in the discourse hierarchy. A translation usually pertains to a complete sentence, paragraph, or even to the entire text, rather than to a word, phrase, or clause within a sentence. This is because the sentence structure of the modern language will often be quite different from that of the text being studied, making it difficult to enter translations for the smallest discourse units that can be concatenated to form useful translations at a higher level. Having said that, it may be useful to enter translations for discourse units that represent words, phrases, or clauses in addition to entering translations of higher-level units in order to ensure that an intelligible translation will be available for a given discourse unit, regardless of its size.

Grammatical Parsing of Discourse Units

Every text, and every epigraphic unit and discourse unit within a text, is represented by a separate database item that can have its own properties, links, and notes like any other OCHRE item. The properties of discourse units, in particular, can be used to represent the grammatical parsing of a text. To do this, the project administrator must first set up a taxonomic hierarchy of morphological properties in the **Taxonomy** category. A project’s taxonomic hierarchy specifies the way in which property names (variables) and values may be used to describe other items and the way in which particular properties (variable-value pairs) can be nested within other properties to represent relationships between specific properties and more general properties (see Chapter 5 on assigning item properties and Chapter 13 on constructing the taxonomic hierarchy).

Here is a partial taxonomic hierarchy of variables and values that could be used to assign morphological properties to a discourse unit that represents a word:



For some texts, it is useful to assign properties to discourse units in order to represent the parsing of each word in the text. In a similar fashion, syntactic properties can be assigned to higher-level discourse units that represent phrases, clauses, and sentences. Parsed texts can then be analyzed automatically and compared to one another by means of queries that search for particular grammatical properties or combinations of properties. Discourse units can be analyzed contextually in relation to other discourse units and in relation to epigraphic units and their properties. Moreover, textual analyses and comparisons can take account of the nongrammatical and nonepigraphic properties of texts such as authorship, language, and provenance.

Like other item properties in OCHRE, discourse properties are not prescribed by the database system. It is up to the project administrator to determine the taxonomic properties that will be available for

describing a text and it is up to the text’s editor to define the scope and nature of the discourse units that make up the text and to assign properties to particular discourse units.

Displaying Texts

The internal structure of a text is represented in a highly atomized fashion by means of item hierarchies in the **Texts** category. Transcriptions and translations of the entire text are then generated from the underlying epigraphic units and discourse units so they can be displayed in a readable fashion. You can display the entire text by selecting the text item in the navigation pane and opening its **View** tab in the item pane.

A text’s **View** tab does not contain a redundant copy of the text. Rather, it shows a dynamically generated view of the text based on its constituent epigraphic and discourse units. This has the advantage of enabling a “live” link between each portion of a text’s transcription or translation and the underlying unit’s properties, notes, and links to other items. For example, individual epigraphic and discourse units can be linked to hotspots in a photograph or facsimile drawing of the original text, thereby connecting each portion of the inscribed writing surface to a particular sign, word, or other unit.

Graphemic, Phonemic, and Hybrid Transcriptions

OCHRE supports three kinds of scholarly transcriptions of texts written in premodern scripts, namely, graphemic, phonemic, and hybrid (graphemic-and-phonemic) transcriptions. All three kinds can be displayed in a text’s **View** tab. Checkboxes at the bottom of this tab allow you to hide or show each kind of transcription. The font used to display each kind of transcription is specified by the project administrator in the **F**onts subtab of the project’s **P**references tab (see Chapter 17). The name used to label each kind of transcription can also be customized; for example, a project may prefer to call its graphemic or hybrid transcriptions simply “transliterations.”

A graphemic transcription is a sign-by-sign rendering of the visual appearance of the text in the form of a character string in which there

is a separate character for each meaningfully distinct graphic sign (grapheme) in the original text. For example, in Egyptology it is common to transcribe hieroglyphic texts by means of standardized hieroglyphic characters (one character per hieroglyph) that mimic the appearance of the original signs but are arranged left to right in horizontal lines, unlike the original texts, in which hieroglyphs may be written vertically or from right to left. This kind of graphemic transcription indicates the individual signs and their sequence, but in a simplified fashion that allows the texts to be more easily cited.

A phonemic transcription uses Latin characters with diacritics or International Phonetic Alphabet (IPA) characters to indicate the spoken sound of the text. For example, in Northwest Semitic philology it is a common practice to transcribe texts written in Ugaritic, Phoenician, Aramaic, Hebrew, and related alphabetic scripts into the Latin alphabet and to include in the transcription either the attested consonants alone or the consonants together with the reconstructed vowels. In many Northwest Semitic scripts only consonants are represented, although the vowels can often be reconstructed with a high degree of confidence. The reconstructed vowels, which are needed to indicate the spoken sound of the text in a phonemic transcription, are not included in a sign-by-sign graphemic transcription, which typically employs either Latin characters with diacritics or Hebrew font characters in such a way that each transcription character corresponds to one alphabetic sign visible in the original text.

Likewise, Egyptologists use the Latin alphabet for phonemic transcriptions that indicate the spoken sound of an ancient Egyptian text. However, phonemic transcriptions of Egyptian texts normally include only the consonants; the vowels are omitted, not because the transcriptions are graphemic, but because the vowels are unknown. By convention, vowels are pronounced as short “e” (IPA ε) when reading an Egyptian text, with a few exceptions. Egyptian phonemic transcriptions without vowels are thus superficially similar to Latin transcriptions of the consonants of Northwest Semitic texts, but the latter are graphemic transcriptions in which the vowels are omitted in order to provide a sign-by-sign representation of the actual physical

appearance of the text, whereas Egyptian transcriptions are essentially phonemic and do not directly indicate the original graphic signs.

Hybrid transcriptions that represent both graphemic and phonemic aspects of a text are common in cuneiform studies, where they are called “transliterations,” in contrast to purely phonemic transcriptions, which are called “normalizations.” The logosyllabic cuneiform writing system, which originated in southern Mesopotamia, was used to write several different languages over a period of more than two thousand years in widely separated geographical locales ranging from present-day Turkey to Iran. A cuneiform transcription (hybrid transcription) consists of Latin characters that indicate the individual signs which make up the text and, in the case of syllabic signs, the pronunciation (phonetic value) of the sign at that point in the text.

For example, the cuneiform sign  has the phonetic value “lum” so a hybrid transcription would indicate the existence of this sign and its pronunciation by writing the syllable in Latin characters as “lum.” The series of syllabic signs that make up a word is indicated by hyphens that separate one syllable from another. For example, the three signs    would be transliterated as the hyphen-separated syllables “a-wi-lum,” the Old Babylonian word for “man.” The normalized phonemic transcription of these syllabic signs, indicating vowel length but obscuring the individual signs, is written *awīlum*.

However, in the case of logographic cuneiform signs (logograms) that represent entire words rather than syllables, the signs are not transcribed phonetically but are indicated via their conventional sign names, which are often written in capital letters to show that they are not pronounced the way they are written. For example, URU is the conventional name of the cuneiform sign , which means “city.” The sign’s name is derived from the sound of the Sumerian word for “city,” *uru*, because it was the Sumerians who first wrote  as a logogram meaning “city.” But the same logogram was pronounced *ālum* in the Old Babylonian dialect of the Akkadian language that eventually displaced Sumerian as the spoken language of Mesopotamia. Thus, a hybrid transcription of an Akkadian text in which the logogram  is indicated by URU is not a fully phonemic

transcription because this sign name does not indicate how the sign was pronounced.

In spite of this, hybrid transcriptions are favored by cuneiformists because a hybrid transcription, unlike a fully phonemic transcription, indicates exactly which signs were written in the original text while providing sufficient information for an expert to determine the spoken sound of the text (i.e., an expert would know the phonetic reading of logograms indicated by sign names like URU). And cuneiform specialists do occasionally use fully phonemic transcriptions, which they call “normalizations.” Using the example given above, the logogram 𒀭 would not be rendered by the sign name URU but phonetically as *ālum* (or a variant with a different case ending) in a phonemic transcription of an Akkadian text.

Viewing a Text

As was noted above in the sections on “Describing Epigraphic Units” and “Describing Discourse Units,” graphemic transcriptions pertain to epigraphic units and phonemic transcriptions pertain to discourse units. In data-entry mode, an epigraphic unit’s graphemic transcription is displayed in the **Signs** area of its **Contents** tab and a discourse unit’s phonemic transcription is displayed in its **Contents** tab.

In addition to displaying a transcription of a particular epigraphic unit or discourse unit in its data-entry tab, you can display the transcriptions and translation of an entire text by selecting the text item in the navigation pane and opening its **View** tab in the item pane. A formatted view of the text will be shown in which the transcriptions of all of its epigraphic units are combined in a single graphemic transcription; the transcriptions of all of its discourse units are combined in a parallel phonemic transcription; the translations of all the discourse units are combined to display a translation of the entire text; and a hybrid transcription is shown, if applicable.

If there is more than one edition of the text, you can choose which one to display by means of a drop-down list at the top of a text’s **View** tab. This list shows the names of the editors or groups of

editors that were entered in the **Editions** tab of the text's item pane (see the section above on “Divergent Editions of the Same Text”).

Checkboxes at the bottom of the text's **View** tab allow you to individually hide or show in separate panes the text's **Hybrid view** (which many scholars may prefer to rename **Transliteration** in their project's **Preferences** tab), its **Graphemic view**, its **Phonemic view**, and its **Translation**. There are also checkboxes that let you hide or show the text's **Description** (its name, aliases, type, properties, and notes), its **Links** to other items, and the **Events** (if any) that affect it. If you hold down the **Shift** key and then click on any of these view panes, the orientation will change from vertical to horizontal, or vice versa.

The screenshot shows the Text tab interface with four view panes:

- RS 15.022+; Transliteration**: Shows the transcription of the text. It includes a column header "recto" and two rows of data:

(1)	[sp]r ¹ l ¹	b ¹ n ¹ s ¹	m ¹ r ¹ k ¹
(2)	[dt./.]	r ¹ b ¹	r ¹ s ¹
- RS 15.022+; Graphemic view**: Shows the graphemic representation of the text. It includes a column header "recto" and two rows of data:

(1)	[s] [p] [r ¹]	b n s	m l r ¹ k ¹
(2)	[d] [t./.]	r ¹ b ¹	r ¹ s ¹
- RS 15.022+; Phonemic view**: Shows the phonemic transcription of the text. It includes a column header "i 1-2" and two rows of data:

i 1-2	sipru	bunu ¹	malki d ¹ t ¹ /d ¹ u
b ¹	ra ² š ¹	¹⁻²⁻²¹ yanhamu	iliba ¹ lu
- RS 15.022+; Translation**: Shows the English translation of the text. It includes a column header "i 1-2" and two rows of data:

i 1-2	Lis ¹ t of the men of the king who are from Ra ² šu[...]	¹⁻²⁻²¹ [Y]a[n] ¹ hamu (or [Mu]n ¹ h ¹ himu)	>iliba ¹ lu,
Abdiyāru ¹ h ¹ u	TTPH, >Artēnu, Yabn ¹ ilu[...], Bur(a)qanu[...], ADR[...], Kurwānu, ARKD ¹ N, >limuna, >Abu ¹ Kani,	Abdiyāru ¹ h ¹ u, Kiliyānu, Hagbānu, HTTN, >Abdimilku, Y[- - -]R ¹ , ¹⁻²⁻²⁴ [c]A ¹ b ¹ d ¹ adat ¹ tu ¹ , Binu ¹ -KT,	

At the bottom of the interface, there are checkboxes for **Description**, **Transliteration**, **Graphemic view**, **Phonemic view**, **Translation**, and **Links**. The **Transliteration** checkbox is checked.

In the top left corner of the **View** tab is a “print” button, which you can use to print information about the text and/or to print the text's transcriptions and translation. Clicking this button will print everything that is currently visible in the **View** tab, including the text's description, links, transcriptions, and translation. Instead of printing the contents of the **View** tab, you can export them to an external file by clicking on the “export as PDF” button or the “export as Excel” button. The checkboxes at the bottom of the **View** tab determine what is included in the printout or exported file.

Standard Text View

In a drop-down list at the top of a text's **View** tab you can choose one of the following view modes: **standard**, **parallel**, and **interlinear**. In the **standard** mode (shown above) the graphemic transcription, phonemic transcription, hybrid transcription, and modern-language translation are constructed by concatenating the text's constituent epigraphic and discourse units, as described above. Line numbers are automatically displayed at the appropriate locations within the transcriptions and translation for discourse units of type "sentence." They are based on the epigraphic structure indicated by epigraphic units of the "line" type. The links between epigraphic units, script units, and discourse units determine which signs and words belong to a particular line of the text. Keep in mind that line numbers, like editorial sigla that indicate damage or emendation, are not part of the text itself. They are generated as needed from the underlying database representation of the text rather than being stored in the actual textual content.

The hybrid transcription draws upon both the epigraphic and the discourse units and conforms to the transliteration conventions for cuneiform texts used in *The Assyrian Dictionary of the Oriental Institute of the University of Chicago*. Hyphens separate the signs in a word. Syllabic signs are indicated by their phonetic readings; however, the phonetic readings of logograms are not shown. Instead, the names of logograms are displayed in capital letters. The names of signs that function as semantic determinatives and thus have no phonetic realization are superscripted using lower case letters.

Outside cuneiform studies these conventions are not widely used so the hybrid transcription can be hidden, if necessary, by unchecking the **Hybrid view** checkbox at the bottom of the **View** tab. On the other hand, cuneiform specialists rarely use a purely graphemic transcription, so they can hide it by unchecking the **Graphemic view** checkbox.

Parallel Text View

In the **parallel** view, the graphemic transcription, phonemic transcription, hybrid transcription (perhaps renamed "transliteration"),

and modern-language translation—or as many of these as have been selected via the checkboxes at the bottom of the **View** tab—are displayed line-by-line in vertical panes that remain horizontally aligned as the user scrolls any of the panes. Clicking or dragging the scroll bar on one of the panes to move a transcription or translation up or down on the screen will cause all of the panes to scroll together, preserving the horizontal alignment of the corresponding lines across the various transcriptions. Note, however, that you can scroll a particular pane independently of the others by using your mouse wheel instead of clicking the scroll bar.

The screenshot shows the 'Text' tab selected in the top navigation bar. Below it are several icons: a magnifying glass, a file folder, 'Abc' (for phonetic transcription), a document icon, a green checkmark, a red X, and a double-headed arrow. The tabs for the three panes are 'Parallel' (selected), 'Prosser, Miller', and 'Display options...'. The left pane, 'RS 15.022+: Transliteration', contains the word 'recto' followed by two columns of numbered entries. The middle pane, 'RS 15.022+: Phonemic view', also contains 'recto' and two columns of numbered entries. The right pane, 'RS 15.022+: Translation', contains the same 'recto' entry followed by a single column of numbered entries. At the bottom of the interface are several checkboxes: 'Description' (unchecked), 'Transliteration' (checked), 'Graphemic view' (unchecked), 'Phonemic view' (checked), 'Translation' (checked), and 'Links' (unchecked).

Transliteration	Phonemic view	Translation
recto	recto	recto
Column 1	Column 1	Column 1
(1) [sp]ʃr ¹ , bn ² , ml ³ k ¹	(1) s̪ipru, bunuš ¹ , malki	[List] of the men of the king who
(2) d ⁴ t ⁵ i, r ⁶ b ⁷ , ri ⁸ s ⁹	(2) dūti, dū bi, raš ² i	are from Raš ⁸ u [...]:
(3) [y/m]ʃn ¹ hamu	(3) yanhamu	[Ya]ʃn ¹ hamu (or [Mu]ʃna ¹ himu),
(4) iłb ¹ c ²	(4) ɔiibac ¹ lu	ɔiibac ¹ lu, ɔabdiyar ¹ ju ¹ , TTPH,
(5) c ¹ b ² d ³ y ⁴ r ⁵ ʃ ⁶	(5) ɔabdiyar ¹ ju ¹	ɔArēnu, Yabni ² ilu[...], Bur(a)qānu
(6) t ⁷ p ⁸ h	(6) t ⁷ p ⁸ h	[...], ɔDR[...], Kurwānu, ɔRKDN,
(7) ārt ¹ n	(7) ɔartēnu	ɔlimuna, ɔabuškani, Yakūnu,
(8) ybni ¹ l ² u	(8) yabni ¹ ilu	Rašap ² abū, Kiliyānu, Hagbānu,
(9) brq ¹ n	(9) burqanu	Httn, ɔAbdimilku, Y[- -]ʃR ¹ ,
(10) ădr []	(10) ădr	
(11) krw ¹ n	(11) kurwānu	
(12) ārkdn	(12) ārkdn	
(13) ilmn	(13) ɔilimuna	
(14) ābškn	(14) ɔabuškani	
(15) ykn	(15) yakūnu	
(16) r ¹ šp ² b ³	(16) rašap ² abu	
(17) k ¹ hn	(17) kilwānu	

Interlinear Text View

The **interlinear** view applies only to hybrid transcriptions. When this option is selected the graphemic and phonemic transcriptions and translation will appear as they do in the **standard** view. However, within the hybrid transcription the lexical meaning and morphological description of each word will be displayed in smaller type on the line below, provided that this information has been entered in a dictionary entry for the word in the current project's **Dictionaries** category (see Chapter 12). If there is no dictionary entry for a given word the corresponding space in the line below will be left blank.

Interlinear ▾ Azzoni, Annalisa ▾

PFAT 001: Transliteration

Obverse

(01) mny ḫš'mh

Lex: Mny: "(personal name; Mannuya)" šm: "name"

Morph: +n.~prop.noun~PN~abs. +n.~m.~sg.~construct+pers.prn.~poss.~3rd~m.~sg.

(02) m⁷r d⁷ɔy

Lex:

Morph:

(03) k⁷p⁷ṛr⁷y⁷ṛ

Lex: kqry: "(gentilic)"

Morph: +n.~prop.noun~GN~Gentilic~abs.

The dictionary information for a selected word can be retrieved within any of the four text view modes and for any of the three kinds of transcription by clicking the “find word” toggle button at the top of the **View** tab. Clicking this button will change the cursor from to . Thereafter, clicking on a word in a transcription will cause a window to appear that contains the selected word’s dictionary entry, if it has one. When you are finished looking up words, click the toggle button again to restore the ordinary cursor.

Restored Text Option

In any of the text view modes described above you can use the “restored” option, which omits from the transcriptions and translation all epigraphic sigla that indicate damage, erasure, or emendation. This makes them more readable.

Highlighting Corresponding Units

Within any of the three text view modes you can display correspondences between epigraphic and discourse units across the various view panes by checking the **Highlight corresponding units** checkbox at the top of the **View** tab. Then, if you click on a word or sign in a transcription or translation to highlight it, the corresponding word or signs in the other panes will be highlighted simultaneously. This

allows you to see the connections between the epigraphic and discourse units that appear in the various transcriptions of the text.

Highlighting corresponding units can be done not just in a text's transcriptions and translation but also in a drawing or photograph of the text. On the right side of the text's **View** tab you will see a drop-down pick-list labeled **Select image** from which you can select an image to display from among the image resources that have been linked to the text.

If you check the **Select from annotated images** checkbox above this pick-list, your choices will be limited to images that have been annotated by linking the text's epigraphic units to shaded "hotspot" regions that demarcate signs in the text. Clicking on a hotspot in such an image will cause the corresponding epigraphic units and discourse units in the transcriptions and translation to be highlighted simultaneously. Likewise, clicking on a word or sign in one of the transcriptions will cause the corresponding hotspot to be highlighted in the image.

Synchronizing a text's image with its transcriptions and translation is possible only if the relevant parts of the image have previously been linked to epigraphic-unit items via hotspot links in the **Image Tools** tab of the image's item pane (see the section above in Chapter 9 on "Embedding Annotations and Hotspot Links within an Image"). In addition, the image resource item must be linked to the text item via the **Links** tab in the text's item pane (see the section on "Adding and Removing Links" in Chapter 6).

The screenshot shows the Text tab interface with three main panes:

- Transliteration:** Shows a list of words with their phonetic transcriptions. For example, 'recto' is listed with '(1) [sp]r̥t̥, bñs̥ ml̥k̥' and '(2) [dt̥/] bñs̥, riᬁ'.
- Phonemic view:** Shows a list of words with their phonetic transcriptions. For example, 'recto' is listed with '(3) [y/m]f̥n̥hm̥' and '(4) ilbñl̥'.
- Translation:** Shows a list of words with their meanings. For example, 'recto' is listed with '[List] of the men of the king who are from Ra²pu[...]'.

On the right side of the interface, there is a 'Select image' dropdown menu set to 'RS 15.022+ rto'. Below the dropdown is a preview area showing a photograph of the tablet fragment with specific signs highlighted in yellow and purple. There are also several checkboxes at the top right of the interface, with the third one, 'Select from annotated images', checked.

A text can be linked to any number of images that depict its physical appearance. The linked images or small “thumbnails” of them can be displayed by checking the **Show associated images** or **Show thumbnails** checkbox on the right side of the **View** tab.

Associating a Text with Its Physical Medium

A text, which is represented by a database item in the **Texts** category, is conceptually distinct from the artifact or surface on which it is inscribed. Thus, it may be useful in some cases to represent a text’s physical medium by means of a separate item in the **Locations & objects** category, especially when the spatial context and the non-epigraphic characteristics of the inscribed object are important for the research at hand. In other cases, the creation of a second database item to represent the inscription as a physical object is unnecessary because it can be represented adequately by an item in the **Texts** category. For example, if the text’s provenance is unknown or irrelevant there may be no reason to represent it by a spatial item in the **Locations & objects** category. By the same token, however, the text’s physical medium should have its own item if its spatial context and nonepigraphic characteristics are important.

The “Associated item” Relational Variable

The connection between a text and its physical medium is indicated by a special relational property of the spatial item in the **Locations & objects** category that represents the inscribed artifact or surface (see Chapter 6 on relational properties). There is a predefined relational variable called “Associated item” in the taxonomy of the OCHRE parent project. If this variable has been copied into the current project’s taxonomy by the project administrator (see Chapter 13) you can use it to establish a connection between an inscribed spatial item and a textual item.

To do so, select the relevant item in the **Locations & objects** category and insert the “Associated item” variable in the **Variable** column of the **Properties** tab. At that point, the corresponding **Value** field will require a link to a target item in the **Texts** category. To enter

the value, double-click on the **Value** field to display **⊕+ “add link”** and **⊖- “remove link”** buttons to the left of the field. After selecting the link target via the **Linked Items** tab of the reference pane, click the **⊕+** button to insert the target text item as the value of the “Associated item” variable.

By explicitly associating a text with its physical medium while treating them as two distinct database items you can describe and query them independently without losing the connection between them. The inscribed artifact or surface has physical properties, contextual relationships, and spatial coordinates like any other artifact. The relational property that associates it with a text item, and thus with an epigraphic hierarchy and a discourse hierarchy, makes it possible to perform queries that simultaneously take account of an inscription’s spatial context, temporal period, physical properties, linguistic structure, and meaningful content.

Variable	Value
Primary classification	Tablet
Museum Number	DO 3912 + 3934 + 3936
Publication	PRU 2: 24
Size	121 x 96 x 29
Joined with, directly	RS 15.048
Joined with, directly	RS 15.052
Associated item	RS 15.022+

An artifact or surface represented by a spatial item in the **Locations & objects** category may have more than one text inscribed on it. These texts may or may not be related to one another in content, writing system, or date. For example, medieval parchment manuscripts were often reused by scraping off the original text in order to write a completely different text on the same parchment, thus creating a “palimpsest.” Using special photographic techniques scholars can often reconstruct the erased text, whose vestiges remain in the parchment. Here, then, is a case in which two different texts occupy the same physical medium. In OCHRE, each text would be represented by a separate text item in the **Texts** category and you would use multiple “Associated item” relational variables to link the spatial item to all of the texts inscribed on it.

Images of a text are normally linked to the text item, not to the associated spatial item. This is done via the **Links** tab in the text’s item

pane, as noted above. The text's images do not need to be linked separately to the spatial item that represents the text's physical medium because the images of any "Associated item" will be retrieved whenever the spatial item is displayed.

Seals and Seal Impressions

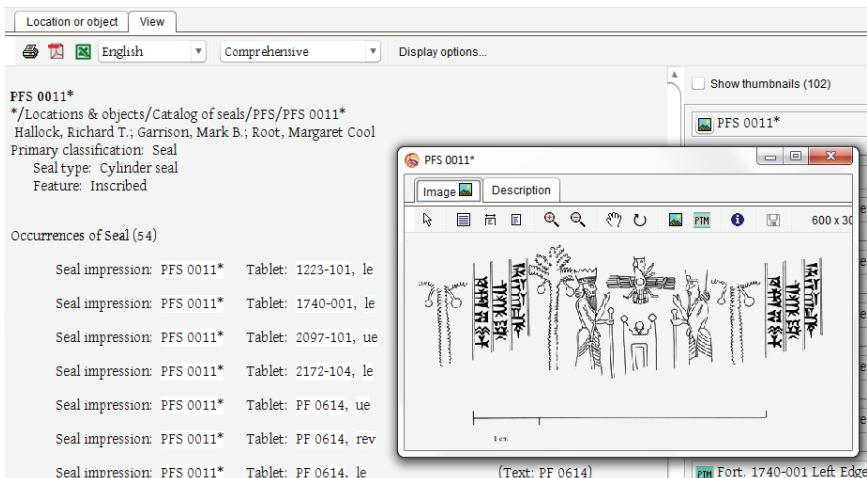
Some texts have seal impressions that may themselves contain texts—for example, the name or title of the person whose seal was used to make the impression. This is a common feature of cuneiform clay tablets, in which seal impressions were often made on the wet clay before it hardened, especially on letters and legal documents. In a similar fashion, seal impressions in wax are found in later periods on texts written in ink on paper or parchment. Modern legal documents make use of seal impressions that are embossed in the paper itself with a special stamping device.

The text contained in a seal impression is separate from the text that was sealed, so they are represented by two different text items even though they share the same physical medium. But there are two added complications here that require some explanation. First, the same seal can be used to create identical seal impressions on many different surfaces, which may or may not already be inscribed. Secondly, the seal itself may be purely pictorial and not inscribed with a text. The challenge is to capture all of the possible variations of seals, seal impressions, and texts by means of interlinked database items and to do so in a logical manner without redundancy.

A text that is incised or embossed on a seal (e.g., the name of the seal's owner) is a mirror image of the text in each of the seal impressions made by that seal. To avoid creating a redundant text item for each identical seal impression made with the same seal, the recommended practice is to create a spatial item in the **Locations & objects** category in order to represent the seal itself and to use the "Associated item" variable to associate the seal with a text item in the **Texts** category that represents the seal's inscription.

It will often be the case that the only evidence of the existence of a seal is the impression it was used to make; the seal itself may have

been lost or destroyed. Nonetheless, a seal should be represented by its own item in the **Locations & objects** category because the impression it has left behind provides information about its size and shape as a physical object and because the seal's own item is the most appropriate owner of the text item that represents the seal's inscription, as opposed to the spatial items that represent the seal's impressions on various surfaces. Thus, the connection between a seal impression and the text it contains is best represented indirectly, via the seal that made the impression.



Moreover, a seal impression as a physical entity should be distinguished from the larger artifact in which it is found. It is a good practice to represent the seal impression as a child of the larger artifact within a spatial hierarchy in the **Locations & objects** category. If the parent artifact is inscribed, its “Associated item” variable would point to the relevant text item. The seal impression (the child item) would point to the seal that made the impression by means of the special “Seal” relational variable. This predefined variable works in exactly the same way as the “Associated item” variable except that it points, not to an item in the **Texts** category, but to an item in the **Locations & objects** category. The seal item referenced as the value of this variable would in turn have an “Associated item” variable to point to the text on the seal itself.

For example, the parent item might be a Mesopotamian cuneiform tablet which has child items that represent the various surfaces of the tablet: obverse, reverse, upper edge, lower edge, left edge, and right edge. Any of these surfaces might in turn have one or more child items that represent seal impressions in that surface, keeping in mind that cuneiform tablets were often impressed repeatedly with the same seal. Each item that represents a seal impression would have a “Seal” relational variable to indicate the seal that made the impression. The seal item in turn would have an “Associated item” variable to indicate the text inscribed on the seal itself.

This arrangement allows for multiple duplicate seal impressions (impressed on the same artifact or on different artifacts) to be linked to the same seal, which in turn is linked to its text. This arrangement works equally well if the seal is purely pictorial and has no text or if the artifact impressed by the seal is itself uninscribed.

Using Queries to Analyze Texts in Their Physical Contexts

OCHRE’s querying capability is discussed in detail below in Chapter 14; however, it is worth mentioning here that queries can be constructed that use complex combinations of textual and contextual search criteria. Analytically powerful queries are made possible by OCHRE’s ability to represent the epigraphic and grammatical structure of a text in a highly atomized way, in as much detail as needed, and to distinguish carefully between texts and their physical media. This allows the construction of queries that simultaneously take into account the linguistic structures, physical characteristics, and spatial and temporal contexts of the texts being studied.

For example, you may wish to find all of the occurrences of a particular grammatical form in texts that use a particular combination of allographs and that are distributed in a particular way both geographically and chronologically. This kind of analysis facilitates research on language change; on the social and political contexts of scribal practices; and on the transmission of those practices over time. Or, to take another example, a prosopographic analysis in the service of economic or diplomatic history could be aided by a query that

finds all occurrences of a particular combination of personal names in texts found within particular archives.

Importing Existing Digitized Texts

Manually constructing the epigraphic and discourse hierarchies of a text and the numerous links between its epigraphic units, script units, and discourse units can be tedious and time-consuming. For this reason, OCHRE lets you construct these hierarchies and links automatically on the basis of existing digitized texts that are stored as ASCII or Unicode character strings. Existing textual content can be imported into OCHRE by viewing it in a word processor or text editor, selecting the lines of text to be imported, and then copying and pasting the selected textual content, as explained below. The punctuation, white space, line breaks, and (in some cases) the formatting of the textual content are used to break it down into its constituent epigraphic and discourse units, which are then automatically created and interlinked to provide an explicit database representation of the text in such a way that its components can be queried and related to other database items in a fine-grained fashion.

The copying of textual content can be done for a whole text at once or for a particular epigraphic or discourse unit within a text. There is also a way to import many texts into a text catalogue in one operation, without having to copy individual texts; this is explained below in the section on “Importing a Batch of Texts.” But first we will explain the procedure for importing one text at a time.

Copying Textual Content from an External Source

Before copying textual content from an external source, you must create an item in the **Texts** category to represent the imported text; select this text item in the navigation pane; and then open its **Import** tab in the item pane—or, if you are importing only part of the text, select the appropriate epigraphic or discourse unit within the text and open its **Import** tab in the item pane. In the **Import** tab, whether it pertains to an entire text or to an epigraphic unit or discourse unit within a text, you will see a **Plain Text** subtab into which you can

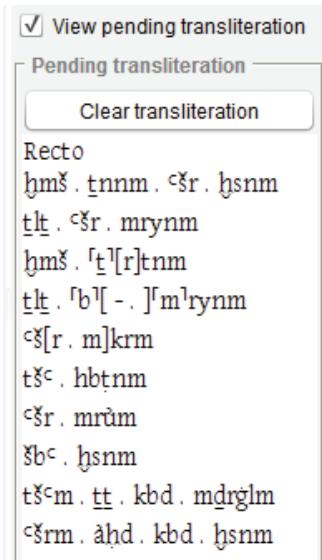
paste textual content that you have copied from another location (e.g., from within a word processor or Web browser).

For a text item, instead of pasting textual content immediately into the **Plain Text** subtab, you can paste it into the **Pending Content** tab (to the right of the **Import tab**) in order to store it temporarily. This is useful if you want to copy and paste a series of digitized texts before going on to import them as OCHRE text items. The **Pending Content** tab has three separate areas: one for a graphemic or hybrid transcription of the text; one for a phonemic transcription of the text; and one for the text's modern-language translation. These areas can be hidden or shown by clicking the checkboxes at the top of the tab. When you are ready to import the textual content, click the **Copy pending content** button in the **Plain Text** subtab of the **Import tab**.

Pasting textual content into the **Plain Text** tab (or the **Pending Content** tab) entails the loss of special formatting such as italics, underlining, superscripts, and subscripts. In some cases, however, the formatting conveys information that is useful for analyzing the textual content into its component epigraphic units and discourse units. Accordingly, **HTML** and **MS Word** tabs are provided next to the **Plain Text** tab. Text-formatting will be preserved when copying HTML and Word documents using the relevant tab.

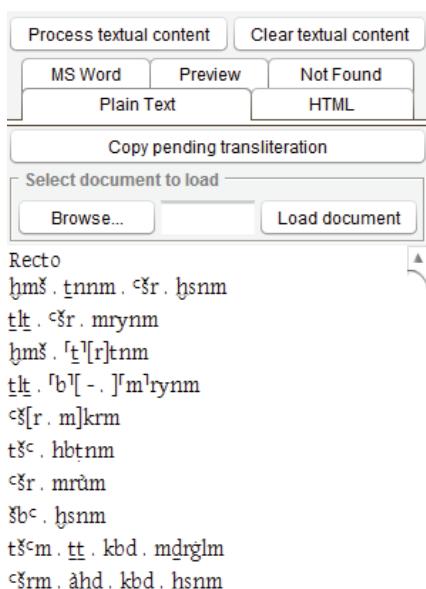
The **HTML** tab is used to copy textual content stored in the “HyperText Markup Language” tagged-text format used by all Web browsers. HTML textual content can be copied from a Web browser or other software that displays this format. The formatting tags embedded in the HTML content will be used when converting it into epigraphic and discourse units.

Alternatively, you can use the **MS Word** tab to copy the contents of a Microsoft Word document. However, this cannot be done by the



copy-paste method. Instead, when you open this tab you will see a **Copy Word document (.docx)** button. Clicking this button will display a window in which you can browse the folders on your computer and choose a Word document. (The document must be in the DOCX file format, not the older DOC format.) The contents of the selected document will then be copied into the **MS Word** tab, preserving italics, superscripts, subscripts, and other formatting.

The textual content you have entered into the **Plain Text**, **HTML**, or **MS Word** tab, either by copy-pasting or by selecting a document, can be manually edited within the tab before it is imported. Note that the other tabs will be automatically cleared whenever textual content is entered into any one of them because only one type of textual content can be used in the import process. There is also a **Clear textual content** button above these tabs that you can click in order to manually erase the contents of the current tab.



It is true that plain-text, HTML, and Microsoft Word documents can all be managed and displayed by OCHRE as external resource items via the **Resources** category (see Chapter 9). However, if such documents contain digitized texts that are objects of study in their own right, they should be imported as items in the **Texts** category so that their internal epigraphic and discourse structures are represented explicitly as hierarchies of database items.

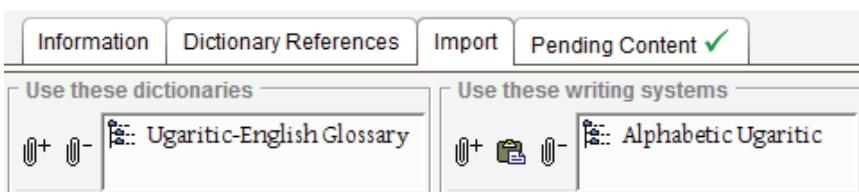
Epigraphic and Discourse Processing Options

To the left of the textual-content tabs you will see a tab entitled **Epigraphic-unit Processing**, if you are in the item pane of an epigraphic unit, or you will see a tab entitled **Discourse-unit Processing**,

if you are in the item pane of a discourse unit. Both tabs will appear if you are in the item pane of a text item.

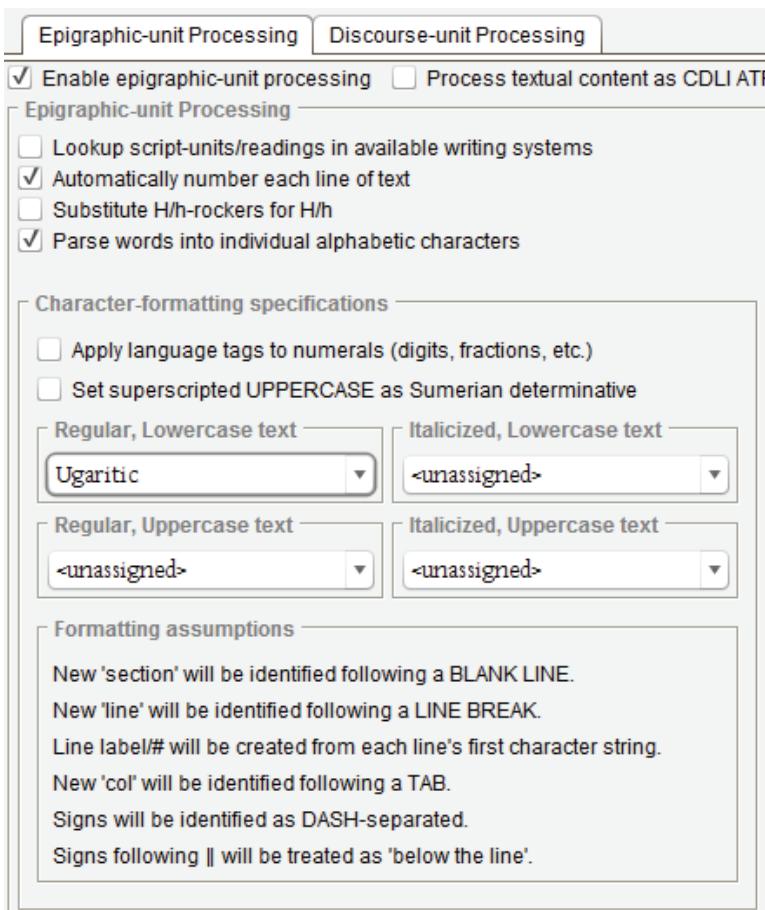
These tabs contain optional settings that control, respectively, how epigraphic units and the epigraphic hierarchy are constructed and how discourse units and the discourse hierarchy are constructed. They also control how epigraphic units are linked to script units and to script-unit allographs, and how discourse units are linked to script-unit readings and to dictionary entries.

At the top of the **Import** tab are two link fields entitled **Use these dictionaries** and **Use these writing systems** which identify the dictionary entries and script units available for linking during the import process. In the **Use these dictionaries** field you can specify one or more hierarchies in the **Dictionaries** category of the current project or of another project to which you have access. In the **Use these writing systems** field you can specify one or more script-unit hierarchies in the **Writing systems** category of the current project or of another project to which you have access. (See Chapter 6 for instructions on entering links using the \oplus “add link” and \ominus “remove link” buttons in conjunction with the **Linked Items** tab in the reference pane.)



Clicking the **Auto-link existing textual content** button above the **Epigraphic-unit Processing** and/or **Discourse-unit Processing** tab will launch an automatic linking mechanism that makes use of the current settings in order to link the existing epigraphic and discourse units of the current text to the relevant script units and dictionary entries. This provides a way to update the links between these units and the script units or dictionary entries that have been entered since the text's component units were created.

The various text-importing and linking options available in the Epigraphic-unit Processing and Discourse-unit Processing tabs are shown below but they are not discussed here. Explanations of these options and step-by-step instructions for importing various kinds of digitized texts can be found on the Web at <http://ochre.uchicago.edu>.



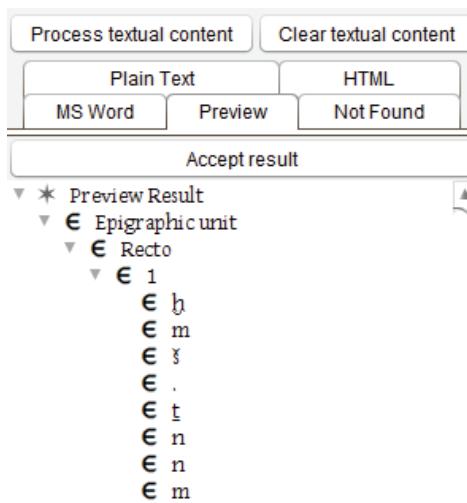
Previewing the Discourse and Epigraphic Hierarchies

When you are ready to import textual content from the Plain Text, HTML, or MS Word tab, first lock the text item into which the content will be added using the “lock” button and then click the Process textual content button. The textual content will then be analyzed and prepared for insertion into the database; however, the new

items and links will not be inserted until you click the **Insert into database** button. Before inserting, it is a good practice to view the **Preview** tab in order to inspect the epigraphic hierarchy and discourse hierarchy that have been automatically generated from the textual content. It may also be helpful to view the **Not Found** tab, which lists all of the epigraphic signs in the textual content for which script units could not be found and all of the words in the textual content for which dictionary entries could not be found.

If there are problems with what has been generated, you can edit or re-enter the textual content in the relevant tab and then click the **Process textual content** button again to generate corrected epigraphic and discourse hierarchies. When you are satisfied that everything is correct, click the **Insert into database** button to complete the text-importing process. The new epigraphic and discourse units will be automatically inserted under the current text item in the navigation pane.

If the current import process pertains to an entire text (i.e., the **Import** tab is in the item pane of a text item), a new epigraphic hierarchy and/or a new discourse hierarchy will be created for the text. However, if the current import process pertains to an epigraphic unit within a text (i.e., the **Import** tab is in the item pane of an epigraphic unit), new epigraphic units will be inserted at the appropriate level of the text's epigraphic hierarchy as children of the epigraphic unit. Likewise, if



the current import process pertains to a discourse unit within a text (i.e., the **Import** tab is in the item pane of a discourse unit), new discourse units will be inserted at the appropriate level of the text's discourse hierarchy as children of the discourse unit.

In summary, the procedure for importing an existing digitized text via the **Import** tab is as follows:

1. Lock the text item into which the textual content will be imported by selecting it in the navigation pane and clicking the  “lock” button in the main toolbar.
2. Add links in the **Use these dictionaries** and **Use these writing systems** fields to specify the dictionary entries and script units that are to be linked automatically to the discourse units (words) and the epigraphic units (signs) generated from the imported textual content.
3. Choose the desired processing options in the **Epigraphic-unit Processing** tab and/or the **Discourse-unit Processing** tab.
4. Load the textual content to be imported into one of the following tabs, either by copying and pasting or by selecting a document file, depending on the textual format: **Plain Text**, **HTML**, or **MS Word**.
5. Edit the textual content, if necessary. To start over, click the **Clear textual content** button and re-load the textual content.
6. Click the **Process textual content** button to generate items that represent the epigraphic and discourse units extracted from the textual content and the epigraphic and discourse hierarchies, and to generate the necessary links between script units, epigraphic units, discourse units, and dictionary entries.
7. View the **Preview** tab to examine the hierarchies of epigraphic and discourse units that have been automatically generated from the textual content.
8. View the **Not Found** tab to see a list of signs in the textual content for which script units were not found and to see a list of words for which dictionary entries were not found.
9. If necessary, edit the textual content and/or adjust the epigraphic or discourse processing options, and then click the **Process textual content** button to regenerate the database items and links.
10. Click the **Insert into database** button to insert the items and links into the database and complete the importing process.

Importing a Batch of Texts

In some cases, you may have a batch of digitized texts that you want to import all at once, rather than importing them one by one. To do so, you must first select a text catalogue in the **Texts** category into which the texts will be imported. The imported texts can be added to an existing catalogue or you can create a new catalogue to contain them. When you select the **Texts** category in the navigation pane and then click the ↗ “insert below” button in the main toolbar, you will be given a choice to create a catalogue of texts or an individual text item.

To import a batch of texts, select the desired text catalogue in the navigation pane and open its **Import** tab in the item pane. At the top of this tab is a field for the **Source file**. The texts to be imported must all be in the same file, stored either in plain-text format (i.e., as ASCII or Unicode characters) or in an XML tagged format. The texts must be separated from one another by a blank line. Details concerning the required textual format and step-by-step instructions for importing batches of texts of various kinds (e.g., cuneiform transliterations) can be found on the Web at <http://ochre.uchicago.edu>.

The path and name of the file that contains the texts can be typed directly into the **Source file** field or selected by means of the **Browse...** button, which opens a window that lets you browse the folders on your computer and choose a file. To control how database items and links will be generated to represent the imported texts, you can add links to hierarchies of dictionary entries and script units via the **Use these dictionaries** and **Use these writing systems** fields; these will specify the dictionary entries and script units to be linked automatically to the discourse units (words) and epigraphic units (signs) generated from the imported textual content. You will also choose the desired processing options in the **Epigraphic-unit Processing** tab and the **Discourse-unit Processing** tab, just as you would do when importing a single text (see <http://ochre.uchicago.edu> for a detailed explanation of these options).

When the source file and the processing options have been specified, click the button labeled **Process texts in source file and**

insert into database to complete the batch-importing procedure. A new text item will be automatically inserted within the text catalogue for each text in the source file. Hierarchies of epigraphic and discourse units will be inserted for each text item. Links will be created from epigraphic units to discourse units, from epigraphic units to script units, and from discourse units (words) to dictionary entries.

The processing and inserting of a batch of texts may take some time, depending on the number and size of the texts in the source file. For information concerning the maximum number of texts that can be imported in one batch and the maximum size of each text, see the documentation on the Web at <http://ochre.uchicago.edu>. In cases where the number of texts to be imported is very large, they can be imported into a single text catalogue in multiple batches, with a separate source file for each batch of texts.

Chapter 12

Constructing and Viewing Dictionary Entries

THE **Dictionaries** category provides a way to represent lexical information about one or more languages. This lexical information can be interlinked with other information in the database, especially with texts and their discourse units in the **Texts** category and with bibliographic data that has been entered in the **Bibliographies** category. You can define your own dictionary entries or you can copy entries from another project by repeating the relevant database items within your project's **Dictionaries** category (see the section in Chapter 4 on "Repeating Items and Hierarchical Branches").

When you select the **Dictionaries** category in the navigation pane and then click the ↲ "insert below" button in the main toolbar, you will be given a choice to create a hierarchy or an individual item. An individual item in the **Dictionaries** category represents a dictionary entry for a particular lemma (headword or lexeme) in the lexicon of a language. A hierarchy of items in this category represents an entire dictionary for a particular language. The order of the items in the hierarchy reflects the order of the entries within the dictionary.

▼ Abc Dictionaries

▼ Old Assyrian Lexicon

- * A
- * B
- * D
- ▼ * E
 - Abc e
 - Abc ebar
 - Abc ebārūtum
 - Abc ebābum
 - Abc ebērum A
 - Abc ebētum
 - Abc ēbišum
 - Abc ebūrum
 - Abc ebūtum
 - Abc edapum
 - Abc edēsum
 - Abc eħħulum
 - Abc ekallum
 - Abc ekēnum
 - Abc ekēpum
 - Abc ela
 - Abc elālum
 - Abc elēn
 - Abc elēnum
 - Abc elēpum
 - Abc eli
 - Abc eli
 - Abc elippum
 - Abc eliš
 - Abc elitum
 - Abc ellum
 - Abc ellutum
 - Abc elūm B
 - Abc elūm E
 - Abc elārum
 - Abc īma
 - Abc emārum

Hierarchies in the **Dictionaries** category are “subordination hierarchies” (see the section on “Containment Hierarchies and Subordination Hierarchies” in Chapter 4). You can use headings within a hierarchy to organize dictionary entries into named groups. For example, the name of each volume in a multivolume dictionary would be indicated by a heading within the hierarchy that represents the entire dictionary. Inserted under each heading would be items that represent the dictionary entries in a given volume.

In data-entry mode, many different data-entry tabs and fields are shown for a dictionary entry, reflecting the potential complexity of this kind of item, in which multiple forms, meanings, and citations may be listed for a given lemma. In view-only mode (and in the **View** tab of the data-entry item pane) these data-entry tabs and fields are not seen. Instead, their contents are combined and formatted to replicate the appearance of the dictionary entry as it would appear in a traditional printed dictionary.

When you select a hierarchy within the **Dictionaries** category the item pane will display data-entry fields for the hierarchy’s creators, creation date, name, abbreviation, description, properties, links, notes, and display preferences. These fields are the same as those used for hierarchies elsewhere (see the section on “Hierarchies of Items” in Chapter 4). But there is an additional field for entering the **Type of dictionary** in which you can specify whether the dictionary entries contained in the hierarchy are “extended” entries or simpler “glossary” entries. An extended entry can be quite complex, with many different meanings and submeanings for a given lemma and detailed morphological, comparative, and semantic information. A glossary entry is simpler; it contains a more restricted set of information about the lemma.

Entering Basic Information about a Lemma

When you select a hierarchy within the **Dictionaries** category and click the ↲ “insert below” button, an individual item will be inserted that represents a dictionary entry for a particular lexical lemma. To construct the entry, select the item in the navigation pane and open

the **Dictionary Entry** tab in the item pane. If the hierarchy's **Type of dictionary** field has been set to "extended," the following overlapping data-entry subtabs will be displayed: **Lemma**, **Forms**, **Comparisons**, **Meanings**, **Bibliography**, **Cross-references**, **Properties**, **Links**, **Notes**, and **Events**.

The tabs for **Properties**, **Links**, **Notes**, and **Events** function in the usual fashion described above in Chapters 5, 6, and 7, reflecting the fact that a dictionary entry is an independent database item like any other OCHRE item. The other data-entry tabs and the fields within them are discussed below. At any time, you can display the current contents of these fields by opening the **View** tab at the top of the item pane, in which you will see a formatted, printable version of the entire dictionary entry that has been produced by combining the data entered in the various fields. The dictionary display options in the **View** tab (and in view-only mode elsewhere) are discussed below in the section on "Displaying Dictionary Entries."

Entering the Citation Forms of a Lemma

The **Lemma** subtab of the **Dictionary Entry** tab contains fields for the forms and meanings of the currently selected lemma (headword). The **Citation form** data-entry area contains two fields: one for the lemma's **Native form** (i.e., a sign-by-sign graphemic transcription)

The screenshot shows a user interface for entering citation forms. At the top, there are two tabs labeled 'Citation form' and 'Citation form'. Below these tabs is a horizontal toolbar with icons for copy (ctrl-c), cut (ctrl-x), paste (ctrl-v), and delete (ctrl-d). Underneath the tabs are two input fields. The first field is labeled 'Native form' and contains the text 'Native form'. The second field is labeled 'Transcription' and contains the text 'pahš-'. Both fields have a placeholder text 'Type here'.

and one for its phonemic **Transcription** (see the section on "Graphemic, Phonemic, and Hybrid Transcriptions" above in Chapter 11). The phonemic transcription of the lemma's citation form normally uses Latin characters with diacritics or International Phonetic Alphabet (IPA) characters, depending on the **Phonemic font** specified by the project administrator in the **Fonts** subtab of the project's **Preferences** tab. The graphemic transcription will be displayed using the non-Latin **Graphemic font** that has been specified in the **Fonts** subtab.

In some dictionaries, there is no graphemic transcription (“native form”) of the lemma but only a phonemic transcription using the Latin alphabet. However, graphemic transcriptions are used in dictionaries in which it is conventional to employ non-Latin characters to display the actual signs used to write the lemma, mirroring the word’s appearance in the texts in which it appears. For example, in philological dictionaries of languages such as Arabic, Chinese, Greek, Hebrew, Hindi, and so on, the citation form of a lemma will normally be presented using a non-Latin writing system, even if there is also a phonemic transcription of the lemma that uses the Latin alphabet. For this reason, if a graphemic transcription has been entered in the **Lemma** tab, it will be displayed as the name of the dictionary-entry item in the **Dictionaries** category of the navigation pane; otherwise the phonemic transcription will be used.

Sometimes there will be multiple citation forms of the same lemma. To add another citation form, click the  “create new” button to the right. This will create a new **Citation form** tab in which you can enter another graphemic and/or phonemic transcription of the lemma. Clicking the  “delete” button will delete the currently selected citation form. Clicking the  “move left” and  “move right” buttons will reorder the sequence of citation forms, determining the order in which they will appear in the formatted dictionary entry. These buttons work in the same way as those for adding, deleting, and reordering free-form notes (see the section on “Entering Free-form Notes for an Item” in Chapter 5).

It is a good practice to indicate explicitly the language of the graphemic and phonemic transcriptions of a lemma’s citation form, even though this is implied by the existence of the lemma within the dictionary of a particular language. The language can be indicated via the **Language** drop-down list on the text-formatting toolbar (see the section on “Specifying the Language or Function of a Selected Portion of Text” above in Chapter 5). Characters, words, or phrases in any of the **Dictionary Entry** fields can be selected by dragging the mouse cursor over them, at which point they can be individually formatted and identified by language and function. This has the benefit of allowing

subsequent queries and displays to make use of this information when searching and presenting data.

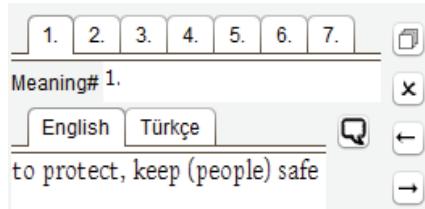
Cross-references ✓	Properties ✓	Links	Notes	Events
Lemma	Forms ✓	Comparisons ✓	Meanings ✓	Bibliography ✓
Additional search terms <input type="text" value="English"/>  <ul style="list-style-type: none"> <input type="checkbox"/> Keep this item private <input type="checkbox"/> All occurrences are cited <input type="checkbox"/> No known occurrences 				
<input type="button" value="Citation form"/> <input type="button" value="Citation form"/>		Primary lemma 		
Native form <input type="text"/>		Root lemma 		
Transcription <input type="text" value="pahš-"/>		Loan word <input type="text"/>		
<input type="button" value="1."/> <input type="button" value="2."/> <input type="button" value="3."/> <input type="button" value="4."/> <input type="button" value="5."/> <input type="button" value="6."/> <input type="button" value="7."/>		Part of speech <input type="text" value="v."/>		
Meaning# 1. <input type="text" value="English"/> <input type="text" value="Türkçe"/> 		Attested writing <input type="text" value="wr. syll. and PAP"/> 		
Discussion:    		Text dating <input type="text" value="from OH"/>		
<input type="button" value="Save entry"/>		<input type="button" value="Preview lemma"/>		<input type="button" value="View full entry"/>

Note that, unlike other data-entry situations in OCHRE, changes you make in the **Lemma** tab and in other tabs are not saved automatically when you leave the tab. You must click the **Save entry** button in the bottom left corner of the tab to save what you have entered. (This is not required for simplified “glossary” entries, where saving is automatic; see below.) At any time, you can click the **Preview lemma** button at the bottom of the tab to display a window that shows how the information that has been entered in the **Lemma**

tab will be formatted. Click the **View full entry** button to see a formatted view of the full dictionary entry, including the data entered in other subtabs of the **Dictionary Entry** tab.

Entering the Meanings of a Lemma

After a new dictionary-entry item has been created, the meanings area in the **Lemma** tab will be empty. To add a meaning for the lemma, click the  “create new” button to the right of the area. This will create a new tab in which you can enter a meaning in the project’s default language and, if you wish, translations of that meaning in other languages. Clicking the  “delete” button will delete the currently selected meaning. Clicking the  “move left” and  “move right” buttons will reorder the sequence of meanings, determining their order in the formatted dictionary entry.



Within each meaning tab is an optional field for entering the **Meaning#**, in case there is more than one meaning. For example, if there are two meanings for the lemma you would create two tabs and would enter “1.” in the **Meaning#** field of the first meaning and “2.” in the **Meaning#** field of the second. These numbers will be inserted before the respective meanings when the dictionary entry is formatted in the **View** tab. The tab headings will also be changed to reflect the meaning numbers.

If multilingual features have been enabled for the project, there will be a tab labeled with the name of the project’s default language (e.g., **English**, **Français**, **Deutsch**, etc.) below the **Meaning#** field. After typing the meaning in this tab you can enter a translation of the meaning in another language, if you wish, by clicking the  “translation” button to the right. You will be given a choice of languages from the list of languages specified by the project administrator in the **Data-entry** subtab of the **Preferences** tab in the project’s item pane (see Chapter 17). A new tab will appear, labeled with the name of the selected language, into which you can enter a translation of the

meaning. Any number of meaning tabs can be created for a lemma and there can be any number of translations for each meaning.

Below the meanings area in the **Lemma** tab is an optional **Discussion** area, which is initially empty. To add a discussion paragraph, click the “create new” button to the right of the area. This will create a new tab in which you can enter a discussion of the lemma and its meanings. Clicking the “delete” button will delete the currently selected discussion paragraph. Clicking the “move left” and “move right” buttons will reorder the sequence of discussion paragraphs, determining the order in which they will appear in the formatted dictionary entry. The discussions, if any, will appear after the list of meanings, which in turn will appear after the lemma’s citation forms. If multilingual features have been enabled, you can also enter one or more translations of a discussion, using the same procedure as you would use for entering a translation of a meaning.

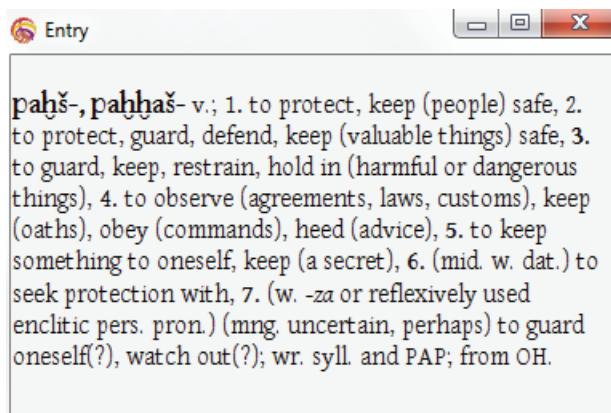
Linking a Lemma to a Primary Lemma or a Root Lemma

The optional **Primary lemma** field in the **Lemma** tab is used to link the current dictionary entry to another dictionary entry for the same word. This is useful when a word has two different forms that have given rise to two different lemmas (headwords) in the dictionary. The two lemmas refer to the same word and thus have the same meanings, so one of them can be treated as the primary lemma and the other as secondary. The data fields of the secondary lemma should be left blank and a link to the primary entry should be entered in its **Primary lemma** field.

The optional **Root lemma** field in the **Lemma** tab is used to link the current dictionary entry to another dictionary entry that contains information about a lexical “root” to which the current lemma is related. For example, most words in Semitic languages can be related to a triconsonantal root and it may be useful to group together the dictionary entries for related words that represent different parts of speech (e.g., nouns, verbs, adjectives, etc.) but share a common root. By linking each lemma to the appropriate root lemma, dictionary entries can be listed, not just alphabetically, but according to their

roots, and queries can easily find all words in texts or dictionary entries that share the same root as a given word.

To enter a **Primary lemma** link or a **Root lemma** link, click the  button in the main toolbar in order to display the **Linked Items** tab in the reference pane; then select the link-target item from the **Dictionaries** category; and then click the  “add link” button to the right of the **Primary lemma** or **Root lemma** field. You can remove the link by clicking the  “remove link” button.



Entering Other Basic Information about a Lemma

The **Lemma** tab also contains optional fields for: (1) its original language, if it is a **Loan word** from another language; (2) its **Part of speech**; (3) its **Attested writing** in the texts in which it appears; and (4) the **Text dating** of the texts in which it appears.

The textual content in the optional **Attested writing** field explains the ways in which the lemma is written (e.g., syllabically, logo-graphically, or both). This is relevant mainly for texts that use a complex nonalphabetic writing system.

There is a checkbox in the **Lemma** tab that allows you to **Keep this item private**, that is, to make the current dictionary entry invisible to anonymous users. Two other checkboxes allow you to indicate whether **All occurrences are cited** (i.e., whether all known textual occurrences of the current lemma have been cited in the various data-entry tabs for the entry) or whether there are **No known occurrences**.

Finally, there is an optional data-entry area in the **Lemma** tab for **Additional search terms**. The words entered here are typically modern-language key words that will be used in queries and other look-up operations when searching for the dictionary entry, in addition to searches based on the lemma's citation forms. The primary search term should be entered in the default tab that is labeled with the name of the project's default language (e.g., **English**). To add another search term, click the  button to the right. You will have the option to create either an **Alias** tab, in which you can enter an alternate term in the default language, or a tab in which to enter a translation of the primary search term in another language.

Entering Morphological Information about a Lemma

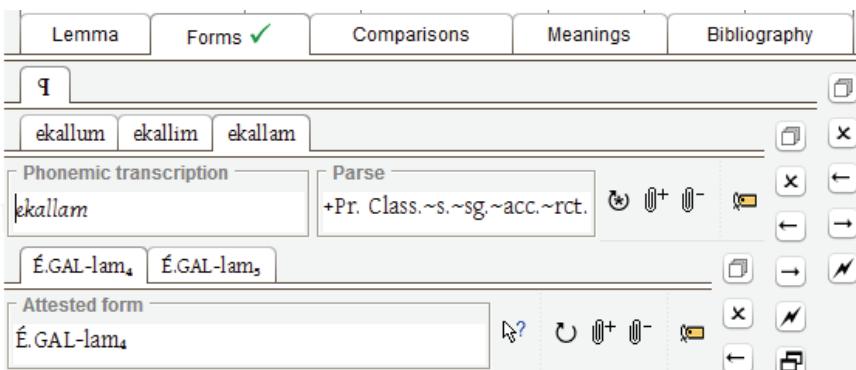
The **Forms** subtab of the **Dictionary Entry** tab is initially empty. It is used to enter a detailed description of the various grammatical forms of the lemma together with references to the textual attestations of each form. Each form of a lemma is parsed on the basis of morphological categories such as person, number, gender, case, tense, mood, voice, etc. The categories depend on the part of speech. OCHRE does not prescribe the morphological categories you must use; each project can devise its own parsing system.

Entering the Grammatical Forms of a Lemma and Their Parses

For some lemmas there may be several grammatical forms, each of which has many attestations. For this reason, forms may be grouped into paragraphs so they can be formatted in a more readable fashion in the printable view of the dictionary entry that is shown in the **View** tab.

To add a paragraph of morphological information, click the  “create new” button on the right side of the **Forms** tab. This will create a new tab labeled ¶. Clicking the  “delete” button will delete the currently selected paragraph tab. Clicking the  “move left” and  “move right” buttons will reorder the tabs, determining the order in which the paragraphs will appear in the formatted view of the dictionary entry.

Within each paragraph of morphological information you can enter any number of grammatical forms. For many lemmas, there will only be one paragraph that contains only one or two forms. Click the “create new” button to the right of the paragraph tab to create a new **Form** tab. Clicking the “delete” button will delete the currently selected **Form** tab. Clicking the “move left” and “move right” buttons will reorder the tabs, determining the order in which the forms will appear within the paragraph in the formatted entry.



A **Form** tab has data-entry fields for the **Phonemic transcription** of the form and for its **Parse**. Whatever you type into the **Phonemic transcription** field will become the title of the current **Form** tab. In this field, as elsewhere in OCHRE, you can use the text-formatting toolbar when entering textual content (see Chapter 5). As with the citation forms of a lemma, it is a good practice to indicate explicitly the language of each form of a lemma via the **Language** drop-down list on the text-formatting toolbar.

You can type a morphological analysis directly into the **Parse** field as ordinary textual content or you can assign morphological properties to this field that will cause it to be filled in automatically with a concatenated string of values (if the field is empty of textual content at the time the properties are assigned). Assigning properties has the advantage of allowing subsequent database queries to find the lemma’s forms automatically based on their person, number, gender, case, tense, voice or other morphological category. Note, again, that

OCHRE does not prescribe the categories you must use; each project can determine its own nomenclature for morphological analysis and can express this in its project-defined properties (see Chapter 5).

To assign properties to the **Parse** field, display the **Linked Items** tab in the reference pane; then click the **Show properties and links** button at the bottom of the **Linked Items** tab of the reference pane (if this area is not already displayed); and then click the “show properties” to the right of the **Parse** field. This will make the **Parse** field active in the **Properties and links for...** field in the bottom area of the reference pane. The morphological properties that you enter and save in the **Properties** tab in the reference pane will be concatenated and inserted into the **Parse** field, using the abbreviated names of the properties’ variables and values (if there are abbreviations).

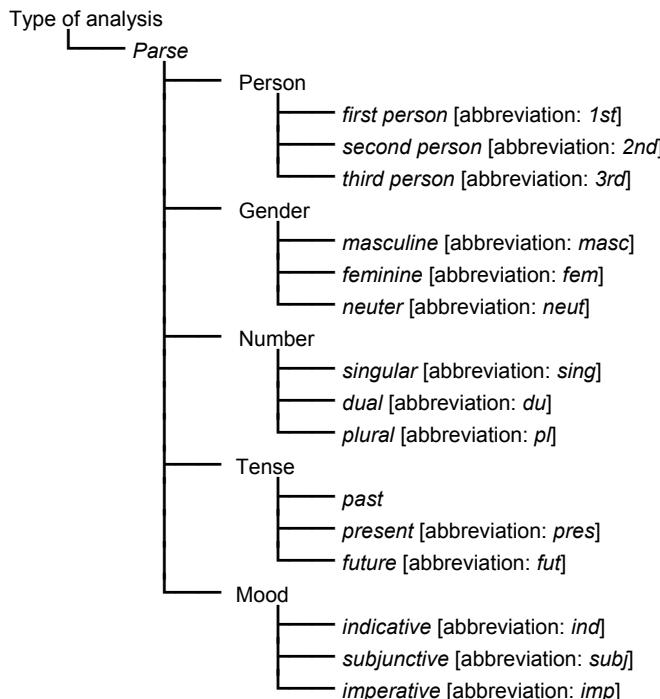
Note that the properties you enter must be explicitly saved to the database by clicking the **Save** button at the bottom of the **Properties** tab. Clicking the “show properties” button at any time will redisplay the properties that have been assigned to this field, no matter what else is currently shown in the **Linked Items** tab of the reference pane.

There is a special nominal value called “parse” that has been predefined in the OCHRE parent project. If this special value is incorporated in your project’s taxonomy as the value of a variable (e.g., “Type of analysis = parse”), then the morphological properties that are hierarchically nested within this variable-value pair in the taxonomy (see the section on “Nested Properties” in Chapter 5) will be displayed in the **Parse** field in a streamlined fashion. You would enter the desired properties in the **Properties** tab

Variable	Value
OARE Primary Classification	Parse
▪ Morphological Form	Substantive
Grammatical Number	Singular
Case	Accusative
State	Status Rectus

in the bottom area of the reference pane as usual, starting with the parent property that uses the “parse” value. But when these properties are saved, the concatenated character string that is automatically inserted into the **Parse** field will consist of only the abbreviated names of the values of the properties that are nested within the special “parse” property, excluding the variable names and excluding the “parse” value itself.

For example, here is a partial hierarchy of variables and values showing how a project’s morphological properties might be nested:



By nesting morphological properties within the “parse” value in the taxonomy and then assigning these properties to the **Parse** field, you will cause this field to be automatically filled with an abbreviated parse string such as “1st-masc-sing-pres-ind.”

Note that if you type the morphological analysis directly into the **Parse** field, any properties that you assign to this field will not be used to fill the field but will be treated as additional information about the

morphology of the form and will be displayed only in the **Properties** tab in the reference pane.

In addition to assigning properties, you can attach to the **Parse** field one or more links to other database items. For example, you might want to link the morphological analysis of a form that has been entered in the **Parse** field to a bibliographic reference represented as a database item in the **Bibliographies** category.

To add a link, click the  button in the main toolbar to display the **Linked Items** tab in the reference pane; click the  “show links” button to the right of the **Parse** field to refresh its **Properties and links for...** field in the bottom area of the reference pane; select the link target in the reference pane; and then click the  “add link” button to the right of the **Parse** field. The link will appear in the **Links** tab in the bottom of the reference pane. Clicking the  “remove link” button will remove the link that is currently selected in the **Links** tab.

Entering the Textual Attestations of a Grammatical Form

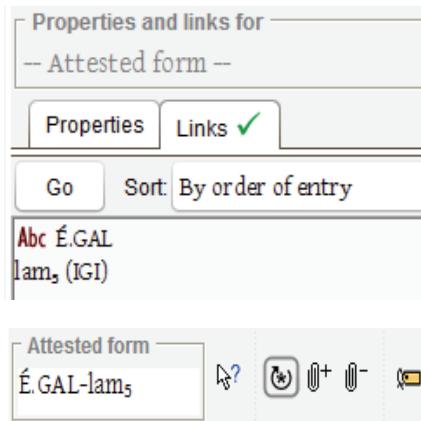
Within each **Form** tab you can also create one or more **Attestation** subtabs for entering the textual attestations of that particular form of the lemma. Click the  “create new” button to the right of the **Form** tab to create a new **Attestation** tab within it. You can delete the currently selected **Attestation** tab by clicking the  “delete” button. Clicking the  “move left” and  “move right” buttons will reorder the tabs, determining the order in which the attestations will appear for the form within the formatted dictionary entry.

An **Attestation** tab has an **Attested form** data-entry field in which to enter a graphemic or hybrid transcription of a particular writing of the form in one or more texts (see the section on “Graphemic, Phonemic, and Hybrid Transcriptions” above in Chapter 11). The transcription of the attested form will be displayed using the **Graphemic font** specified by the project administrator in the **Fonts** subtab in the **Preferences** tab of the project’s item pane (see Chapter 17). If it is a hybrid transcription, the project administrator should ensure that the graphemic font is able to display the appropriate characters.

Whatever is entered into the **Attested form** field will become the title of the current **Attestation** tab. It is a good practice to add links from this field to the script-unit readings that correspond to the attested form (see the section on “Entering the Allographs and Readings of a Script Unit” above in Chapter 11). This has the benefit of integrating the dictionary entry more tightly with the database representation of the writing system (or systems) employed to write the texts in which the lemma is used. A side-effect of this is the ability to perform queries that search for attested forms written in one writing system as opposed to another, in cases where a language was written using more than one writing system (e.g., alphabetic versus logosyllabic).

You can leave the **Attested form** field empty and add script-unit readings one by one via the \oplus “add link” button; or, to save time, you can type the attested form and click the F “find script units” button to the right of the field in order to find all the script-unit items in the project’s **Writing systems** category that match what you have typed in the field and create automatic links to these items. To see the links, display the **Linked Items** tab in the reference pane (if it is not already displayed) and click the U “show links” button to the right of the **Attested form** field in order to refresh this field and make it active in the **Properties and links for...** area in the bottom of the reference pane.

To manually add a script-unit link to an **Attested form** field, select as the link target the desired script-unit reading in the **Writing systems** category and click the \oplus “add link” button to the right of the **Attested form** field. The link will appear in the **Links** tab in the bottom of the reference pane. Clicking the \ominus “remove link” button will remove the link that is currently selected in the **Links** tab. If the



Attested form field was left blank, the name of the form will be derived by concatenating the relevant script unit names.

To assign properties to an attested form, click the  “show properties” button to the right of the **Attested form** field in order to make this field active in the **Properties and links for...** area in the bottom of the reference pane; then enter properties in the **Properties** tab in the usual fashion. Unlike the properties assigned to the **Parse** field, the properties assigned to the **Attested form** field are not used to fill in the field but simply provide additional information about the contents of the field. Thus, properties are used to fill in the content of the **Parse** field and provide additional information about the **Attested form**, whereas links are used, conversely, to fill in the **Attested form** field and provide additional information about the **Parse**.

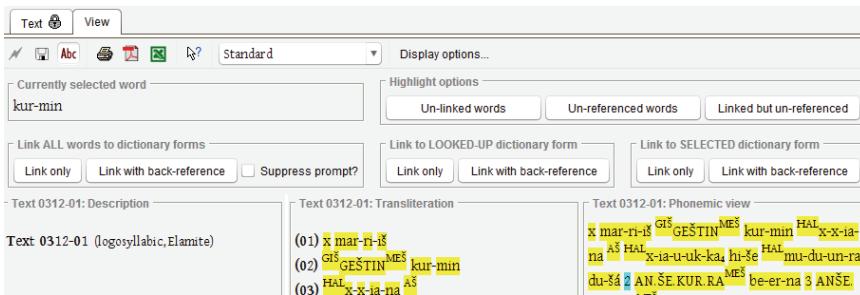
Entering Textual References for an Attested Form

Below the **Attested form** field in each **Attestation** tab is an area in which to enter a table of textual references for the attested form. If the texts on which the dictionary is based are represented as database items in the project’s **Texts** category, this table can be filled in automatically by linking the words in the texts to attested forms in the project’s dictionary. However, only the project administrator can perform this operation.

To link the words in a text to attested forms in the dictionary, select the text item in the navigation pane and lock it by clicking the  “lock” button in the main toolbar. Then open its **View** tab in the item pane. If you are the project administrator, you will see an  “add dictionary links” button. When you click this button you will see several buttons in the text’s **View** tab above the transcriptions of the text. As you move the mouse cursor over the transcriptions, the cursor will change into a pointing hand when it is positioned over a word that is capable of being linked to an attested form in the dictionary; if you click the mouse this word will appear in the **Currently selected word** field in the top left corner of the tab. You



have the option to **Link to LOOKED-UP dictionary form** or **Link to SELECTED dictionary form** by clicking the appropriate button. Alternatively, you can link all of the words in the text to dictionary forms, not just the currently selected one, by clicking the **Link ALL words to dictionary forms** button.

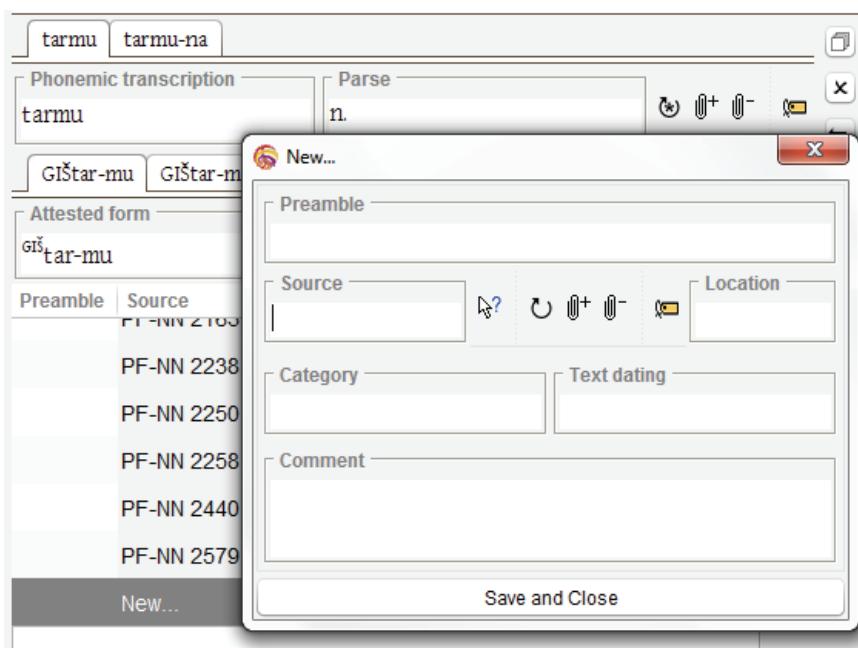


The **Link to LOOKED-UP dictionary form** button (for the currently selected word) or the **Link ALL words to dictionary forms** button will search the project's dictionary entries in the **Dictionaries** category for matches between the word (or words) in the text and identical attested forms in one of the dictionary entries. If a match is found, a link will be created—both a “forward” link, from the word (discourse unit) to the attested form, and a “reverse” link, from the attested form to the word. You can click one of the buttons in the **Highlight options** area to see which words in the text have been successfully linked.

In cases where no match was found for a word, you can manually select an attested form to which to link the word and then click the button labeled **Link to SELECTED dictionary form**. The form must first be selected as the link target in the reference pane. To select it, click the **Ø** button in the main toolbar in order to display the **Linked Items** tab in the reference pane (if it is not already displayed) and then choose the **Dictionaries** category, the desired dictionary entry, and the appropriate attested form within the entry. A dictionary entry may contain several different kinds of components, including meanings, citations, grammatical forms, and attested forms (i.e., attestations of the grammatical forms), so to be sure that you have selected an

attested form, move the mouse cursor over the form in the link-target area of the reference pane and wait a second to make sure an “Attested form” label pops up.

If the texts on which the dictionary is based are not represented as database items in the project’s **Texts** category, you can enter textual references for the attested form directly into the table below the **Attested form** field. To do so, click the “create new” button on the right side of the **Attestation** tab in order to create a new row in the table. Then double-click on the new row (in the **Source** column) to display a window with data-entry fields for a **Preamble**, **Source**, **Location**, **Category**, **Text dating**, and **Comment** that pertain to the textual reference.



The **Source** field contains the name of the text. After you type in the name, you can click the “find text” button to try to link it automatically to a text item in the project’s **Texts** category. This may be feasible, even when the automatic reference mechanism described above is not feasible, provided that the relevant text is represented in

some fashion as a database item in the **Texts** category, even if the words in the text are not represented as individual (and thus automatically linkable) discourse units in the text's discourse hierarchy.

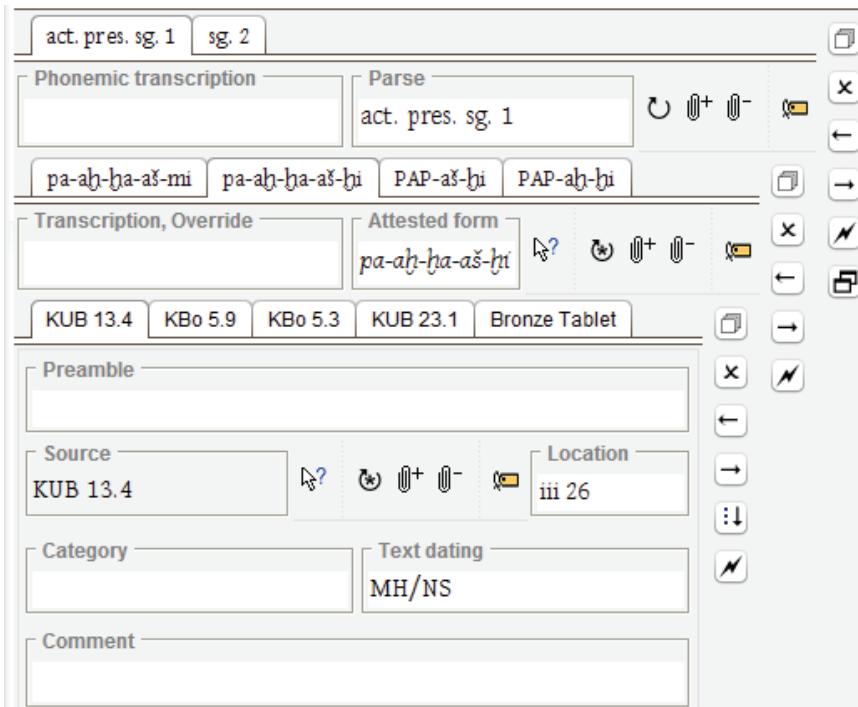
The **Location** field contains the page, column, line, or other location where the attested form appears within the text that is named in the **Source** field. The **Preamble**, **Category**, **Text dating**, and **Comment** fields are optional and will often be left blank.

After you have finished entering the information for the new textual reference, click the **Save and Close** button in the window to insert what you have entered into the table of references for the attested form. You can delete the currently selected row in the table by clicking the **x** “delete” button on the right side of the **Attestation** tab. Clicking the **↑** “move up” and **↓** “move down” buttons will reorder the references in the table, determining the order in which they will appear within the formatted dictionary entry.

Preamble	Source	Location	Category	Dating	Comment
1679-101		01			
2127-101		01			
Fort. 0013-102		02			
Fort. 1250-101		01			
Fort. 1806-006		01, 03, 05			
Fort. 2119-106		01			

Alternatively, a project may choose to use a tabbed format instead of a tabular format to display the textual references for an attested form (see the example on the facing page). The data-entry fields

would be the same (i.e., **Preamble**, **Source**, **Location**, **Category**, etc.) but they would appear on separate tabs instead of in a table. Tabs should be used when you wish to be selective, rather than exhaustive, in your listing of textual attestations. Tabs work well if there are not many textual attestations of a form whereas tables are better if there are many of them.



As in the **Lemma** tab, any data you enter in the **Forms** tab is not saved automatically when you leave the tab. You must click the **Save entry** button in the bottom left corner of the tab to save what you have entered. At any time, you can click the **Preview forms** button at the bottom of the tab to display a window that shows how the information that has been entered in the tab will be formatted. Click the **View full entry** button for a formatted view of the full dictionary entry, including the data entered in other subtabs of the **Dictionary Entry** tab (see the example on the next page).

The screenshot shows a digital dictionary application window. At the top, there are tabs for 'Dictionary' and 'View'. Below the tabs is a toolbar with icons for search, file operations, and language selection ('English'). A dropdown menu labeled 'Full entry' is open. To the right of the entry field is a 'Display options...' button. The main content area displays the following text:

pahš-, pahhaš- v.; 1. to protect, keep (people) safe, 2. to protect, guard, defend, keep (valuable things) safe, 3. to guard, keep, restrain, hold in (harmful or dangerous things), 4. to observe (agreements, laws, customs), keep (oaths), obey (commands), heed (advice), 5. to keep something to oneself, keep (a secret), 6. (mid. w. dat.) to seek protection with, 7. (w. -za or reflexively used enclitic pers. pron.) (mng. uncertain, perhaps) to guard oneself(?), watch out(?); wr. syll. and PAP; from OH.

act. pres. sg. 1 *pa-ah-ha-aš-mi* KUB 29.1 i 19 (OH/NS), *pa-ah-ha-aš-ht* KUB 13.4 iii 26 (MH/NS), KBo 5.9 i 24, 27, ii 13, KBo 5.3 i 38, iv 27, 28 (Supp. I), KUB 23.1 ii 29, Bronze Tablet ii 40, 42, 72, 74 (Tudh. IV), PAP-*aš-ht* KUB 40.38:7 (Tudh. IV or Supp. II), KBo 14.112:4 (Supp. II), KUB 23.44 iii 7, 11 (Supp. II), PAP-*ah-ht* KUB 26.33 iii (22), 23 (Supp. II); sg. 2 *pa-ah-ha-aš-ti* KBo 5.3 ii 10 (Supp. I), KUB 26.37 obv. 13 (Supp. I), KBo 5.13 ii 14, iii 20 (Murš. II), KBo 5.4 obv. 44 (Murš. II), KUB 21.1 iv 38 (Muw. II), KBo 4.10 rev. 6, 8 (Hatt. III or Tudh. IV), PAP-*aš-ti* KUB 19.55 rev. 40 (NH), KUB 48.123 i 24.

Entering Comparative Information about a Lemma

In some dictionary entries there may be a discussion of the lemma that compares it to other words in the same language or to words or phrases in other languages. This should be entered as textual content in the **Comparisons** tab. Topics in the comparative discussion are grouped into paragraphs so they can be formatted in a more readable fashion in the printable view of the dictionary entry that is shown in the **View** tab. Any topic that does not belong in the **Forms** or **Meanings** tabs should be included here, even if it does not pertain to lexical comparisons.

To add a paragraph to the comparative discussion, click the “create new” button on the right side of the **Comparisons** tab. This will create a new tab labeled . Clicking the “delete” button will delete the currently selected paragraph tab. Clicking the “move left” and “move right” buttons will reorder the tabs, determining the order in which the paragraphs will appear in the formatted dictionary entry.

Within each paragraph tab you can enter free-form textual content using the text-formatting toolbar and the keypad of special characters (see the section on “Entering Textual Content in Notes

and Elsewhere” above in Chapter 5). Note that you can embed properties and hyperlinks to other database items within selected portions of the textual content. Embedded hyperlinks are useful for (but not limited to) establishing cross-references to other dictionary entries in the **Dictionaries** category, to text items in the **Texts** category, to bibliographic references in the **Bibliographies** category, and to scholars who are represented as persons in the **Persons & organizations** category.

Data entered in the **Comparisons** tab is not saved automatically. You must click the **Save entry** button in the bottom left corner of the tab to save what you have entered. Click the **Preview comparisons** button to display a window that shows how the information in the tab will be formatted. Click the **View full entry** button to see a formatted view of the full dictionary entry, including the data entered in other subtabs of the **Dictionary Entry** tab.

(Akk.) ina irtikī ina erşetim uşrinnī KUB 1.16 iv 71-72 = (Hitt.) nu=mu tagga
 [niya=ti] taknaz pa-ah-[š]i “[On your] breast protect me from the earth” KUB
 1.16 iii 72-73 (edit of Hatt. I, NS), ed. HAB 16f, cf. Götz, ZA 34:183 and
 Melchert, Diss. 182f.; (Akk.) u abuya "Az[ir]a qadu KUR-šu ittašaršu KUB 3.14
 obv. 7-8 = (Hitt.) [ABU=YA "Azir]an QADU KUR=ŠU pa-ah-ha-aš-ta-at “My
 father protected Aziru together with his land” KUB 21.49 obv. 6 (treaty w.
 Duppi-Tessub, Murš. II), ed. SV 1:6f.

Entering Semantic Information about a Lemma

The meanings of a lemma are listed briefly in the **Lemma** tab. However, a full discussion of the lemma’s meanings and submeanings can also be entered in the **Meanings** tab together with textual citations that illustrate the lemma’s use in various contexts. Note that the data you enter in this tab is not saved automatically. You must click the **Save entry** button in the bottom left corner to save what you have entered. Clicking the **Preview meanings** button will display a window in which you can see how the data that has been entered in the **Meanings** tab will be formatted.

To enter information about one of the lemma's meanings, click the  “create new” button on the right side of the **Meanings** tab. This will give you a choice to create a **Meaning** tab, or to create a **Discussion** tab that pertains to a previously entered meaning or meanings. Clicking the  “delete” button will delete the currently selected tab. Clicking the  “move left” and  “move right” buttons will reorder the tabs, determining the order in which the meanings and related discussions will appear in the formatted dictionary entry.

Entering Meanings, Submeanings, and Related Discussions

The first tab to be created will normally be a **Meaning** tab. Any number of **Meaning** tabs can be created at the top level within the **Meanings** tab, representing the primary meanings of the lemma. There are fields in each **Meaning** tab for the meaning number (#) and a **Gloss** in the project's default modern language. These are typed in as ordinary textual content. Whatever you enter in the # field (e.g., “1.”, “2.”, “3.”, etc.; or, at a lower level, “a.”, “b.”, “c.”, etc.) will be placed before the gloss in the formatted, printable version of the dictionary entry and will also become the title of the tab. Each project can determine its own numbering system for meanings and sub-meanings.

Within each **Meaning** tab you can create subtabs for one or more subordinate meanings, textual citations, and discussions. To do so, click the  “create new” button inset on the right side of the **Meaning** tab. This will give you a choice to create a **Meaning** tab in which to enter a submeaning of the parent meaning; or a **Citation** tab in which to enter a textual quotation that includes a form of the lemma and illustrates the parent meaning; or a **Discussion** tab in which to enter a discussion of the parent meaning and related citations.

At any level of the tab hierarchy, clicking the  “delete” button will delete the currently selected tab and clicking the  “move left” and  “move right” buttons will reorder the tabs, determining the order in which the meanings, citations, and related discussions will appear in the formatted dictionary entry.

The screenshot shows the 'Meanings' tab of the Dictionary Entry Constructor. At the top, there are tabs for 'Cross-references', 'Properties', 'Links', 'Notes', and 'Events'. Below these are sub-tabs for 'Lemma', 'Forms', 'Comparisons', 'Meanings', and 'Bibliography', with 'Meanings' being the active tab. A horizontal bar numbered 1 through 7 is present. The main area displays three levels of meaning hierarchy:

- Level 1:** One entry under '#', labeled 'Gloss', containing '1.' followed by 'to protect, keep (people) safe'. It includes edit icons (copy, delete, move) and a sub-tab for 'a.'.
- Level 2:** An entry under 'a.', labeled 'Gloss', containing 'a.' followed by 'subj. gods (most exx. mid.)'. It includes edit icons and a sub-tab for '1'.
- Level 3:** An entry under '1', labeled 'Gloss', containing '2.' followed by 'in letters:'. It includes edit icons.

Below the main area are tabs for 'Citation', 'Text', and 'Publication'. At the bottom, there are sections for 'Context' (containing the quote "Let all be well with you") and 'Transcription' (containing the text "nu=tta DINGIR.MEŠ TI-an harkandu nu SAG." and "DU=KA pa-ah-ša-an-da-nu"), each with its own set of edit icons.

To summarize: in each **Meaning** tab you can create any number of subordinate **Meaning** tabs, which can in turn have any number of subordinate **Meaning** tabs. Thus, you can represent any number of meanings, submeanings, and sub-submeanings for a given lemma, using as many hierarchical levels as are needed. Furthermore, for each meaning, at any level, you can enter one or more textual quotations and discussions, which are entered in **Citation** tabs and **Discussion** tabs that belong to a parent **Meaning** tab.

A **Discussion** tab is quite simple. It functions exactly like a note tab (see the section on “Entering Textual Content in Notes and Elsewhere” in Chapter 5). It allows you to enter textual content in the

project's default language together with translations of that content into other languages, if multilingual features have been enabled for the project. Click the “translation” button on the right-hand side of the **Discussion** tab and choose the desired language. A new tab will appear beside the **English** tab (or the tab labeled **Français**, **Deutsch**, **Italiano**, **Español**, etc., depending on the project's default language) in which you can enter the translation.

Entering Textual Citations for a Meaning or Submeaning

A **Citation** tab is more complex. It contains sub-tabs for the **Text** of the citation, which is usually a brief quotation from a longer text, and for the quotation's **Publication** references. Any number of textual citations can be adduced for a particular meaning by creating additional **Citation** tabs in the relevant **Meaning** tab.

The screenshot shows a software interface for managing citations. At the top, there are three tabs: "Citation", "Citation", and "Citation". Below them is another set of tabs: "Text" (which is highlighted in blue) and "Publication". Under the "Text" tab, there are two sections: "Context" and "Transcription".

- Context:** Contains the quote "Let all be well with you" followed by a link icon (a small square with a circle and a line) and a plus sign (+).
- Transcription:** Contains the transcription "nu=tta DINGIR.MEŠ TI-an harkandu nu SAG." followed by a link icon and a plus sign (+). Below it is another transcription "DU=KA pa-ah-ša-an-da-su" followed by a link icon and a minus sign (-).
- Translation:** Contains the English translation "May the gods keep you alive (and) protect your person (lit. head)" followed by a link icon and a plus sign (+). Below it is another translation "x-x" followed by a link icon and a minus sign (-).

The **Text** subtab within a **Citation** tab has data-entry areas for the **Context**, **Transcription**, and **Translation** of the quoted excerpt. In each of these areas, you can type textual content directly into the field on the left or you can add one or more links on the right that point to discourse units of a text item in the **Texts** category.

Wherever possible, it is a good practice to integrate dictionary entries with the texts on which the dictionary is based—in this case, by adding links to discourse units. This avoids redundant duplication of information and eliminates discrepancies between the cited textual excerpts and the underlying texts from which they are taken. Any changes or corrections made to discourse units will be automatically visible in the dictionary entries in which those discourse units are used.

To add a discourse-unit link, click the \oplus button in the main toolbar in order to display the **Linked Items** tab in the reference pane (if it is not already displayed); choose the **Texts** category; select a link target within the discourse hierarchy of the relevant text; and then click the \oplus “add link” button in the **Context** or **Transcription** or **Translation** data-entry area. Clicking the \ominus “remove link” button will remove the link that is currently selected.

In these three areas, any textual content that is directly entered on the left will override the discourse-unit links on the right. In the formatted view of the citation or of the larger dictionary entry, which you can see if you click the **Preview citation** button at the bottom of the **Citation** tab or if you click the **Preview meanings** or **View full entry** buttons, the textual content that has been entered in the fields on the left side of the **Context**, **Transcription**, and **Translation** areas will be shown, rather than the discourse units (if any) that have been added on the right. However, if the text-entry field on the left is empty, the discourse units will be displayed.

The **Context** area would normally contain a paraphrase or summary, written in the project’s default modern language, of the preceding part of the text leading up to the quoted excerpt, or it would contain some other explanation of the context of the quotation. This information is optional.

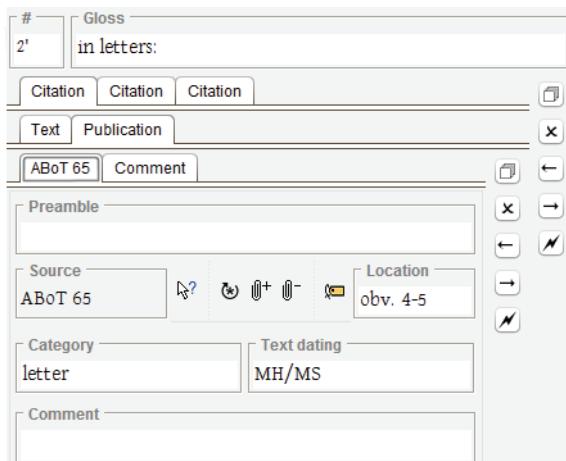
The **Transcription** area contains the actual textual quotation in its original language. You can type the quotation directly into the field on the left or you can represent it via one or more links to discourse units, as noted above. If the discourse units you have entered are not continuous but represent fragments of a larger text that you do not wish to quote in full, you can insert an ellipsis (“...”) after the most recently added discourse unit by holding down the “Shift” key when clicking the \oplus “add link” button.

If the field on the left side of the **Transcription** area is empty, the concatenated phonemic transcriptions of the linked discourse units will be displayed in the formatted view of the dictionary entry. However, it is often desirable to display the word in the quotation that corresponds to the lemma of interest in a sign-by-sign graphemic or

hybrid transcription, rather than in a phonemic transcription (see the section on “Graphemic, Phonemic, and Hybrid Transcriptions” above in Chapter 11). To do this, select the relevant discourse-unit link in the list of links on the right side of the **Transcription** area and click the ✕ “show signs” toggle button. This will cause the selected discourse unit to be displayed in a sign-by-sign transcription.

The **Translation** area contains a translation of the textual quotation in the project’s default modern language. The translation can be entered by means of one or more discourse-unit links, in which case the translations of the discourse units will be concatenated in the formatted view of the dictionary entry.

In addition to a quotation **Text** tab with its **Context**, **Transcription**, and **Translation** areas, each **Citation** tab also contains a **Publication** tab in which to enter one or more references for the quotation. Multiple references for the same quotation are allowed because the textual content that is being quoted might appear in more than one manuscript copy or published edition. When you click the ☒ “create new” button inset on the right side of the **Publication** tab, you will be given a choice either to create a **Ref** tab in which to enter a textual reference or to create a **Comment** tab in which to enter a comment about the reference. Any number of **Ref** tabs and **Comment** tabs can be created within the **References** tab.



Clicking the ✕ “delete” button will delete the currently selected tab. Clicking the ← “move left” and → “move right” buttons will reorder the tabs, determining the order in which the references and comments will be displayed in the formatted view of the dictionary entry.

A **Ref** tab has data-entry fields for a **Preamble**, **Source**, **Location**, **Category**, **Text dating**, and **Comment** that pertain to the textual reference. The **Source** field contains the name of the text. After you type in the name, you can click the  “find text” button to try to link it automatically to a text item in the project’s **Texts** category. The contents of the **Source** field will become the title of the **Ref** tab after you click **Save entry** to save what you have entered in the tab. The **Location** field contains the page, column, line, or other location where the quotation appears within the text named in the **Source** field. The **Preamble**, **Category**, **Text dating**, and **Comment** fields are optional and will often be left blank.

You can assign properties to a reference. You can also attach links to other database items. To assign properties, click the  button in the main toolbar in order to display the **Linked Items** tab in the reference pane (if it is not already displayed) and then click the  “show properties” button to the right of the **Source** field to make this field active in the **Properties and links for...** area in the bottom of the reference pane. At that point, you can enter properties for the reference in the **Properties** tab and save them by clicking the **Save** button in that tab.

To assign links, click the  “show links” button to the right of the **Source** field; select the link-target item in the reference pane; and then click the  “add link” button to the right of the **Source** field. The link will appear in the **Links** tab in the bottom of the reference pane. Clicking the  “remove link” button will remove the currently selected link from the **Links** tab.

Entering Bibliographic References and Dictionary Cross-references

The **Bibliography** subtab within the **Dictionary Entry** tab contains bibliographic references for the current entry. These references can be entered as links to items in the **Bibliographies** category or as free-form textual content.

To add a link to a bibliographic-entry database item, click the  button in the main toolbar in order to display the **Linked Items** tab in the reference pane (if it is not already displayed). Then select a link

target within the **Bibliographies** category and click the \oplus “add link” button to the left of the **Bibliographic entries** field. Clicking the \ominus “remove link” button will remove the link that is currently selected in this field. Clicking the \uparrow “move up” and \downarrow “move down” buttons will reorder the links, determining the order in which the bibliographic references will appear within the formatted dictionary entry.

To enter a bibliographic reference as free-form textual content, click the  “create new” button on the right side of the **Bibliography** tab. This will create a new paragraph tab labeled  in which you can type a bibliographic reference (see the section on “Entering Textual Content in Notes and Elsewhere” above in Chapter 5). Any number of paragraph tabs can be added in this way; it is a good practice to use a separate tab for each bibliographic reference. Clicking the  “delete”



button will delete the currently selected tab. Clicking the \leftarrow “move left” and \rightarrow “move right” buttons will reorder the tabs, determining the order in which the bibliographic references will be displayed in the formatted view of the dictionary entry.

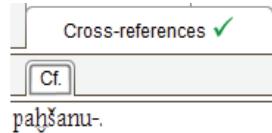
Click the **Save entry** button in the bottom left corner to save what you have entered. Clicking the **Preview bibliography** button will display a window in which you can see how the data that has been entered in the **Bibliography** tab will be formatted.

The **Cross-references** subtab within the **Dictionary Entry** tab contains cross-references from the current entry to other dictionary entries. These cross-references can be entered as links to items in the **Dictionaries** category or as free-form textual content.

To add a link to another dictionary entry, click the \ominus button in the main toolbar in order to display the **Linked Items** tab in the reference pane (if it is not already displayed). Then select a link target within the **Dictionaries** category and click the \oplus “add link” button to the left of the **Dictionary entries** field. Clicking the \ominus “remove link”

button will remove the link that is currently selected in this field. Clicking the ↑ “move up” and ↓ “move down” buttons will reorder the links, determining the order in which the dictionary cross-references will appear within the formatted dictionary entry.

To enter a dictionary cross-reference as free-form textual content, click the  “create new” button on the right side of the **Cross-references** tab. This will create a new paragraph tab labeled Cf. in which you can type a cross-reference (see the section on “Entering Textual Content in Notes and Elsewhere” above in Chapter 5). Any number of paragraph tabs can be added in this way; it is a good practice to use a separate tab for each cross-reference. Clicking the  “delete” button will delete the currently selected tab. Clicking the ← “move left” and → “move right” buttons will reorder the tabs, determining the order in which the cross-references will be displayed in the formatted view of the dictionary entry.



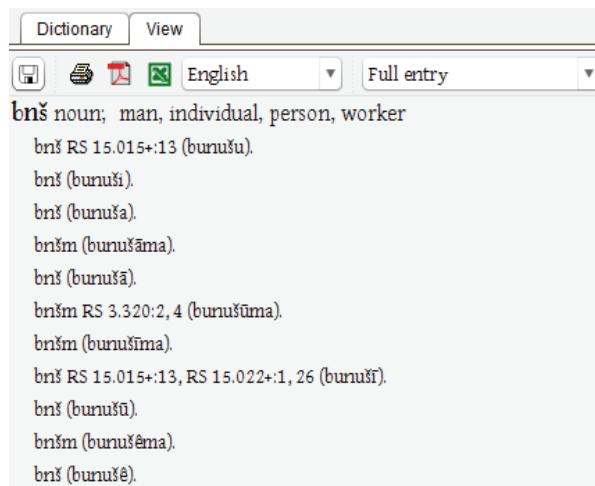
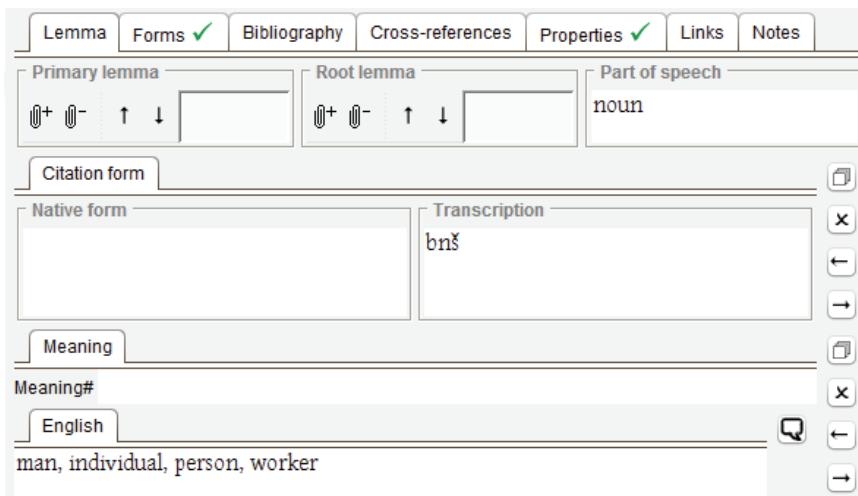
Click the **Save entry** button in the bottom left corner to save what you have entered. Clicking the **Preview dictionary cross-references** button will display a window in which you can see how the data that has been entered in the **Cross-references** tab will be formatted.

Constructing Simplified Glossary Entries

Each item hierarchy in the **Dictionaries** category contains items that represent the entries within a particular dictionary. If you select such a hierarchy in the navigation pane you will see a field for the **Type of dictionary** in the item pane, which can be “extended” or “glossary.” A dictionary hierarchy can consist of items that represent extended entries or items that represent simpler glossary entries, but not both.

The procedure for constructing an extended dictionary entry has been described above. A glossary entry contains a subset of the tabs available for an extended entry. If you select an item within a glossary-type hierarchy in the **Dictionaries** category and then open the **Dictionary Entry** tab in the item pane, the following overlapping data-entry subtabs will be displayed: **Lemma**, **Forms**, **Bibliography**, **Cross-references**, **Properties**, **Links**, **Notes**, and **Events**. The **Comparisons** and

Meanings tabs that are displayed for an extended entry will not be available. Furthermore, the **Lemma** tab will not contain fields for **Loan word**, **Part of speech**, **Attested writing**, or **Text dating**. Otherwise, the various tabs and fields in a glossary entry will function in the manner described in the preceding sections in this chapter.



Displaying Dictionary Entries

The individual components of a dictionary entry, which may be quite numerous and may be organized in a complex way within the entry,

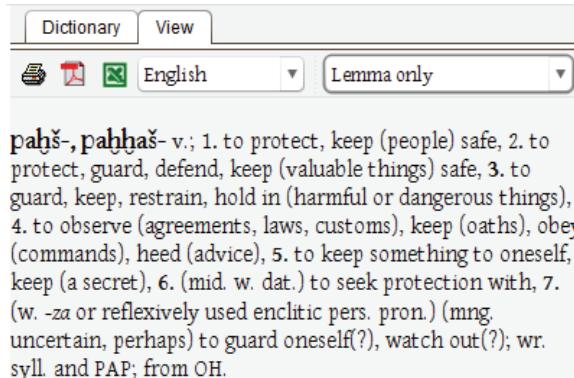
are presented to end-users in a readable, formatted display that mimics the appearance of a printed dictionary article. When you are in data-entry mode you can see the formatted display of the currently selected dictionary entry by opening its **View** tab in the item pane. In view-only mode, this is the only way in which a dictionary entry will be displayed; the underlying data-entry fields and tabs will be hidden from anonymous view-only users who are not logged into the project. You can use the special tools available for items and hierarchies in the **Presentations** category to construct an easy-to-use public presentation of your project's dictionary entries for such users, making use of alphabetical indexes and other navigation aids (see Chapter 16).

Although the complexity of the underlying data fields is hidden from end-users of the dictionary, each dictionary entry's internal structure remains visible and accessible in the formatted display of the entry. In other words, the formatted view of an entry has a "live" connection to the database items from which it has been generated. The different structural components of an entry can be displayed in different colors and typefaces, providing visual cues concerning their function. A pointing-hand cursor will appear as the mouse cursor rolls over various parts of the entry, indicating that by clicking at that spot the user can open a window to display information about the selected component.

Moreover, the dictionary entry as a whole can be displayed in various abbreviated forms, hiding details that the current user does not want to see. When you display the readable, formatted view of a dictionary entry—either in the entry's **View** tab in data-entry mode or in view-only mode—you will see two drop-down pick-lists at the top of the tab or window in which the entry is displayed: one that lets you select the modern language in which the entry is to be displayed and another that lets you select a display mode. The language selected in the first drop-down list determines which translation to use when displaying the entry, in cases where an entry or part of an entry has textual content that has been entered in more than one modern language. If a particular part of the entry has not been translated into the selected language, even though another part of the entry has been

translated, then whatever has been entered in the project's default modern language will be shown in place of a translation.

The possible display modes in the second drop-down list are **Full entry**, **Forms**, **Lemma only**, and **Outline**. In the **Full entry** mode, all of the information that has been entered for the entry will be displayed. In the **Forms** display mode, the detailed semantic analysis that has been entered in the **Meanings** tab will be omitted; only the morphological analysis entered in the **Forms** tab will be displayed, along with the basic information about the lemma that has been entered in the **Lemma** tab. In the **Lemma only** mode, only this basic information will be shown.



In the **Outline** mode, all of the information entered in all of the data-entry tabs is available, but this information will be displayed in a hierarchical tree (see the example on the facing page). To expand a branch of the tree, click the ► icon to the left of the branch. To collapse a branch, click the ▼ icon. The **Outline** mode provides an easy way to see at a glance the grammatical forms of a lemma and, by expanding a branch, to see the textually attested variants of each grammatical form; likewise, you can see at a glance the meanings of a lemma and the submeanings within each meaning.

In each of the display modes, words or phrases within the dictionary entry may be highlighted or color-coded. Highlighting indicates the presence of an embedded hyperlink from the highlighted portion of the textual content to another database item (see the

section on “Embedding Properties and Hyperlinks within Textual Content” in Chapter 5). When the mouse cursor rolls over a highlighted hyperlink, the cursor will change to a pointing hand and clicking the mouse will display a window that contains information about the linked item. Color-coding indicates the language or function of the word or phrase. To enable hyperlink-highlighting or color-coding, click the “display options” button in the View toolbar. This will pop up a window with checkboxes for customizing the current display.

English

Outline

▼ Abc pa-h̥-j-

- ▼ act. pres. sg. 1 ...
 - ▼ act. pres. sg. 1
 - Abc pa-ab-h̥-ha-aš-mi
 - Abc pa-ab-h̥-ha-aš-ji
 - Abc PAP-aš-ji
 - Abc PAP-ab-ji
 - sg. 2
 - pl. 1 ...
- ▼ pret. sg. 1 ...
 - ▼ pret. sg. 1
 - Abc pa-ab-h̥-ha-[aš-]ju-u]n
- ▼ sg. 3
 - Abc pa-ab-h̥-ha-aš-ta
 - Abc PAP-aš-ta
- pl. 3
- imp. sg. 2 ...
- pl. 2 ...
- mid. pres. sg. 1 ...
- pl. 1 ...
- pret. sg. 1 ...
- imp. sg. 3 ...
- pl. 3 ...
- part ...
- iter. imp. sg. 3 ...
- Unlike takš-, which has two alternative stem forms
- ▼ 1. to protect, keep (people) safe
 - ▼ a. subj. gods (most exx. mid.)
 - 1' in treaties:
 - 2' in letters:
 - 3' in rituals:
 - 4' in festivals:

Dictionary View

English Full entry Display options...

GIS pahhiša, GIS pahiša Luw. n. neut.; (a noisemaker); from MH/NS.†
 sg. nom.-acc. GIS pa-ah-hi-ša KBo 5.1 iv 32 (MH/NS), KBo 22.135 i 3, Bo 4951 rev. 4, 5 (StBoT 29:124), [GIS] pa-a-hi-ša Bo 4951 rev. 3 (StBoT 29:124).

Obj. of *walh-* “to beat”: (In the morning, the sacrificers purify the child) GIS pa-ah-hi-ša = ya = šši šarā walhanzi “and beat a *p.* over him” KBo 5.1 iv 32-33 (rit. of Papanikri, NH/NS) ed. Pap. 12*f. w. n. 4, StBoT 29:120; Sommer/Ehelolf read GIS PA-ah-hi-ša in KBo 5.1 iv 32 (Pap. 76); (earlier apparently they beat a stick; then) [... GIS] p]a-a-hi-ša UL walh[zi ...] nu = ššan 1 GIS pa-ah-hi-ša x [... (x x-anzi)] / [mahhan=ma=aš] (var. GIM-[an ...]) parkuišzi GIS pa-ah-hi-ša dān wal[(han)zi] “[...] he does not beat the *p.*; [...] ... thereupon they ... one *p.*, [but] when [he(?)] is pure, they beat the *p.* a second time” Bo 4951 rev. 3-5 (rit. for birth) w. dupl. KBo 17.68:2, ed. StBoT 29:124f. On the basis of Bo 4951, Helolof in OLZ 32:322f. corrected his earlier reading of the word as GIS PA”-ah-hi-ša to GIS pa-ah-hi-ša.

GIS *p.* in KBo 22.135 i 3 occurs in a rit. performed by a LU AZU “exorcist”; the other refs. occur in birth rituals. In view of the use of *walh-* elsewhere for playing percussion instruments, one should consider the possibility that *p.* was a noisemaker of some kind, made (partly?) of wood. Starke (StBoT 31:208) posits a Luwian stem *pahhi-* and translates “Trommel” or “Gong.”

Sommer/Ehelolf, Pap. (1924) 76; Helolof, OLZ 32 (1929) 322f; Friedrich, HW (1952) 153 (“Stock (?), Gerte (?);”); Beckman, StBoT 29 (1983) 122, 288 (tr. “stick(?)” employed only provisionally).

Citation URL
<http://ochre.lib.uchicago.edu/ochre.jsp?id=6eaae070-b138-0d03-1adf-b244f4fdc18b>

Searching Dictionary Entries

When displaying a dictionary entry in view-only mode, or in the entry’s **View** tab in data-entry mode, there will be a “search” button above the entry. You can type a word or phrase into the field beside this button and then click the button in order to highlight in bright yellow all occurrences of that word or phrase within the current entry. The number of matches will be displayed beside the search field.

More elaborate searches of dictionary entries can be done by means of named queries. Queries specify the scope and criteria of a

database search. They can be quite complex, depending on how many factors are to be considered when performing the search. Once a query has been defined, it is typically saved for repeated use by the same user or by other users. There is a way for the project administrator to construct complex queries relevant to the data in his or her project and to save these queries in a simplified form, making it easy for other users to obtain quick answers to commonly asked types of questions. The procedure for constructing and using queries is discussed below in Chapter 14. Examples of queries that analyze the content of dictionaries can be found there.

The screenshot shows a digital dictionary interface with the following elements:

- Top Bar:** Includes tabs for "Dictionary" and "View", and a search bar containing the word "gods".
- Search Results:** A list of 11 matches. The first entry is highlighted in yellow. It includes the Akkadian source (Akk.), the Hittite translation (ina trtiki ina ersetim us̄irini), the KUB reference (KUB 1.16 iv 71-72 = (Hitt.)), the full text (nu-mu tagga[niya-ti] taknaz pa-ab-[š]), and a detailed description: "[On your] breast protect me from the earth" (KUB 1.16 iii 72-73 (edict of Hatt. I, NS), ed. HAB 16f., cf. Götze, ZA 34:183 and Melchert, Diss. 182f.; (Akk.) u abuya ^mAz̄i[ra qadu KUR-šū: ittas̄a-šū KUB 3.14 obv. 7-8 = (Hitt.) [ABU=YA ^mAz̄i]an QADU KUR-šū pa-ab-ja-aš-ta-at "My father protected Aziru together with his land" KUB 21.49 obv. 6 (treaty w. Duppi-Teššub, Murš. II), ed. SV 1:6f.).
- List of Matches:** Below the first entry, a numbered list provides definitions for the term "gods":
 1. to protect, keep (people) safe — a. subj. gods (most exx. mid.) — 1' in treaties: nu mān kiš līngāuš(!) pa-ab-ja-aš-du-ma šumāš-a DINGIR.MEŠ-éš pa-ab-ša-an-da-nu (var. pa-ab-ša-an-ta-nu) "If you keep these oaths may the gods protect you too" KBo 8.35 ii 14-15 (treaty w. Kaška people, MH/MS) w. dupl. KUB 40.36 + KUB 23.78b ii 10-11 + KUB 26.6:11-12, tr. Kalkær 111, cf. also 4 c, below; nu-ata kiš NIŠ DINGIR.MEŠ ... SILIM-!-li pa-ab-ša-an-[t]a-ru KBo 4.10 + KBo 50.60 (ZA 63:86) rev. 9-10 (treaty w. Ulm-Teššub, Hatt. III or Tudj. IV), ed. van den Hout, Diss. 36f., see CHD *līngāt*-1 e (now w. join 1548/u this passage belongs s.v. *līngāt*-2 b 1'), cf. KBo 5.3 ii 11-12 (treaty w. Ḫuqqana, Supp. I), etc. (see index of forms in SV 2:202).

Chapter 13

Organizing Terms in a Taxonomy and Thesaurus

Each project in OCHRE has a special hierarchy in its **Taxonomy** category that specifies the way in which terms defined by the project can be used to describe database items. This hierarchy consists of database items that represent the taxonomic terms themselves—that is, the project’s variables and values. The hierarchy prescribes the list of possible values for each qualitative variable. It also prescribes how particular variable-value pairs (properties) can be nested within other, more general, variables.

Apart from the project’s taxonomic hierarchy, each variable can be semantically related to other variables by means of thesaurus relationships. Likewise, qualitative values can be semantically related to other values. Thesaurus relationships are entered via the **Thesaurus** tab of each variable or value item. You can display all existing thesaurus relationships by clicking the **Property variables** or **Property values** category heading in the navigation pane and then opening the category’s **Thesaurus** tab in the item pane. This will show you an alphabetical list of the currently defined variables or values, as the case may be, together with their synonyms. To see a complete thesaurus of both variables and values, intermingled in a single list of terms, select the **Thesaurus** category heading in the navigation pane and open its **View** tab in the item pane.

This chapter explains the procedure for constructing a project’s taxonomy. It also explains the procedure for establishing thesaurus relationships among the terms used in the taxonomy and for displaying these relationships in an alphabetically arranged thesaurus. A project’s taxonomy and thesaurus are powerful tools that enable researchers to define their own scholarly nomenclature and to exploit the semantic relationships among terms in the course of describing and searching items of interest to them.

Constructing a Project’s Taxonomy

Before anyone can use properties (variable-value pairs) to describe a project’s database items, the project administrator must construct a taxonomic hierarchy. When a project is first created, its **Taxonomy** category will be empty. Select the category heading in the navigation pane and click the ↗ “insert below” button in the main toolbar in order to create a taxonomic hierarchy. Only one hierarchy is permitted in the **Taxonomy** category.

The item pane of the taxonomic hierarchy will contain data-entry fields for entering its **Name**, **Abbreviation**, and **Description**. It will also have a tab for entering the hierarchy’s **Creators** and the date and time when it was created. At the bottom of the pane will be tabs for the hierarchy’s **Preferences**, **Properties**, **Links**, and **Notes**. These fields and tabs function in the same way as they do for any other hierarchy (see the sections on “Inserting New Items” and “Display Preferences for Hierarchies and Categories” above in Chapter 4).

Before you are permitted to insert anything into the taxonomic hierarchy, you will first be required to lock the “Inbox” hierarchies in the **Property variables** and **Property values** categories (see the section on “Using an ‘Inbox’ to Hold New Items” above in Chapter 4). The “Inbox” hierarchies may be hidden if they are empty. To display them, select the **Property variables** and **Property values** categories in the navigation pane and turn on the **Show inbox even if empty** checkbox in the **Preferences** data-entry area of the item pane. You can then lock each “Inbox” hierarchy by selecting it in the navigation pane and clicking the  “lock” button in the main toolbar.

To start building the taxonomic hierarchy, select the hierarchy in the navigation pane and click the ↗ “insert below” button to insert a new variable item at the top level of the hierarchy. The same variable item will be inserted automatically in the “Inbox” of the **Property variables** category. Any number of variables, of any type, can be inserted at the top level of the hierarchy.

As an alternative to creating a new variable item you can insert an existing variable in the taxonomic hierarchy. To do so, click the  toggle button in the main toolbar to display the **Linked Items** tab of

the reference pane (if it is not already displayed); then select an existing variable item in the **Property variables** category as the link target; and then click the  “repeat link target” toolbar button to copy that variable into the taxonomic hierarchy. However, it is often more convenient to create a variable in both places at once by inserting it in the taxonomic hierarchy via the  button, rather than first having to create it in the **Property variables** category and then copying it into the taxonomic hierarchy via the  button.

Keep in mind that a variable inserted or copied into the taxonomic hierarchy in the **Taxonomy** category is not a different item from its counterpart in the **Property variables** category. It is merely another manifestation of the same database item; thus, any changes to the variable will be reflected automatically in both places. The same is true for each qualitative value in the taxonomic hierarchy, which is a manifestation of the same item in the **Property values** category.

Creating a New Variable

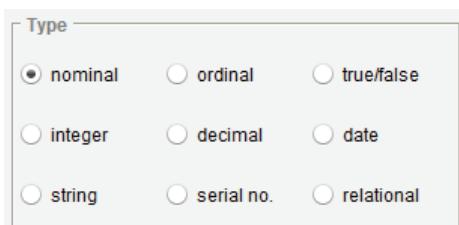
After a new variable item has been inserted in the taxonomic hierarchy (or has been inserted in the **Property variables** category, to be repeated later within the **Taxonomy** category) the item pane will display data-entry fields for the variable’s **Name**, **Abbreviation**, and **Description**. You would enter textual content in these three fields in the way you would for any other item, perhaps also entering aliases for the name and translations of the name and description in languages other than the project’s default language (see the section on “Item Names and Aliases” in Chapter 4).

In the **Type** area below the **Description** field you have a choice of nine different variable types. The variable’s type determines the kind of value it can be paired with to form a property (see Chapter 5 on assigning properties to items as variable-value pairs). The possible types are:

- **nominal**, which means that the variable must be paired with a character-string value that is represented as a database item in the **Property values** category, and the variable’s values consist of names or classifications that have no inherent order;

- **ordinal**, which means that the variable must be paired with a character-string value that is represented as a database item in the **Property values** category, and the variable's values consist of names that do have an inherent order;
- **true/false**, which means that the variable must be paired with a Boolean value representing “true” or “false”; this choice between two and only two values is represented in the interface by means of a checkbox, that is, a check entered in the checkbox represents the value “true”;
- **integer**, which means that the variable must be paired with a numeric integer value;
- **decimal**, which means that the variable must be paired with a numeric decimal value;
- **date**, which means that the variable must be paired with a value that represents a calendar date;
- **string**, which means that the variable must be paired with a character-string value that may contain letters or numbers, or both, but does not exist as a database item in the **Property values** category;
- **serial no.**, which means that the variable must be paired with an integer value that is automatically incremented and is guaranteed to be unique for that variable;
- **relational**, which means that the variable must be paired with a value that is a link to another database item in the current project or in a different project (see Chapter 6).

The variable's type determines the kinds of statistical operations that can be performed on its values. Statisticians distinguish four basic levels or scales of measurement. At the nominal (or categorical) level of measurement, values consist of names or classifications that have no inherent order. For example, the variable “Gender” can take the values “male” and

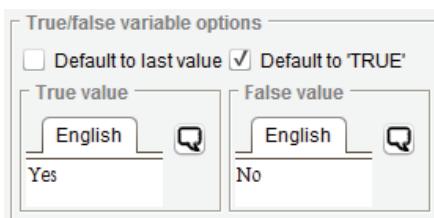


“female.” At the ordinal level of measurement, values have a relative order but the differences between the values are not meaningful. For example, the variable “Size” can have the ordered values “small,” “medium,” or “large.” At the interval level of measurement, values have a relative order and are measured on a constant scale, so that the differences between adjacent values represent equivalent intervals, but the scale has no zero point. This means that there can be no meaningful arithmetic ratios of values. For example, calendar dates are interval values because the ratio between the year A.D. 2000 and the year A.D. 1000 (for instance) is meaningless. Finally, at the ratio level of measurement, values have a relative order and are measured on a constant scale that does have a zero point, permitting meaningful arithmetic ratios of values. For example, the variable “Weight” would have numerical values whose ratios are meaningful, allowing you to say, for example, that an object that weighs 20 kilograms is twice as heavy as an object that weighs 10 kilograms.

In keeping with this distinction between levels of measurement, OCHRE distinguishes **nominal** and **ordinal** variables from other types of variables and from one another. They belong to the nominal (or categorical) and the ordinal levels of measurement, respectively. Variables of the **true/false**, **string**, **serial no.**, and **relational** types are special kinds of nominal variables, whereas **date** variables belong to the interval level of measurement and **integer** and **decimal** variables belong to the ratio level.

Additional data-entry fields will appear to the right of the **Type** field. The fields shown will depend on the type. All variable types have a **Default to last value** checkbox which, if it is checked, causes the last value assigned to the variable to be used as its default value the next time the variable is employed.

True/false variables have a **Default to ‘TRUE’** checkbox which, if it is checked, causes the value of the variable to be set to “true” by default whenever it is used to describe an

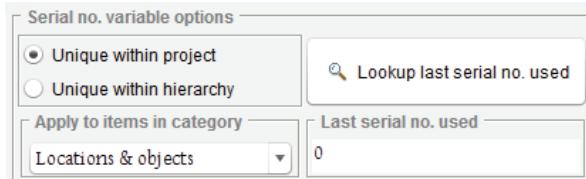


item. You can customize the appearance of the “true” and “false” values by entering textual content in the **True value** field and/or the **False value** field. For example, you might want to display the “true” value as “Yes” and the “false” value as “No.”

For an **integer** or **decimal** variable there will be optional fields for entering the variable’s units of measurement, the minimum and maximum values allowed for the variable, and its default value. These fields are discussed in detail below in the section on “Working with Quantitative Properties.”

For a **date** variable there will be a checkbox that lets you specify that the value of the variable should automatically **Default to current date** whenever it is used to describe an item.

For a **serial no.** variable you can specify whether the integer value of the variable must be **Unique within project** or merely **Unique within hierarchy**, that is, whether users will be prevented from entering the same number twice when using this variable to describe two different items, regardless of which hierarchies the two items are in, or whether duplicate numbers could be entered using this variable as long as the items being described are



in different hierarchies. The integer value of a **serial no.** variable is automatically incremented every time it is used by adding one to the previous value assigned to it, starting from zero; but you can force the internal counter to start at a different number or to skip a range of numbers by typing a number in the **Last serial no. used** field. You can click the **Find last serial no. used** button to update the field to the last number used. The uniqueness restrictions and the internal counter of a given **serial no.** variable pertain to the use of that variable to describe items within a particular item category; thus the **Apply to items in category** field is used to specify which category. Note that regardless of whether the variable is **Unique within project** or **Unique within hierarchy**, its serial number sequence is independent of the sequences used by other **serial no.** variables.

For a **relational** variable there will be a drop-down pick-list entitled **Category of related item** in which to specify the category of items that can serve as link targets for the variable (see the section on “Relational Properties” above in Chapter 6). You can also enter the **Name of inverse relationship**. For example, if you enter “Teacher of” in the variable’s **Name** field you might enter “Student of” in the **Name of inverse relationship** field. If you do not enter an inverse name, the name of the inverse relationship will default to the variable’s name.

Below the **Type** area, at the bottom of the item pane, are tabs for **Preferences**, **Thesaurus**, **Links**, and **Notes**. The **Preferences** tab contains a checkbox to **Keep this item private** (i.e., make it invisible to anonymous users) and a drop-down pick-list to determine the **Sorting method for child items** (see the section on “Display Preferences for Hierarchies and Categories” above in Chapter 4 for a description of the available sorting methods).

The sorting method pertains to **nominal** variables. It determines the display order of the values that are the children of the variable in the taxonomic hierarchy. If the sorting method is set to **None** the values will be displayed in the order in which they were inserted as children of the variable. Variables of the **ordinal** type can also have child values in the taxonomy; however, the sorting method does not apply to them. The values of an **ordinal** variable are not sorted automatically because their order is significant, indicating their relative positions in an ordered series. For example, the ordinal variable “Size” might have the value “small”, “medium,” or “large,” and the fact that these values are listed in that order in the taxonomy indicates which is bigger and which is smaller.

A variable’s **Preferences** tab also contains a checkbox entitled **Inherited by contained items**. If you check this box, any property (variable-value pair) that uses this variable will appear automatically among the properties of every item that is hierarchically contained within the item to which it was originally assigned. This is explained in more detail in the section on “Inherited Properties” above in Chapter 5.

The **Thesaurus** tab is discussed below in the section on “Entering Thesaurus Relationships for a Variable or Value.” The **Links** and **Notes** tabs function as they would for any other item.

Prescribing the Values Allowed for a Nominal or Ordinal Variable

At the top level of the taxonomic hierarchy in the **Taxonomy** category you can insert only variables (which are also repeated in the **Property variables** category). At the next level of the hierarchy you can insert, not variables, but only values (which are automatically repeated in the **Property values** category).

Moreover, only **nominal** and **ordinal** variables can have child items. Their child items prescribe the qualitative values that are allowed for the parent variable. Other types of variables cannot have child items; thus, there is no further taxonomic branching beneath variables of the **integer**, **decimal**, **string**, **date**, **true/false**, **serial no.**, or **relational** types. However, a taxonomic branch can be constructed beneath a child value of a **nominal** or **ordinal** variable because such values can have child items of their own consisting of variables of any type. And the **nominal** and **ordinal** variables that are children of a value can, in turn, have child values of their own in the next level of the hierarchy. Thus, you can construct a taxonomic hierarchy of arbitrary depth by alternating variables and values at successive levels within the hierarchy.

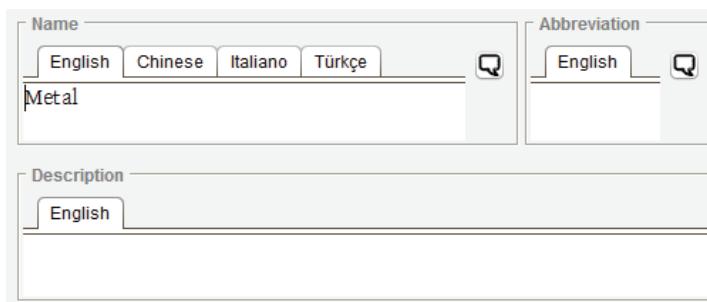


If you select a **nominal** or **ordinal** variable in the taxonomic hierarchy and click the ↗ “insert below” button, a new value item will be inserted as a child of the variable. During subsequent data entry, only the values entered as children of a variable can be paired with that variable to form a property (see the section on “Assigning Properties to an Item” above in Chapter 5). If you assign a **nominal** or **ordinal** variable to an item during data-entry, you will be provided with a drop-down pick-list of the character-string values allowed for that variable based on the variable’s current children in the taxonomy. In contrast, values must be entered directly for **integer**, **decimal**, **string**, **date**, **true/false**, **serial no.**, and **relational** variables. As noted above, the order of an **ordinal** variable’s child items as they are listed in the taxonomic hierarchy indicates the relative positions of that variable’s possible values within an ordered series and this ordering can be used in queries (e.g., a query can understand that “small” is less than “medium,” which is less than “large”).

As an alternative to inserting a new value item directly into the taxonomic hierarchy, click the ⌂ toggle button in the main toolbar to display the **Linked Items** tab of the reference pane (if it is not already displayed), then select an existing value item in the **Property values** category as the link target and click the ↗ “repeat link target” toolbar button to insert that value into the taxonomic hierarchy. However, it is often more convenient to create a value in both places at once by inserting it in the taxonomic hierarchy using the ↗ button, rather than having to create it first in the **Property values** category and subsequently link it into the taxonomic hierarchy using the ↗ button.

Keep in mind that a value inserted into the taxonomic hierarchy in the **Taxonomy** category is not a different item from its counterpart in the **Property values** category. It is merely another manifestation of the same database item; thus, any changes to the value will be reflected automatically in both places. The same is true for each variable that appears in the taxonomic hierarchy, as has already been noted above. A variable in the taxonomy is a separate manifestation of the same item in the **Property variables** category.

After a new value item has been inserted in the taxonomic hierarchy (or has been inserted in the **Property values** category, to be repeated later within the taxonomic hierarchy), it will be selected in the navigation pane and the item pane will display data-entry fields for the value's **Name**, **Abbreviation**, and **Description**. You would enter textual content in these three fields in the way you would for any other item, perhaps also entering aliases for the name and translations of the name and description in languages other than the project's default language (see the section on "Item Names and Aliases" in Chapter 4). The value item's **Name** field contains a character string that can be paired with the contents of a variable item's **Name** field to form a property (see Chapter 5).

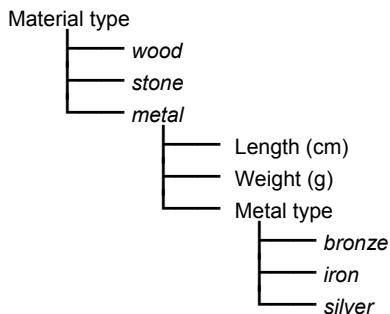


At the bottom of the item pane are tabs for the value item's **Preferences**, **Thesaurus**, **Links**, and **Notes**. The **Preferences** tab contains a checkbox to **Keep this item private** (i.e., make it invisible to anonymous users) and a drop-down pick-list to determine the **Sorting method for child items** (see the section on "Display Preferences for Hierarchies and Categories" above in Chapter 4 for a description of the available sorting methods). The sorting method determines the display order of the variables, if any, that are the children of the value in the taxonomic hierarchy. If the sorting method is set to **None** the variables will be displayed in the order in which they were inserted as children of the value.

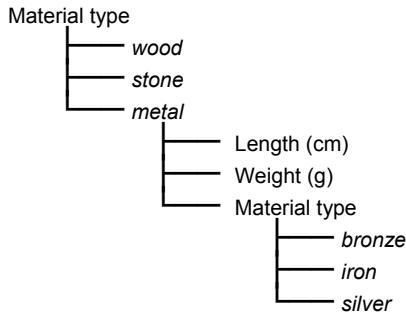
The **Thesaurus** tab is discussed below in the section on "Entering Thesaurus Relationships for a Variable or Value." The **Links** and **Notes** tabs function as they would for any other item.

Representing Relationships between General and Specific Terms

The main purpose of allowing you to construct a deep taxonomic hierarchy—one in which the values of variables contain lower-level variables, which in turn contain values that themselves contain variables, and so on—is to enable you to specify relationships of logical subordination between more general terms and more specific terms. For example, the property “Material type = metal” might have the subordinate property “Metal type = bronze” nested within it in the description of a particular item. To enable this in the taxonomy, you would insert a value item named “metal” as a child of a variable item named “Material type”; then insert a variable named “Metal type” as a child of the “metal” value; and then insert a value named “bronze” as a child of “Metal type” (see the diagram below). The procedure for using taxonomically organized variables and values to assign subordinate properties to items is described in the section on “Nested Properties” in Chapter 5.



Alternatively, to avoid proliferating variables, the same variable can be nested recursively within itself. Using the example given above, “Material type” would still contain “metal” as one of its values, but that value would in turn contain the same variable “Material type,” within which would be the more specific values “bronze,” “iron” and “silver” at a lower level in the taxonomic hierarchy (see the diagram below). This provides a way to represent simultaneously the fact that both metal and bronze are types of material and the fact that bronze (or iron, or silver, etc.) is a more specific type.



In either case, the taxonomic hierarchy is an alternating hierarchy of variables that contain values, which in turn contain variables, and so on. For each **nominal** or **ordinal** variable you *must* enter one or more child values. For each of these values you *may* enter a child variable, recognizing that some variables are quite specific and are needed only within the context of another, more general, variable; or perhaps only in the context of a particular value of the more general variable. The same variables and values can be reused at different locations in the taxonomic hierarchy, depending on what is needed in a particular context.

Organizing a project's descriptive terms (i.e., its property variables and values) in a hierarchical taxonomy has two main benefits: it facilitates the entering of item properties and it represents explicitly the logical relationships between general and specific properties in a way that enables more powerful queries. The taxonomy constrains data entry, preventing errors and automatically generating context-specific pick-lists of permitted variables and values.

The taxonomy also allows automatic query expansion in order to include more specific values of a recursively nested variable. Using the example illustrated above, a database query whose search criteria included "Material type = metal" would automatically retrieve items with the more specific properties "Material type = bronze" or "Material type = iron" or "Material type = silver," as well as items with the general property "Material type = metal." The recursive taxonomic hierarchy formally represents the fact that bronze, iron, and silver are specific types of metal.

Borrowing from another Project’s Taxonomy

For some portions of your project’s taxonomy, you may wish to borrow a taxonomic arrangement of terms that has previously been defined by another OCHRE project in order to save yourself the trouble of creating it on your own. Projects within the same discipline will often use a standard set of terms to describe particular kinds of material. For example, philological projects that deal with similar languages and texts may use a common grammatical taxonomy. Likewise, a number of different archaeological projects may share the same pottery typology. Rather than laboriously re-entering the same variables and values, you can borrow a variable, a value, or an entire taxonomic branch of nested values and variables by using OCHRE’s “bound branch” mechanism (see the section on “Repeating Items and Hierarchical Branches” in Chapter 4).

Note that the project administrator of the lending project must first make that project’s taxonomy available to the person who is doing the borrowing by creating a user account for the borrower that allows view-only access to the lending project’s **Taxonomy** category (see Chapter 17). Whatever is copied by the borrower from the lending project’s taxonomy will become available to all other users of the borrower’s project for their data entry and querying, even though they do not have user accounts in the lending project.

To borrow a variable, value, or a branch of nested variables and values from another project’s taxonomy, select the appropriate parent in your own project’s taxonomic hierarchy; then use the **Linked Items** tab of the reference pane to select the variable or value that is to be borrowed from the **Taxonomy** category of the lending project, making it the link target; and then click the  button “repeat link target with bound branch” on the main toolbar. This will insert a copy of the target variable or value and all its descendants into your project’s taxonomic hierarchy.

Note that if you have borrowed, not just a single item, but a taxonomic branch, the internal organization of the borrowed branch cannot be changed. Its child items cannot be moved, deleted, or inserted within your project because the branch structure is still

connected to the original branch in such a way that any changes made in the original branch by the lending project will be automatically shown in the copied branch. As a matter of good practice, therefore, you should borrow only from projects whose taxonomies you can trust to remain unchanged or to change only in ways that are beneficial to your own project. Wherever it is feasible, taxonomic branches that represent standardized terminologies of general utility will be set up within the “OCHRE” parent project from which all projects descend. Any project can borrow from the parent “OCHRE” project at any time with the assurance that its taxonomy will not change in a way that causes problems for the borrowing projects.

Constructing a Project’s Thesaurus

Among researchers in humanistic and social-scientific disciplines—and in some branches of the natural sciences—there is often no standardized terminology for describing entities of interest. Different researchers frequently use different terms to describe the same thing. This hinders the automated integration of data across research projects because database software has no way of knowing when two different terms are synonymous. To solve this problem, there have been many attempts over the years to devise and impose standardized terminologies that are intended to be used by all researchers in a given domain. But this solution has usually proved to be unworkable, not just because many scholars are (perhaps egotistically) wedded to their own terminologies, but because their differing terminologies often reflect divergent research agendas and interpretive perspectives. Legitimate disagreements about what to call things and what distinctions ought to be made should not be suppressed by a computer database system.

For this reason, OCHRE does not prescribe a standardized terminology. Each project can define its own terms and organize them in its own taxonomy, as explained above. Furthermore, there is no standardized thesaurus of related terms. Each project can define its own thesaurus relationships to permit the automatic matching of terms. This reflects the fact that a thesaurus is itself a work of

scholarship that should be credited to the scholars who created it and should be used for automatic term-matching at the discretion of each scholar who is searching and comparing data derived from different researchers. Thus, when performing a query, an OCHRE user can decide which project's thesaurus, if any, should be used to match different but potentially synonymous terms.

To put it another way: meaning depends on context. The meanings of terms and their relationships to other terms cannot be formalized once and for all and then applied in an anonymous, automated fashion without regard to the specific context in which they are being used. This is especially true in humanistic disciplines, which are characterized by a wide range of scholarly traditions and interpretive perspectives. As a matter of principle, then, OCHRE does not replace the scholarly choices of individual researchers with a single built-in taxonomy and thesaurus, or with "intelligent" search algorithms that match terms in an obscure (to the user) and potentially misleading manner. To do so would be to deprive scholars of the ability and the responsibility to describe and interpret entities in their own way—and in a manner that is transparent to other scholars, so that their choices can be discussed and contested. As is explained in more detail below in Chapter 19, the design of the OCHRE system is based on the principle that many of the semantic distinctions important to scholars cannot be adequately formalized in advance and therefore should not be built into the internal structure of a database system but should be a matter of ongoing scholarly interpretation and debate. What a database system can do is make it easy for scholars to record all of their observations and distinctions in a reproducible matter, including their interpretations of the meanings of terms, thus greatly facilitating the sharing of information.

Aliases versus Thesaurus Relationships

Thesaurus relationships serve to link pairs of variables and pairs of qualitative values, indicating that two variable names or two values are synonymous or are semantically related in some other way. This mechanism is used mainly to integrate the terminologies of different

projects with the goal of enabling database queries to search across multiple projects and find similar items that have been described using different, but synonymous, terms. It is a manageable task to link the terms used in one project to those in another because most projects use a limited number of different variables and values—numbering in the hundreds, not the thousands. In any case, this task needs to be performed only once for each pair of projects whose terminologies are integrated.

Thesaurus relationships are often unnecessary for relating pairs of terms employed *within* your own project because you can use instead the alias and translation mechanism described above in Chapter 4 (in the section on “Item Names and Aliases”). The **Name** field of any database item, including the **Name** field of a variable or value item, can contain one or more aliases in the project’s default language, as well as translations into other languages. An **Alias** entered in the **Name** field of a variable item or a value item serves the same purpose as a thesaurus “synonym” relationship.

Different project teams who are in communication with one another also have the option of agreeing on a common set of terms, which can be incorporated (upon request) into the master taxonomy of the “OCHRE” parent project, from which variables and values can be borrowed by any other project (see the section above on “Borrowing from another Project’s Taxonomy”). If the participating projects use different default languages, translations entered in the **Name** fields of the variables and values would allow each project to use the appropriate names in its own default language to refer to the same taxonomic items, avoiding the need for thesaurus relationships.

Before anyone can build thesaurus relationships for variables and values, the project administrator must construct a thesaurus hierarchy. When a project is first created, its **Thesaurus** category will be empty. Select the category heading in the navigation pane and click the ↳ “insert below” button in the main toolbar in order to create a thesaurus hierarchy. Only one hierarchy is permitted in the **Thesaurus** category.

Entering Thesaurus Relationships for a Variable or Value

To enter thesaurus relationships for a variable or a qualitative value, select the relevant database item in the **Property variables**, **Property values**, or **Taxonomy** category in the navigation pane and then open the selected item's **Thesaurus** tab in the item pane. There are two columns within this tab: one for **Synonyms** and one for **Related terms** (i.e., terms that are semantically related in some way but not actually synonymous). In each column you can enter one or more links to other variables or other values, as the case may be. A linked variable or value can be selected from within your own project or from another OCHRE project to which you have access.

A link can be added to the **Synonyms** or **Related terms** column by selecting the link target in the **Linked Items** tab of the reference pane and then clicking the \oplus “add link” button at the top of the column (see Chapter 6). If there are multiple links, they can be reordered by means of the \uparrow “move up” and \downarrow “move down” buttons. A link can be removed by selecting it in the column and clicking the \ominus “remove link” button at the top of the column. The number of links in the column is indicated at the top of the column.

The screenshot shows the OCHRE interface with the Thesaurus tab selected for a database item named "1. Proclamation of Anitta, King of Kussara".

- Name:** English, Français, Deutsch. Abbreviation: CTH 1. Count: 0.
- Description:** English.
- Preferences:** Thesaurus (checkmark).
- Synonyms:** 1 item. List: Anitta Text.
- Related terms:** 0 items.

Chains of Thesaurus Relationships

Thesaurus relationships are bidirectional (i.e., symmetric). Thus, a project that does not define its own thesaurus relationships can benefit from relationships created by another project that involve its own variables and values, as long as the other project's thesaurus is selected in the query. For example, if Project A uses the property "Ceramic shape = bowl" while Project B uses the equivalent property "Pottery form = bowl," only one of the two projects needs to create thesaurus relationships between the synonymous terms "Ceramic shape" and "Pottery form." If Project A has indicated that its term "Ceramic shape" is a synonym of Project B's term "Pottery form," then Project B does not have to define this relationship. Project A's thesaurus relationships can be used in a query performed within Project B.

Thesaurus relationships are also transitive. This means that a query that uses thesaurus relationships to expand its search criteria can follow a chain of relationships, from the terms used in Project A to those used in Project B, and then from the terms used in Project B to those in Project C, and from Project C to Project D, so on. As a result, a project's thesaurus does not need to contain all the possible thesaurus relationships from its own terms to synonymous terms in every other project that a user might wish to search. The thesaurus relationships created in other projects can be used instead.

For example, consider a situation (illustrated below) in which Project A's variable "Ceramic shape" is linked as a synonym to Project B's variable "Pottery form" via a thesaurus relationship defined in Project A. Meanwhile, Project B's variable "Pottery form" is linked to Project C's variable "Ceramic form" via a thesaurus relationship defined in Project B. A query performed in Project A that refers to the variable "Ceramic shape" could use the thesaurus relationships defined in both Project A and Project B, allowing it to retrieve the synonyms "Pottery form" in Project B and "Ceramic form" in Project C. If someone performing a query in Project A did not wish to match terms in Project C that are not included in Project A's thesaurus, he or she would simply omit Project B's thesaurus from the query.



There are many possible semantic mappings from one set of terms to another. It is unrealistic to advocate the use of a single standard thesaurus for all projects, especially in light of the fact that researchers often disagree about the semantic relationships among terms. The choice of which thesaurus relationships are to be used to find similar items when searching across multiple projects is itself a scholarly decision that will depend on the question being asked and how the scholar asking the question understands the terminologies of the various projects.

Having said that, there may well be situations in which a single thesaurus would be used by common agreement among a particular group of related research projects, even though these projects could not manage to agree on a common terminology expressed in a single taxonomy. For example, the leaders of archaeological field projects working in the same geographical region might agree to integrate their divergent (but much cherished) terminologies—not by adopting a common set of terms, but by adopting a common thesaurus to relate those terms to one another.

Such a thesaurus could be implemented within a separate OCHRE project created solely for that purpose. This project would contain, as its only database items, the variables and values of all the participating projects. The project administrator of this dedicated thesaurus project—let us call it Project T, for the sake of convenience—would first copy these projects' variable and value items into Project T's **Property variables** and **Property values** categories. Thesaurus relationships between pairs of variables and pairs of values would then be defined in Project T using the procedure described above. At that point, a query performed from within any of the participating projects could invoke Project T's thesaurus in order to search across all of the projects for similar items that have been described using different but synonymous terms.

Working with Quantitative Properties

In the item pane of a variable of the **integer** or **decimal** type there are optional data-entry fields for entering the quantitative variable's units of measurement, the minimum and maximum values allowed for the variable, and a default numeric value that will be automatically entered whenever the variable is first assigned in a property.



The number you enter in the variable's **Minimum value** field is used during data entry to prevent anyone from entering a smaller value for the variable than the one you specify. Likewise, the value you enter in the **Maximum value** field is used to prevent anyone from entering a larger value than the one you specify. These fields provide a useful mechanism for preventing data-entry errors; however, they are optional. You can leave them empty or use just one of them, if there is no need for the other. The value you enter in the **Default value** field will be inserted automatically during data entry to save time.

Chapter 14

Using Queries to Retrieve Information

Each item in the **Queries** category represents a named query definition that contains search criteria and instructions about the scope of the search. Query definitions are saved as database items so they can be used repeatedly and shared with other users. A project administrator will normally construct a variety of predefined queries that are intended to be used frequently by members of the project team in data-entry mode and by view-only users who will study the project's data. The project administrator can choose predefined queries to display in a **Queries** tab in the navigation pane, for the convenience of view-only users. This is done via the **View-only** subtab in the **Preferences** tab of the project's item pane (see Chapter 17).

The list of database items retrieved by a query can be inspected on the spot or saved for future use as a named set of items in the **Sets** category. The query item's **View** tab in the item pane shows a list of the items found by the most recent performance of the query. You can click on any individual item in the list to display its information or you can view all of the items together using various display formats.

The scope of a query is defined in terms of the category of items to be searched and the project (or projects) to be searched. A query can span multiple projects but it will retrieve only those items to which the user has access. Items marked "private" by a project will not be retrieved unless the user has been given a user account by that project and has access to the relevant item category. Items that are public and thus visible to anonymous users will be retrieved by a query regardless of the user's access privileges.

A query can search for particular properties (variable-value pairs) and combinations of properties. Parentheses and logical operators can be used to construct Boolean algebraic expressions of considerable complexity which incorporate multiple properties. With respect to textual content, a query can search for a character string (a word or phrase) within the names, descriptions, and free-form notes of items.

The search criteria of a query make use of the intrinsic properties of database items, their textual content, and the events that affect them. However, queries can also be combined to create compound queries that search, not just the intrinsic characteristics of items, but also the extrinsic spatial and temporal containment relationships among items and the links between pairs of items.

Extrinsic relationships between items are represented in OCHRE by various means. They can be represented by containment hierarchies in certain item categories: for example, spatial containment hierarchies in the **Locations & objects** category and temporal containment hierarchies in the **Periods** category. A compound query can use these hierarchies in combination with intrinsic item characteristics to search for items that existed before, after, or at the same time as items of a certain kind; or that spatially contain, or are contained within, certain items (see the section below on “Using Containment Hierarchies in a Compound Query”).

Extrinsic relationships between items can also be represented by relational properties (see Chapter 6) and by simple links between pairs of items. Relational properties can be searched via a query’s item-property criteria.

In addition, thesaurus relationships can be used by a query to find items that have been described using different but synonymous or semantically related terms (see the discussion of thesaurus relationships above in Chapter 13). When performing a query, you can decide whether to use synonyms to expand the scope of the query and you can choose which project’s thesaurus relationships to use.

Organizing and Describing Queries

When you select the **Queries** category in the navigation pane and then click the ↲ “insert below” button in the main toolbar, you will be given a choice to create a hierarchy or an individual item. Hierarchies in the **Queries** category are “subordination hierarchies” (see the section on “Containment Hierarchies and Subordination Hierarchies” in Chapter 4). Thus, you can use headings within a query hierarchy to organize your queries into named groups. This is useful

for keeping track of the many queries that a typical project will create over time. If a query hierarchy is selected in the navigation pane, the item pane will display data-entry fields for the hierarchy's creators, creation date, name, abbreviation, description, properties, links, notes, and display preferences. These fields are the same as those used for hierarchies elsewhere.

When you select an individual query item in the navigation pane, the **Query** tab in the item pane will contain several data-entry areas. At the top are three overlapping tabs for the query item's **Information**, **Contexts**, and **Creators**.

The **Information** tab contains data-entry fields for the query's **Name** and **Abbreviation**. In either of these fields you can click the “translation” button to enter an alias in the project's default language or to enter textual content in another language. The **Information** tab also contains a drop-down pick-list for specifying the query's **Type**, which determines how the query will be presented to view-only users (see the section below on “Constructing a Generic Query to be Customized by Other Users”).

The **Contexts** tab works in the usual way to show all of the hierarchical contexts in which the query item appears (see Chapter 4). Each context is displayed as a slash-separated “path” that shows the item's location within a particular hierarchy. The first part of each path indicates the project in which the item is located, recognizing that items can be copied from other projects (the current project is indicated not by name but simply by an asterisk). Clicking the **Go to** button to the right of the **Contexts** area will scroll to the hierarchical location of the selected path in the navigation pane. Clicking the **Refresh** button will refresh the list of context paths in case the list is empty or out of date.

In the **Creators** tab is a field in which you can enter one or more links to items in the **Persons & organizations** category; these represent the creators of the query. To the right of this field is a data-entry field for entering the creation date (click the button to the left of this field to insert the current date; the button to the right displays a calendar from which you can select a date).

Below the **Information**, **Contexts**, and **Creators** tabs is an area in which is displayed a quasi-English summary of the current query's search criteria and scope. This summary is generated automatically from what is entered in the **Scope** and **Criteria** tabs at the bottom of the item pane. Below the query summary is a button labeled **Perform query**. When you have finished specifying the scope and criteria, you would click this button to perform the query and produce a list of query results.

The screenshot shows the Zincirli interface with several search fields and a generated query summary:

- Name:** English
Coins
- Abbreviation:** English
- Type:** <Select type>

FIND ALL { items } FROM Zincirli
WITHIN { Locations & objects [Area 1] }
DURING { Periods [Zincirli chronology] }
WHERE
 Location or object type is Registered item
 Object type is Coin

Perform query

Scope Criteria **Description XQuery Links Notes Save Results**

Projects: 1

- Ashkelon Expedition
- Electronic Chicago Hittite
- Marathi Online
- Persepolis Fortification A
- Zincirli Expedition

Category: Locations & objects

Limit search to items within: Zincirli Expedition

- Locations & objects
 - Area 0
 - Area 1
 - Area 2

Periods: Limit search to items within Zincirli Expedition

- Periods
 - Zincirli chronology
 - Zincirli Area 5 phases
 - Zincirli Area 3 phases

Beside the **Scope** and **Criteria** tabs at the bottom of the item pane are overlapping tabs for the query item's **Description**, **Links**, and **Notes**. In the **Description** tab you can enter free-form textual content to describe the query. For a complex query, it is a good practice to enter a detailed description in order to document the choices made concerning the query's scope and criteria in such a way that other scholars can understand the thinking which lies behind these choices. The **Links** and **Notes** tabs function as they would for any other item (see Chapters 5 and 6).

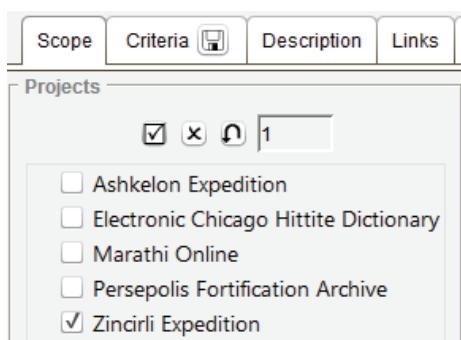
Defining a Query's Scope

The **Scope** tab is used to specify which database items are to be considered when you perform the query. It contains three columns: one for specifying the **Projects** to be searched; one for the **Category** of database items to consider; and (optionally) a column in which to limit the search to particular chronological **Periods**.

Selecting the Projects to Include in a Query's Scope

In the **Projects** column you will see a list of all the OCHRE projects to which you have access, either because you have a user account in the project or because the project has made its data public for anonymous users to see. A query can search multiple projects but it will retrieve only those database items to which the person performing the query has access. Items marked “private” by a project will not be retrieved unless the user has been given a user account by that project and has access to the relevant item category. Items that are public and thus visible to anonymous users will be retrieved by a query regardless of the user’s access privileges.

To select a project to include in the query’s scope, check the checkbox to the left of the project’s name. To select all of the projects at once, click the “select all” button at the top of the column. To de-select all the projects, click the “de-select all” button. To invert the current selections (i.e., select what is currently not selected and de-select what is currently selected), click the “invert selection” button.



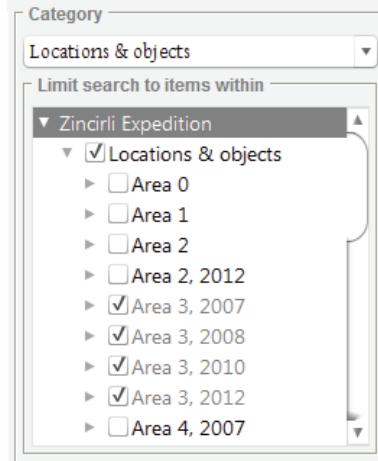
Selecting the Category of Items to Include in a Query's Scope

In the **Category** column you will see a drop-down pick-list of item categories from which to choose. There will be a few options not normally available when selecting a category, including the option to

restrict the query's scope to only the attested forms or only the citations contained in **Dictionary** items, and to restrict the scope to only the epigraphic units or only the discourse units of **Text** items.

You can also limit a query's scope to the contents of a pre-existing set of items. This is useful for constraining the current query by the set of results obtained by a previous query.

After you have chosen a category you can further limit the query's scope to particular groups of items within the category by checking the appropriate checkboxes to the left of the item names in the hierarchy.



Limits a Query's Scope to Particular Periods

In the **Periods** column you will see all the hierarchies of items found within the selected project's **Periods** category. You can limit the query's scope to particular periods or subperiods by checking the appropriate checkboxes to the left of the period names.

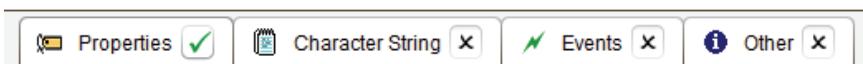
Limiting the scope to particular periods means that items which are considered for inclusion in the query results will be inspected to see if they are linked in some way to one of the selected period items and they will be included in the results only if such a link exists. Links to period items may be entered either by means of a simple link (which could be a general link or a named link from a **Period** data field) or by means of a relational property—see the section on “Period Links” above in Chapter 6.

Defining a Query's Search Criteria

The **Criteria** tab contains the search criteria that database items must match in order to be included in the query results. There are subtabs entitled **Properties**, **Character String**, **Events**, and **Other** for entering different kinds of criteria. Any or all of these four tabs may be used in

a query. Some queries will be quite simple, with search criteria entered in only one of the tabs. Other queries will be more complex, with search criteria entered in two or more of these tabs.

Note that the criteria which have been entered in a particular tab will be used in the query only if the tab's checkbox has been checked and shows a green checkmark. This allows you to ignore previously entered criteria in one or more of the tabs and to perform a query without them, without having to delete the criteria, which you may want to re-use in a later query. In order to be included in the query results, a database item must match *all* of the criteria entered in *all* of the tabs that have been activated by checking their checkboxes. In other words, it is assumed that there is an “intersection” (logical AND) of the various active criteria, not a “union” (logical OR).



Searching Based on the Properties of Items

The **Properties** tab allows you to search for particular item properties (variable-value pairs) or combinations of properties that have been used to describe items which are being considered for inclusion in the query results. One or more rows may be entered in this tab using the normal procedure for entering the variables and values of properties (see Chapter 5 and the image on the next page).

However, in addition to a variable and a value, there is an operator in each row that indicates the relationship between the property value being sought and the value specified in the row. Double-click in the **Operator** column to the right of the **Variable** column to display a drop-down pick-list of operators. The particular operators available in the pick-list will depend on the variable's type. If no operator is entered for the row, the equality operator **is** will be used by default.

For any type of variable, you can search for a value of the variable that **is** or **is not** the value specified. For **ordinal**, **date**, **integer**, and **decimal** variables, you can search for a value that **is greater than**, **is greater than or equal to**, **is less than**, or **is less than or equal to** the

value specified. For **string** variables, you can search for a value that **starts with** or **contains** the character string specified in the **Value** column.

Sometimes you may simply want to find items that have been described using a particular variable, regardless of the variable's value. To do this, use the operator **is present** or **is not present**, which will cause the query to determine whether the variable specified has been assigned (or has not been assigned) to the items being considered for inclusion in the query results.

To indicate that you want to find items that possess a particular property which was marked as “uncertain” during data entry, click the checkbox in the **?** column at the right end of the row (see Chapter 5 on “Indicating Uncertainty about a Property”).

Each variable-operator-value row in the **Properties** tab functions as a logical expression that evaluates to TRUE or FALSE. If you enter more than one row, you must insert a logical operator in front of the second, third, and subsequent rows in order to combine each successive row with the previous rows in a single logical (Boolean) expression. Double-clicking in the leftmost column will display a drop-down pick-list of the following operators: **not**, **and**, **or**, and **skip**.

Enter properties						
	+/-x	(Variable	Operator	Value) ?
skip	Location or object type	is	Pottery	<input type="checkbox"/>		
	▼ Object type	is	Vessel	<input type="checkbox"/>		
skip	▼ Form	is	Open vessel	<input type="checkbox"/>		
	Form	is	Bowl	<input type="checkbox"/>		
or	Form	is	Plate	<input type="checkbox"/>		
or	Form	is	Platter	<input type="checkbox"/>		

The first three of these function as normal logical operators. The **not** operator negates the truth-value of the expression to which it is prefixed, changing TRUE to FALSE and FALSE to TRUE. The **and**

operator combines the preceding expression (variable-operator-value row) with the expression to which it is prefixed, causing the pair of expressions to evaluate to TRUE only if both are individually TRUE. The **or** operator also combines two expressions, but it causes them to evaluate to TRUE if either of them is TRUE.

The **skip** operator is unique to OCHRE and requires some explanation. The variables and values entered in the **Properties** tab must conform to a taxonomic hierarchy (see Chapter 13), requiring you in some cases to enter a parent property in order to enter a variable that is nested within that property in the taxonomy. But you will often wish to ignore the parent property when performing a query. The **skip** operator causes the query processor to skip over the row to which this operator is prefixed in order to do a search based on a lower-level property without regard to its parent property.

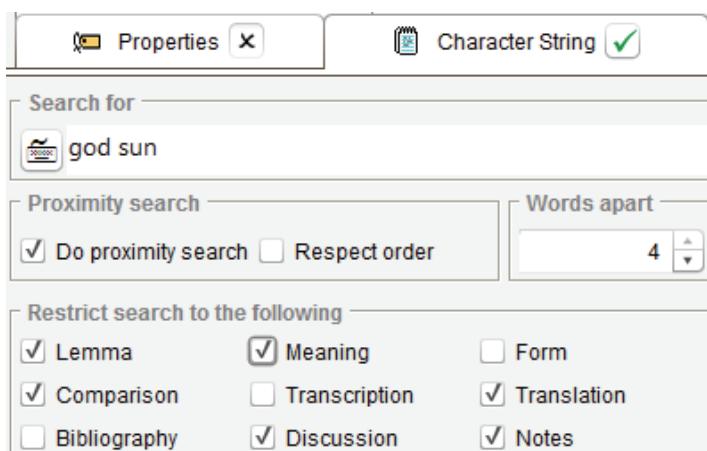
You can control the order of evaluation of the logical expressions represented as rows in the **Properties** tab by inserting parentheses to create nested expressions. Double-clicking in the column immediately to the left of the **Variable** column or in the column immediately to the right of the **Value** column will display a drop-down pick-list from which you can choose the correct number of left- or right-parentheses. By using parentheses together with logical operators, you can construct a Boolean algebraic expression of considerable complexity that incorporates multiple variables and values.

The search criteria entered in the **Properties** tab can be expanded automatically using thesaurus relationships to find synonyms or related terms for the variable names and qualitative values specified in the **Properties** tab. Synonymous or related variables and values would be treated by the query as if they themselves had been entered as search criteria in the **Properties** tab. At the top of this tab is a **Thesaurus** field in which to enter one or more links to projects whose thesauruses are to be used. This is described in more detail below.

Searching for a Character String in the Textual Content of Items

At the top of the **Character String** tab is a field labeled **Search for** in which to enter a string of characters (usually a word or phrase) that

will be sought within the textual content of the items being considered for inclusion in the query results based on the query's scope. To the left of this field is a  toggle button that you can click to display the keypad of special characters, facilitating the entry of special characters in the search string (see the section on "Using the Keypad of Special Characters" above in Chapter 5).



You can refine the search by doing a proximity search in which a given character string is sought within a certain distance from another character string. To do a proximity search, enter two character strings in the **Search for** field, surrounding each string with quotation marks and separating the two strings by a blank space; check the **Do proximity search** checkbox; and enter the desired number of **Words apart** on the right side of the tab. When the query is performed, the first character string in the **Search for** field will be sought within the specified number of words from the second character string. If you check the **Respect order** checkbox, the order of the two strings in the **Search for** field will be respected, so that if they are found in the reverse order, this will not be treated as a match. Otherwise, finding them in either order within the specified distance from one another will be treated as matching the search criteria.

Normally, a character-string query would search every kind of textual content that belongs to the items which are being considered

for inclusion in the query results. To allow you to restrict the search to particular kinds of textual content, there are checkboxes at the bottom of the **Character String** tab under the heading **Restrict search to the following**. The checkboxes shown here will depend on the item category specified in the query's **Scope** tab. You can check one or more of the checkboxes to restrict the search to the designated kinds of textual content.

For most item categories, the only options will be to restrict the search to the **Name** fields, **Description** fields, or **Notes** of the items being considered. However, for some categories there will be additional checkboxes because items in those categories possess more kinds of textual content. For example, if the **Dictionaries** category is specified in the query's scope, there will be additional checkboxes allowing you to restrict the search to each dictionary item's lemma, meanings, forms, comparisons, bibliography, etc.

Searching Based on Events that Affect Items

In some item categories, it is possible to record events that affect an item (see Chapter 7). You may therefore wish to search for items, not on the basis of their intrinsic properties or textual content, but on the basis of their events. To do so, you would enter search criteria in the **Events** subtab of the query's **Criteria** tab. You can search based on the type of event and/or the location, person, or chronological period associated with the event. Instead of a period link, some events have specific dates (or starting and ending dates), which you can enter as search criteria.

The **Event** area at the top of the **Events** subtab contains a drop-down pick-list of all the events that have been defined by your project via the values allowed for the special "Events" variable in your project's taxonomy (see the section on "The 'Events' Variable and Its Project-defined Values" above in Chapter 7). Below this pick-list is a checkbox labeled **Consider 'last' events only**. If you check this box, only the latest event of the type you specify will be considered. For example, you may wish to find items that have the event "Moved to" but you want to consider only the last such event for each item.

To the right of the **Event** area is an **Event fulfillment** area with the option to **Restrict to fulfilled events** or **Restrict to unfulfilled events**. As was explained in Chapter 7, a value of the “Events” variable can be designated as the “event fulfillment” for another value of that variable. For example, the value “To photograph” represents an unfulfilled event whose counterpart is the value “Photographed.” You can restrict the search to fulfilled events, which means that an item will meet the search criteria only if two events have been recorded for it: one representing an unfulfilled event (e.g., “To photograph”) and the other the fulfillment of the first event (e.g., “Photographed”). Similarly, you can restrict the search to unfulfilled events, which means that an item will meet the search criteria only if an unfulfilled event has been recorded for the item but no corresponding fulfillment event has yet been recorded.

Below the **Event** area are fields for entering additional search criteria that pertain to the **Location**, **Person**, **Period**, or **Other** entity associated with the events recorded for the items being considered. Use the $\mathbb{0}+$ “add link” and $\mathbb{0}-$ “remove link” buttons to the left of each of these fields to add or remove links to target items in the relevant categories, which you would select in the usual way via the **Linked Items** tab of the reference pane (see Chapter 6).

Event <input type="text" value="To draw"/> <input type="button" value="▼"/> <input type="checkbox"/> Consider 'last' events only	Event fulfillment <input type="checkbox"/> Restrict to fulfilled events <input checked="" type="checkbox"/> Restrict to unfulfilled events
Location $\mathbb{0}+$ $\mathbb{0}-$ <Locations & objects>	Person $\mathbb{0}+$ $\mathbb{0}-$ Reczuch, Karen
Period $\mathbb{0}+$ $\mathbb{0}-$ <Periods>	Other $\mathbb{0}+$ $\mathbb{0}-$ <Select category>
Date (MM/dd/yy) <input type="text" value="<Select>"/> <input type="button" value="▼"/> <input type="button" value="Calendar"/>	End date (MM/dd/yy) <input type="text" value="<Select>"/> <input type="button" value="▼"/> <input type="button" value="Calendar"/>

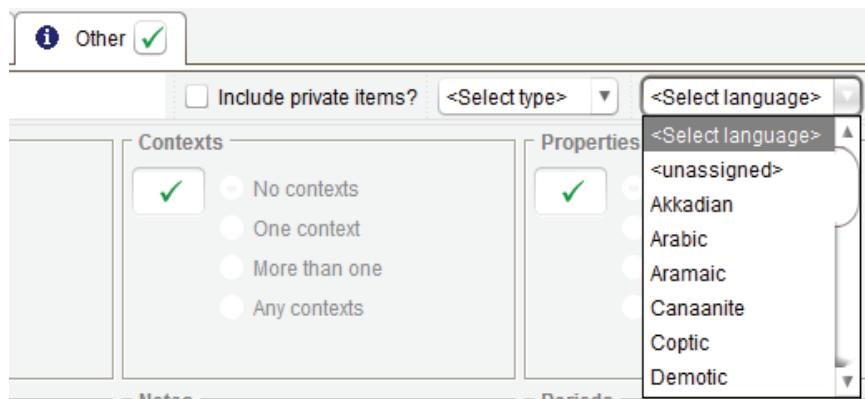
You can also search based on the calendar dates of events via the **Date** and **End Date** fields (note that if an ending date were entered for

an event, its **Date** field would contain its starting date). The drop-down pick-lists to the left of the date fields allow you to specify whether you want the query to find matches for the exact date specified or for an earlier or later date.

Searching Based on Other Criteria

The **Other** subtab of the **Criteria** tab contains miscellaneous additional search criteria. The **Date last modified** field at the top of the tab allows you to do a search based on the most recent date and time when information about an item was changed in the database. The drop-down pick-list in this field lets you specify whether the database modification date should be equal to, earlier than, or later than the date and time you have entered.

The format for entering the date will depend on the date-format preference chosen for the project by the project administrator in the **Data-entry** subtab of the **Preferences** tab in the project's item pane (see Chapter 17). If you want to specify not just the date but also the time of day, you have the option of appending a time in the following format: *T**hh:mm:ss*, where *T* is the letter "T" and *hh* is the hour according to a 24-hour clock; *mm* are the minutes; and *ss* are the seconds. For example, the time of 2:15 p.m. would be entered as "T14:15:00."



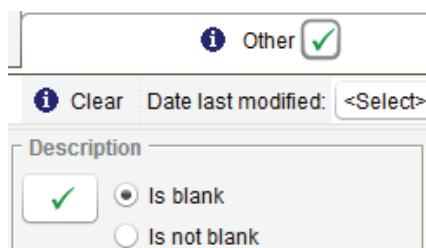
Another option shown at the top of the **Other** tab is a checkbox that lets you **Include private items**. If this is checked, the query will

consider items in the current project that have been marked as “private” and to which you have access, based on your user privileges. Otherwise, “private” items will not be included in the query.

In addition, you can search based on the **Type** fields of items, provided that the category of items selected in the query’s scope possesses a **Type** field (otherwise the **<Select type>** option will not appear in the **Other** tab). You can also search based on the language of text items if the **Texts** category is specified in the query’s scope.

The remaining options in the **Other** tab are “radio-button” choices that can be activated by clicking the checkbox to the left of the choice. The options available will vary depending on the item category that has been specified in the query’s scope. For example, if you are searching **Locations & objects**, you can search for the absence or presence of observations, which do not apply to other item categories. The following search options are available:

- The **Description** option lets you find items based on whether their **Description** fields are blank.
- The **Abbreviation** option lets you find items based on whether their **Abbreviation** fields are blank.
- The **Contexts** option lets you find items based on how many times an item is repeated in different hierarchical contexts.
- The **Properties** option lets you find items based on how many properties (if any) have been assigned to an item.
- The **Links from** option lets you find items based on how many links from an item have been created to link it to other items; you can restrict this search to one or more types of resource items, in particular, to which the item has been linked, by using the buttons on the right side of this option.
- The **Links to** option lets you find items based on whether any links have been created to an item from another item.



- The **Notes** option lets you find items based on how many notes (if any) have been created for an item.
- The **Events** option lets you find items based on how many events (if any) have been recorded for an item.
- The **Sub-items** option lets you find items based on whether an item has child items in a hierarchy of items.

You can use more than one of these search options in the same query, but all of the criteria entered in the **Other** tab must be met in order for an item to be included in the query results. Only those items in the query's scope that match every criterion entered in all of the active subtabs of the **Criteria** tab will be included in the list of query results, and a subtab is activated by clicking the checkbox at the top of the tab so that a green checkmark appears.

Expansion of Search Criteria via Thesaurus Relationships

As was noted above, a query's search criteria can be expanded automatically during the performance of the query by means of thesaurus relationships that allow the query processor to include terms that are synonymous with, or semantically related to, the variables and values specified in the search criteria.

For example, if you enter the expression "Color is purple" in a row of the query's **Properties** tab, the query processor would automatically add the expression "Color is magenta," yielding the compound logical expression "Color is purple OR Color is magenta," provided that the value "magenta" had been related to "purple" as a synonym in a predefined thesaurus relationship.

At the top of the **Properties** tab is a data-entry area labeled **Thesaurus** in which to enter one or more links to projects whose thesaurus relationships you wish to use in the query (for a detailed discussion of thesaurus relationships, see Chapter 13). To the right of the **Thesaurus** link field are two checkboxes. Checking the **Synonyms only** checkbox will cause only thesaurus relationships of the "synonym" type to be used; if you leave it unchecked, both synonyms and related-term relationships will be used. Checking the **Identical names**

as synonyms checkbox will cause identical variable names or identical values that are found in different projects within the scope of the query to be treated as synonyms, even if there is no explicit thesaurus relationship between them.

Working with Query Results

After the query's scope and criteria have been entered, click the **Perform query** button. The database items found by the query will be listed in the **Query Results** tab of the reference pane on the right side of the OCHRE window. If this tab is not currently displayed, click the  “show query results” button in the main toolbar in order to display it. The **Query Results** tab will be cleared when you close your OCHRE session or log out of the project, but the list of results from the most recent performance of the query will be saved and can be viewed subsequently by opening the query item's **View** tab.

Viewing Items in a List of Query Results

When you click on any item in a list of query results, in either the **Query Results** tab or the query's **View** tab, information about the selected item will be displayed. This provides a convenient way to inspect the items that have been retrieved by the query. At the top of the **Query Results** tab and at the top of the **View** tab are shown the **Date and time** when the query was performed and a **Count** of the number of items found by the query. There is also a drop-down pick-list entitled **Sort by** that lets you choose the order in which to display the list of query results (see the section on “Display Preferences for Hierarchies and Categories” in Chapter 4 for an explanation of the various sorting methods).

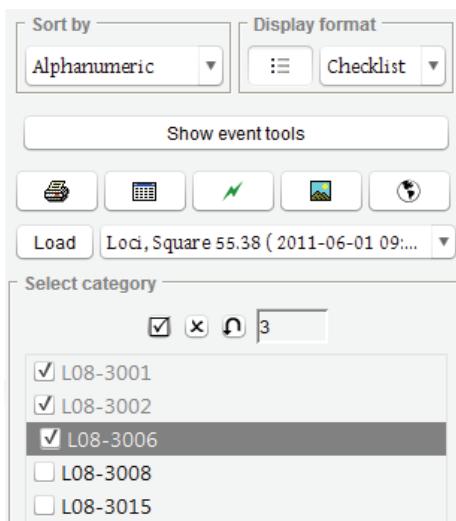
At the top of the **Query Results** tab there is an additional drop-down pick-list entitled **Display format** that allows you to choose one of the following ways of displaying the query results:

- **List** format, which simply lists the names of the items found by the query and allows you to click on an item name in order to display information about it in a separate window.

- **Checklist** format, which provides a checkbox in front of each item name in the list. Clicking on an item does not display information about it but selects it for inclusion in a printed report, table, or other view of the query results (see the discussion of these viewing options below). The checkboxes also provide a way for an event to be assigned to selected items among the query results rather than assigning the event to all items in the list (see the section below on “Assigning Events to Items in a List of Query Results”).
- **Outline** format, which applies only to dictionary items. The contents of each dictionary entry in the list of query results are outlined by displaying the lemma’s forms and attested forms, and its meanings and submeanings, in a hierarchical fashion (see Chapter 12 concerning the forms and meanings in a dictionary entry).

Just to the left of the **Display format** pick-list is a toggle button that lets you choose between working with the current list of query results or a set of saved results. If you choose to work with a set of saved results you will be given a pick-list of previously saved sets. Select the set of interest and click the **Load** button to retrieve its items. The current list of query results and the list of saved results are treated independently and can be active simultaneously. Each can be displayed in any of the available formats.

In the query item’s **View** tab you can select the mode of display for query results by clicking on the “list view,” “table view,” “form view,” “image view,” or “map view” button. These view modes work in exactly the same way as the modes of



viewing sets of items in the **Sets** category (see “Viewing a Set of Items” above in Chapter 4). Note, however, that only the “table view,” “image view,” and “map view” modes are available in the **Query Results** tab in the reference pane.

Assigning Events to Items in a List of Query Results

Chapter 7 explains how events can be recorded for individual items. But sometimes it is useful to be able to assign an event automatically to an entire group of items that have been retrieved by a query. For example, if a group of artifacts has been moved from one location to another, you will want to add a “Moved to” event to each item that represents an artifact in the group. You would first perform a query to retrieve the relevant items and then assign the same event to all of them at once in order to update their inventory locations. Events are most often recorded for items that belong to the **Locations & objects** and **Persons & organizations** categories but they can also be recorded for **Concepts**, **Resources**, and **Texts**.

To assign an event to all of the items in a list of query results, click the **Show event fields** toggle button in the **Query Results** tab. (This button is visible only to project administrators because only they can assign an event to a list of query results.) Clicking this toggle button will display a field for entering an **Event** (via a drop-down pick-list of the event names which are predefined for the project as values of the special “Events” variable) and it will also display link fields for entering the **Location**, **Person**, and **Period** associated with the event. You can also enter the calendar **Date** (starting date)

The screenshot shows a dialog box titled "Assign event ...". It contains the following fields:

- Event:** A dropdown menu showing "Drawn".
- Location:** A link field labeled "<Locations & objects>" with plus and minus buttons.
- Person:** A link field showing a person icon and the name "Reczuch, Karen" with plus and minus buttons.
- Period:** A link field labeled "<Periods>" with plus and minus buttons.
- Date (MM/dd/yy):** A calendar input field showing "07/01/10".
- End date (MM/dd/yy):** A calendar input field with a clear button.
- Comment:** A large text input field.

and **End date** of the event, as well as a descriptive **Comment**. For more details about these data-entry fields, see the section on “Adding and Removing Events for an Item” above in Chapter 7.

After you have entered the necessary information in these fields, click the **Assign event** button to assign the specified event to all the items in the list of query results. If you are using the **Checklist** display format, the event will be assigned only to those items you have selected by clicking their checkboxes; otherwise, the event will be assigned to all of the items listed in the **Query Results** tab.

To show the events that have already been assigned to the items in the list of query results, click the “show events” button in the **Query Results** tab. This provides a useful way to determine whether or not a new event needs to be assigned to the items, in case there is any doubt about this.

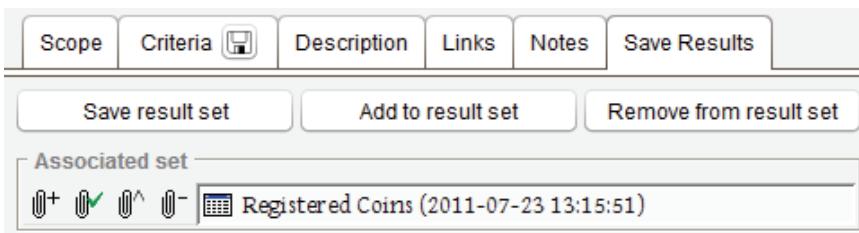
Saving a List of Query Results as a Named Set of Items

The current list of query results can be saved as a named set of items via the **Save Results** tab in the item pane of the query item. If you click the **Save result set** button in this tab, you will create a new item in the “Inbox” hierarchy of the **Sets** category. The name of the new set item will consist of the query’s name followed by the date and time when the query was performed. The items in the set will consist of the items listed in the query results. You can then remove items from this set or add items to it, as usual, but keep in mind that this will cause the set to differ from the query results.

If you perform the same query again and open its **Save Results** tab, you will see that the name of the previously created set of query results is shown in the **Set of query results** area. If you want to create an entirely new result set for the query, you should use the button to remove the current set (which will still exist in the **Sets** category) and then click the **Save result set** button to save a new set whose name reflects the date and time of the most recent performance of the query. In other words, if the **Set of query results** area is empty, clicking the **Save result set** button will create a new result set whose name incorporates the most recent query-performance date and time. The

newly created set will be added to the “Inbox” hierarchy of the **Sets** category and will be automatically linked as the saved result set.

However, you may wish to keep the existing set and simply update it with the most recent query results, which may have changed if the query’s criteria were modified or the data being queried was changed. In that case, do not remove the set shown in the **Set of query results** area but simply click the **Save result set** button to update this set and replace its contents. Alternatively, you can click the **Add to result set** button to add any new items in the current query results to the existing set, or you can click the **Remove from result set** button to remove from the existing set any items found in the current query results—this can be used to “subtract” from an existing set the results of a different query.



Note that the name of the set shown in the **Set of query results** area will not be changed automatically when you update its contents. This might be misleading if the name contains the date and time of a previous performance of the query and thus does not reflect the most recent query performance. In that case, you may wish to edit the set’s name, which you can easily do on the spot by double-clicking on the name in the **Set of query results** area in order to display the set’s item pane in a floating window. But note that the corresponding item label in the **Sets** category of the navigation pane will not be updated to reflect the change until you collapse and re-expand that category or until you click the **↻** “refresh” button in the main toolbar.

Using Queries in View-only Mode

The preceding sections of this chapter have dealt with the creation and use of queries in data-entry mode. Queries are constructed in

data-entry mode by the project administrator and by other members of the project team to whom the project administrator has given user accounts and has granted access to the **Queries** category.

The present section deals with queries that have been constructed in data-entry mode and have then been made available for use in view-only mode by anonymous users who do not have user accounts in the project but want to view and retrieve data that the project has made public. (See the section on “Levels of Access” above in Chapter 3 concerning the distinction between data-entry mode and view-only mode.) View-only users can customize and combine predefined queries in flexible ways to meet their needs but they are shielded from the complexity of constructing the queries. Thus, they can use complex search criteria previously entered by an expert user without having to learn how to enter these criteria themselves.

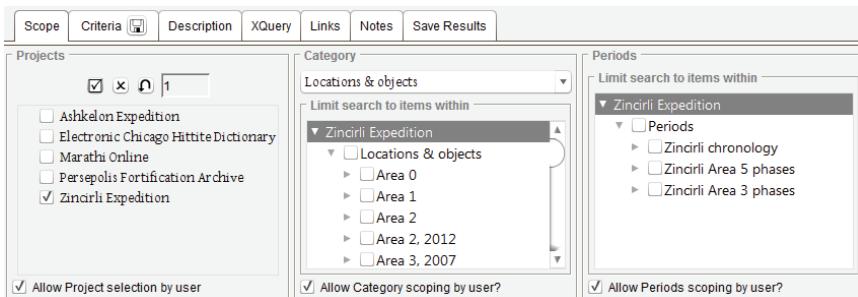
Displaying Queries for Use in View-only Mode

The only queries in a project’s **Queries** category that are visible to view-only users are those which have been made available by the project administrator for public use. This is accomplished by inserting or copying query items into an item hierarchy in the **Queries** category that is designated as the project’s public query hierarchy. The project administrator can designate a hierarchy as the public query hierarchy by adding a link to it in the **Public query hierarchy** field of the **View-only** subtab in the project’s **Preferences** tab (see the section on “View-only Preferences” in Chapter 17).

If a public query hierarchy has been designated for the project, the names of all the queries contained in this hierarchy will be displayed to view-only users in the navigation pane in a tab entitled **Queries**. The public **Queries** tab will also be visible in data-entry mode, in addition to the project’s private **Queries** category.

The headings in the public query hierarchy will be used to group the queries. Clicking on a query name will cause the predetermined options available for customizing the selected query to be displayed in the item pane. The kind and format of the user-customization options are determined by the query’s **Type** and by the settings

entered by the project administrator when constructing the query in data-entry mode (see the section below on “Constructing a Generic Query to be Customized by Other Users”).

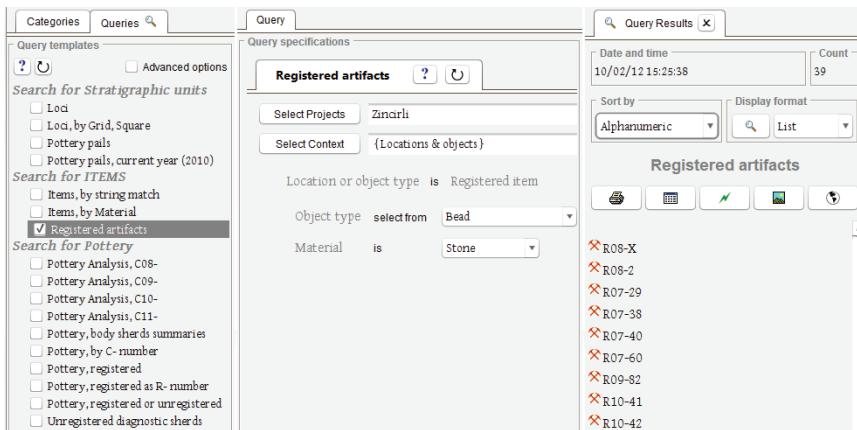


If the creator of the query has allowed view-only users to determine its scope (this is done via the **Allow ... by user** checkboxes in the **Scope** pane, shown above), the user of the query will see buttons in the item pane to **Select Projects**, **Select Context**, and **Select Period**. Clicking the **Select Projects** button will display a window in which the user can select the projects to be included in the query. The **Select Context** button will display a window in which the user can select one or more hierarchies or hierarchical branches in the item category specified for the query in order to search among only those items. The **Select Period** button will display a window in which the user can select a period or periods from the selected projects’ **Periods** categories in order to search among only those items that are linked to the selected periods.



The user can read information about the query by clicking the ? “query information” button in the query’s tab heading in the item pane. This will display the textual content that was entered in the query’s **Description** field. When a project administrator creates a query intended for subsequent use by others, it is a good practice to enter a detailed description that explains the query’s scope and criteria in an understandable way.

After choosing a query in the **Queries** tab and customizing it in the item pane, the user would click the button labeled **Perform single query** at the bottom of the item pane in order to display a list of the items retrieved by the query. Above the list of query results will be the **Date and time** when the query was performed and a **Count** of the number of items found by the query. The user can choose display options in the **Sort by** and **Display format** fields (see the section above on “Viewing Items in a List of Query Results”). The user can click on an individual item in the list of query results to display information about that item. The user can also display information about all of the items at once by clicking the “table view” button or (if available) the “image view” or “map view” button.



Performing Multiple Queries

By default, you will select and perform a single predefined query. However, you can also perform a multiple query that makes use of two or more predefined queries to produce a list of query results. Clicking the **Show query options** checkbox at the top of the **Queries** tab in the navigation pane will let you choose one of these options:

- **Single**, the default single-query option described above, in which you select one query from the **Queries** tab; customize it as needed; and then click the **Perform single query** button to produce the list of query results.

- **Multiple (list)**, in which you can select more than one query; customize each query as needed; specify how the results of each query are to be combined with the results of all preceding queries; and click the **Perform multiple query** button to perform the queries and produce a final list of query results.
- **Multiple (tab)**, which is the same as the **Multiple (list)** option except that the various queries to be combined and performed are shown in overlapping tabs rather than in a vertically scrolling list of tabs.
- **Contextual**, in which you select multiple queries and customize each query as needed but do not simply list the queries sequentially; instead, the queries are nested hierarchically to limit the query results based on the containment relationships among items. This allows you to indicate whether you want to find items that contain items of a certain kind or are contained within items of a certain kind. For example, you can use a contextual query to find archaeological layers of a particular type that contain artifacts of a certain kind. This option is explained in more detail below in the section on “Performing Contextual Queries.”

In the “multiple” mode, each query you select in the **Queries** tab in the navigation pane will be shown in the item pane in its own tab. These tabs will be arranged in a vertically scrolling list if you choose the **Multiple (list)** option and they will be displayed as overlapping tabs if you choose the **Multiple (tab)** option. The heading of a query’s tab will show the query’s name followed by four buttons.

Immediately to the right of the query’s name is the “perform this query” button; clicking it will cause just that query to be performed as a single query, without regard to the other queries that have been selected. The number of items found by this query will be shown in the second button to the right of the query’s name; clicking it will display the list of query results in the reference pane. The “query information” button is the third button; clicking it will display the contents of the query’s **Description** field. The fourth

button is the “restore default settings” button; clicking it will restore the default values of the predefined query that you may have changed when customizing it with your own values.

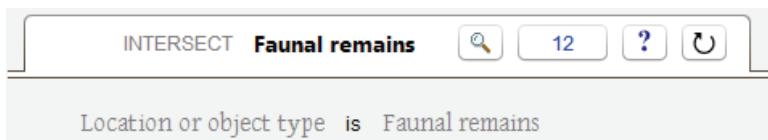
In the **Multiple (list)** example shown below, of the twenty-four items in the “Registered artifacts” query results that met the criteria of being “Worked Bone” and of the twelve items in the “Faunal remains” query results that were “Mammalian Tarsals,” there was an intersection, or overlap, of nine items, which are displayed in the multiple query’s results list.

Query Specification	Results
Registered artifacts	Date and time: 11/03/12 14:52:43 Count: 9 Sort by: Alphanumeric Display format: List
Faunal remains	Date and time: 11/03/12 14:52:43 Count: 9 Sort by: Alphanumeric Display format: List
INTERSECT	Date and time: 11/03/12 14:52:43 Count: 9 Sort by: Alphanumeric Display format: List

In the tab headings of the second and subsequent queries in a multiple query there is a toggle field to the left of the query’s name. By default, this field will be set to **COMBINE**, meaning that the items found by the query shown in the current tab will be combined with the results of the preceding queries; that is, there will be a union of the query result set of the current query with the cumulative result set of the preceding queries. If you click on **COMBINE** it will change to **INTERSECT**, which will restrict the query results to the intersection of the current query’s result set and the cumulative result set of the

preceding queries so that only items found in both sets will be retained. If you click on **INTERSECT** it will change to **EXCLUDE**, which means that the items found by the current query will be removed from the cumulative result set of the preceding queries. Clicking on **EXCLUDE** will change the field back to **COMBINE**.

By means of their toggle fields, the results of the various queries that make up a multiple query can be combined. The operators specified in these fields are the equivalent of the Boolean operators AND (**INTERSECT**), OR (**COMBINE**), and NOT (**EXCLUDE**).



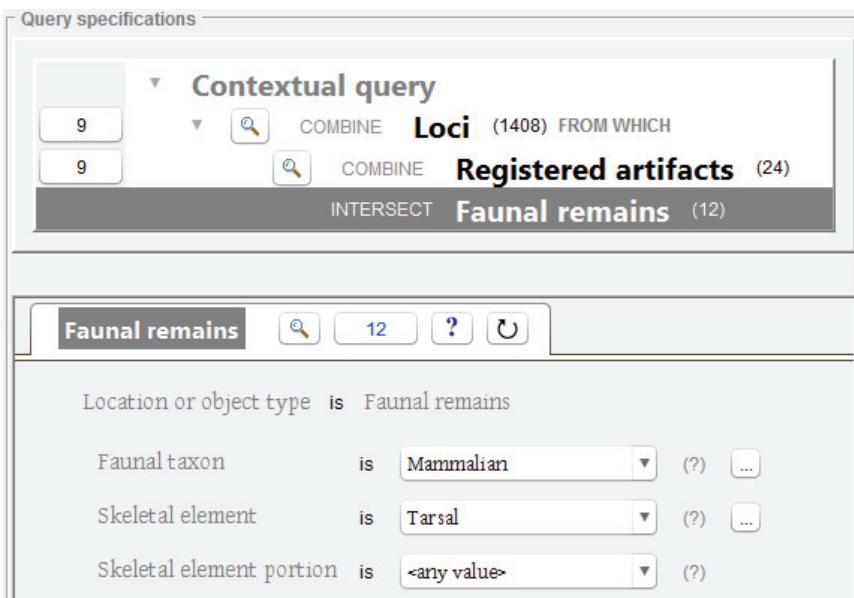
In order to use the same query more than once in a **Multiple query**—presumably because you want to customize it with different search values—you must click the **+** button at the top right of the **Queries** tab in order to add another instance of the query that is currently selected in the navigation pane. To remove a duplicated query, select it in the navigation pane and click the ***** button.

After you have selected, customized, and combined the queries, click the **Perform multiple query** button to generate the final list of query results. This list will appear in the reference pane on the right. Above the list of query results will be the **Date and time** when the query was performed and a **Count** of the number of items found by the query. You can select display options in the **Sort by** and **Display format** fields (see the section above on “Viewing Items in a List of Query Results”). Click on an individual item in the list of query results to display information about that item; or display information about all of the items by clicking the **table view** button, or, if available, the **image view** or **map view** button.

Performing Contextual Queries

In the “contextual” mode, you can select multiple queries in the **Queries** tab in the navigation pane and display them in the item pane,

just as in the “multiple” mode. However, the queries will be arranged, not in a vertically scrolling list or in a series of overlapping tabs, but hierarchically. Each query you select in the navigation pane will become the child of whatever is currently selected in the query hierarchy in the item pane. This allows you to list as many queries as you wish at each level of the hierarchy and to construct as many hierarchical levels as are needed.



In a contextual query, the item pane is divided into two parts. The upper part shows the hierarchy of queries that you have selected and the lower part shows a query tab for the query that is currently selected in the hierarchy. The heading of the query tab will contain the query's name and the same four buttons as are shown in each query tab in a multiple query. Immediately to the right of the query's name is the “perform this query” button; clicking it will cause just that query to be performed as a single query, without regard to the other queries that have been selected. The number of items found by this query will be shown in the second button to the right of the query's name; clicking it will display the list of query results in the reference pane. The “query information” button is the third button;

clicking it will display the contents of the query's **Description** field. The fourth button is the  "restore default settings" button; clicking it will restore the default values of the predefined query that you may have changed when customizing it with your own values.

If multiple queries are listed as siblings at the same level in the query hierarchy, there will be a toggle field to the left of the query's name for the second and subsequent queries in the list, just as in the "multiple" mode. By default, this toggle field will be set to **COMBINE**, meaning that the items found by the current query will be combined with the results of the preceding queries in the list of siblings; that is, there will be a union of the query result set of the current query with the cumulative result set of the preceding queries.

If you click on **COMBINE** it will change to **INTERSECT**, which will restrict the query results to the intersection of the current query's result set and the cumulative result set of the preceding queries so that only items found in both sets will be retained. If you click on **INTERSECT** it will change to **EXCLUDE**, which means that the items found by the current query will be removed from the cumulative result set of the preceding queries. Clicking on **EXCLUDE** will change the toggle field back to **COMBINE**.

By means of their toggle fields, you can combine the results of the various queries in a list of queries at a given level of the contextual query hierarchy. The operators specified in the toggle fields are the equivalent of the Boolean logical operators AND (**INTERSECT**), OR (**COMBINE**), and NOT (**EXCLUDE**).

In a contextual query, there is a second toggle field that is not present in a multiple query. This toggle field displays an operator that pertains to the queries that are hierarchically nested within the current query. The operator will appear to the right of the query's name for the last query listed at a given level of the hierarchy. By default, it will be set to **THAT CONTAIN**. This means that the query results will be restricted to those items which have been found by the queries at the current and higher levels of the hierarchy but that also contain items which have been found by the queries at lower levels of the hierarchy.

For example, if the queries at the current and higher levels of the query hierarchy produce a cumulative result set that consists of items representing archaeological layers of a certain kind, you can further limit the query results to layers that contain artifacts of a certain kind. The final list of query results is drawn from the cumulative result set produced by the queries at the current and higher levels of the hierarchy, not from the queries at the lower levels, which serve only to limit the results of the higher level queries.

If you click on **THAT CONTAIN**, the operator will change to **FROM WHICH**, which means that the query results will be restricted to those items found by the queries at lower levels of the hierarchy that were also found by the queries at the current and higher levels of the query hierarchy. For example, you might want to find all artifacts of a certain kind among a larger group of artifacts that were selected according to different criteria. In contrast to queries which use **THAT CONTAIN**, the final list of query results in a **FROM WHICH** query will come from the queries at the lower levels of the hierarchy, not from the queries at the current and higher level, which serve only to limit the results of the lower level queries.

After you have selected, customized, and combined the queries in a hierarchy, click the **Perform contextual query** button to produce the final list of query results. This list will appear in the reference pane on the right. Above the list of query results will be the **Date and time** when the query was performed and a **Count** of the number of items found by the query. You can select display options in the **Sort by** and **Display format** fields (see the section above on “Viewing Items in a List of Query Results”). Click on an individual item in the list of query results to display information about that item; or display information about all of the items by clicking the  “table view” button or, if available, the  “image view” or  “map view” button.

A final note about contextual queries: hierarchical containment relationships among database items can be exploited by combining multiple predefined queries in view-only mode to form a contextual query, as described above. But item hierarchies can also be exploited within an individual predefined query. This is done by choosing

particular hierarchical branches in the **Category** subtab of a query's **Scope** tab in order to limit the query results to the items contained in a particular branch of a containment or subordination hierarchy. For example, you could limit the search to artifacts found in a particular building at an archaeological site, provided that the containment relationships between the artifacts and the building were represented by an item hierarchy. However, the benefit of using multiple pre-defined queries in a contextual query is that it allows you to perform a much more complex search that takes advantage of hierarchical relationships, not just by limiting the scope of an individual query, but by using the results of one query to limit the scope of another.

Item hierarchies are only one of the ways by which inter-item relationships are represented in OCHRE. They are also represented by relational properties (see the section on "Relational Properties" in Chapter 6 above). Inter-item relationships represented by relational properties can also be exploited by queries. This is done by entering the relevant relational variables and target items as query criteria in the **Properties** subtab of the query's **Criteria** tab (see the section on "Searching Based on the Properties of Items" above).

Constructing a Generic Query to be Customized by Other Users

The previous section has dealt with the customization and combination of predefined queries that have been displayed as public queries in the **Queries** tab of the navigation pane. The customization options for a query are determined when the query is constructed in data-entry mode. The project administrator (or whoever creates the query) will usually specify some of the search criteria and leave other criteria blank so they can be filled in by users at the time the query is performed. It is up to the project administrator to predefine as many generic, easily customizable queries as are needed by end-users of the project's data, shielding those users from the complexity of constructing powerful queries with complex search criteria.

The format used to display a predefined query to end-users is determined by what is entered in the query's **Type** field and in its **Criteria** tab. This format is automatically generated by the OCHRE

software from the query settings, employing ordinary English phrases instead of cryptic symbols or technical terms.

The various possible query types and the procedure for determining what an end-user will see and can customize are not discussed here but are documented on the Web at <http://ochre.uchicago.edu>. The OCHRE query types and related procedures are constantly being elaborated, so current step-by-step instructions that are addressed to project administrators and other expert users are provided online. Some of the query types pertain to specialized queries used to search resources, texts, and dictionaries, for which additional search options are needed.

PRESENTING INFORMATION

Chapter 15

Displaying Hierarchies, Tables, and Maps

THIS chapter deals with the different ways of presenting the information that is stored for an individual database item or for a group of items organized in a hierarchy or set (see Chapter 4). There are several different display modes for individual items and item hierarchies. In addition, information about all of the items in a hierarchy or a set can be displayed in a table or a geographical map. In a future version of the OCHRE software, information about items will be able to be displayed also in chronological timelines and semantic network “graphs” (diagrams that show the relationships among items).

Display Modes for Individual Items and Item Hierarchies

When you select an item or a hierarchy in the navigation pane and open its **View** tab in the item pane, you will see a drop-down list at the top of the tab that lets you determine the display mode. The **Information** and **Images** modes show only data pertaining to the item or hierarchy itself, not the items it contains. The various **Descendants...** display modes show all the descendant items of the current item or hierarchy in different formats and levels of detail. The **Linked items** mode shows not just the descendant items but all other database items that have been linked to them.

Displaying Information about an Item or Hierarchy

The **Information** display mode is the default option when viewing an item or hierarchy. In this mode the **View** tab will show the name, translated names and aliases (if any), abbreviation, description, properties, periods, links, and notes of the selected item or hierarchy. It will also show the number of descendant items currently contained within the selected item or hierarchy, as well as the category of those items, but it will not show these descendants.

In the **Images** display mode the **View** tab will show all the images that have been linked to the currently selected item or hierarchy. The options available in this display mode are described above in Chapter 9 in the section on “Browsing Multiple Images.”

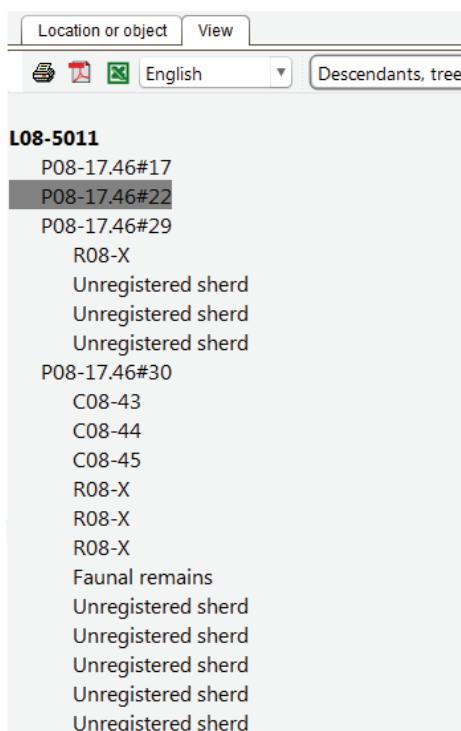
Displaying Descendant Items Contained in an Item or Hierarchy

Often you will want to see, not just the selected item or hierarchy itself, but all the items that are hierarchically contained within it. If you select the **Descendants (flat)** display mode, the **View** tab will show a vertically scrolling list of information about the current item or hierarchy and each of the descendant items contained within it. For each descendant you will see its name, translated names and aliases (if any), abbreviation, description, properties, periods, links, and notes. A descendant item’s links to other items (including links to image resources) will appear as shaded hyperlinks; clicking a hyperlink will display a window that contains information about the linked item.

In the **Descendants (tree)**

display mode, only the names of the descendant items will be shown. Their names will be arranged in a hierarchical list. However, each name will be a hyperlink; clicking it will pop up a window that contains the information about the descendant item.

In **Descendants (tabbed)** mode, information about the descendant items will be displayed in a series of overlapping tabs. The names of the first generation of descendants will be shown in the tab headings, arranged vertically along the left edge of the **View** tab.



You can display the information about any of these top-level descendants by clicking the appropriate tab heading. If a top-level descendant has descendant items of its own, information about these lower-level descendants will be shown in a “flat” list within its tab.

The screenshot shows a software interface with a toolbar at the top containing icons for location, object, and view, followed by a dropdown menu labeled "View". Below the toolbar is a toolbar with icons for file operations (New, Open, Save, Print, Find, Copy, Paste) and language selection (English). A dropdown menu labeled "Descendants, tabbed" is open, and a "Display options..." button is visible. The main area displays a hierarchical list of objects on the left and detailed information for the selected object on the right. The selected object is "P08-17.46#30 Boyd, Samuel". The detailed information includes:

- Pail no.: 30
- Location or object type: Pottery pail
- Date excavated: 07/24/08
- Top elevation (m): 519.78
- Bottom elevation (m): 519.67
- Number of debris pails: 260
- Sift ratio: 1:5
- Total number of sherds: 244
- Number of diagnostic sherds: 64

Below this, there is a section for "C08-43" which includes a thumbnail icon and the text "[C08-43 profile]". The description for this item is: "Common ware jar with simple, thickened rim." The observation date is listed as 08/11/08, and the observer is listed as "Soldi, Sebastiano". The object is categorized as a "Registered item" and identified as a "Vessel" with a "Closed vessel" form.

In **Descendants (images)** mode, the **View** tab will show all the image resources that are linked to descendant items contained in the currently selected item or hierarchy. See the section on “Browsing Multiple Images” in Chapter 9 for the options available in this mode.

In **Descendants (table)** mode, the **View** tab will show a tabular view of the descendant items contained within the currently selected item or hierarchy. The options available in this mode are described below in the section on “Displaying Tables of Data.”

Finally, in **Descendants (map)** mode, all of the geospatial mapping data stored for descendant items contained within the currently selected item or hierarchy will be displayed in the **View** tab. The options available for viewing maps are described below in the section on “Displaying Geographical Maps.”

The screenshot shows a digital catalog interface with two main panes. The left pane displays a hierarchical tree view of artifacts. The root node is L08-5011, which contains P08-17.46#17, P08-17.46#22, P08-17.46#29, and P08-17.46#30. P08-17.46#30 further contains C08-43, C08-44, C08-45, R08-X, R08-X, R08-X, Faunal remains, Unregistered sherd, Unregistered sherd, Unregistered sherd, Unregistered sherd, and Unregistered sherd. The right pane shows four artifact profiles: C08-43 profile, C08-44 profile, C08-45 profile, and another unlabeled profile. Each profile includes a thumbnail image and a scale bar indicating 5 cm.

The screenshot shows a digital catalog interface for item 0030-301. The top section displays basic item details: Observation 2 (06/25/2007) by Azzoni, Annalisa. Below this, it lists primary classification (Tablet), editorial series (Unedited), script/language (Aramaic), group (X), condition (Bad), reading (No good), writing technique (Ink), associated item (PFAT 520), and reporting category (PFAT). The bottom section, titled "Linked Item(s)", shows a table with two rows. The first row is for PFAT 520: Transliteration, listing obverse readings: (1) [] w^y? and (2) pt^p zy m^sr. The second row is for PFAT 520: Phonemic view, showing the phonetic values wy pt^p zy m^sr qmh g 5.

Displaying Items that are Linked to an Item or Hierarchy

In the **Linked items** display mode (shown above) the **View** tab will show a vertically scrolling list of information about the currently selected item or hierarchy and a similar scrolling list for each of the

items linked to it. For each linked item you will see the item's name, translated names and aliases (if any), abbreviation, description, properties, periods, links, and notes. In each case, the information will appear as it would in the **Information** mode for an individual item. The **Linked items** display mode is useful when you want to see in one place all of the information about items that are related to the main item of interest but are not hierarchically contained within it.

Displaying Tables of Data

Unlike individual items and item hierarchies, tables are not an underlying structural feature of an OCHRE database (see Chapter 19). A table is simply one possible view of a group of items that have been organized in a hierarchy or set. Other possible views of a hierarchy or set of items include geographical maps, chronological timelines, and semantic network graphs of inter-item relationships.

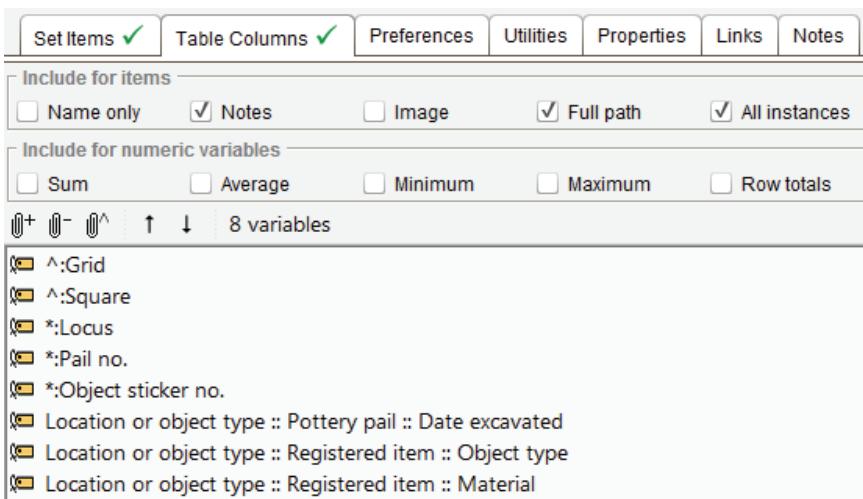
To display the table view of a set of items, select the set in the navigation pane, open its **View** tab, and then click the  “table view” button (see the section on “Viewing a Set of Items” in Chapter 4). In many cases, a set will consist of a list of items found by a query that has been named and saved for repeated use. If you have performed a query but do not wish to save its results as a set, you can nonetheless see a table view of the items found by the query by clicking the  “table view” button in the **Query Results** tab (see the section on “Viewing Items in a List of Query Results” in Chapter 14).

To display a table view of the items contained in a hierarchy that has been constructed within any of the OCHRE item categories, select the hierarchy in the navigation pane (or select a heading or containing item within the hierarchy), open its **View** tab, and choose the **Descendants (table)** display mode (see the section above on “Displaying Descendant Items Contained in an Item or Hierarchy”).

Customizing the Table View of a Set of Items

As discussed above in Chapter 4 in the section on “Viewing a Set of Items,” the table view displays the items in a tabular form in which each item is shown as a row in the table and the items’ properties are

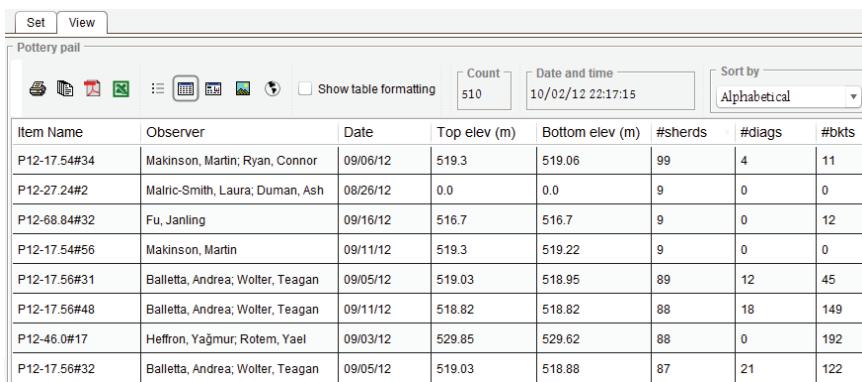
shown as columns. The table view can be customized via the **Table Columns** data-entry tab that is visible in the item pane of the set item. This tab contains links to variables in the **Property variables** category. The variables entered here determine which properties are shown as table columns in the table view of the set. The order of the variables in the tab determines the left-to-right order of the table columns. If no variables are entered in the **Table Columns** tab, a table will be displayed whose columns correspond to all of the variables currently used by items contained in the set.



Column variables can be added or removed individually by selecting them as link targets in the **Linked Items** tab of the reference pane and then clicking the \oplus “add link” or \ominus “remove link” buttons (see Chapter 6). You can also specify which variables are to be used to group and sort the rows in the table. To group by a variable, select it in the list and click on the \wedge button. To sort by a variable, select it in the list and shift-click on the \wedge button.

In addition to the list of variables that define the table columns, other options are available in the **Table Columns** tab to customize the table view. In the **Include for items** area you can specify that any or all of the following be included for each item as additional columns in the table: **notes** (i.e., the item’s free-form notes as opposed to its

properties), **image** (i.e., the first image resource that is linked to the item), **full path** (i.e., the full hierarchical context of the item), and **all instances** (i.e., all observations of locations and objects and all interpretations of concepts). In the **Include for numeric variables** area you can specify that any or all of the following summary statistics be included as additional rows with values for each column that represents a numeric property: **sum**, **average**, **minimum**, **maximum**, and **row totals**.



The screenshot shows the Archaeology Data System (ADS) interface. At the top, there are two tabs: "Set" and "View". The "View" tab is selected. Below the tabs, the title "Pottery pail" is displayed. The main area contains a table with the following data:

Item Name	Observer	Date	Top elev (m)	Bottom elev (m)	#sherds	#diags	#bkts
P12-17.54#34	Makinson, Martin; Ryan, Connor	09/06/12	519.3	519.06	99	4	11
P12-27.24#2	Malric-Smith, Laura; Duman, Ash	08/26/12	0.0	0.0	9	0	0
P12-68.84#32	Fu, Janling	09/16/12	516.7	516.7	9	0	12
P12-17.54#56	Makinson, Martin	09/11/12	519.3	519.22	9	0	0
P12-17.56#31	Balletta, Andrea; Wolter, Teagan	09/05/12	519.03	518.95	89	12	45
P12-17.56#48	Balletta, Andrea, Teagan	09/11/12	518.82	518.82	88	18	149
P12-46.0#17	Heffron, Ya'gmar; Rotem, Yael	09/03/12	529.85	529.62	88	0	192
P12-17.56#32	Balletta, Andrea; Wolter, Teagan	09/05/12	519.03	518.88	87	21	122

Printing and Exporting a Table of Data

There are buttons in the **View** tab of a set or hierarchy that allow you to print out its table view or export it to an external file. Click the “print summary” button to print out a condensed version of the current view or click the “print details” button to print a full report that contains all available information. Click the “export as PDF” button to export this full report as an Adobe PDF document. Click the “export as Excel” button to export the data table as a Microsoft Excel spreadsheet in the XLSX format. After a table has been exported as a spreadsheet file, Excel or other external software can be used to format the data table, calculate statistics, and display the results in pie-charts, histograms, and the like.



Displaying Geographical Maps

To display the map view of a set of items, select the set in the navigation pane, open its **View** tab, and then click the  “map view” button (see the section on “Viewing a Set of Items” in Chapter 4). In many cases, a set will consist of a list of items found by a query that has been named and saved for repeated use. If you have performed a query but do not wish to save its results as a set, you can nonetheless see a map view of the items found by the query by clicking the  “map view” button in the **Query Results** tab (see the section on “Viewing Items in a List of Query Results” in Chapter 14).

To display a map view of the items contained in a hierarchy that has been constructed within any of the OCHRE item categories, select the hierarchy in the navigation pane (or select a heading or containing item within the hierarchy), open its **View** tab, and choose the **Descendants (map)** display mode (see the section above on “Displaying Descendant Items Contained in an Item or Hierarchy”).

The map view of a group of items shows the geospatial data associated with all of the items in the set or hierarchy being displayed. If none of the items has such data, a message will be shown informing you that the view cannot be generated. Otherwise, a map will be displayed in the **View** tab with a special pane to the left of the map that lists each item as a separate map layer.

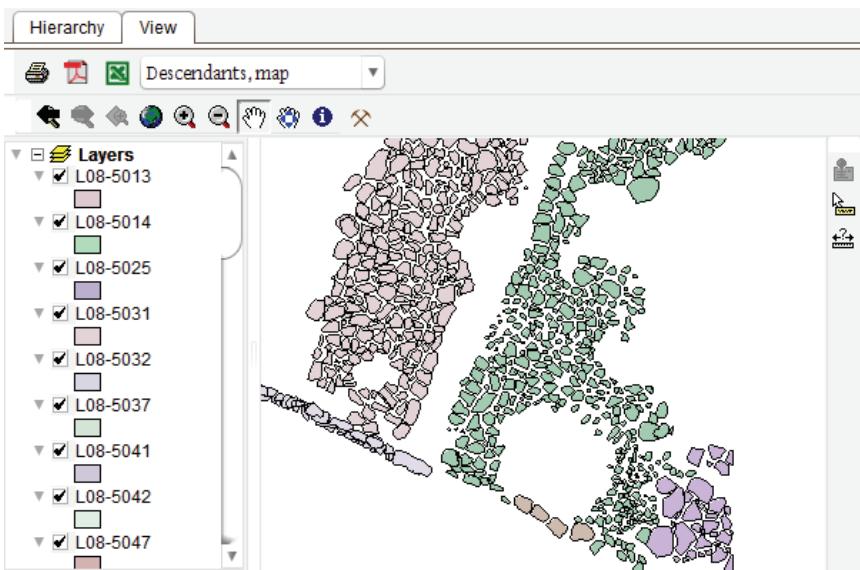
Any database item can be linked to any number of geospatial resource items. A geospatial resource will contain either a group of vector map shapes or a raster image (see Chapter 9). The vector shapes that are linked to an item collectively constitute a map layer that can be uniquely colored or shaded and independently hidden or shown. Each raster image constitutes its own map layer.

A geospatial resource item can represent a vector shapefile consisting of subfiles with SHP, SHX, DBF, PRJ, and SBN extensions, following the shapefile standard established by ESRI, Inc. A shapefile contains the coordinates for one or more points, polylines, or polygons, depending on the shapefile’s type.

Alternatively, a geospatial resource item can represent a raster file: either a georeferenced raster in the GeoTIFF or GRID format or an

ordinary raster image in TIFF or JPEG format paired with a “world” projection file that allows the image to be displayed within a planar coordinate system (e.g., *raster_1.tif* paired with a *raster_1.tfw* world file or *raster_1.jpg* paired with *raster_1.jpw*).

When you display a map view of a set or hierarchy of items, all of the shapefiles that have been linked to a particular item will be combined to form a single vector map layer that corresponds to that item. This vector map layer will consist of all of the item’s points, polylines, and polygons. However, each raster resource that has been linked to the item will be treated as a separate raster map layer.



Customizing the Map View of a Set or Hierarchy of Items

At the top of the map view are toolbar buttons that allow you to pan or zoom the map and to display information about the items and map shapes displayed in it. If you click the “pan” button, the cursor will change to a hand icon to indicate that you are in panning mode. You can then pan the map by holding down the mouse button and dragging the mouse in the desired direction. If you click the “zoom in” or “zoom out” button, the cursor will change to the appropriate icon, after which you can click your mouse button to

enlarge or reduce the scale of the map. Clicking the  “full extent” button will zoom the map to show its full extent.

The current scale of the map is shown in the bottom of the map view; the scale will change as you zoom in and out. The map coordinates of the current cursor position are shown in the bottom left corner of the map view. The X coordinate is the location of the cursor on the horizontal axis of the map’s coordinate space (e.g., the “easting” coordinate in a geographical coordinate system) and the Y coordinate is the location on the vertical axis (e.g., the “northing” coordinate in a geographical coordinate system). The coordinate space is determined by the coordinates used in the spatial data files that have been linked as resources to the items for which the map view has been generated.

To measure the distance between two points on the map, click the  “measure” button in the toolbar and then click on the starting point and hold the mouse button down as you drag the cursor to the ending point; then release the mouse button to display the distance.

To display information about a particular item portrayed in the map, click the  “item information” button in the toolbar and then position the cursor over the relevant part of the map and click the mouse button. Keep in mind that a particular item may have multiple vector shapes (points, polylines, and polygons) that collectively constitute the vector map layer that corresponds to the item. To display information about a particular shape, click the  “shape information” button in the toolbar and then position the cursor over the relevant shape and click the mouse button.

The various map layers are listed in a special pane to the left of the map. Each vector layer corresponds to one item in the group of items whose spatial data is being displayed. An item’s raster files are listed as separate map layers, one layer per raster. The order in which map layers are drawn (i.e., which ones are on top and which are in the background) is determined by the order in which the corresponding items occur in the set or hierarchy being viewed. To reorder the map layers you must reorder the items. In cases in which an item has multiple raster layers, the order of the raster layers is determined

by the order in which the corresponding geospatial resource items have been linked to the item being displayed. A map layer can be individually shown or hidden by clicking its checkbox to select or de-select it in the pane to the left of the map.

Using Images and Maps to Browse a Project's Data

You can click on any item displayed in an image view or map view in order to pop up a floating window that contains information about the selected item. But this window will in turn allow you to select additional views of other items. Thus, you can click repeatedly to open a series of view windows in order to browse through a wide range of database items in a flexible manner.

In addition to ad hoc browsing of this sort, there is a more structured way to present a wide variety of your project's data to end users. Using the **Presentations** category, you can create public presentations of data that enable anonymous users who have not logged into the project to browse and query diverse information while shielding them from the complexity of the underlying database items and relationships. The construction of public presentations is described in detail in the next chapter.

Chapter 16

Public Presentations of a Project’s Data

THE preceding chapters have explained the procedure for entering and viewing a project’s data by members of the project team, that is, by registered users who have logged into the database with their usernames and passwords (see Chapter 17 for an explanation of how to create user accounts). However, the project administrator can also expose a project’s data to public view by anonymous users who do not have usernames and cannot log into the database in data-entry mode but are nonetheless allowed to browse and query the project’s data (or selected portions of the data) in a predetermined fashion.

This is done via the **Presentations** category. Items in this category are used to specify how a project’s data should be presented to the public. This category provides a way to customize a project’s appearance for public use in view-only mode without displaying the more complex data-entry mode.

Each item hierarchy in the **Presentations** category represents an entire presentation. The items within a hierarchy are components of that presentation. The sequence of components within the hierarchy and the nesting of components within the hierarchy’s headings and subheadings determine their order of presentation.

Presentations might be created for pedagogical purposes, to be used as scripted lessons, or as a form of scholarly publication of information for use by researchers and advanced students. A presentation can include database items from any category. The creator of a presentation has a great deal of flexibility in determining what to show and how to show it.

The project administrator can specify in the project’s **Preferences** tab whether a presentation hierarchy is to be used automatically to display the project’s data to any anonymous user who opens the project (see the “View-only” subsection of “Project-wide Preferences” in Chapter 17). The project administrator can also specify which presentation to use, in case more than one presentation hierarchy has been created.

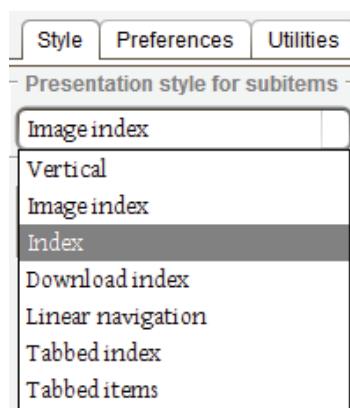
Presentation Styles and Components

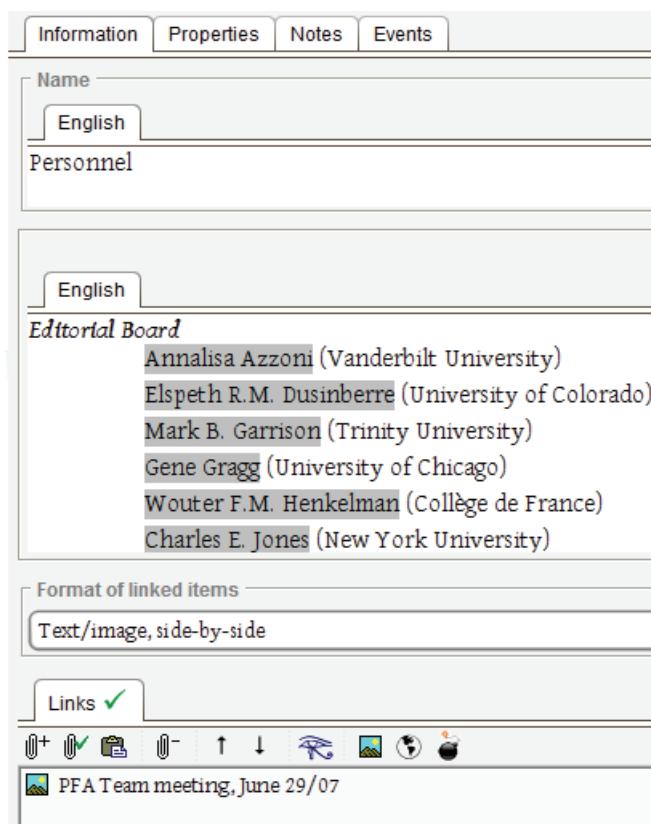
When you select the **Presentations** category in the navigation pane and click the ↲ “insert below” button in the main toolbar, a new hierarchy will be created. A hierarchy in this category is the organizing framework for a presentation. Each presentation consists of separate components that are represented as individual **Presentation Component** items within the hierarchy.

Presentation hierarchies are “subordination hierarchies,” not “containment hierarchies” (see the section on “Containment Hierarchies and Subordination Hierarchies” above in Chapter 4). Thus, when you select a hierarchy in the **Presentations** category and click the ↲ “insert below” button in the main toolbar, you will be given a choice to create a heading or a new **Presentation Component** item. A heading may in turn contain components or additional headings. The order of components at each level in the hierarchy and the nesting of components within headings and subheadings determine their display order in the presentation.

As elsewhere in OCHRE, the item pane of a presentation hierarchy contains data-entry fields for the hierarchy’s **Name**, **Abbreviation**, and **Description**. There is a tab at the top of the item pane for entering the presentation’s **Creators** and the date and time it was created. There are other tabs for entering **Preferences**, **Properties**, **Links**, and **Notes** that pertain to the presentation as a whole.

A presentation hierarchy has an additional tab, not found in other kinds of hierarchies, entitled **Style**. This tab is used to specify how the presentation’s components are to be displayed to an anonymous view-only user. At the top of the **Style** tab there is a data-entry field entitled **Presentation style for subitems** that contains a drop-down pick-list with a number of available styles.





The various presentation styles are explained below. But first it is necessary to explain how an individual presentation component is used to select data that will be displayed in accordance with the chosen style. If a **Presentation Component** item is selected in the navigation pane, you will see a tab at the top of the item pane in which to enter the component's **Creators** and the date and time it was created. There will also be tabs for entering optional **Properties**, **Notes**, and **Events** that pertain to the component (see Chapters 5 and 7). However, the most important information about a presentation component is shown in its **Information** tab (see the example above). This tab contains data-entry fields for entering the component's **Name**, **Abbreviation**, and **Description**. There is an additional text-entry field below the **Name** field for entering content for the component that will be displayed to users who view the presentation.

Additional content for a presentation component is specified in its **Links** tab, which is shown at the bottom of the **Information** tab. One or more links to database items from any category in the project can be entered in the **Links** tab in the usual way (see Chapter 6). These links determine what will be seen when a user selects the component as part of a presentation. Above the component's **Links** tab is a data-entry field entitled **Format of linked items**. This field contains a drop-down pick-list with the following formatting options:

- **Linear sequence**
- **Single item, view**
- **Single item, form view**
- **Single item, data entry**
- **Show links in sidebar**
- **Text/image, side-by-side**
- **Image/text, side-by-side**
- **Show in separate window**

The formatting option selected for a presentation component's linked items interacts with the presentation's style to determine how the items will be displayed. Many variations are possible. For a detailed discussion of the formatting options together with examples of how they can be used in conjunction with the various presentation styles, see the additional documentation concerning OCHRE presentations that is available on the Web at <http://ochre.uchicago.edu>.

Presentation component View

Personnel

Editorial Board

Annalia Azzoni (Vanderbilt University)
Elspeth R. M. Dusinberre (University of Colorado)
Mark B. Garrison (Trinity University)
Gene Grage (University of Chicago)
Wouter F.M. Henkelman (Collège de France)
Charles E. Jones (New York University)
Matthew W. Stolper (University of Chicago)

Technical Staff

Lec Maj (University of Chicago)
John Sanders (University of Chicago)
Sandra Schloen (University of Chicago)

Collaborators

Pierre Briant (Collège de France)
Robert Englund (University of California at Los Angeles)
Marilyn Lundberg (University of Southern California)
Bruce Zuckerman (University of Southern California)



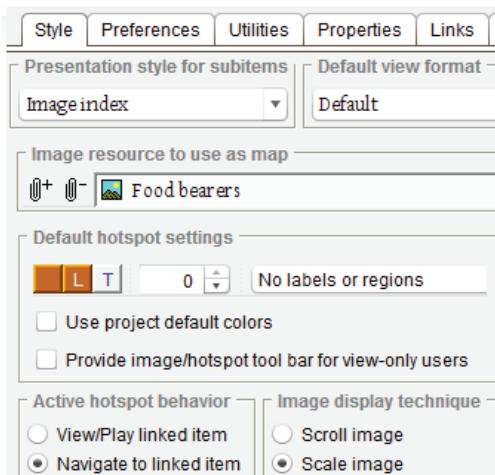
As you construct a presentation hierarchy and format its components, you can stop at any time to see how it will be displayed by opening the hierarchy's **View** tab. This shows how the presentation would appear to an anonymous view-only user. In the example on the facing page, generated from the presentation component illustrated on the previous page, the **Text/image, side-by-side** format was used.

The “Vertical” Presentation Style

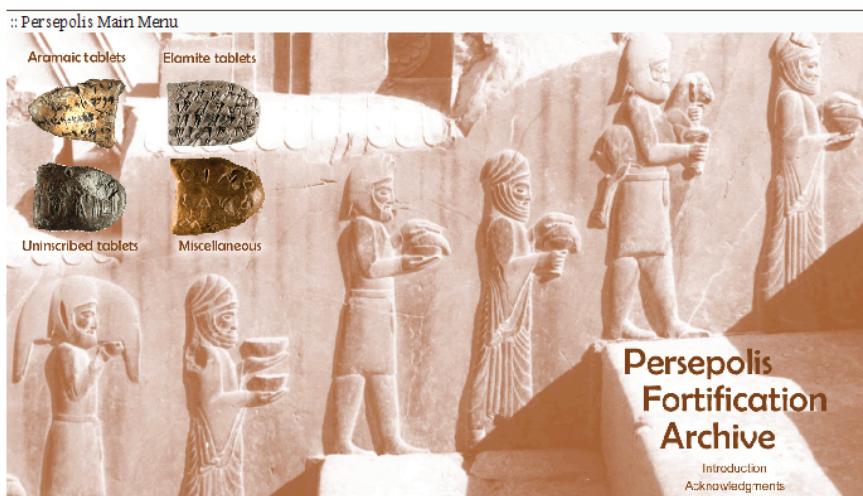
The **Vertical** style is the first choice in the **Presentation style** pick-list found at the top of a presentation hierarchy's **Style** tab. In this style, the presentation's components are stacked one on top of another in a vertically scrolling sequence. This style is useful when the number of components is limited and the presentation hierarchy is flat, with no nesting of components inside other components by means of subordinate headings.

The “Image index” Presentation Style

A particularly effective style is the **Image index**, which makes use of a bit-mapped image (e.g., a photograph or map) in which are embedded live links that a user can click to display other database items. The user can then navigate a project's data visually by clicking on the image and drilling down to lower levels of detail.



When you select the **Image index** option in the **Presentation style for subitems** pick-list of the hierarchy's **Style** tab, a link field entitled **Image resource to use as map** will be displayed in which you must specify the image to be used. An example of this presentation style is shown at the top of the following page.



Index of Miscellaneous Tablets



- [Akkadian Text](#)
- [Coin-impressed Tablet](#)
- [Enigmatic Text](#)
- [Erased Tablet](#)
- [Greek Text](#)
- [Old Persian Text](#)
- [Phrygian Text](#)

 [Abc](#) [Index](#) [Help](#) [Standard](#)

[Akkadian text: Description](#)

Akkadian text (logosyllabic, Akkadian)

Primary classification: Text

The “Index” Presentation Style

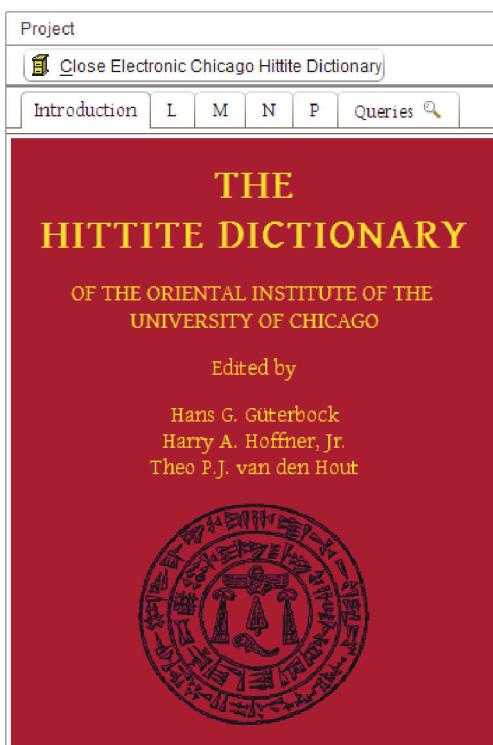
In the **Index** presentation style (shown above) the presentation’s components are listed by name in an index that serves as a navigation pane on the left side of the screen. A user can select a particular component to display by clicking on its name in the index. The

- ▼  [Miscellaneous Tablets](#)
- ▶ ★ [Overview](#)
- ▼ ★ [Index of Miscellaneous Tablets](#)
 -  [Akkadian Text](#)
 -  [Coin-impressed Tablet](#)
 -  [Enigmatic Text](#)
 -  [Erased Tablet](#)
 -  [Greek Text](#)
 -  [Old Persian Text](#)
 -  [Phrygian Text](#)

currently selected component is highlighted in the index and its contents (based on its linked items and the chosen display format for linked items) are shown in the main window pane to the right of the index. Note, however, that this style is not well suited to a nested hierarchy of components that have been grouped under subordinate headings. In that case, the **Tabbed index** style would be more suitable.

The “Tabbed index” Presentation Style

The **Tabbed index** style is similar to the **Index** style; however, it is better suited to displaying presentation components that have been grouped under headings in the presentation hierarchy. The top-level headings in the hierarchy are displayed as overlapping tabs in a navigation pane on the left side of the screen. Within each tab is an index list of the components that have been grouped under a particular heading in the hierarchy. A user can select a component to display by clicking on its name in the appropriate tab. The currently selected component is highlighted in the tab’s index and its contents (based on its linked items and the chosen display format for linked items) are shown in the main window pane to the right of the tabbed index.



The “Tabbed items” Presentation Style

In the **Tabbed items** presentation style, the components and headings at the top level of the presentation hierarchy are shown in a

series of overlapping tabs arranged horizontally across the screen (see the example below). This style works best when the presentation's components are not nested hierarchically under subordinate headings, in which case the **Tabbed index** style would be better. The **Tabbed items** style also works best with a wide monitor that has sufficient space for several overlapping tabs. If monitor width is limited, the **Vertical scrolling** style would be more suitable.

The screenshot shows a presentation interface titled "Elamite Tablets". At the top, there are four tabs: "Aramaic Epigraphs on Elamite Tablets", "Catalog of Seals", "Glossary, Elamite-English", "Overview", "Index of Elamite Texts", and "Index of Elamite Texts, by Category". The "Overview" tab is selected. Below the tabs, there is a section titled "Achaemenid Elamite" featuring a large image of an Elamite tablet and the following text:

The largest component of the Persepolis Fortification Archive consists of tablets and fragments bearing texts in Achaemenid Elamite language, written in cuneiform script. Most of the Elamite tablets also bear impressions of seals. The tens of thousands of Elamite cuneiform tablets and fragments are the remains of about 15,000–18,000 original documents. These originals in turn were

On the right side of this section, there are four smaller images labeled PF 0998 left edge, PF 0998 rev, PF 1000 top edge, and PF 0998 rev, each with a scale bar.

The “Linear navigation” Presentation Style

In the **Linear navigation** presentation style the presentation's components are displayed one by one as a series of pages. At the top of the screen are **◀ Back** and **▶ Forward** buttons that allow the user to go from component to component in a linear fashion according to the order of components specified in the presentation hierarchy. Alternatively, the user can go directly to a particular component by selecting it in the drop-down pick-list to the right of the navigation buttons.

The screenshot shows a presentation interface titled "General introduction". At the top, there are navigation buttons: "Back" (with a left arrow), "Forward" (with a right arrow), and a dropdown menu currently set to "Persepolis". Below the navigation, there is a section titled "Persepolis" containing the following text:

Pārsa, called Persepolis by later Classical Greek authors, is a complex of monumental buildings on a massive stone platform, accompanied by the tombs of kings cut into a slope overlooking the platform. The site is in the heart of ancient Persia, about 40 km northeast of modern Shiraz, in southern Iran.

At the bottom of this section, there is a link: "The Achaemenid Persian king Darius I (522–486 BC) began to build the platform and the first palaces early in his reign over an empire that reached from Central Asia to the Nile and from the

To the right of the text, there is a box titled "Persepolis Aerial View" containing an image of an aerial view of the archaeological site.

Web Links to OCHRE Database Items and Presentations

Information stored in an OCHRE database can be cited by means of a Web link (a Uniform Resource Locator or URL) to a particular database item. Some items represent a group of other items, as is the case with items in the **Sets** and **Presentations** categories. Clicking a Web link to an OCHRE item may therefore display data for an individual item or for a group of items; for example, a list of items found by a query or a more complex sequence and arrangement of items prescribed in a presentation.

A Web link to a database item in the University of Chicago's OCHRE database takes the form of a typical Web URL in the following format: `http://ochre.uchicago.edu/ochre.jsp?id=UUID` where *UUID* is a character string that serves as the “universally unique identifier” of the item. The following is an example of a universally unique item identifier of the kind used in OCHRE: `bd431eb5-0179-27d2-21b2-f3784124e520`. Every database item, no matter how small, has its own unique identifier.

You can obtain the URL for a database item by selecting the item in the navigation pane and opening its **View** tab in the item pane. At the bottom of the **View** tab the item's **Citation URL** will be displayed in the format described above. This URL can be copied and pasted into a Web page or some other external document as an ordinary Web link. This provides a simple mechanism for referencing and displaying a particular OCHRE item.

Citation URL

`http://ochre.lib.uchicago.edu/ochre.jsp?id=7d1859aa-2423-0965-f895-73be13fb71f5`

A public presentation of OCHRE data of the kind described in the first part of this chapter has a citation URL and thus can be launched by means of a Web link. This is possible because an entire presentation is itself represented by a single database item, namely, the hierarchy item in the **Presentations** category that contains the presentation's components. To obtain the URL of a presentation, select the relevant hierarchy in the navigation pane and then open the

View tab in the item pane. At the bottom of that tab you will see the presentation's **Citation URL**, which can be copied and pasted into an external document as a Web link. Clicking on the link will launch the OCHRE Java software and display the presentation. Embedding a Web link to a presentation in an ordinary Web page or other external document thus provides a simple way to display a public presentation of OCHRE data to anonymous view-only users (i.e., users who do not have a username and password assigned by the project).

Chapter 17

User Accounts, User Preferences, and Project Preferences

THERE are two kinds of OCHRE user: (1) the anonymous user who has not logged into a project and thus is limited to viewing whatever data the project has made public for such users to see; and (2) the named user who has logged into a project with a username and password assigned to him or her by the project administrator and thus is able not only to view but also to modify the project's data, to the extent allowed by the privileges granted to that user.

This chapter explains the procedure to be followed by a project administrator when creating user accounts for a particular project. It then explains the preferences which users can themselves choose to determine how information will be displayed to them and the project-wide preferences that the project administrator can establish for all users of the project.

The project administrator's own user account is created by the OCHRE database administrator when the project is initially set up. The project administrator can then add as many user accounts as the project needs without intervention from the database administrator. For information about setting up a project in the University of Chicago's OCHRE database and obtaining a project administrator account, please contact the OCHRE Data Service by e-mail at ochre@uchicago.edu.

Creating User Accounts and Controlling What Users Can See and Do

If you are logged in as a project administrator, you will see a **Users** category in the alphabetical list of item categories in the navigation pane. Database items in this category represent persons who have been given usernames and passwords by the project administrator that allow them to log into the project as named users. If you are logged into a project but you are not the project administrator, you

will see only one item in the **Users** category. This item represents your own user account. If you select it, you will be able to see your data-entry privileges and modify your display preferences (see the section below on “Individual User Preferences”).

To be designated as a user in the **Users** category, a person must first be represented as a database item in the **Persons & organizations** category (see the section above in Chapter 8 on “Describing and Organizing Persons and Organizations”). Below the **Type** field in the item pane of an item in this category is a checkbox labeled **Authorized user**. If the “current person” option is selected in the **Type** field and if you are logged in as the project administrator, you will be able to check this box to create a user account for the person. Note that you may need to click the  “refresh” button in the main toolbar to update the list of authorized users in the **Users** category.



Assigning Usernames and Passwords

When a user is selected by the project administrator in the **Users** category, there will be a data-entry field at the top of the item pane for the **Username**. After the username has been entered, the project administrator will check the appropriate checkboxes in the **Privileges** tab to specify the user’s data-entry privileges and will then click the **Activate account** button.

The user’s initial password will be a default password that the project administrator should communicate to the user in a secure fashion. The first time a user logs into OCHRE he or she should change the initial password to a personal password. The project administrator will not have access to any personal passwords but will be able to reset a user’s password to the default password at any time by selecting the user in the **Users** category in the navigation pane and clicking the **Reset password** button in the item pane.

The project administrator can deactivate a user’s account by clicking the **Deactivate account** button. This does not remove the user from the list of the project’s authorized users in the **Users** category

but simply prevents the user from logging into the project; the user account can later be reactivated by clicking the **Activate account** button. To remove a user from the project altogether, the project administrator would click the **Remove user from project** button. In either case, the user's underlying person item will not be deleted from the **Persons & organizations** category.

A person who has a user account in another OCHRE project can be registered as a user in the current project, allowing that person to use the same username and password to log into both projects. This can be done if the relevant person item in the other project's **Persons & organizations** category is accessible from the current project (contact the OCHRE Data Service by e-mail at *ochre@uchicago.edu* for assistance in gaining access to other projects' users). If so, the person can be selected in the **Linked Items** tab in the reference pane and copied into the current project's **Persons & organizations** category (see the section on "Repeating Items and Hierarchical Branches" in Chapter 4). The copied person will then appear automatically as a user in the current project's **Users** category with the same username as in the other project. The username and password cannot be changed by the current project's administrator; however, the copied user's privileges in the current project can be specified and the copied user can be removed at any time by clicking the **Remove user from project** button.

Specifying User Privileges

Privileges for viewing and editing the project's data are specified by the project administrator via a grid of checkboxes in the **Privileges** tab. These checkboxes determine the user's ability to insert, modify, or even to view various categories of database items. Data-entry privileges can be quite narrowly defined for each user. For example, a research assistant might be allowed to modify items in the **Locations & objects** category but not to modify items in other categories.

The various OCHRE item categories are listed alphabetically down the left side of the **Privileges** tab. Seven columns of checkboxes are arranged horizontally across this tab, allowing the project

administrator to specify the user's level of access to items in each category. The access levels are arranged from left to right in decreasing order. Checking the box in a particular column automatically confers upon the user the lower levels of access indicated in the columns to the right of the selected column.

The screenshot shows a user management interface with the following details:

- User View** tab is selected.
- Name**: Schloen, Robbie
- Username**: xxxxxxxx
- Action buttons: Activate account ..., Deactivate account ..., Reset password ..., Remove user from project ...
- Privileges tabs: Privileges (selected), Preferences, Properties, Links, Notes.
- Update user privileges ...** table:

Access	Admin	Insert/Delete	Insert only	Edit content	View/Link	View only	No access
Bibliographies	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Concepts	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Dictionaries	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Locations & objects	<input checked="" type="checkbox"/>						
Periods	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Persons & organizations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Presentations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

The access levels are:

- **Admin**, which is full access to all items in the category—this level of access is reserved for the project administrator and is not available to other users;
- **Insert/Delete**, which lets the user insert and delete items in the category and also view and link items and edit their content;
- **Insert only**, which lets the user insert new database items but not delete them and also lets the user view and link items and edit their content;
- **Edit content**, which lets the user view and link items and edit their content (i.e., their data fields and textual content);
- **View/Link**, which lets the user view and link items;
- **View only**, which lets the user view items but nothing else;

- **No access**, which prevents the user from seeing the items.

The project administrator always has **Admin** level access to all item categories for the project. The highest level of access that other users can be granted for any category is **Insert/Delete**. Whenever the project administrator changes a user's privileges the changes will not take effect until the **Update user privileges** button has been clicked.

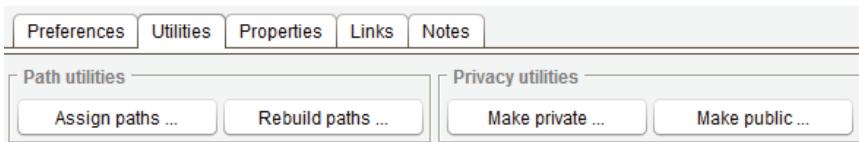
Keeping Items Private and Making Them Public

Specifying different privileges for different users is one way to control what people can see in a project's data. However, customized user privileges of this sort can be specified only for named users who have user accounts. To keep items hidden from anonymous users, you can mark items, hierarchies, and entire categories as "private." This is done during data entry by means of the checkboxes in the **Preferences** tab of the item pane that are labeled **Keep this item private**, **Keep this hierarchy private**, and **Keep this category private**. The label of the checkbox will depend on what is currently selected in the navigation pane. In addition, there is a **Private?** checkbox in the **Notes** tab that lets you make a note invisible to anonymous users.

The entire project can be made totally invisible to anonymous users by checking the **Keep project data private** checkbox in the **View-only** subtab of the project's **Preferences** tab. If this box is checked, the project will not appear in the list of public projects visible to anonymous users when they start OCHRE. If this box is not checked, the project will appear in the list of public projects; but any items, hierarchies, or categories that have been marked as private will not be visible.

In many cases, items that have been marked as private will not remain private forever but will be made public when the project's data-entry and editing phase has been completed. To make the data public the project administrator needs only to uncheck the relevant checkbox in the **Preferences** tab of each private item, hierarchy, or category. This is easy to do for an entire category or hierarchy but if individual items have been marked as private it may be time-consuming to find them all and uncheck their privacy checkboxes.

one by one. Thus, the project administrator can uncheck the privacy checkboxes of all private items in a given hierarchy all at once by selecting the hierarchy in the navigation pane, opening its **Utilities** tab in the item pane (this tab is visible only to the project administrator), and clicking the **Make public** button. Conversely, all of the public items contained in hierarchy can be marked as private by clicking the **Make private** button in the **Utilities** tab.



Overriding Database Locks

In data-entry mode it is sometimes necessary to “lock” a database item while you are modifying its information in order to prevent anyone else from trying to modify it at the same time (see the section on “Locking Items While You Modify Them” in Chapter 3). If a user forgets to unlock an item after he or she has finished modifying it, the project administrator may have to override the database lock so that other users can modify the item. To do so, the project administrator will select the item in the navigation pane and click the “unlock” button. This should be done carefully and only after making sure that the user who locked the item is not still editing it. The “whose lock?” button displays the name of the user who locked the currently selected item.

Individual User Preferences

Each user can establish personal preferences for his or her user account. This is done by selecting the relevant user account in the **Users** category and then opening the **Preferences** tab in the item pane. Only one item will be visible in the **Users** category, namely, the current user’s own account, unless the current user is the project administrator.

Password

Users can change their passwords by clicking the **Change password...** button at the top of the **Preferences** tab. This will display a window in which to enter the new password.

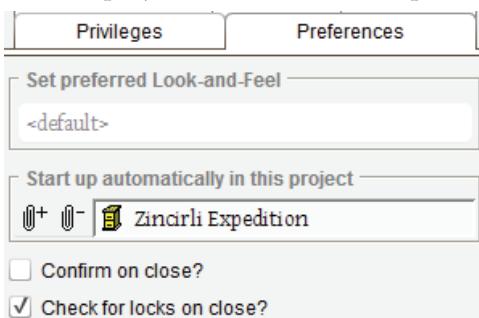
Look-and-feel

In the **Look-and-Feel** drop-down pick-list of the **Preferences** tab users can choose either a **Native** look-and-feel or the **Nimbus** look-and-feel. If the native option is chosen, the style and color of the windows, tabs, scrollbars, checkboxes, data-entry fields, and other elements by which the OCHRE user interface is displayed will be derived from the operating system on the computer that is currently being used.

For example, if you are using a Windows computer, the interface will be displayed using the Windows look-and-feel. If you are using a Macintosh computer, you will see the Macintosh look-and-feel. Likewise, a Linux computer will show its native look-and-feel. If you select a different look-and-feel, the change will not take effect until the next time you log into OCHRE.

Project to Open at Log-in

The **Preferences** tab also contains an optional link field entitled **Start up automatically in this project**. A user would specify a project in this field to avoid having to select the project from a list and open it manually every time he or she logs into the database. This option is especially relevant to users who work with only one project. The specified project will be selected and opened whenever the user logs in.



Confirm on close?

The **Confirm on close?** option allows users to specify whether they wish to be prompted to confirm their intention to close OCHRE whenever the program is about to close. This prevents the user from closing OCHRE inadvertently.

Check for locks on close?

The **Check for locks on close?** option allows users to specify whether a check should be done to see if they have left database items locked when they are about to close a project. If any items are locked, they will be listed and the user will be given three options: (1) **Unlock and Close**, which will unlock all of the locked items and go on to close the project; (2) **Unlock**, which will unlock all of the locked items but not close the project; and (3) **Close project**, which will close the project without unlocking the locked items.



Project-wide Preferences

The project administrator can establish project-wide preferences that control how information is displayed to all users of the project's data. This is done by selecting the project item in the navigation pane (i.e., the item at the top left of the navigation pane that is the parent of all the item categories) and then opening the project's **Preferences** tab in the item pane. This tab has several subtabs in which various kinds of preferences can be entered.

Data-entry Preferences

The **Data-entry** subtab of a project's **Preferences** tab contains several options that determine how the project's data will be displayed to logged-in, named users in data-entry mode. At the top of the **Data-entry** subtab is a checkbox to **Display unused categories**. Checking this box will cause all item categories to be shown in the navigation pane even if the project has no database items in a particular category.

If the box is unchecked, unused categories will not be shown in the navigation pane, thus simplifying the user interface. For example, if a project does not use the **Dictionaries** category, it can be hidden.

There is another checkbox at the top of the **Data-entry** subtab to **Enable multilingual features**. Checking this box will allow users to add translations of textual content. If the box is unchecked, the various data-entry fields used to enter textual content will be displayed in a more streamlined fashion because there will be no need to allow space for multiple tabs that contain translations of names, abbreviations, descriptions, notes, and other textual content.

Below this checkbox is a drop-down pick-list in which the project administrator can specify the project's **Default modern language**. This is the language in which primary names and other untranslated textual content are entered. Note that there is no requirement that the default language be English or any other European language. OCHRE supports Arabic and Hebrew, which are written right-to-left, as well as nonalphabetic East Asian languages and scripts such as Chinese, Japanese, and Korean.

Below the default language pick-list is a list of checkboxes in which the project administrator can specify the available **Languages for translation**. The languages that are checked in this list will appear in a drop-down pick-list whenever a user clicks the  “translation” button to create a new translation tab while entering textual content (see the section on “Item Names and Aliases” in Chapter 4 for an explanation of how to enter translations).

There are two optional link fields in the **Data-entry** subtab that pertain to the automatic importing of texts (see the section in Chapter 11 on “Importing Existing Digitized Texts”). The field labeled **Dictionaries for automated text entry** allows the project



administrator to add one or more links to hierarchies in the **Dictionaries** category. The field labeled **Writing systems for automated text entry** is for links to hierarchies in the **Writing systems** category.

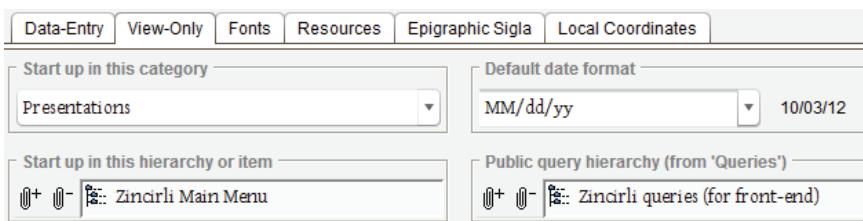
Any dictionaries and writing systems specified in the **Data-entry** subtab will be used to look up discourse units (in the dictionaries) and epigraphic units (in the writing systems) during the text-importing process. They will be used to supplement the dictionaries and writing systems that are specified by users when importing texts. If no matches are found in what a user has specified (or if the user has not specified a dictionary or writing system), matches will be sought in the default dictionaries and writing systems specified by the project administrator. They will be searched in the order in which they are listed in the link fields and the search will cease as soon as a match is found.

View-only Preferences

The **View-only** subtab of a project's **Preferences** tab contains fields and checkboxes that determine how information will be displayed to anonymous view-only users. The view-only display of a database item seen by such users can be inspected in data-entry mode by opening the item's **View** tab in the item pane. The **View** tab provides a convenient way for logged-in, named users to see how the data they are entering will be formatted for anonymous users.

The project administrator can specify which item category should be displayed automatically to an anonymous user who opens the project and which hierarchy or item within that category should be shown. The category is chosen from the pick-list labeled **Start up in this category**. The starting hierarchy or item within the chosen category is specified by adding a link to the field labeled **Start up in this hierarchy or item**.

To use a predefined presentation to control what an anonymous user will see, the project administrator would select **Presentations** as the start-up category and then select a hierarchy within that category as the start-up hierarchy. (See Chapter 16 concerning public presentations of a project's data.)



At the top of the **View-only** subtab is a pick-list for choosing the **Default date format**. The selected format will be used to display dates and times. Below this pick-list are two link fields that pertain to the public queries a project may make available to anonymous users. The **Public query hierarchy** field determines which of the project's pre-defined queries an anonymous user will see. The query hierarchy specified in this field will be displayed in the navigation pane in a tab entitled **Queries** (see the section on "Displaying Queries for Use in View-only Mode" above in Chapter 14). The optional **Public query tutorial** field can be used to specify a presentation hierarchy that explains how to use a project's public queries.

The **View-only** subtab also contains the following checkboxes that allow the project administrator to control how the project's data will be displayed:

- **Keep project data private**, i.e., make the entire project invisible to anonymous users until it is ready to be made public (see the section above on "Keeping Items Private and Making Them Public").
- **Show reference pane on start-up**, i.e., display the reference pane on the right side of the OCHRE window by default. If this box is not checked, the reference pane will be hidden and the user must click the \emptyset or Q toggle button on the main toolbar to show it (see the section on "The Reference Pane" above in Chapter 3).
- **View document resources in separate windows**, i.e., open a new window to display each document resource item that a user selects, permitting multiple documents to be shown at one time in separate windows. If this box is not checked, the

same window will be used for each document that the user opens, automatically hiding the previously opened document.

- **View image resources in separate windows**, i.e., open a new window to display each image resource item that a user selects, permitting multiple images to be shown at one time in separate windows. If this box is not checked, the same window will be used for each image that the user opens, automatically hiding the previously opened image.

At the bottom of the **View-only** subtab is a data-entry field for the project's **Copyright notice** and a field for a **Watermark on images** to be overlaid on all images displayed to anonymous users (see the section on "Image Watermarks and Preferred Audio and Video Formats" in Chapter 9). The **Copyright notice** pertains to all of the project's resources, saving the trouble of entering a notice for each resource individually. However, a copyright notice which has been entered in the **Copyright** field of an individual resource item will override the project-wide copyright notice. This is necessary in case there are multiple copyright holders for a project's resources.

If the **Use copyright notice as watermark** checkbox is checked, the textual content entered by the project administrator in the **Copyright notice** field will be used as the watermark overlaid on the project's images. This provides a way to indicate who owns the images and to discourage unauthorized copying.

If the copyright notice is not going to be used as the watermark, the project administrator can enter textual content in the **Watermark text** field that will be overlaid on the project's images. The watermark will not appear in data-entry mode. It is intended to prevent an anonymous user who does not have a user account for the project from creating unmarked copies of the project's images to be used without attribution to the project.

The radio-button **Options** to the right of the **Watermark text** field determine whether the watermark specified in the **View-only** subtab of the project's **Preferences** tab will override watermarks that have been specified elsewhere in the project. Watermarks can be specified

at the item level for the image resource, at the resource hierarchy level, and at the project level. If the **Lowest** option is chosen, the watermark displayed on an image will be the one entered at the lowest level (i.e., an item-level watermark will override a hierarchy-level watermark, which will override the project-level watermark). If the **Current** option is chosen, the project-level watermark will override watermarks entered at a lower level. If the **SUPPRESS** option is chosen, no watermark will be displayed on the project's images regardless of what has been specified at lower levels.

Font Preferences

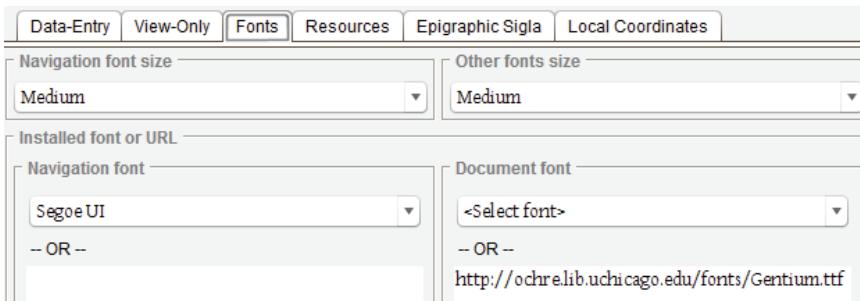
The **Fonts** subtab of a project's **Preferences** tab allows the project administrator to specify the fonts that will be used to display various kinds of textual content. All of the fonts specified here must be Unicode fonts. Four different fonts can be specified, as follows:

- A **Navigation font** that will be used in the navigation pane on the left side of the OCHRE window.
- A **Document font** that will be used in item names, abbreviations, descriptions, and notes; in resources of the “internal document” type; in translations of text items; and in pick-lists and other data-entry fields.
- A **Graphemic font** that will be used to display graphemic transcriptions of text items (e.g., a hieroglyphic font for Egyptian hieroglyphic texts).
- A **Phonemic font** that will be used to display phonemic and hybrid transcriptions of text items (see the section in Chapter 11 on “Graphemic, Phonemic, and Hybrid Transcriptions”).

In each case, the project administrator can use a pick-list to choose one of the fonts installed on his or her computer, on the assumption that the same font is available on each user's computer. Alternatively, if this is not a valid assumption, the project administrator can enter a URL to a font file stored on a Web server, from which the font will be downloaded as needed. The latter option is normally recommended in order to ensure that the specified font is

available to all users. (Note that a downloaded font is used only during the OCHRE session and is not permanently installed on the user's computer.) If no font is specified for a given type of content, a default font will be used.

The project administrator can also specify the display size of the fonts. At the top of the **Fonts** subtab are pick-lists in which to specify the **Navigation font size** and the **Other fonts size**. In both cases there are five different size options: **largest**, **larger**, **medium**, **smaller**, and **smallest**. It is often desirable to make the navigation font larger than the other fonts, which is why there are two font-size pick-lists.



Resource Preferences

The **Resources** subtab of a project's **Preferences** tab contains a field labeled **Resource root folder (high-resolution)** in which the project administrator can specify the Internet location of the full-size versions of the project's external resource files (see the section above in Chapter 9 on "File Locations for External Resources"). In the case of images, in particular, there is often a distinction between the uncompressed high-resolution version of an image and a much smaller "preview" or "thumbnail" version of the same image. The high-resolution version contains more detail but may require a long time to be transmitted to the user's computer.

For this reason, there is a second field in the **Resources** subtab entitled **Preview root folder (low-resolution)**. In this field the project administrator can specify the Internet location of the smaller, compressed versions of the project's external resource files. When viewing an image in OCHRE the low-resolution version will be downloaded

first, allowing the user to view it before deciding whether to download the high-resolution version.

In either case, the root folder is a character string that specifies only the first part of the URL for a given resource. Each resource item has a **File URL** data-entry field in its item pane in which to enter the second part of the URL, including the actual file name. This is concatenated to the root folder to complete the URL. The resource root folder is distinguished from each resource's relative path in this way in order to allow the project's external resources to be moved en masse to a different server location (i.e., to a different root folder) without disrupting the hierarchical organization of the files into folders and subfolders.

The high-resolution and low-resolution versions of the same image resource can be distinguished by means of file suffixes (see the section in Chapter 9 on "Image File Suffixes and File Formats"). If a suffix is specified in the **Image resource file suffixes** area of the **Resources** subtab, it will be used to find the appropriate file, depending on which version is to be displayed.

High-resolution and low-resolution versions of the same image resource can also be distinguished by means of their file formats. The file extensions used to indicate the differing formats of a project's high-resolution and low-resolution files can be entered in the **Image resource file formats** area of the **Resources** subtab. Only the file extension should be entered, not the dot that separates the file name from the file extension (e.g., enter "tif" for TIFF and "jpg" for JPEG).

At the bottom of the **Resources** subtab are two pick-lists in which the project administrator can specify a **Preferred audio format** for audio resources used by the project and a **Preferred video format** for video resources. This is useful in cases where an audio or video clip is stored in two different versions in two files that use different formats (e.g., Microsoft's AVI format and Apple's MOV QuickTime format). Both versions of the resource can be stored in the same folder because only the preferred format will be used if there is a choice of more than one format.

Epigraphic Sigla

In the **Epigraphic Sigla** subtab of a project's **Preferences** tab the project administrator can customize the epigraphic sigla that are used in transcriptions of texts to indicate various kinds of damage, erasure, and emendation (see the section on "Indicating Damage, Erasure, Gaps, and Emendation" in Chapter 11). The project administrator can also customize the sigla that are used to indicate sign placement and as separators of words, syllables, logograms, etc.

The project's current sigla are shown in the **Epigraphic Sigla** subtab. To change a particular siglum, the project administrator must move the cursor and click the mouse to select an existing siglum and then type in the desired character or characters where the flashing text-editing cursor appears. The keypad of special characters can be used to enter special characters that cannot be entered via the keyboard (see the section on "Using the Keypad of Special Characters" above in Chapter 5). Changes to the epigraphic sigla can be canceled and the default sigla restored by clicking the **Restore default sigla** button.

Local Coordinates

Database items in the **Locations & objects** category have a **Coordinates** tab for entering their map coordinates, which can be either geographical coordinates expressed in degrees of latitude and longitude or local Cartesian coordinates that specify a position in

Separators	
Word:	-
Syllable:	-
Logogram:	.
Numeral:	+
Fraction:	/
Single line:	-
Double line:	=
Sign Placement	
Written within:	x
Written as ligature:	+
Written over erasure:	
Written above the line:	^
Written below the line:	
Other	
Scribal error, emended:	(I)
Collated reading:	*
Uncertain reading:	(?)
Unusual form:	

two-dimensional space (x, y) or in three-dimensional space (x, y, z). Many archaeological projects record the positions of architectural features, debris layers, and individual artifacts in terms of local coordinates rather than geographical coordinates.

In the **Local Coordinates** subtab of a project's **Preferences** tab the project administrator can indicate how to convert the project's local coordinates to geographical coordinates (see the section on "Entering the Local Coordinates of a Location or Object" in Chapter 8). If this conversion information is available, items that have local coordinates can be displayed on the same map as items that have geographical coordinates.

If a project's local coordinate space is a planar projection of geographical coordinates using the Universal Transverse Mercator (UTM) projection, the project administrator should enter the appropriate zone number in the **UTM zone** field of the **Local Coordinates** subtab, indicating also whether the zone is in the northern (**N**) or southern (**S**) hemisphere. This is all the information that is needed to convert local UTM coordinates to geographical coordinates and vice versa.

At the top of the **Local Coordinates** subtab is a drop-down pick-list in which to specify the units of measurement of the project's local coordinates, which can be one of the following: **meters**, **centimeters**, **kilometers**, **feet**, **inches**, or **miles**. If a UTM zone is specified in the **UTM zone** field, the **Units** field must be set to **meters**. In the UTM coordinate system the **X (easting)** value is the distance in meters east of the central meridian of the projected planar zone, which is arbitrarily assigned an easting value of 500,000 meters; and the **Y (northing)** value is the distance in meters north of the equator.

If the local coordinates are not UTM coordinates, the project administrator must enter the geographical coordinates (latitude and longitude) of three points in the local coordinate system. These values will be used to convert local coordinates to geographical coordinates. In the data-entry areas labeled **Lat-long of (0,0)**, **Lat-long of (0,1000)**, and **Lat-long of (1000,0)** there are subfields for entering the latitude and longitude of the (0,0) origin and two other points in the local

Cartesian coordinate space. Latitudes and longitudes must be entered in decimal degrees with up to eight decimal places, ranging from -90 to +90 for latitude and from -180 to +180 for longitude.

Chapter 18

Importing and Exporting Data

OCHRE is an online database system. It provides a central place where all of a project's data can be stored and viewed. In many cases, data will be entered directly into the database by end-users. But there are circumstances in which an external data file must be imported into the database or exported from it. For example, data tables created using Microsoft Excel, FileMaker, or similar software can be automatically imported. Likewise, digitized texts that have been previously created using text-editing software can be uploaded all at once into the OCHRE database.

Conversely, selected information in the form of data tables and texts can be exported from the central database for use with other software. Data can even be "round-tripped" by repeatedly exporting and re-importing the same table with the aid of an automatic table-synchronization mechanism that matches the table's structure to OCHRE's internal data structures. This is useful when you need to record or edit data offline using a computer that is not connected to the Internet and then go online to upload your changes.

In addition, all or part of a project's data can be exported as a stand-alone archive consisting of self-describing XML files. Funding agencies often now require long-term preservation of research data in a standards-compliant and Web-accessible archive. A stand-alone OCHRE archive uses a "denormalized" XML format that is much less atomized than the underlying database format. However, an archive is only a snapshot of the dynamic database content at a particular time; thus, it will become out of date whenever your data is modified in the central database, requiring you to regenerate the archive. The advantage is that it stores the data in a simpler form that does not require the use of specialized database software but can be displayed by Web browsers and searched by Web search engines.

Contact the OCHRE Data Service at ochre@uchicago.edu for information about importing and exporting data files, synchronizing external data tables with the central database, and archiving data.

Importing and Exporting Tables and Texts

The procedure for importing a data table is not described in this manual. Importing a table can be quite complex and is part of the data-conversion service offered by the staff of the OCHRE Data Service, who can be reached by e-mail at *ochre@uchicago.edu*.

The procedure for importing a text is described in detail above in Chapter 11, in the section on “Importing Existing Digitized Texts.” A text to be imported must be represented as a series of character strings separated by white space and line breaks. OCHRE uses the punctuation, white space, and line breaks to construct hierarchies of epigraphic units and discourse units that represent the text and to establish cross-cutting links between the text’s epigraphic units and its discourse units.

Data tables and texts are both exported in the same way, by clicking the  “export as Excel” button or the  “export as PDF” button in the **View** tab of the relevant database item. The procedure for exporting a table of data is described in Chapter 4, in the section on “Viewing a Set of Items”; and also in Chapter 15, in the section on Printing and Exporting a Table of Data.” The procedure for exporting a text is described in Chapter 11, in the section on “Viewing a Text.”

Only two file formats are available for exporting a table or text: the Microsoft Excel XSLX format and the Adobe PDF format. When a text is exported in the Excel format, different spreadsheet columns are used for the various transcriptions of the text (graphemic, phonemic, and hybrid); for its translation; and for its name, description, notes, and properties. The Excel format was chosen as the primary export format for OCHRE texts and tables because many different kinds of software can work with this format.

Chapter 19

A Generic Ontology for Scholarly Research

IN computer science, an ontology is a formal description of concepts and relationships relevant to a particular domain of knowledge.¹ Taxonomies, thesauruses, and other kinds of “controlled vocabularies” can be regarded as partial ontologies. However, a full ontology goes beyond a controlled vocabulary by specifying the relationships among various kinds of entities and not just the terms used to describe the entities.

Most ontologies are quite specific and thus are limited in their application, being restricted to a particular domain of knowledge. For example, an archaeological project will typically make use of a set of terms and conceptual distinctions that are applicable to the region and period of the archaeological site being excavated but are not applicable to all sites or to other kinds of cultural and historical research.

Some ontologies are more generic, however. They are defined on a higher plane of abstraction in order to facilitate the sharing of information between specific knowledge domains. Such ontologies are sometimes called “upper ontologies” or “foundation ontologies.” OCHRE employs an upper ontology that is well suited to scholarly research in archaeology, history, and textual studies, and could easily be applied more broadly in many other academic disciplines.

¹ In philosophy, ontology is the study of “being,” or existence, and of the basic categories of being. The term was borrowed by artificial-intelligence researchers in the 1980s. Tom Gruber, then at Stanford University, defined it this way: “A body of formally represented knowledge is based on a *conceptualization*: the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold among them. . . . An *ontology* is an explicit specification of a conceptualization.” See Thomas R. Gruber, “A Translation Approach to Portable Ontology Specifications,” *Knowledge Acquisition* 5 (1993) 199–220.

Cultural disciplines are characterized by a wide variety of domain-specific ontologies. This creates a need for a generic ontology that can subsume specific and often quite heterogeneous ontologies while preserving all of their idiosyncratic terms and conceptual distinctions. OCHRE's generic ontology provides the basis for integrating, within a common semantic framework, the heterogeneously structured data produced by many different research projects, each of which has its own nomenclature and recording system. It enables the automated analysis of large amounts of information of diverse origin.

Note, however, that an ontology is merely a description of concepts and relationships; it is not a technical specification for any particular data-retrieval system. It must be implemented in a working computer system by means of a “schema” which defines logical data structures that represent what is described in the ontology.

The distinction between ontologies and schemas is sometimes ignored but is nonetheless important. A given ontology can be implemented in a variety of different database systems that are based on different data models and thus employ different logical schemas. We have implemented the OCHRE ontology in an XML database system based on the semistructured data model, as opposed to a system based on the more familiar relational data model (see Chapter 20). Our implementation of the OCHRE ontology makes use of hierarchies of XML data objects that are retrieved by means of XQuery, the querying language specifically designed to work with XML data.

However, as we explain in the next chapter, the OCHRE ontology could also be implemented in a relational database system using data objects defined as relational “tuples” (table rows) and retrieved by means of the SQL query language. Likewise, it could be implemented in a so-called graph database using data objects defined as subject-predicate-object “triples,” following the RDF specification, and retrieved by means of the SPARQL query language.²

² XML is an acronym for Extensible Markup Language, which defines a widely used tagged-text format for transmitting structured information via the Internet (see <http://www.w3.org/xml>). RDF is an acronym for Resource Description Framework, which uses XML to specify how one can represent information in terms of subject-

We have found that an XML database system using XQuery is the most efficient way to implement the OCHRE ontology using currently available technology. But the ontology itself is independent of any particular implementation. The decision about how best to implement it is a matter of software engineering and depends on the software tools and standards that are available at the time. What matters most in the long run is the ontology, not the implementation. Converting data from one database implementation to another—from an XML schema to a relational schema or an RDF-graph schema—is a relatively simple matter, provided that the underlying ontology remains the same. Thus, the OCHRE ontology is a stable and sustainable basis for storing and integrating cultural information over the long term, even as database software and schemas change.

An Item-based Approach

As the preceding chapters have made clear, the basic structural elements in OCHRE’s ontology are individual items of information that are of interest to particular researchers and are defined in relation to their research needs. Each item is a unit of information that has its own name, properties, and links to other items. Items of interest can be defined by researchers at any spatial or temporal scale. An item can correspond to any kind of entity, including physical objects, spatial locations, periods of time, written works, digital resources, abstract concepts, and active agents (who may be individual or collective; living or dead; historical or fictional). Hierarchies of items are used to represent in a concise and intuitive manner inter-item relationships of spatial containment, temporal sequence, and logical subordination. Cross-hierarchy links between items represent other relationships.

predicate-object “triples” (see <http://www.w3.org/rdf>). SPARQL is a query language designed for retrieving information from large collections of RDF triples (see <http://www.w3.org/TR/rdf-sparql-query>). XML, XQuery, RDF, and SPARQL are all nonproprietary standard specifications published by the World Wide Web Consortium, which is responsible for the HyperText Markup Language (HTML) standard on which the Web itself is based.

The OCHRE ontology does not predetermine how items are defined or interrelated. It does not predetermine the terminology to be used to describe them. Instead, it provides a generic framework within which researchers can organize their units of information and describe them using their own taxonomies. A small number of high-level semantic categories are used to classify items as units of space, time, agency, discourse, and so on (see Chapter 2). But beyond this high-level categorization there is no requirement that an item be assigned to a predetermined class of items.³

In this respect, the OCHRE ontology differs from the class-based ontologies long prevalent in commercial and governmental database systems, in which the basic structural elements are predefined classes of items, usually represented in tabular form with one table per class. In a class-based ontology, individual items of interest correspond to rows in a table and the items in a particular table are assumed to have a common set of properties, which correspond to the table's columns. For example, the class of customers for a particular business would be represented by a table in which each row represents a different customer and the columns represent separate pieces of information about each customer.

The class-based design so widely used in the business world is not well suited for scholarly research in archaeology, history, or textual studies. Tables are easy to work with using commercial database software, which helps to explain the popularity of this approach even among scholars who study human culture and history. In archaeology, for example, the usual approach has been to define a group of classes and to create a table for each class; for example, one table each for debris layers, architectural features, ceramic artifacts, metal artifacts, stone artifacts, faunal remains, botanical remains, etc. Each

³ The hierarchical, item-based approach employed in OCHRE is described in David Schloen's article "Archaeological Data Models and Web Publication Using XML" in *Computers and the Humanities* 35 (2001) 123–152. This article is out of date in some respects. It was written before the OCHRE project was begun and before XML databases were widely available, but it explains the basic design principle that has informed the development of OCHRE.

observed entity is represented as a row in one of the tables. The table columns represent the properties used to describe the entities.

However, class-based ontologies of this sort require every item of interest to be placed within a predefined structure. The properties available to describe an item are thus limited to the columns predefined in its table. It is not easy to add or delete a property for an individual item because this would affect an entire table column and hence an entire class of items. Scholarly observations are forced into a rigid mold, relegating information about idiosyncrasies or unusual properties to unstructured (or unpredictably structured) notes.

As a result, a class-based ontology makes it difficult to represent the variability and complexity of the information encountered in academic disciplines that deal with human culture. These disciplines have few common standards that prescribe how items of interest ought to be distinguished and described. In such an environment the fundamental problem with a class-based ontology is that the classification of items is determined in advance by the creator of the database system instead of allowing multiple overlapping classifications to emerge from the ongoing analysis and comparison of many individual units of information.⁴

In an item-based ontology, on the other hand, any number of properties can be used to describe a particular item and new properties can be added as needed without affecting other items. Classes are not built into the database ahead of time but are generated by queries based on the properties of individual items. Queries yield sets of items that are considered (at least by the person doing the query) to be in some way similar or worthy of being grouped together. Classifications are determined by the researchers themselves as they work with data. No classification is treated as absolute or universal.

⁴ It is important not to confuse the class-based organization of data described here with the relational data model, which is discussed in Chapter 20. A hierarchical, item-based ontology can be implemented in a relational database system, as we did in an earlier database system that served as a prototype for OCHRE. However, in practice, most people use relational tables in a “flat” manner to represent predefined classes of similar items.

This is in keeping with the assumptions and practices of critical scholarship in the humanities and social sciences, in which semantic authority does not belong to the head of a bureaucratic organization or to a discipline-wide standards committee, but to the individual scholar—or, in the case of collaborative research, to the individual research team and its leaders. Accordingly, OCHRE treats taxonomies as works of scholarship to be credited to particular researchers.

OCHRE can accommodate as many taxonomies as there are researchers. Having said that, OCHRE also makes it easy to share taxonomies, in whole or in part, among groups of scholars who agree on a common approach. Indeed, a standardized taxonomy, if it is available and useful, can easily be employed. However, OCHRE does not enforce such standards. The choice of whether to use them is left up to the individual scholar or research team.

Semantic authority and standardization are discussed in more detail below in the section on “Ontology Alignment and Data Integration.” But first it will be useful to consider two key features of OCHRE’s generic ontology, namely, its atomization of information and its use of recursive hierarchies.

Atomization and Dynamic Recombination

Information is broken down into small units within a database that implements the OCHRE ontology. The many different entities and relationships that scholars distinguish in the course of their research—and the divergent observations and interpretations of the same entities by different scholars—are all represented as separately addressable units of information with unique “keys.” Indeed, scholars themselves are represented by distinct units of information in order to attribute observations and interpretations to particular people.

For example, an archaeological artifact is represented as a spatially situated unit of observation in a hierarchy with other such units and its properties are represented by means of links to taxonomic units (variables and values), which are themselves organized in a taxonomic hierarchy that constrains the possible pairs of variables and values allowed in particular contexts (see Chapter 5). Likewise, a text that is

an object of study in its own right is represented as a hierarchy of epigraphic units broken down to the level of individual graphic signs and the text's epigraphic hierarchy is cross-linked with a separate hierarchy that represents the text's discourse units, down to the level of individual words and perhaps even morphemes (see Chapter 11).

This high degree of atomization prevents inefficient and error-prone duplication of the same information at different places in the database. Moreover, it exposes in an explicit manner all the relationships and dependencies among separate units of information. As a result, it permits great flexibility in the use of the information. Atomized units of information can be dynamically recombined and presented in many different ways, depending on the needs of a particular user at a particular time.

Atomization and dynamic recombination of information have long been the hallmarks of well-designed relational database systems. The process of structuring relational table schemas and the linkages between them in a way that maximizes the benefits of atomization is called "normalization." In the case of XML databases, which are well suited to implementing the OCHRE ontology, a similar process yields an atomized design that achieves the same purpose. Thus, the XML document schemas discussed below in Chapter 20 specify the structure of a normalized database—using this term, which is derived from relational database theory, in a nontechnical sense.

However, a high degree of atomization requires specialized software that can retrieve and recombine information from the database to meet particular needs. An OCHRE database is atomized to the point where each descriptive property that is used in the database is itself an independent, addressable unit of information that exists in only one place and is linked to all the entities that it describes. Thus, for example, the character string "diameter" that names a descriptive property exists in only one place in the database. Any change to this character string will be instantly available in any view of the data in which this property is used.

This has many advantages with respect to maintaining the accuracy, integrity, and flexibility of the database. However, a specialized

database client application, such as the Java user interface described in this book, is needed to follow the links in order to reconstitute from the atomized units a readable version of the information in response to an end-user's actions.

In some cases, people may wish to search and display OCHRE data without the aid of a specialized database client application, relying only on a Web browser and a string-matching search engine. For this reason, it is possible to export from an OCHRE database a “denormalized” archive consisting of plain-text XML files, in which the information is much less atomized than it is in the database (see Chapter 18). To generate such an archive from the database, the software will follow the links among atomized units in order to reconstitute the information in a human-readable form that is easy to display but is characterized by a great deal of redundancy and a lack of explicit, machine-readable internal structure.

Hierarchies and the Power of Recursion

OCHRE makes extensive use of hierarchies to represent, in a compact and efficient manner, relationships of spatial containment, temporal sequence, linguistic structure, and logical subordination (see the section on “Hierarchies of Items” in Chapter 4). However, these hierarchies are not rigid structures into which information must be forced in an unnatural way. The contents of each hierarchy are determined, not by the software, but by end-users, for whom hierarchies are an intuitive means of representing the entities and relationships of interest to them. A given entity may appear in more than one hierarchy or it may appear in different locations within the same hierarchy; thus each hierarchy functions as a model of the relationships among a set of entities without preventing the same entities from being modeled in the database in other ways. Different scholars will often employ different hierarchical structures to model the same entities, reflecting different conceptions of the relationships among them, and OCHRE is designed to facilitate this.

Furthermore, the number of levels within a hierarchy is up to the end-user. A hierarchy can be highly ramified, with as many levels as

are needed to represent the relationships someone has identified. At the other extreme, a hierarchy might have only one level, yielding a “flat” list rather than a tree structure. Thus, OCHRE users are not forced to organize their information hierarchically when a simple list would suffice.

Finally, OCHRE does not rely solely on hierarchies to represent relationships among entities. Users can create direct links from one entity to another, cutting across hierarchies or connecting entities within the same hierarchy in a nonhierarchical way. (See Chapter 6 for an explanation of the various ways of linking entities.)

Having said this, the benefits of a hierarchical organization of information should not be underestimated. Hierarchies are used wherever possible in OCHRE because they provide a predictable structure for representing complex relationships. Recursive hierarchies are especially useful in this regard because they can be manipulated using powerful programming techniques that simplify the software and make it easier to maintain. A recursive hierarchy is one in which entities contain other entities of the same type, repeating the same structure at successive levels of the hierarchy. Such a hierarchy therefore exemplifies the “divide and conquer” approach in which complex problems are broken down into simpler problems whose solutions are then combined to give the solution to the original problem.

In the earlier chapters of this book, recursive hierarchies are called “containment hierarchies” to distinguish them from nonrecursive “subordination hierarchies” (see Chapter 4). For example, spatial hierarchies in the **Locations & objects** category are recursive because they represent the spatial containment of units of observation at successively smaller scales of measurement. Likewise, temporal hierarchies in the **Periods** category are recursive because they represent the containment of subperiods within longer periods, and discourse hierarchies in the **Texts** category represent recursive linguistic structures.

Many linguists, following Noam Chomsky, believe that recursion is a fundamental property of human language and, indeed, is what distinguishes human language from other kinds of animal communi-

cation.⁵ Whether or not this is true, it is clear that recursive hierarchies are easily understood by most people and serve as an intuitive mechanism for organizing information, as well as being easy to work with computationally. Thus, they play a major role in OCHRE's ontology and the internal database structures that implement it.

Beyond Gutenberg: Escaping the Digital Library Paradigm

The metaphor of the “digital library” is widespread. This metaphor has its uses but, unfortunately, it has infected software development for the humanities in a way that limits the potential benefits of digitization. Structural concepts derived from the technology of printed books are too often used as the conceptual basis for digital data structures rather than being used merely as convenient ways to present information to human readers who have been trained to read printed books. The slavish digital imitation of predigital structures extends to the imitation of the printed library catalogues traditionally used to retrieve physical books, leading to a problematic structural distinction between “metadata” and “data.”

An example of the use of a predigital structural concept in the digital representation of texts, even of ancient texts that are objects of specialized study, is the concept of a text as a linear sequence of characters arranged to form lines, which are themselves arranged in sequences to form pages, which in turn are arranged in sequences to form books. When such sequences are the primary means of encoding a text’s internal structure, the result is a “flat” representation that fails to capture adequately in digital form the many different entities and relationships apparent to any expert reader of the printed text.

However, if we break free of this sequence-dependent “Gutenberg ontology,” we can use digital data structures to distinguish the epigraphic structure of a text from its discourse structure while capturing

⁵ See “The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?” by Marc D. Hauser, Noam Chomsky, and W. Tecumseh Fitch, in *Science* 298 (2002) 1569–1579. See also “Evidence and Argumentation: A Reply to Everett (2009)” by Andrew Nevins, David Pesetsky, and Cilene Rodrigues in *Language* 85/3 (2009) 671–681.

the complex interplay between these two aspects of a written work. Moreover, we can very efficiently represent scholarly disagreements about how to read a particular text in a way that allows conflicting interpretations to be readily disseminated and debated.

Epigraphic structure is a function of the medium of the work whereas discourse structure is a function of the meaning of the work. They are related but logically separate. But to separate them requires breaking free of the limiting paradigm in which computer software merely mimics a predigital medium—namely, the printed book—rather than reaching its full potential to represent the phenomena being described in a more abstract and complex manner.

Indeed, a major intellectual benefit of digitization is that it allows scholars to make explicit, and thus amenable to automated comparison and analysis, the many conceptual distinctions they employ in the course of their research. Such distinctions are necessarily made in accordance with a historically situated and culture-bound perceptual and interpretive grid. Digitization exposes the interpretive habits and choices inherent in scholarly discourse in a way that suits the scientific ideal of transparency of method and repeatability of results.

But this salutary effect of digitization is nullified when the interpretive choices embedded in digital data structures and the algorithms that manipulate them are hidden from scholars or are imposed on them in the form of alien conventions of representation derived from other domains (see the section below on “Ontological Heterogeneity and Semantic Authority”). Likewise, the benefits of digitization are diminished when the data structures created by a computer programmer are hobbled by a simple Gutenberg ontology incapable of representing the conceptual distinctions that scholars routinely make.

With this in mind, OCHRE’s item-based design breaks free of the rigidity of text-representation schemas that treat as primary the concepts of “line” (character sequence), “page” (line sequence), and “book” (page sequence), just as it breaks free of the class-based concept of a rigid “table” of data. These are not suitable concepts for the fundamental digital representation of the entities and relationships encountered in scholarly research. They rely too much on relative

position in one or two dimensions and thus make it difficult to represent multiple, overlapping interpretations of an object of study. Linear sequences of characters and tabular arrays of rows and columns are useful for displaying information to human readers who have been conditioned by the flat printed pages of the Gutenberg era, but they are not suitable structures for a general-purpose computer database from which information must be retrieved and displayed in many different forms.

By a similar logic, OCHRE makes no structural distinction between data and metadata. The concept of metadata (i.e., “data about data”) emerged from libraries, which used card catalogues and printed indexes to keep track of physical books. Before digitization these catalogues were necessarily kept separately from the information to which they referred. Thus a few selected properties of the indexed information sources—such as author, title, publisher, publication date, and so on—were chosen as keys by which scholars could locate information of interest to them.

But in a database system like OCHRE, all of the properties of an item, including so-called metadata, are stored in the same way. There is no need to make an internal structural distinction between data and metadata even if such a distinction is made by a human user. Metadata is like any other kind of information that can be extracted from the database and displayed in a particular form in order to serve a particular purpose.

Having said that, OCHRE can easily extract and display selected properties of database items to serve as metadata about the entities those items represent. For example, some of the properties recorded for items that represent external resources (e.g., images, scanned documents, video clips, etc.) or that represent physical objects (e.g., artifacts in a museum) will correspond to standardized metadata attributes, such as “creator” and “creation date.” These properties can easily be embedded within a stand-alone OCHRE archive in a format that obeys metadata standards such as the Dublin Core, allowing software based on those standards to search for information in the archive (see Chapter 18).

Ontology Alignment and Data Integration

OCHRE's highly atomized design enables it to integrate data from many different research projects while preserving the terminology and conceptual distinctions of each project. OCHRE does not impose a common nomenclature or recording system. Instead, it provides a common integrative structure at the level of basic spatial, temporal, and logical relationships that apply to all projects. Integrating data of diverse origins entails the alignment of heterogeneous ontologies, which is accomplished in OCHRE by subsuming project-specific ontologies within a more general “upper” ontology.

This is made possible by OCHRE's item-based organization of information. It is difficult to integrate data from heterogeneous class-based databases because the number and types of the classes (usually represented as tables) vary from one database to the next. Even when there are similar classes in two different databases, the predefined properties (table columns) of the classes do not match. Integrating such databases requires special-purpose software that is laboriously programmed to take account of the idiosyncrasies of each database.

In OCHRE this problem is avoided because class-based tabular structures—and sequential textual structures, for that matter—are decomposed into smaller entities with their own properties. Classes comprised of entities imported from heterogeneous datasets may then be constructed, as needed, by comparing the individual properties of the entities, without requiring that each entity be preassigned to a group that is assumed to have the same set of properties. And because OCHRE represents spatial, temporal, linguistic, and logical relationships as flexible hierarchies of entities, the hierarchies of one project can easily be merged with the hierarchies of another project, resulting in a larger database that contains the data from both projects while remaining coherently organized in a predictable fashion.

Integration without Forced Standardization

Many previous attempts at data integration in cultural disciplines have involved the imposition of a standardized ontology intended to

be adopted by all researchers. For archaeological data, in particular, national governments often try to impose standards in the name of efficiency. But this usually proves to be unworkable, not just because researchers are selfishly wedded to their own familiar taxonomies, but because their divergent terminologies reflect disparate research traditions and interpretive perspectives—not to mention the different languages spoken by researchers from different countries.

For this reason, OCHRE does not impose a standardized terminology or recording system. Instead, it provides a generic framework within which each project's terminology and conceptual distinctions can be expressed. Property names (attributes) are not built into the structure of the database as table column headings, as they are in class-based data-retrieval systems. In OCHRE, property names are themselves treated as project-defined data within a more abstract database structure in which taxonomic descriptors are manipulated by end-users as entities in their own right. Each project can define its own descriptive property names and values and can organize them into a meaningful hierarchy, yielding a project-specific taxonomy.

Furthermore, OCHRE users can easily specify the semantic relationships between the property names and qualitative property values that are found in different taxonomies—that is, they can specify thesaurus relationships indicating synonyms, broader terms, narrower terms, and related terms. These user-created thesaurus relationships can then be saved and used in database queries to find similar items that have been described by different projects using different terms.

But, just as there is no standardized taxonomy in OCHRE that applies to all projects, there is no universal, standardized thesaurus. A thesaurus is itself a work of scholarship to be credited to a particular researcher and shared with other researchers, who can decide whose thesaurus to use when querying the database or whether to devise their own. This reflects the fact that meaning depends on context. The meanings of terms and their relationships to other terms cannot be formalized once and for all and then applied in an anonymous, automated fashion without regard to the specific intellectual context

in which the terms are being used, especially in disciplines that are characterized by a wide range of scholarly traditions and interpretive perspectives.

Ontological Heterogeneity and Semantic Authority

As a matter of principle, then, OCHRE does not attempt to replace the scholarly choices of researchers with a built-in taxonomy and thesaurus. To do so would be to deprive scholars of the ability and the responsibility to describe and interpret the data in their own way, and in a way that is exposed to scrutiny by other scholars so that interpretations can be contested and discussed. OCHRE's design is based on the belief that many of the semantic distinctions important to researchers cannot be adequately formalized in advance and therefore should not be built into the internal structure of a database system but should be a matter of ongoing scholarly debate.

OCHRE is designed to facilitate this work of interpretation, not to replace it. It does so by making it easy to construct and to share both taxonomies and taxonomy-integrating thesauruses. These are not assumed to be universal. They are human works which are created in particular research settings and reflect the interpretive perspectives of embodied human agents, who themselves are rooted in particular historically and linguistically mediated traditions.⁶

⁶ For a philosophical discussion of the dependence of meaning on context in relation to the limitations of artificial intelligence, see *What Computers Still Can't Do: A Critique of Artificial Reason* by Hubert L. Dreyfus (Cambridge, Mass.: MIT Press, 1992). Following Heidegger, Dreyfus argues that our understanding of the world emerges from the background of our historically situated human embodiment. This background can never be fully articulated; thus our understanding of the world (including our scholarly judgments, in this case) cannot be adequately emulated in formal symbols and algorithms of the kind used by digital computers. Steven Horst, writing on "The Computational Theory of Mind" in the online *Stanford Encyclopedia of Philosophy* (2009), summarizes his argument as follows: "Dreyfus argued that most human knowledge and competence—particularly *expert* knowledge—cannot in fact be reduced to an algorithmic procedure, and hence is not computable in the relevant technical sense. Drawing upon insights from Heidegger and existential phenomenology, Dreyfus pointed to a principled difference between the kind of cognition one might employ when learning a skill

Moreover, a thesaurus that connects taxonomic terms is just as much a work of scholarship as the taxonomy itself or descriptions made using the taxonomy. Researchers must decide whose thesaurus to employ when they use OCHRE to retrieve data from more than one project, or whether they should create their own thesaurus relationships. The responsibility for this is in the hands of the person asking the question, not the person who recorded the data and certainly not the person who wrote the software. This is because, as we have said, semantic relationships among different taxonomies cannot be established once and for all and employed without regard to the current intellectual context.

This does not rule out semi-automated ontology alignment—that is, the use of language-processing algorithms that propose thesaurus relationships to a human expert, who then makes the final decisions about how to relate one taxonomy to another. OCHRE will be continuously enhanced to incorporate the latest algorithms developed for this purpose in order to speed up thesaurus construction. But, even so, the semantics of an OCHRE database are determined by its users rather than by an anonymous semantic authority. OCHRE users can easily scrutinize the interpretive choices of the named scholars who created the taxonomies they use when entering data and the thesaurus relationships they invoke when performing a query. They are not at the mercy of a predefined taxonomy, nor are they dependent on an automated search engine or other software whose algorithm for matching similar terms is hidden from them, usurping their semantic authority in a different way.⁷

and the kind employed by the expert. . . . more often than not, argues Dreyfus, it is not possible to capture expert knowledge in an algorithm, particularly where it draws upon general background knowledge outside the problem domain.”

⁷ Concerning semi-automated ontology alignment, see “Why Your Data Won’t Mix” in *Queue* 3/8 (2005) 50–58 by Alon Halevy, a computer scientist who is an expert on data integration. He states (pp. 54–55): “Some argue that the way to resolve semantic heterogeneity is through standard schemas. Experience has shown, however, that standards have limited success and only in domains where the incentives to agree on standards are very strong. . . . Resolving schema heterogeneity is inherently a heuristic, human-assisted process. Unless there are very

In its approach to data integration and to the representation of information in general, OCHRE conforms to long-established practices that characterize critical scholarship. In this regard, it is different from database systems that conform to practices of institutional control and standardization which have emerged in commercial and governmental contexts—in which, of course, the vast majority of existing database systems have been developed, with the result that most software is designed to meet the needs of bureaucratic organizations in which semantic authority is imposed from the top down.

In contrast, most cultural and historical researchers today would agree that how one describes an object of study and how one determines whether it is similar to something else reflect a historically situated understanding of the world. There is no absolute and universal way of describing the cultural and social world, and many would argue that the same is true for the natural world. For this reason, scholarly practices have been developed to identify who said what, and when and where they said it, in order to encourage individual interpretive autonomy and to discourage anonymous and unquestioned semantic authority. This is done by crediting descriptions and interpretations to particular named scholars and by providing mechanisms for widely disseminating scholarly work so it can be used, criticized, and perhaps replaced by other scholars.

Accordingly, unlike many database systems whose design has been borrowed from nonacademic domains, OCHRE does not try to change these longstanding scholarly practices but rather to facilitate them. It does so by means of a database which is sufficiently well structured that it can be efficiently searched but which does not impose a predefined taxonomy or hide from view the logic by which information is retrieved. Instead, each OCHRE project team can construct their own taxonomy (or can borrow a taxonomy created by

strong constraints on how the two schemas you are reconciling are different from each other, one should not hope for a completely automated solution. The goal is to reduce the time it takes human experts to create a mapping between a pair of schemas, and enable them to focus on the hardest and most ambiguous parts of the mapping.”

another project) and each OCHRE end-user can decide whose thesaurus to employ when matching terms in one project's taxonomy with those in another. Interpretive choices are exposed to view, both in the description of particular items and in the retrieval of what are deemed to be similar items. Scholars who use OCHRE are not forced to conform to the "computer" but are themselves responsible for the semantics of the database, which is not an interpretive agent in its own right but merely a tool to facilitate the human work of interpretation.

Of course, even OCHRE's highly abstract ontology is not timeless or absolute. Embedded in it are historically situated assumptions about how to organize knowledge. But we have found that this ontology, which relies on very general concepts such as space, time, and agency, can represent a very wide range of scholarly descriptions and interpretations.

To put it another way: ontological heterogeneity is not a vice to be suppressed but a defining virtue of critical scholarship. Computerization of data should not be used as an excuse to suppress this heterogeneity, imposing a false consensus by institutional fiat. Instead, database software should make it easy for scholars to disagree with others and defend their own interpretations—or to reach agreement, as the case may be—thereby facilitating a productive "conflict of interpretations."

Interoperability with TEI, CIDOC CRM, and Other Standards

Although OCHRE can cope very well with ontological heterogeneity and does not rely on standards, it is able to import and export data files structured in accordance with the standardized ontologies that have been developed in recent years to facilitate the exchange of cultural and textual data. This is not an endorsement of these ontologies, which have various limitations, but rather a way to exploit the fact that people have taken the trouble to store their data in a standardized form. The ontologies in question are less generic than OCHRE's abstract ontology, so all the concepts and relationships they specify can be imported without loss of information.

TEI: A standardized ontology for representing written texts has been implemented in an XML tagging scheme by the Text Encoding Initiative (TEI), “a consortium which collectively develops and maintains a standard for the representation of texts in digital form” (see <http://www.tei-c.org>). The TEI schema is widely used in academic circles and has many useful features. In the future, OCHRE will have the ability to import digitized texts that conform to this schema, mapping the TEI’s entities and relationships onto its own database entities and relationships.

However, in contrast to OCHRE’s mode of representing texts, the TEI schema is not well suited for representing ancient texts written in complex logosyllabic scripts that have many allographic variants (e.g., Mesopotamian cuneiform and Egyptian hieroglyphs). It cannot easily represent as distinct yet interconnected structures the overlapping epigraphic and discourse hierarchies that constitute a text, or, more precisely, a particular reading of a text (see the section entitled “Representing a Text as Physical Signs and as Meaningful Discourse” above in Chapter 11). This is especially problematic in cases where a single text is subjected to multiple epigraphic analyses and discourse analyses by different scholars whose competing interpretations need to be easily recorded and displayed.

For this reason, OCHRE will use the TEI schema as a convenient exchange format for importing and exporting digitized texts but will not use it in its underlying representation of texts. The TEI schema is too “flat” for this purpose because it follows what we have called the “Gutenberg ontology” (see the section above entitled “Beyond Gutenberg: Escaping the Digital Library Paradigm”). In other words, TEI takes a document-based approach to software development, imitating the structure of printed books, rather than a database approach in which all of the conceptual distinctions and relationships apparent to an expert human reader are represented explicitly within digital data structures so they can be readily shared and analyzed. Digitizing a text in accordance with the TEI schema is a matter of annotating character strings with embedded tags, leaving it to human readers—or to parsing algorithms that attempt to simulate human

expertise—to make explicit the implicit relationships indicated by the linear sequence of the characters.

It is true that the TEI schema is less “flat” than some other digital text-representations because it does explicitly represent a text’s hierarchical structure, up to a point. The linear sequence of characters and embedded tags in a TEI text-representation reflects an abstract model of the text as an “ordered hierarchy of content objects,”⁸ which is certainly useful, as far as it goes. However, the TEI schema cannot easily represent multiple overlapping hierarchies that all model the same text. In practice, TEI text-encodings conflate a text’s epigraphic structure with its linguistic discourse structure instead of distinguishing them via different hierarchical models. But many texts, especially ambiguous or physically damaged texts, are susceptible to multiple readings on both the epigraphic and discourse levels. Each reading results in a different hierarchy that will, as a rule, share many elements with other hierarchical models of the same text. A fully adequate schema for the digital representation of texts must be able to integrate multiple overlapping hierarchies of textual “content objects.”

This is in keeping with the notion that any digital representation or transcription of a text that goes beyond a visual facsimile entails a context-dependent interpretation of what the marks on an inscribed surface might signify. There is no single, absolute representation of a text, and thus of its hierarchical structure, but only different readings, which correspond to different hierarchical models of the text’s components.

The database-oriented alternative to the document-based software paradigm is to decompose a text into a highly atomized collection of independently addressable epigraphic and linguistic entities whose interrelationships are explicitly represented by means of multiple hierarchies and cross-hierarchy links. Flat, human-readable character strings can then be generated as needed, in whatever configuration is

⁸ See the chapter on “Text Encoding” by Allen H. Renear (esp. pp. 224–225) in *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens and John Unsworth (Malden, Mass.: Blackwell, 2004).

desired, from a more flexible underlying representation. This is the approach taken in OCHRE.

CIDOC-CRM: A quite different standardized ontology for cultural research has been created by the International Committee for Documentation (Comité international pour la documentation, or CIDOC) of the International Council of Museums. This ontology is called the “Conceptual Reference Model” (CRM).

Unlike the TEI tagging scheme, the CIDOC CRM is not a working schema but simply a formal specification of concepts and relationships. Moreover, it is not designed to represent the content of texts but is focused on representing information about cultural artifacts, especially information of interest to museum curators and archivists. The CIDOC ontology is quite complex, consisting of hundreds of concepts and relationships. Unfortunately, there is no generally available data-retrieval schema in which it is implemented. When such a schema becomes available, the ability to import and export data files that conform to this schema will be added to OCHRE. This is feasible because the many different concepts and relationships specified in the CIDOC ontology can be easily mapped onto the more abstract OCHRE ontology.⁹

In the meantime, researchers can implement the CIDOC ontology within their own OCHRE projects, if they wish to do so.

⁹ The CIDOC CRM ontology is described online at <http://www.cidoc-crm.org>, from which the following definition is taken: “The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the ‘semantic glue’ needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives.” See also Martin Doerr’s article “The CIDOC Conceptual Reference Module [sic]: An Ontological Approach to Semantic Interoperability of Metadata,” *AI Magazine* 24/3 (2003) 75–92.

CIDOC concepts and relationships can be used as the basis for project-defined items and item hierarchies, for taxonomies of item properties (including relational properties), and for item-to-item links. In general, any cultural or scientific ontology or taxonomy can be represented in OCHRE, regardless of whether it has been made explicit in a formal ontology like the CIDOC Conceptual Reference Model or is implicit in a functioning database.

A Data Warehouse versus On-the-fly Mediation

There are two main approaches to integrating heterogeneous data sources. The data can be left in its original form and combined “on-the-fly” when a user submits a query, or it can be combined ahead of time in a “data warehouse.” A warehouse is a single comprehensive database structured in accordance with a global schema, within which a variety of other databases that conform to disparate local schemas can be merged.¹⁰

OCHRE uses the warehouse approach. An OCHRE database serves as a repository in which heterogeneously structured data from many different sources is integrated according to a global schema that reflects a single, highly generic ontology. This allows large amounts of scholarly data to be manipulated and searched efficiently.

Merging heterogeneous data within a single, well-indexed database structure often permits faster and more sophisticated querying of

¹⁰ A third approach involves “federating” independent databases by implementing a connection between each pair of databases whose data is to be combined. See pp. 1041–1049 in *Database Systems: The Complete Book* by Hector Garcia-Molina, Jeffrey D. Ullman and Jennifer Widom (2d ed.; Upper Saddle River, N.J.: Pearson Prentice Hall, 2009). These authors note (p. 1042) that: “The problem with this architecture [i.e., federating databases] is that if n databases each need to talk to $n - 1$ other databases, then we must write $n(n - 1)$ pieces of code to support queries between systems.” In contrast, they explain (p. 1043) that “in the *data warehouse* integration architecture, data from several sources is extracted and combined into a *global schema*. The data is then stored at the warehouse, which looks to the user like an ordinary database.” On-the-fly mediation is similar in that it creates “a *virtual database*, which the user may query as if it were *materialized* (physically constructed, like a warehouse)” (p. 1042).

a kind that is conducive to comprehensive analysis and interpretation. For this reason, data warehouses are widely used in large commercial and governmental organizations. In the scholarly world, a data warehouse makes even more sense because scholars have less need than do business users for instant access to rapidly changing data sources, which is what usually motivates the adoption of a data-integration strategy that uses on-the-fly mediation of local databases rather than a global data warehouse. Furthermore, many scholars lack the technical and financial resources required to keep their data accessible on the Internet and thus available for on-the-fly mediation, so some kind of separate online repository is desirable. Finally, the very process of importing one's data into an OCHRE database helps to ensure its correctness because the decomposition and normalization that this entails exposes inconsistencies and errors.

Among scholarly research tools that are comparable to OCHRE, the Digital Archaeological Record (tDAR) is an example of an on-the-fly mediation system as opposed to a data warehouse (see <http://www.tdar.org>). Although tDAR is also intended to serve as a repository for archaeological data from many different projects, the heterogeneous tables, images, maps, and documents uploaded by contributing researchers are maintained in their original form (i.e., according to their original schemas) rather than being decomposed and merged into a single database. Data files are uploaded to tDAR to ensure that they will be preserved and will remain accessible on the Internet, and thus amenable to on-the-fly mediation. They are not uploaded in order to merge them within a data warehouse. Data integration is accomplished instead by extracting information from the original files as needed in response to a user's query. In our view, however, a central data warehouse based on OCHRE's global schema allows for more powerful and reliable querying of data from multiple projects while preserving all of the terms and distinctions contained in the original data files.

Chapter 20

Implementing the OCHRE Ontology in XML

OCHRE's generic ontology is designed to be implemented in a high-performance database system that can accommodate a large number of different projects with many simultaneous users. The majority of such systems today are based on the relational data model. They make use of data tables ("relations") which are searched and manipulated by means of the SQL querying language.¹ However, OCHRE has been implemented in a database system based on the semistructured data model instead of the more familiar relational model. A relational implementation would certainly be possible, but a semistructured database is better suited for storing and querying flexible hierarchies of data objects of the kind used in OCHRE.²

The Extensible Markup Language (XML) and a related query language called XQuery have emerged in the last decade as well-supported and nonproprietary technical standards that provide an excellent vehicle for representing and manipulating hierarchically organized semistructured data. The World Wide Web Consortium that is responsible for HTML is also responsible for XML and the XML Schema and XQuery standards which underlie semistructured database systems (see <http://www.w3.org/xml>).³ In recent years, leading vendors of database software such as IBM and Oracle have added to their software the ability to implement robust "native XML"

¹ On the relational data model and SQL, see pp. 17–65 in *Database Systems: The Complete Book* by Hector Garcia-Molina, Jeffrey D. Ullman and Jennifer Widom (2d ed.; Upper Saddle River, N.J.: Pearson Prentice Hall, 2009).

² For a technical definition of the semistructured data model and how it differs from the relational model, see *Data on the Web: From Relations to Semistructured Data and XML* by Serge Abiteboul, Peter Buneman and Dan Suciu (San Francisco: Morgan Kaufmann, 2000). See also pp. 483–515 in *Database Systems: The Complete Book* by Hector Garcia-Molina, Jeffrey D. Ullman and Jennifer Widom (2d ed.; Upper Saddle River, N.J.: Pearson Prentice Hall, 2009).

³ A good introduction to XQuery is *XQuery: The XML Query Language* by Michael Brundage (Boston: Addison-Wesley, 2004).

databases in which very large collections of XML data objects (“documents”) are stored, indexed, and searched efficiently without reference to relational tables.⁴

The Semistructured Data Model and XML Database Systems

The semistructured data model occupies the middle ground between the strongly structured relational model and the weakly structured graph (or network) data model that characterizes the World Wide Web. Although it is not usually described this way, the Web as a whole can be regarded as a very large database with a very simple structure. It is distributed across the Internet and is not under the control of any one person or organization, but it nonetheless functions as a single data-retrieval system. In popular parlance, the Web is often confused with the Internet, but the Web is merely one way of using the Internet to retrieve information.

The Web’s structure is very simple and flexible in comparison to the structure of a relational database, which consists of data tables that are relatively rigid and complex. The Web is defined solely in terms of generic units of information (“pages”) that are linked to other units of information. The internal semantic structure within these units of information is not specified. Data sources with specified semantic structures are of course accessible on the Internet—in many cases via links from Web pages—but they have no universal predictable structure and thus they require the use of special-purpose

⁴ XML data objects are called “documents,” not because they necessarily resemble ordinary documents or are intended to be read by people, but because they can be easily displayed as human-readable text. They can be displayed as human-readable documents because they are encoded at the level of binary digits using the Unicode standard for representing text characters, of which the older ASCII standard is a subset. More specifically, XML documents employ the UTF-8 or UTF-16 encoding forms (see <http://www.unicode.org>). All modern computers support the non-proprietary Unicode character-encoding standard; thus XML documents can be readily transmitted from one computer to another, regardless of the computer’s operating system. For this reason, XML has become an important vehicle for transmitting structured information over the Internet. However, XML is equally important as a powerful notation for expressing the semistructured data model.

software in each case to interpret them. The Web per se, regarded as a predictable data-retrieval system, is very simple, consisting of inter-linked HTML files and other data objects whose type and location can be described in HTML. The Web's simple HTML-defined data schema was originally designed for human browsing and reading of information. Thus it specifies how information should be displayed but does not specify its semantic structure except to say that certain units of information are linked to other units in some fashion.

Relational databases are based on mathematical set theory and consist of sets of “tuples” that constitute “relations.” A relation is usually depicted as a two-dimensional table whose rows are its tuples. The Web, in contrast, is an example of a graph database. Graph databases are based on mathematical graph theory and consist of networks of “vertices” (also called nodes) connected by “edges” (also called arcs). The enormous success of the Web as a means of storing and retrieving information has attracted renewed attention to the graph data model, which had been neglected for many years.

The OCHRE ontology is closely aligned with the semistructured data model. As a result, it can be more efficiently implemented in an XML database using XQuery than in a relational database or a graph database. It could certainly be implemented in such databases; in this case, however, the semistructured data model has notable advantages over both the strongly structured relational data model and the weakly structured graph data model used by the Web.⁵

Advantages over the Strongly Structured Relational Data Model

In a database that conforms to the relational data model, the structure of the database (its logical schema) is defined by a set of two-dimensional tables, each of which represents a mathematical relation between a set of entities and their attributes. More precisely, a relation is a set of tuples that have the same attributes. A tuple is an

⁵ XML databases are technically also graph databases that depend on graph theory rather than the set theory that underlies relational databases, but they are optimized to work with a particular kind of graph structure consisting of vertices and edges that represent hierarchies of data elements and attributes.

ordered list of elements, in this case, a list of the values of a set of attributes that pertain to an entity of interest. A tuple is depicted as a row in a table whose columns correspond to attributes. The fundamental logical unit in a relational database is the table row (tuple).

A relational table's rows usually correspond to a set of entities of interest and its columns correspond to attributes used to describe those entities. Each cell of a table, at the intersection of a row and a column, contains the value of an attribute for a particular entity, as in this schematic example:

ENTITY ID NO.	TYPE	COLOR	LENGTH
1	Type C	red	6 cm
2	Type A	green	9 cm
3	Type A	blue	2 cm
4	Type D	red	5 cm

The relational model is very effective if the information to be represented has a predictable structure, as is often the case in commercial contexts, in which entities of interest (e.g., employees, customers, products, etc.) are described in a standardized way. The power of the relational model is particularly apparent when data tables are decomposed and interconnected via overlapping sets of attributes (“keys”) in order to create a normalized database in which data can be recombined and presented in different ways and in which there is no redundant duplication of information.

However, this mode of organizing data is not a good way to represent complex hierarchies and loosely structured textual information. In principle, such information can be represented in a relational database, but not easily or efficiently. Scholarly research in cultural, social, and historical disciplines is characterized by spatial, temporal, linguistic, and logical hierarchies of various kinds. Scholars in these disciplines need a data model that suits their research needs instead of being forced to squeeze their data into an inappropriate tabular mold. The semistructured data model, for which XML provides a formal notation, is just such a model.

XML is particularly good at representing the semistructured data found in written texts and other kinds of hierarchies, but it can also represent highly structured data tables. XML therefore plays an important role in information integration. It defines a common data format that can represent a wide range of heterogeneous information that has been organized in different ways, ranging from highly structured tables to less structured textual documents.

The fundamental logical unit in an XML database is an XML document, just as the fundamental logical unit in a relational database is a table row (tuple). An OCHRE database consists of a large number of small XML documents that represent information in a highly atomized fashion. The OCHRE ontology described in the preceding chapters is implemented by means of XML documents that each contain hierarchies of information, are themselves organized into hierarchies of documents, and also contain links that cut across the intra-document and inter-document hierarchies to connect units of information in nonhierarchical ways.

Each document in an OCHRE database conforms to one or another of the eighteen different XML document types that collectively implement the OCHRE ontology (see the section below entitled “XML Document Schemas”). The structure of each type of document is prescribed in accordance with the XML Schema standard published by the World Wide Web Consortium. The eighteen document schemas together constitute the full logical schema for a semistructured database that can be implemented in any database management system that supports the XML Schema and XQuery standards (e.g., Oracle and IBM’s DB2, among others).

Semistructured XML databases can index and retrieve hierarchically organized data more efficiently than comparable relational databases. Keep in mind that although XML is a text-based notation, XML documents are not stored internally as verbose strings of text characters within an XML database. XML documents are logical data objects that are converted internally to an underlying database format that permits them to be indexed and retrieved efficiently. This conversion is invisible to the application programmer, who deals only

with the logical XML structure, just as relational tables are themselves logical data objects that are converted to lower-level data structures that the application programmer does not see.

Advantages over the Weakly Structured Web Data Model

The simplicity of the Web data model has made it easy for people with limited technical expertise to share information on the Internet. Scholars in cultural and historical disciplines have benefited because the Web has encouraged the digitization and dissemination of vast quantities of primary research material—literary texts, historical archives, maps, photographs, audio and video clips, architectural and artifactual descriptions, etc.—as well as secondary literature in the form of monographs, dictionaries, and journal articles.

But the very simplicity of the Web means that it lacks a predefined structure that would enable scholars to disseminate their detailed analyses of information in a way that is conducive to automated data retrieval and analysis, as opposed to human browsing and reading. The complex conceptual distinctions that experts make as they study and discuss their sources are not represented in a predictable digital form on the Web, making it difficult to find and compare specific scholarly observations and interpretations in an efficient and comprehensive manner.

Querying by means of a Web search engine such as Google is not an adequate solution to this problem. A search engine is a software tool that uses algorithms of various kinds to match the character strings (e.g., keywords) entered by a user against the documents accessible on the Web. A search engine returns a list of matches ranked in order of their presumed relevance to what the user wants to find. This is very helpful as a rough-and-ready way to locate information but, as every Web user knows, search engines are limited by their inability to understand human language in the way a human being does. They often return irrelevant information or fail to find relevant information.

The semantic vacuity of the Web, which has fostered this brute-force method of searching by matching character strings, prompted

the development of the Extensible Markup Language (XML) in the late 1990s. XML was developed to provide a standardized format for representing complex semantic structures on the Web. However, the semistructured data model that XML expresses is quite different from the loose graph model that underlies the Web and has proved to be quite powerful in its own right. Thus XML has given rise to full-blown databases that are structurally different, not just from relational databases, but from the Web itself.

To clarify this point, it should first be said that XML specifies a plain-text notation for linearizing hierarchically structured data. An XML document represents a semantic hierarchy of information by means of character strings annotated with embedded “tags” that are set apart by angle brackets. Data stored in this linearized textual form can be easily transmitted to different kinds of computing devices regardless of which operating system they use. Moreover, information of all kinds, whether or not it is originally hierarchical, can be represented by means of XML hierarchies.

As tagged-text documents, XML files are similar to the HTML files on which the Web is based.⁶ But HTML provides a notation only for indicating how data should be displayed, not for indicating the semantic structure of the data itself. Tags marked off by angle brackets are inserted around character strings to indicate whether they should be displayed as headings, paragraphs, lists, tables, and so on, and to indicate which fonts, styles, and colors should be used when formatting the character strings. Tags can also be used to indicate where and how an image, video clip, or other nontextual information should be displayed. Very importantly, HTML tags can indicate “hyperlinks” that connect locations in the current document to other

⁶ HTML is an acronym for HyperText Markup Language. It was introduced by Tim Berners-Lee in 1990, giving birth to the World Wide Web. HTML was based on an earlier standard called SGML (Standard Generalized Markup Language), which in turn was derived from the Generalized Markup Language developed by IBM in the 1960s to facilitate the sharing of machine-readable documents such as instruction manuals. See *HTML and XHTML: The Definitive Guide* by Chuck Musciano and Bill Kennedy (6th ed.; Sebastopol, Calif.: O'Reilly Media, 2007).

locations in the same document or in documents stored elsewhere on the Internet. Web browser software allows users to follow such links from one document to another, displaying information that has been assembled automatically from a variety of sources.

XML was developed because HTML tags are limited to indicating how data should be displayed rather than indicating the meaning and structure of the data. XML documents bear a superficial resemblance to HTML documents because they, too, consist of plain-text character strings with embedded tags.⁷ But XML is “extensible” in a way that HTML is not. HTML has a limited set of predefined tags that specify how information should be presented to end-users in a Web browser and how it should be linked to other information. XML, on the other hand, provides a mechanism for defining new tags to describe the structure and meaning of digital data without regard to how it is presented. Thus, despite its name, the Extensible Markup Language is not itself a markup language (i.e., a particular tagging scheme) but rather a way to define new markup languages.

The XML tagged-text format is very powerful because it can be used as a general-purpose mechanism for describing the structure of any kind of data, ranging from relatively unstructured textual documents intended to be read by human beings to highly structured databases. For this reason, XML and related standards such as XQuery are now used, not just to exchange data on the Web, but to store and retrieve data of all kinds within general-purpose database systems. XML documents can be easily transmitted across the Internet, just like HTML documents, but they are not limited to the

⁷ “Plain-text” here simply means that XML files are encoded at the level of binary digits using the Unicode Transformation Format (UTF) character-encoding scheme, whose UTF-8 version subsumes the older ASCII standard developed in the 1960s, which was based on even older telegraphic codes. UTF-8 is an encoding standard promulgated by the Unicode Consortium. It allows for thousands of characters beyond the original 128 ASCII characters, including not just Latin-based alphabetic characters but characters in non-Latin and nonalphabetic writing systems such as the Chinese, Japanese, and Korean scripts. It does so by encoding each of the additional characters with 16, 24, or 32 bits (2, 3, or 4 bytes) instead of the 8 bits (1 byte) sufficient for the original ASCII characters.

Web. HTML documents are designed to be displayed on the receiving end by Web browser software but there is no need for an XML-based system to be restricted by the limited set of features a Web browser can provide. For this reason, the OCHRE client application described in this book bypasses the user's Web browser and sends XML documents (including XQuery queries, which are themselves XML documents) directly to and from a database server.

A semistructured XML database can therefore transcend the weakly structured Web data model, providing a predictable way to transmit complex semantic structures, including relational table structures, while still accommodating more loosely structured sources of information such as free-form texts. Indeed, many apparently unstructured texts contain inherent structural hierarchies that are best represented by means of XML (see the discussion of textual hierarchies above in Chapter 19, in connection with the Text Encoding Initiative's hierarchical schema for representing texts). The OCHRE ontology relies on ordered hierarchies to represent spatial, temporal, linguistic, and logical relationships in a concise yet flexible manner. Thus, this ontology can be implemented more efficiently in a semi-structured XML database than in a nonhierarchical web of data.

Having said this, there is nothing to prevent the OCHRE ontology that is described in this book from being expressed in the Web Ontology Language (OWL) or some other ontology format designed to facilitate the so-called Semantic Web, which aims to extend the Web by adding machine-readable semantic structure to it. The OCHRE ontology could be implemented in a Web graph database using RDF “triples,” OWL, and the SPARQL querying language.⁸

⁸ On the Web Ontology Language, see <http://www.w3.org/TR/owl2-overview>. RDF is an acronym for Resource Description Framework, which uses XML to represent information as subject-predicate-object “triples”; see <http://www.w3.org/rdf> and <http://www.w3.org/TR/rdf-schema>. SPARQL is a querying language designed to work with RDF; see <http://www.w3.org/TR/rdf-sparql-query>. Support for RDF, SPARQL, and OWL has now been incorporated into high-performance database software sold by leading vendors such as Oracle. For information about the Semantic Web in general, see <http://www.w3.org/standards/semanticweb>.

However, it is not clear what would be gained by doing this. It would amount to implementing a hierarchical design within a nonhierarchical graph data model. This is certainly possible; but in comparison to an XML database implementation using XQuery, which was designed to work with hierarchical data, an RDF graph database will be less efficient for data retrieval, more cumbersome to program, and more difficult to maintain.

Indeed, proponents of XML and XQuery have pointed out that an XML database can easily represent both relational tuples and graph data structures stored as subject-predicate-object triples (3-tuples). Likewise, relational SQL queries and SPARQL graph queries can be easily translated into XQuery, but not vice versa. Thus an XML database provides all the benefits of the relational and graph data models, wherever these are useful for modeling information, without losing the ability to do highly optimized querying of semi-structured hierarchical data.

XML Document Schemas

The OCHRE ontology has been implemented in a set of eighteen XML document schemas. These schemas do not define metadata tags to be used simply for annotating data, unlike XML schemas designed only to facilitate data exchange. Instead, the OCHRE schemas collectively constitute the logical schema of a fully fledged XML database, in the same way that a set of interconnected table schemas collectively constitute the logical schema of a relational database.

Previously, we called our eighteen XML document schemas the “Archaeological Markup Language,” abbreviated as ArchaeoML, reflecting our initial focus on archaeological data. However, the schemas and the generic ontology they express have proved to be applicable to a much wider range of research data, so we have dispensed with this name. Moreover, the designation “markup language” was confusing for people who assumed that our XML schemas merely define metadata tags for marking up existing documents and data tables rather than representing scholarly data of all kinds, including texts and tables, in a more fundamental way.

Ontological Concepts and XML Document Types

The OCHRE ontology includes the following fourteen basic concepts, each of which corresponds to an XML document type. Each of these concepts is derived from the more abstract concept of an “item” that has its own name, properties, notes, events, and links to other items.

- *Project item*: An OCHRE project that is the “owner” of a collection of database items of various kinds.
- *Spatial item*: A spatially situated unit of observation, however it be defined, on any scale (see Chapter 8).
- *Temporal item*: A period of time of any duration (see Chapter 8).
- *Agency item*: A person or organization—real or fictional, current or historical—to whom a reference is made or an action or role is attributed (see Chapter 8).
- *Resource item*: Information stored in a separate file or data source for which a URL link is stored in the database (see Chapter 9).
- *Bibliographic item*: Information about a published work, which may include a link to a digital version of the work, if available (see Chapter 10). Digital versions of published works can be represented by resource items.
- *Script item*: A character or sign in a writing system (see Chapter 11).
- *Epigraphic item*: A physical unit of inscription in any written medium, however it be defined (see Chapter 11).
- *Discourse item*: A linguistic unit of discourse, however it be defined (see Chapter 11).
- *Lexical item*: A dictionary or glossary entry (see Chapter 12).
- *Taxonomic variable item*: A qualitative or quantitative property used to describe an item (see Chapter 13). Special relational properties represent relationships between items.
- *Taxonomic value item*: A value of a qualitative property measured on a nominal or ordinal scale of measurement.

- *Taxonomic predefinition item*: An ordered list of variable-value pairs that can be used repeatedly to describe items (see the section on “Predefined Properties” in Chapter 5).
- *Concept item*: A concept that cannot be represented by any of the other item types.

In addition to these fourteen basic concepts, the following four concepts are used to organize instances of the basic ontological concepts into groups for various purposes. These four grouping concepts are also descended from the general “item” concept; thus instances of any of them can also have names, properties, and links of their own and can in turn be organized into groups. Each of them corresponds to an XML document type.

- *Hierarchy*: A hierarchy of items of the same type. A hierarchy can be a recursive “containment” hierarchy or a nonrecursive “subordination” hierarchy (see the section on “Hierarchies of Items” in Chapter 4).
- *Set*: An ordered or unordered list of items of one or more types (see the section on “Sets of Items” in Chapter 4). Set items can be used to store manually created lists of items of various kinds that represent entities of interest. Set items can also be used to store automatically generated result sets produced by database queries. Tabular views of information, spatial maps, and printed reports can be generated from sets.
- *Query*: A definition of the scope and criteria for retrieving items of interest by means of an automated database query (see Chapter 14).
- *Text*: A written text represented by a hierarchy of epigraphic items and a hierarchy of discourse items (see Chapter 11).

Each of the fourteen basic concepts and four grouping concepts corresponds to an XML document schema, for a total of eighteen document types. In addition, there are three ontological concepts that correspond to particular kinds of hierarchies defined in a way that constrains the composition of the hierarchy and gives it a particular

meaning. However, these additional concepts do not have their own document types but are all represented by the hierarchy concept:

- *Writing system*: A hierarchy of script items that represents a writing system (see Chapter 11).
- *Taxonomy*: A hierarchy of taxonomic variable and value items that represents the project’s taxonomy (see Chapter 13).
- *Presentation*: A hierarchy of items that prescribes a particular mode and sequence for presenting public information to anonymous users (see Chapter 16).

Normalized Database Form and Denormalized Archival Form

An OCHRE database stores information in a highly atomized fashion (see the section on “Atomization and Dynamic Recombination” above in Chapter 19). Individual items of information are stored in many small, independently addressable XML documents. These documents are interlinked in such a way that information can be flexibly searched and recombined as needed. Thus, the eighteen XML schemas that define the OCHRE document types constitute a “normalized” database design, if we take the liberty of borrowing this term from the realm of relational databases, where an analogous design strategy leads to the decomposition and interlinking of data tables in order to achieve the most efficient representation of information in a form that is suitable for many different purposes.

Project administrators can export a “denormalized archival form” of their data derived from the underlying “normalized database form” that is manipulated by the database client application described in this book. The denormalized archival form consists of character strings assembled from the many atomized units of information in the database, yielding a “flattened” but more readable version of the information that is suitable for display in a Web browser and for retrieval by search engines without requiring the use of the OCHRE database client application (see Chapter 18). Denormalized archives contain embedded HTML stylesheets that specify how the information should be displayed in a Web browser.

An OCHRE database's XML documents contain both the normalized database form of the data and the denormalized archival form. The denormalized character strings that comprise a human-readable version of information about a particular entity are stored within the particular document that uniquely identifies this entity as an atomized component of the normalized database.

Consider, for example, an XML document that represents an archaeological artifact. This document will contain links to other atomized documents that represent the artifact's spatial context, chronological period, intrinsic properties, and other information about it (e.g., photographs and bibliography). Denormalized character strings that provide a readable version of all the information about the artifact are automatically assembled by following these links and are stored within the XML document that represents the artifact itself.

In this way, a close connection is maintained within the database between the normalized database form of the data and its more readable, albeit highly redundant, denormalized archival form. In fact, a stand-alone archive exported from an OCHRE database simply consists of a plain-text file which contains all of the XML documents that belong to a particular project (or a selected subset of those documents). These exported documents will contain both the normalized database form of the data and, wherever necessary, the denormalized archival form suitable for displaying and searching the data in a Web browser. An exported OCHRE archive therefore has the added benefit of providing a full backup copy of the project's information that preserves both the normalized and denormalized versions of the data.

Bibliography

- Abiteboul, Serge, Peter Buneman, and Dan Suciu. 2000. *Data on the Web: From Relations to Semistructured Data and XML*. San Francisco: Morgan Kaufmann.
- Babbie, Earl. 2010. *The Practice of Social Research*. 12th ed. Belmont, Calif.: Wadsworth.
- Bailey, Kenneth D. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Thousand Oaks, Calif.: Sage Publications.
- Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, Mass.: MIT Press.
- Brundage, Michael. 2004. *XQuery: The XML Query Language*. Boston: Addison-Wesley.
- Burnard, Lou, Katherine O'Brien O'Keeffe, and John Unsworth, eds. 2006. *Electronic Textual Editing*. New York: Modern Language Association of America.
- Doerr, Martin. 2003. "The CIDOC Conceptual Reference Module [sic]: An Ontological Approach to Semantic Interoperability of Metadata." *AI Magazine* 24/3: 75–92.
- Dreyfus, Hubert L. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, Mass.: MIT Press.
- Garcia-Molina, Hector, Jeffrey D. Ullman, and Jennifer Widom. 2009. *Database Systems: The Complete Book*. 2d ed. Upper Saddle River, N.J.: Pearson Prentice Hall.
- Gruber, Thomas R. 1993. "A Translation Approach to Portable Ontology Specifications." *Knowledge Acquisition* 5: 199–220.
- Halevy, Alon. 2005. "Why Your Data Won't Mix." *Queue* 3/8: 50–58.
- Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" *Science* 298: 1569–1579.
- Ling, Tok Wang, Mong Li Lee, and Gillian Dobbie. 2005. *Semistructured Database Design*. New York: Springer.
- Lock, Gary. 2003. *Using Computers in Archaeology: Towards Virtual Pasts*. London: Routledge.

- McCarty, Willard. 2005. *Humanities Computing*. New York: Palgrave Macmillan.
- Musciano, Chuck, and Bill Kennedy. 2007. *HTML and XHTML: The Definitive Guide*. 6th ed. Sebastopol, Calif.: O'Reilly Media.
- Schloen, J. David. 2001. "Archaeological Data Models and Web Publication Using XML." *Computers and the Humanities* 35: 123–152.
- Schreibman, Susan, Ray Siemens, and John Unsworth, eds. 2004. *A Companion to Digital Humanities*. Malden, Mass.: Blackwell.
- Schreibman, Susan, and Ray Siemens, eds. 2007. *A Companion to Digital Literary Studies*. Malden, Mass.: Blackwell.
- Taylor, Arlene G., and Daniel N. Joudrey. 2009. *The Organization of Information*. 3d ed. Westport, Conn.: Libraries Unlimited.

Index

- aliases, 13, 28, 51, 55–56, 153–54, 165, 247, 254, 259–60, 327
allographs, 165, 222
ArchaeoML (Archaeological Markup Language), 372
atomization, 344–45, 375
attested form, 223
bibliographies, 9, 25–26, 35, 57, 58, 124, 133, 151–52, 159, 169, 183, 209, 221, 229, 235–36
chronological periods, 9, 19, 35, 58, 59, 92, 100, 105, 113–24, 266, 270, 347
citation URL, 45, 317–18
closing a project, 8, 42
concepts (as items), 9, 19–20, 35, 76, 79, 91–97, 282
data models, 364–72
data-entry mode, 42
dictionaries, 9, 26–27, 35, 133, 169, 191, 202, 209–43, 327–28
discourse units, 22–24, 162, 180–88
epigraphic units, 22–25, 162, 163–66, 172, 181, 188
events (that affect items), 19, 46, 97–105, 120, 125, 170, 189, 211, 237, 270, 275–76, 279, 281–82, 311
file formats, 130, 134, 333
AIFF, 136
AVI, 132, 136, 333
BMP, 134
DOC, 132–36, 201
GIF, 134
HTML, 132–36, 200–5, 369–70, 375
JPEG, 20, 123, 130, 134, 137, 304, 333
MP3, 136
MPEG-4, 136
PDF, 20–25, 67, 79, 103, 128–29, 132–36, 140, 145, 147, 151, 155, 158, 161, 175, 189, 303, 338
PNG, 134, 142
PPT, 132–36
PTM, 126, 138–39
RTF, 132, 135
TIFF, 20, 123, 130, 134, 137, 304, 333
TXT, 132, 135
WAV, 136
XLS and XLSX, 132–36
general links, 87
geospatial data, 137
grammatical form, 221
hierarchies (of items), 9, 16, 25, 27, 35, 58–59, 109–14, 121, 170, 203, 207, 210, 246, 251, 254, 266, 280, 297, 310, 341, 346–47, 374
containment hierarchies, 27, 210, 266
discourse hierarchies, 22–23, 170–72, 180, 183, 195, 199, 202, 204–5, 226

- epigraphic hierarchies*, 22–23, 170–76, 202–4, 345
subordination hierarchies, 18, 21–23, 27, 31, 36, 58, 61, 121, 124, 161, 210, 266, 310, 347
- hotspot links, 87, 134, 140, 149, 193
- hyperlinks, 46, 82, 87–88, 140, 182, 240
- image tools, 88–89, 126, 134, 140–49, 193
- importing data, 132–34, 163, 169, 171–75, 180, 199, 206, 327, 337–38
- Inboxes, 57–58, 63, 91, 154–55, 246, 283
- inventories, 101–2, 113
- item pane, 9–13, 19, 27, 31–32, 38, 42–59, 62–64, 69, 74, 76, 78, 80, 82, 83–85, 88–103, 108–13, 117–21, 124–40, 145–47, 152, 158–60, 165, 168, 173, 177, 180–82, 185, 188, 193, 196, 199, 201, 204, 206, 210–14, 221–23, 237–39, 245–47, 251, 254, 261, 264–68, 277, 283–91, 297, 302, 310, 311, 317–18, 320, 323–24, 326–28, 333
- JRE (Java Runtime Environment), 6–7
- lemma, 210–17, 227–40
- links subtab, 46, 88–92
- locations and objects, 9, 13–25, 34–35, 55–59, 62, 74–79, 89–92, 97, 100–1, 105–9, 112, 118, 121–24, 140, 145–47, 152, 194–97, 266, 282, 321, 334, 347
- locking items, 43, 203–5, 223, 246, 324
show all locks, 43
unlock, 43, 324
whose lock?, 43, 324
- logging in, 8, 42
- logging out, 8, 42
- multilingual features, 56, 78, 85, 127, 153, 157, 214–15, 232, 327
- named links, 87
- navigation pane, 9–13, 31–45, 49–64, 90–173, 180–88, 199, 204–12, 223, 237, 245–46, 254, 260–61, 265–67, 284–94, 297, 301, 304, 310–31
- observations, 76
- opening a project, 8, 41
- persons and organizations, 9, 13, 17–19, 35, 56–57, 88, 94, 97, 100–1, 105, 118–24, 154–60, 168, 229, 267, 282, 320–21
- presentations, 9, 35, 39, 42, 239, 307–10, 317, 328
- properties, 28, 45–50, 56, 69–95, 105–115, 120, 125–26, 140, 152, 165–67, 170, 173, 181–83, 194, 211, 219–23, 235–37, 240, 246, 250–55, 264, 270–73, 278–79, 294, 310–11, 373
- queries, 9, 36–37, 57, 111, 117, 198, 242, 265–66, 284–85, 287–90, 294, 317, 329, 343
- query results, 46–48, 90, 147, 280–93, 301–4

- reference pane, 9, 44–48, 52–53, 62, 65, 76–78, 82–84, 89–90, 94–95, 100, 109, 119–21, 141–47, 154, 157, 163, 166, 174, 179, 181–83, 195, 202, 216, 219–24, 233–36, 247, 253, 257, 261, 276, 280, 288, 290–93, 302, 321, 329
resources, 9, 20–21, 25, 36, 45, 57, 91–97, 119, 123–32, 140–45, 151–55, 201, 282, 332–33
reverse links, 92
scope, 111, 117, 268–70, 275, 294
script units, 23, 174–76, 181
sets, 9, 36–38, 57, 61, 63–64, 109, 147, 265, 282–84, 374
special characters, 81, 274, 334
taxonomy, 9, 30, 32–36, 57, 69, 74, 97–99, 112, 183, 245–47, 252–53, 257, 260–61, 375
texts, 9, 22–26, 36, 57, 81, 97, 133, 161–64, 167–69, 180–81, 185, 194–99, 201, 206, 209, 223, 225–26, 232–33, 235, 275, 278, 282, 338, 347
thesaurus, 30–31, 245, 251–63, 273, 279
translations, 55–56, 78, 84–86, 153, 157, 165, 171, 182–83, 189, 214, 232–33, 260, 327
users, 3, 9–13, 17, 36, 40, 120, 172, 267, 286, 294, 319–25
anonymous users, 41
named users, 41
user accounts, 319
variables and values, 69
view tab, 27, 38, 45–46, 60, 64–67, 85, 89, 93, 103, 134, 138, 145–47, 160, 185, 188–94, 210–11, 214, 217, 223, 228, 239–42, 265, 280–81, 297–304, 313, 317–18, 328, 338
view-only mode, 42, 284–85, 329
watermarks, 131, 330
Web links (URLs), 317
writing systems, 9, 23–25, 36, 81, 161–64, 171, 174–75, 179, 202, 222, 328
XML, 4–6, 11, 206, 337, 340, 341, 345–46, 357, 363–72
Zotero, 26, 153–59