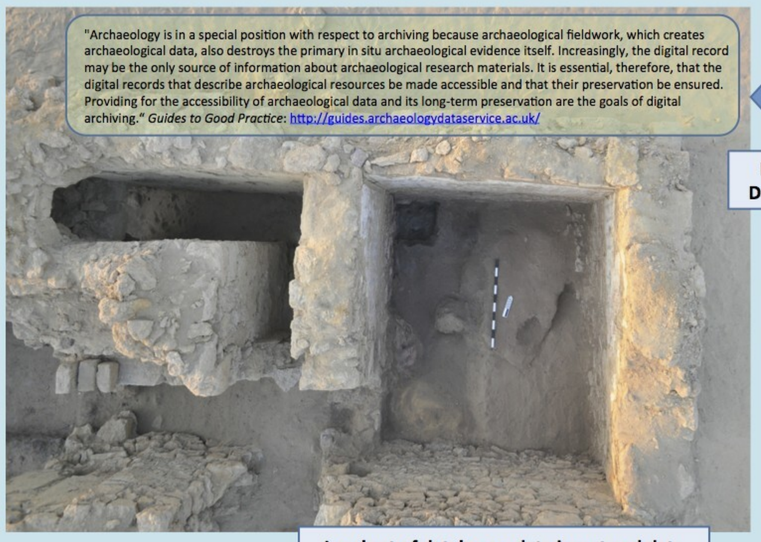
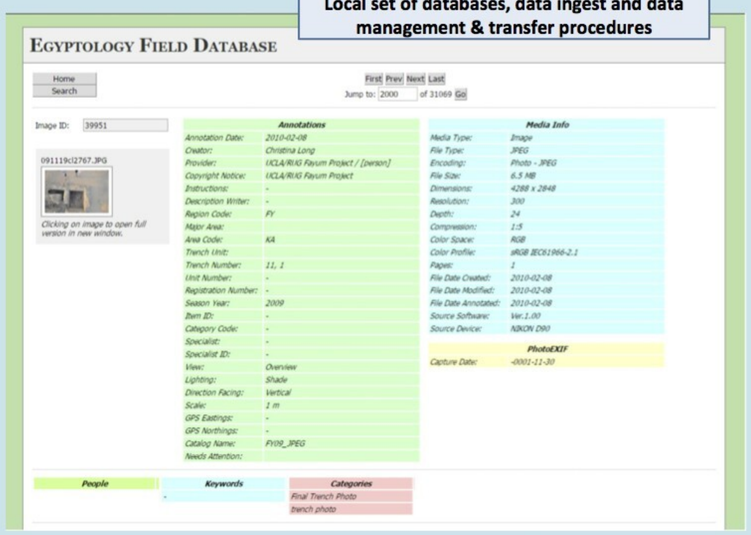
From Field to Footnote: A Case Study in the Management of Archaeological Data



Fayum Database

Local set of databases, data ingest and data management & transfer procedures

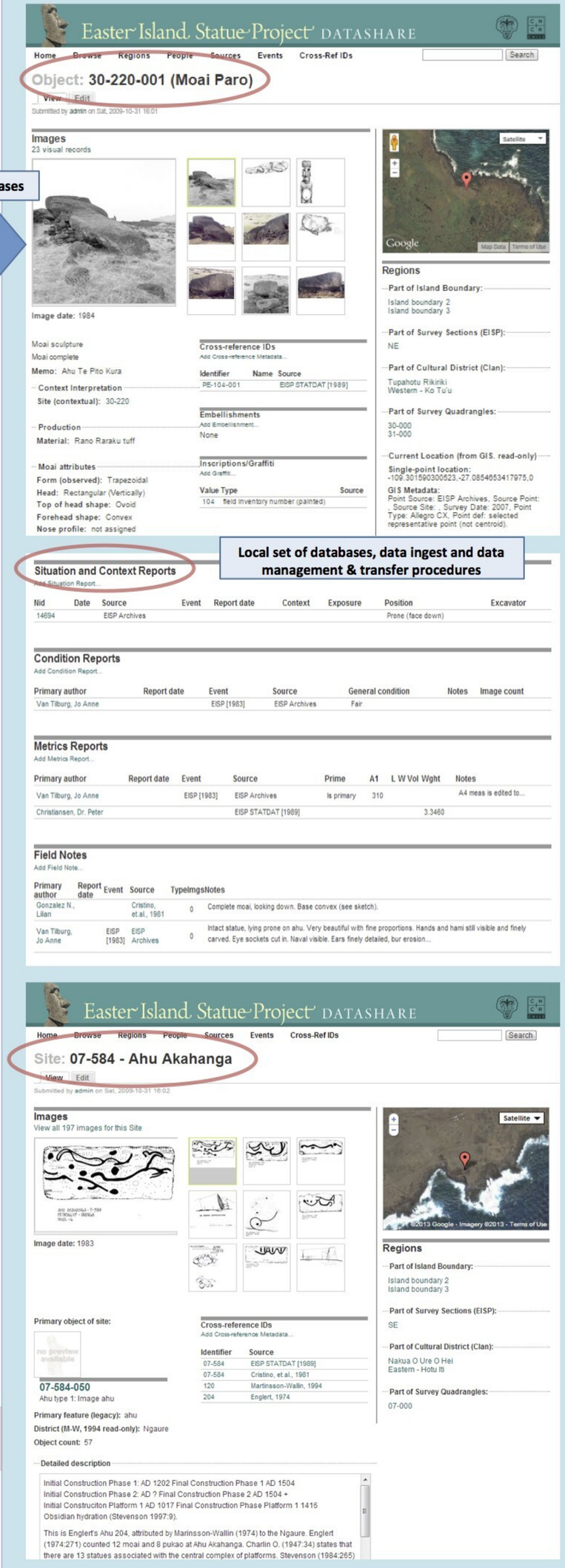


Target Projects:

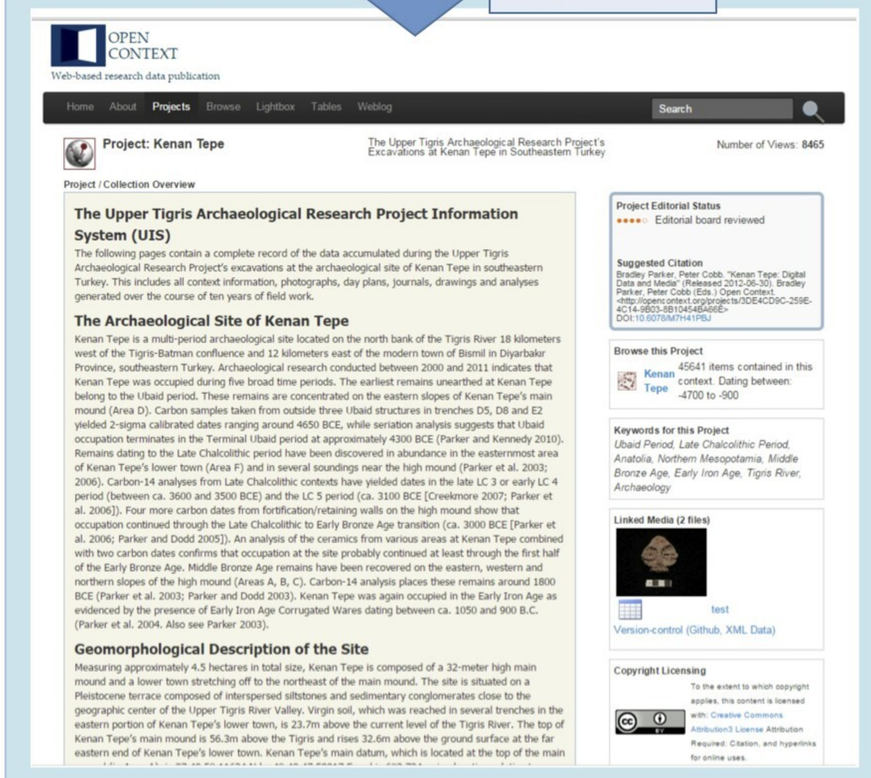
- **Fayum Greco-Roman excavations (UCLA/Rijkuniversiteit Groningen) - http://www.archbase.com/fayum/**
- Fayum Neolithic survey and excavations (UCLA/University of Auckland) - http://www.archbase.com/fayum/
- Prehistoric Sitio Drago, Bocas del Toro, Panama (Tom Wake, UCLA) - http://antiquity.ac.uk/ProjGall/wake/
- **Easter Island Project (Jo Anne van Tilburg, UCLA) - http://www.eisp.org/**
- **Upper Tigris Archaeological Research Project (Bradley Parker, Univ. Utah) - http://opencontext.org/projects/3DE4CD9C-259E-4C14-9B[illegible]104**

EISP Databases

UTARP Project Data in Open Context



Local set of databases, data ingest and data management & transfer procedures



Introduction

"Since 1975, the Cotsen Institute of Archaeology Press (formerly the Publications Unit) has served to preserve cultural heritage through the documentation and publication of scholarly archaeological research. Specializing in producing high-quality academic volumes, the Press publishes about 10 titles per year in seven different series..." (http://www.ioa.ucla.edu/publications/introduction)

The Institute plans to include two new services into its publication program:

- Web-based online publications
- Data hosting and citation services

These services are an expansion of existing digital publication activities at the Cotsen and it is not the intention to replicate services that are already available.

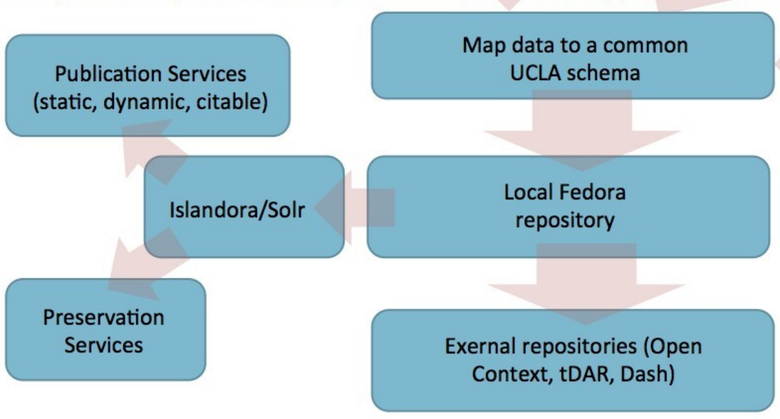
Online Publication

The Cotsen Institute currently publishes books in print, and as PDFs through the University's eScholarship repository: http://escholarship.org/uc/cioa.

The Digital Library currently hosts a number of web sites with narrative, contextual, and/or scholarly content, including Carolingian Culture at St Gall and Reichenau (http://stgallplan.org), the David Livingstone Spectral Imaging Project (http://livingstone.library.ucla.edu), UCLA Preserved Silent Animation (http://animation.library.ucla.edu), and the Los Angeles Aqueduct Digital Platform (http://digital.library.ucla.edu/aqueduct). A number of other scholarly projects—some already using the Digital Library's repository services--are in the process of being migrated into the Digital Library context: the UCLA Encyclopedia of Egyptology (http://uee.ats.ucla.edu), Ancient Egyptian Architecture Online (http://dai.aegaron.ucla.edu), Digital Karnak (http://dlib.etc.ucla.edu/projects/Karnak/) and the Digital Roman Forum (http://dlib.etc.ucla.edu/projects/Forum/).

In order to provide sustainable online publication services the Digital Library is standardizing on a few publication frameworks and components. One such framework will be based on the existing UCLA Encyclopedia of Egyptology (UEE), and will provide publication of texts from XML/TEI. Looking forward, the UEE publication framework can be used to provide publication services for the Cotsen Institute of Archeology Press. There are a number of objectives:

1. Capture a SNAPSHOT of the data at strategic points in time.
2. Create STATIC subsets of for publication purposes.
3. Create DYNAMIC services around the snapshot for users to explore the data snapshot.
4. Deposit static snapshots to external repositories as required or desired by project managers.



Data Hosting and Citation

There are currently three candidate data hosting services:

Dash
http://dash.cdlib.org

- A California Digital Library Service
- Content hosted in Merritt and replicated through the San Diego Supercomputer Center
- Intended as a general purpose data preservation and access service
- Not subject specific
- Does not include any peer review or data validation services
- Not currently in production mode

Open Context
http://opencontext.org

- Maintained by a 501(c)(3) Non-profit corporation, The Alexandria Institute Archive, http://alexandriaarchive.org
- Content hosted in Merritt
- Is a publication service for archeological data, including editorial and peer review, citation tracking, discovery and download services
- Not intended as a preservation repository
- Fully formed, mature service
- Assigns ARKs to individual items; assigns DOIs to datasets and projects
- Option of working through OpenContext to archive data with the UC Curation Center

tDAR: The Digital Archaeological Record
http://www.tdar.org

- Operates under the umbrella of Digital Antiquity, a multi-institutional organization that has been explicitly designed to ensure the long-term financial, technical, and social sustainability of tDAR. Currently being incubated at Arizona State University
- Designed to serve the needs of a wide range of archaeologists, researchers, organization, and institutions who use or manage archaeological resources
- Hosts documents, data sets, images, reports, bibliographies, etc.
- Extensive data dictionary: https://dev.tdar.org/confluence/display/TDAR/Data+Dictionary

More generic, less discipline specific

REPOSITORY SERVICES

More discipline specific, less generic

Project Questionnaire

The purpose of this questionnaire is to provide the UCLA Digital Library Program with information to inform a discussion about appropriate data curation and publication services that the Library can provide or broker.

General information about the project

- Name(s) and contact information for the Project director(s)
- Name(s) and contact information of the best person/people to answer questions about the project data and technologies

Project description

- What was the start date of the project (or data collection)?
- Is data collection ongoing, or is the data static?
- If data is static, what was the date on which data collection ceased?
- If data collection is ongoing, how often is new data gathered or updated?
- How many concurrent projects under the umbrella? How many databases/data stores managed by co-directors or graduate students?

Data storage and maintenance

- What type of database or data storage software is used for the project?
- What other software is used for data management, publication, etc.?
- Please provide a description of the database structure: tables and fields, relationships between tables
- If using a data store other than a traditional relational database please provide a description of the data structures in that store.

Data attributes

- Is the data currently available online (private, public)?
- How comprehensive is the data?
- What external references exist to data not in the database?
- If there are sub- or related projects, how do they relate to each other? Is data periodically aggregated?
- What are the main object types that your data describes?

Objectives

- What are your future objectives for the data? Archiving? Publication? Normalization? Sharing? Reuse?
- What role do you see for the Library, or other data repositories? What do you need to meet your objectives?
- Do you have contractual or permit-related needs for your data?
- Are there other archiving needs that you have? PDFs. Images. etc.

Stephen Davison & Willeke Wendrich
University of California, Los Angeles

With thanks to: Stephanie Simms, Deidre Whitmore and Lisa McAulay