

Unified Digital Format Registry (UDFR) User's Guide

Version 1.0.0 – 2012-06-04

UC Curation Center California Digital Library University of California











Copyright © 2012, The Regents of the University of California All rights reserved

This document is licensed under the Create Commons Attribution-Share Alike 3.0 United States license. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/3.0/us/ or request a copy from Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Acknowledgements

The Unified Digital Format Registry (UDFR) is a product of the University of California Curation Center (UC3), http://www.cdlib.org/uc3, at the California Digital Library (CDL). The project was funded by the Library of Congress as part of its National Digital Information Infrastructure Preservation Program (NDIIPP).

The UDFR project team gratefully acknowledges the support of the UK National Archives (TNA) and the Agile Knowledge Engineering and Semantic Web (AKSW) research group at the University of Leipzig.

UC3 project team	Library of Congress	UK National Archives	AKSW, Univ. Leipzig
Stephen Abrams	Martha Anderson	Tim Gollins	Philipp Frisch
Lisa Dawn Colvin	Leslie Johnston	Tracey Powell	Norman Heino
John Kunze		Spencer Ross	Sebastian Tramp
Margaret Low			
Mark Reyes			
Abhishek Salve			
Marisa Strong			

Version history

Version	Date	Note
1.0.0	2012-06-04	Initial publication

UDFR User's Guide Page 1 of 48







Contents

Acknowledgements	1
Version history	1
Contents	2
Introduction	4
1 Getting started	5
1.2 User roles	5
1.3 User interface	5
1.3.1 OntoWiki pane	6
1.3.1.1 User menu	7
1.3.1.2 Extras menu	9
1.3.1.3 Help menu	10
1.3.1.4 Debug menu	10
1.3.2 Knowledge Base pane	11
1.3.3 Navigation pane	12
1.4.4 Login pane	
1.3.5 Workspace pane	
1.3.6 Contextual menu	14
2 Initial data loads	18
2.1 MIME media types	18
2.2 PRONOM	18
3 Search and browse	20
3.1 Browse	20
3.2 Search	20
4 Format representation information	21
4.1 Review status	22
4.1.1 Reviewing properties	22
4.2 Provenance history	23
4.3 Community comments	24
4.4 RDF source	24







4.5 Instance linkir	ng	. 25
5 User contribution	on	. 26
5.1 Create a reso	urce	. 26
5.2 Editing a reso	ource	. 29
5.3 Deleting a res	source	. 31
5.4 Deleting a pro	operty	. 31
6 Advanced featur	res	. 34
6.1 File upload		. 34
6.2 SPARQL queri	ies	. 35
6.2.1 Query form	n	. 35
6.2.2 SPARQL en	dpoint	. 37
Appendix A RDF ar	nd the Linked Data	. 42
Appendix B UDFR	implementation	. 43
Appendix C UDFR	class structure	. 44
Deferences		16







Introduction

A deep understanding of digital formats is necessary to support the long-term preservation of digital assets, as it facilitates the preservation of the *information content* of those assets, rather than just their *bit stream representations*. A format is the set of syntactic and semantic rules that govern the mapping between information and the bits that represent that information. The Unified Digital Format Registry (UDFR) is a new open source, semantically-enabled platform for the collection, long-term management, and dissemination of the significant properties of formats of interest to the preservation community [5]. The UDFR builds upon and "unifies" the function and holdings of two existing registry solutions: PRONOM, from the UK National Archives [13]; and GDFR (Global Digital Format Registry), from Harvard University [8]. While these services rely on older relational database technology, the UDFR uses a semantic database in which all information is represented in RDF form [23] and exposed as Linked Data [11]. Use of the UDFR is open to the public, although contribution or editing of information requires prior self-service account registration.

The UDFR was developed by the University of California Curation Center (UC3) [6] at the California Digital Library (CDL) [3] with funding from the Library of Congress as part of its National Digital Information Infrastructure Preservation Program (NDIIPP) [10].

UDFR User's Guide Page 4 of 48







1 Getting started

The UDFR portal is found at http://udfr.org (see Figure 1). This page provides links to the registry itself at http://udfr.org/onto/onto.rdf, and the online UDFR documentation at http://udfr.org/docs.

Unified Digital Format Registry (UDFR) a semantic registry for digital preservation The UDFR is a reliable, publicly accessible, and sustainable knowledge base of file format representation information for use by the digital preservation community. A format is a set of semantic and syntactic rules governing the mapping between abstract information and its representation in digital form. While many worthwhile and necessary preservation activities can be performed on a digital asset without knowledge of its format, that is, merely as a sequence of bits, any higher-level preservation of the underlying information content must be performed in the context of the assets format. The UDFR seeks to "unify" the function and holdings of two existing registries, PRONOM and CDFR (the Global Digital Format Registry), in an open source, semantically enabled, and community supported platform. The UDFR was developed by the University of California Curation Center (UC3) at the California Digital Library (CDL), funded by the Library of Congress as part of its National Digital Information Infrastructure Preservation Program (NDIIPP). The service is implemented on top of the OntoWiki semantic wiki and Virtuoso triple store. Registry - Start Here - UDFR ontology (RDF) - Documentation Copyright © 2012, The Regents of the University of California | Powered by OntoWiki | Terms of use | Privacy policy UDFR

Figure 1 – UDFR portal page, http://udfr.org/

1.2 User roles

The UDFR defines four distinct user roles:

- Consumer. An anonymous read-only user of the UDFR.
- *Contributor*. A registered user of the UDFR permitted to add and edit format representation information.
- *Reviewer*. A contributor who is additionally permitted to review contributor-supplied information for technical veracity.
- Administrator. A UDFR system administrator, with all privileges.

This document provides a general introduction to the UDFR features available to consumers and contributors.

1.3 User interface

The general layout of the UDFR user interface, http://udfr.org/registry is shown in Figures 2 and 3. There are five main screen areas, or panes.

UDFR User's Guide Page 5 of 48







- OntoWiki pane. The OntoWiki pane (uppermost-left, labeled "O" in the Figures) contains menu options of general utility throughout the use of UDFR. (OntoWiki is the underlying semantic wiki platform on which the UDFR is based [19]. For more implementation details, see Appendix B.)
- Knowledge Base pane. The Knowledge Base pane (upper-left, "K") offers a choice of knowledge bases (KBs), or RDF models [23], for selection.

NOTE The terms "knowledge base", "model", and "graph" are used synonymously within the semantic web community.

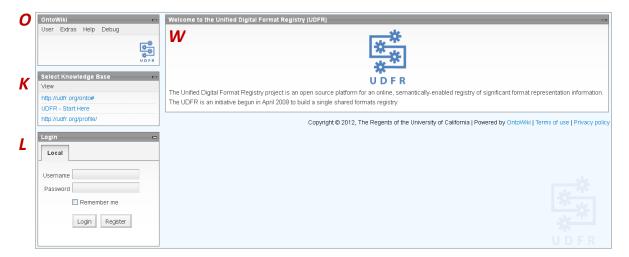


Figure 2 - Initial UDFR user interface layout

- Navigation pane. The Navigation pane (lower-left, "N") provides an RDF class browser for navigating the UDFR format representation information. This pane is not displayed until a knowledge base is selected in the Knowledge Base pane as described in Section § 1.3.2, Knowledge Base pane.
- Login pane. The Login pane (bottom-left, "L") lets contributors login to their user account. Potential contributors can also follow a link for new account registration. Once logged in, this pane is no longer displayed.
- Workspace pane. The Workspace pane (right, "W") is function dependent. Its appearance will vary based on the operation being performed.

1.3.1 OntoWiki pane

The OntoWiki pane contains menu options of general utility throughout the use the UDFR. There are four main menus:

- User.
- Extras.

UDFR User's Guide Page 6 of 48







- Help.
- Debug.

Once a knowledge base has been selected, as described in Section § 1.3.2, , the OntoWiki pane also displays a search box.

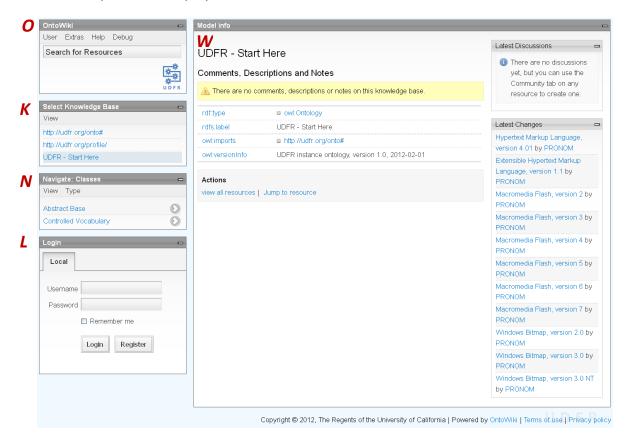


Figure 3 – UDFR user interface layout following knowledge base selection

1.3.1.1 User menu

The "User / Register New User" menu option permits contributors to register for user accounts (see Figure 4).

NOTE Account registration is necessary *only* for the contribution of new data or editing of existing data in the UDFR. No account registration is necessary for read-only access to the UDFR.

The details of user account registration are provided in Section § 1.3.4, Login pane.

UDFR User's Guide Page 7 of 48





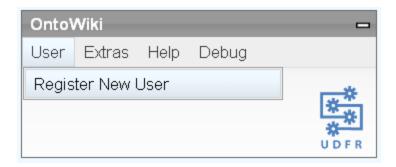


Figure 4 – OntoWiki pane User menu

Once a contributor is logged in, the "User" menu will also provide options for setting user preferences and logging out (see Figure 5).

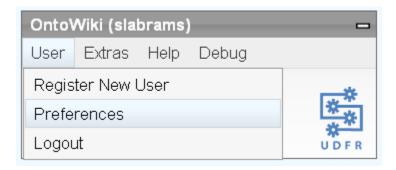


Figure 5 – OntoWiki pane User menu after login

The "User / Preferences" option permits a contributor to update his or her user account preferences, including email address and password (see Figure 6).



Figure 6 – User preferences

NOTE Although the username appears on the form, it is *not* editable.

UDFR User's Guide Page 8 of 48







1.3.1.2 Extras menu

The "Extras" menu offers three options:

- File Manager, for uploading sample files.
- Queries, for issuing SPAQRL queries [16].
- Select Language, for selecting the user interface language.

As advanced topics, file upload and SPARQL queries will be discussed later on in Sections § 5.2 and 5.2.

The "Extras / Select Language" menu option selects the language used in the UDFR user interface. Customized language is available for:

- Chinese.
- English.
- French.
- German.
- Russian.

Additional language customizations may become available in the future.

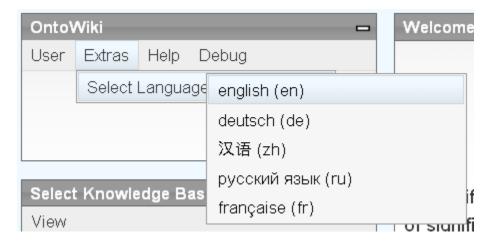


Figure 7 – OntoWiki pane Extras menu

Changing the user interface language does *not* affect the format representation information displayed by the UDFR, only the user interface. For example, selecting the German language would change the appearance of the OntoWiki pane as follows:

UDFR User's Guide Page 9 of 48







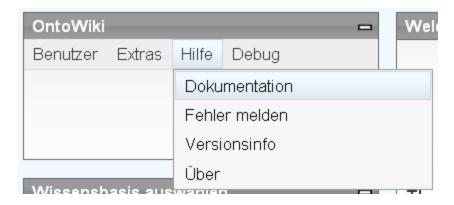


Figure 8 – OntoWiki pane Extras menu in German

1.3.1.3 Help menu

The "Help" menu provides links to help information and documentation about the OntoWiki semantic wiki underlying the UDFR (see Figure 9).

1.3.1.4 Debug menu

The "Debug" menu provides options for clearing various internal caches and restarting a UDFR session, which clears all caches and logs out the user (see Figure 10).

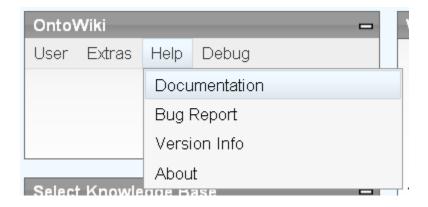


Figure 9 – OntoWiki pane Help menu

UDFR User's Guide Page 10 of 48





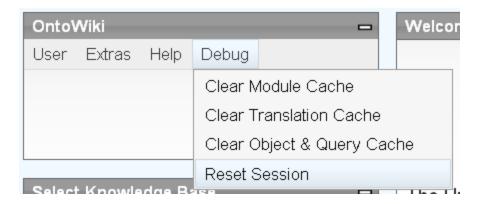


Figure 10 - OntoWiki pane Debug menu

1.3.2 Knowledge Base pane

The Knowledge Base permits the selection of the knowledge base (KB), or RDF *model*, as the focus for subsequent searches and edits (see Figure 11). The UDFR defines three main knowledge bases:

- UDFR data. The UDFR data knowledge base holds the public format representation information. To select this KB click on the "<u>UDFR – Start Here</u>" link. Hereinafter this KB will be referred to as the "data" KB.
- UDFR ontology. The UDFR ontology knowledge base defines the RDF ontology, or schema, for expressing format representation information. To select this KB click on the "http://udfr.org/onto#" link.
- UDFR user profiles. The UDFR user profile knowledge base holds public information about UDFR users, including name, title, and organizational affiliation. To select this KB click on the "http://udfr.org/profile/" link.

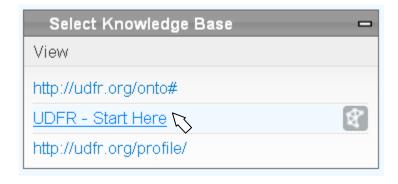


Figure 11 - Knowledge Base pane

Selecting a knowledge base causes the display of the Navigation pane as described in Section § 1.4.3.

UDFR User's Guide Page 11 of 48







1.3.3 Navigation pane

The Navigation pane provides an ontology browser for navigating the UDFR format representation information (see Figure 12). This pane is not displayed until a knowledge base is selected as described in Section § 1.4.2, *Knowledge Base pane*.

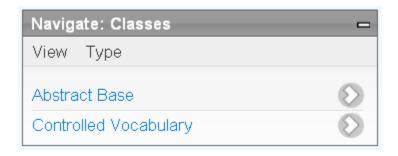


Figure 12 – Navigation pane for data knowledge base

The UDFR classes are defined in a hierarchical structure. The full hierarchical structure of the data knowledge base is shown in Appendix § C, UDFR class structure. In the Navigation pane the left and right arrow icons and ascend and descend one level of the class hierarchy, respectively. The smaller double left arrow icon returns to the root class.

Clicking on any class name will display all of the *instances* of that class in the Workspace pane (see Figure 13).

1.4.4 Login pane

The Login pane permits contributors to register for and login to their user accounts (see Figure 14).

NOTE Account registration and login are *only* necessary for the contribution of new, or editing of existing, representation information. Read-only access to all representation information can be performed anonymously.

To register for a user account, click the "Register" button and complete the form in the Workspace pane (see Figure 15). Form fields marked with a red asterisk * are required; all other fields are optional.

UDFR User's Guide Page 12 of 48







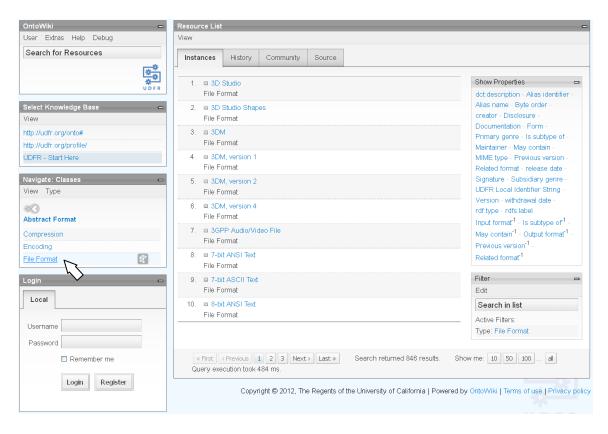


Figure 13 – Displaying class instances in the Workspace pane

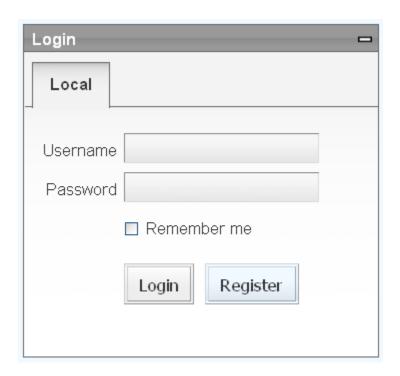


Figure 14 - Login pane

UDFR User's Guide Page 13 of 48







Note By completing the user account registration form you agree to contribute format representation information to the UDFR under the terms of the Creative Commons Attribution license (CC-BY). That attribution is fulfilled by your further agreeing that your full name, title, institutional affiliation, and additional profile information and website URL, if specified, *will* be publicly displayed in the "History" tab for all representation information that you contribute or edit. This information is made public in furtherance of the legitimate UDFR business need to provide strong provenance information for all contributed content.

Passwords are *never* publicly displayed and are stored by the UDFR *only* in a hashed form that is opaque to UDFR administrators. Passwords can be reset with the "User / Preferences" menu option of the OntoWiki pane, as described in Section § 1.3.1.1, *User menu*.

A forgotten password can be recovered by completing the "Login" operation (see Figure 14) and clicking the "Forgot your password?" link (see Figure 16) and supplied a account username or email address (see Figure 17).

The registration form incorporates a Captcha that *must* be entered into the form in a case *sensitive* manner (see Figture 15). If the Captcha is too indistinct or ambiguous, a new one can be generated by clicking the "refresh" icon .

Once a contributor is logged in this pane is no longer displayed. Once logged in, a contributor can logout via the "User / Logout" or "Debug / Reset Session" menu options in the OntoWiki pane, as described in Section § 1.4.1.

1.3.5 Workspace pane

The Workspace pane holds the results of UDFR search and browse operations. The appearance of its displayed information is operation dependent. Many of these operations will result display an ordered *list* of results. The Workspace pane supports paged browsing for lists that cannot be accommodated on a single screen (see Figure 18).

Buttons at the bottom of the pane provide direct access to the first, previous, next, last page of results, as well as to an arbitrarily-numbered page. Additional buttons control the maximum number of results displayed on any given page (see Figure 19).

1.3.6 Contextual menu

Whenever mousing over any active link in the UDFR user interface, a contextual menu becomes available as indicated by the graph icon . Clicking the graph icon will

UDFR User's Guide Page 14 of 48







display a context-sensitive menu offering appropriate options.

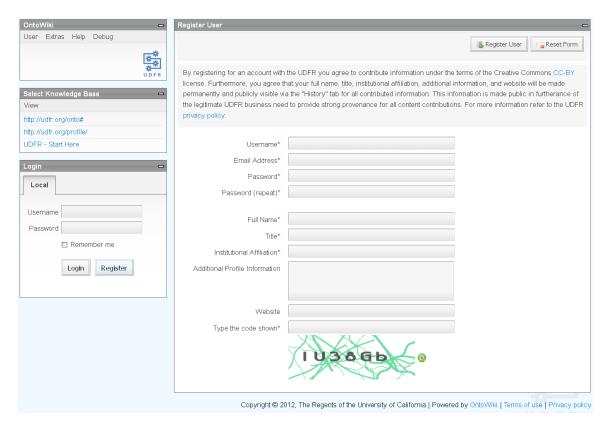


Figure 15 – User registration form



Figure 16 – Bad password

UDFR User's Guide Page 15 of 48







Figure 17 – Password recovery

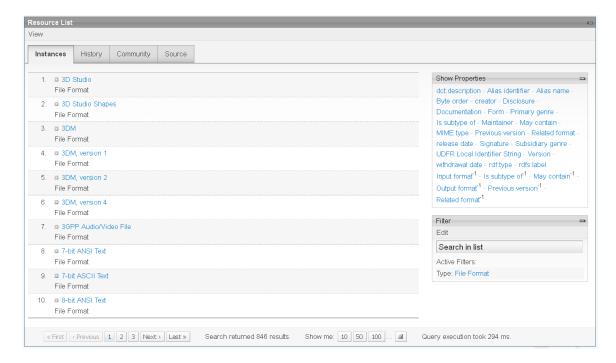


Figure 18 – File format list

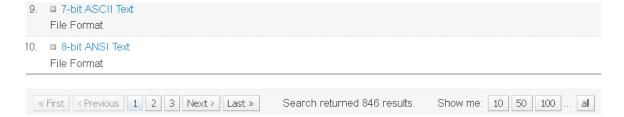


Figure 19 – Paged results buttons

UDFR User's Guide Page 16 of 48







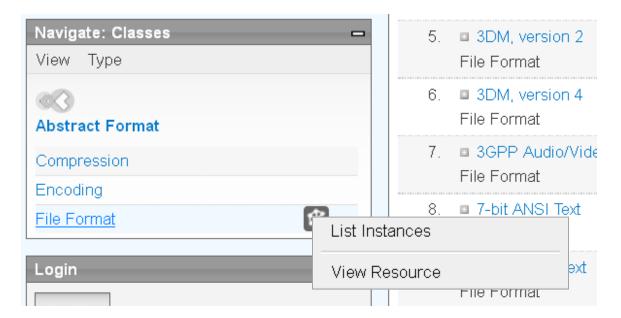


Figure 20 – Contextual menu

UDFR User's Guide Page 17 of 48







2 Initial data loads

The UDFR was initialized with data loads from two external data sources:

- MIME media types [9].
- PRONOM technical registry [13].

2.1 MIME media types

MIME types were loaded from http://mediatypes.appspot.com/ [12], which itself manages data "routinely scrapped form IANA [Internet Assigned Numbers Authority]". The data export was taken on February 22, 2012, and encompassed the following types:

Count	Media type
809	application/*
125	audio/*
39	image/*
14	message/*
14	multipart/*
51	text/*
56	video/*

2.2 PRONOM

PRONOM data were loaded from

1,127

http://nationalarchives.gov.uk/PRONOM/Default.aspx, the technical registry operated by the UK National Archives [13]. The data export was taken on February 23, 2012, and encompassed the following entities:

Table 1 – Initial MIME media types

Count	Resource
846	File formats
28	Character encodings
17	Compression algorithms
1,237	Identifiers
1,006	External signatures

UDFR User's Guide Page 18 of 48





Internal signatures
MIME media types (not already loaded from Appspot)
Agents
Software packages
Software processes
Intellectual property rights (IPR) statements
Relationships

8,274

Table 2 – Initial PRONOM resources

The PRONOM data export is made freely available under the terms of the Open Government License (OGL), http://www.nationalarchives.gov.uk/doc/open-government-licence/. Unless otherwise specified, all other contributed data all made freely available under the terms of the Creative Commons Attribution license (CC-BY), http://creativecommons.org/licenses/by/3.0/.

Additional data may have been contributed since the time of these initial data loads.

UDFR User's Guide Page 19 of 48







3 Search and browse

UDFR format representation information can be retrieved via search or browse.

3.1 Browse

Use the ontology browser in the Navigation pane to navigate to the class of interest, then click the class name to display all instances of that class in the Workspace pane, as described in Section § 1.3.3, *Navigation pane* (see Figure 13).

3.2 Search

All textual data in the UDFR is indexed for search. Enter terms in the search box in the OntoWiki pane and press the Return or Enter key to display matching search results in the Workspace pane (see Figure 21).

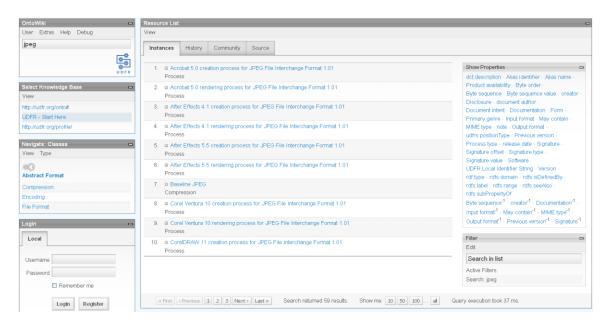


Figure 21 – Search results

UDFR User's Guide Page 20 of 48







4 Format representation information

As described in Section § 3, UDFR format representation information can be retrieved via search or browse. The unit of display in the UDFR is the *resource*, which corresponds to an RDF *class* (see Appendix C, *UDFR class structure*.). The Workspace pane display for a given format (or other resource) will be similar to that shown in Figure 22.

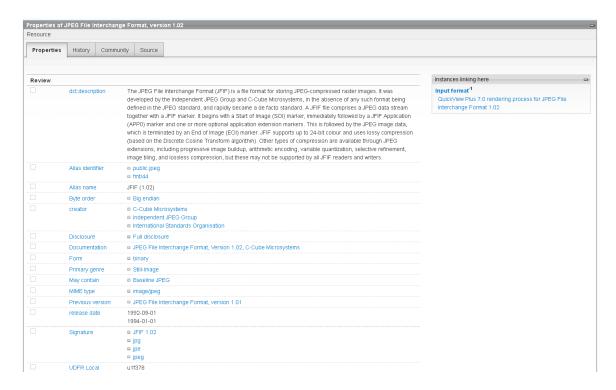


Figure 22 – Representation information for JPEG JFIF 1.02

The representation information takes the form of a list of property names and values associated with the resource in alphabetical order. Note that all of the property names, and many of the values, are actionable links. Clicking the link will display the properties pertaining to that property or value.

In addition to the property names and values, the following additional information is made available:

- Review status.
- Provenance history.
- Community comments.
- RDF source.
- Instance linking.

UDFR User's Guide Page 21 of 48







4.1 Review status

Every representation information property is associated with a review status set by UDFR users with the Reviewer role. Reviewed properties are indicated by the presence of a green checkmark icon before the property name (see Figure 23).

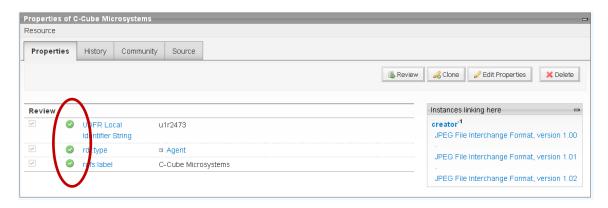


Figure 23 – Review status

4.1.1 Reviewing properties

UDFR users holding the Reviewer role will see active review checkboxes on all representation information displays (see Figure 24).

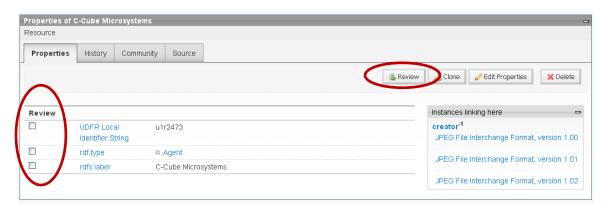


Figure 24 – Review checkboxes enabled

Once the appropriate boxes are checked, the review status can be updated by clicking the "Review" button.

UDFR User's Guide Page 22 of 48







4.2 Provenance history

UDFR tracks the full provenance of every representation information property. This information is made available via the "History" tab (see Figure 26).



Figure 26 – History tab

All actions related to properties are tracked:

- · Adding, editing, and deleting.
- Reviewing.

The "History" display shows the type of action (and its underlying RDF form), the user who performed the action, and the timestamp of its performance. User names are actionable links. Clicking this link will display additional information about the user in a popup window so that an informed decision can be made regarding the veracity of the contributed information (see Figure 27).

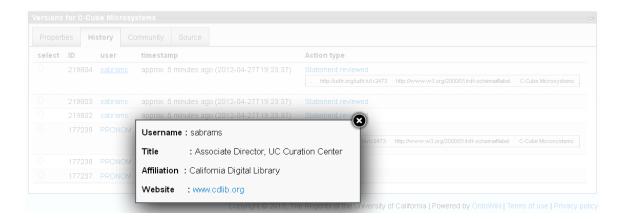


Figure 27 – User profile information

UDFR User's Guide Page 23 of 48







4.3 Community comments

The "Community" tab displays comments contributed by register users holding the Contributor (or Reviewer) role (see Figure 28).



Figure 28 - Community tab

4.4 RDF source

The RDF underlying the displayed properties can be seen in the "Source" tab (see Figure 29).

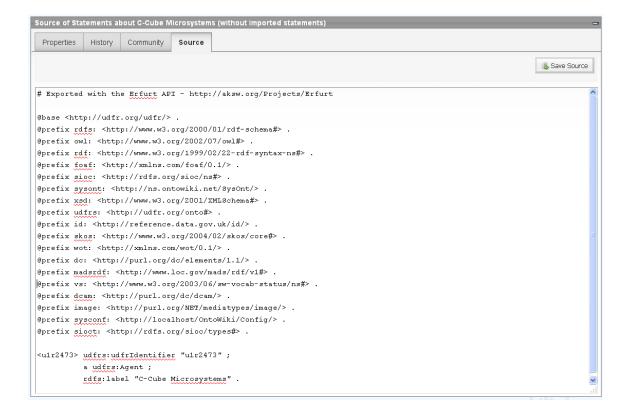


Figure 29 - Source tab

UDFR User's Guide Page 24 of 48







4.5 Instance linking

The "Instance linking" sub-pane of the Workspace displays active links to other properties whose values are the property being shown. For example, Figure 30 displays the representation information properties defined for the C-Cube Microsystems corporate agent. The "Instances linking here" sub-pane to the right of the display indicates that C-Cube is the creator of the JPEG JFIF 1.00, 1.01, and 1.02 formats; as well as indicating the presence of a user contributed comment.

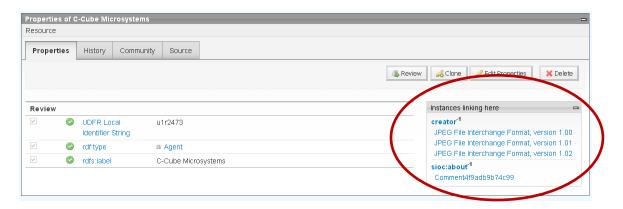


Figure 30 – Instance linking

UDFR User's Guide Page 25 of 48







5 User contribution

UDFR users holding the Contributor (or Reviewer) role are permitted to add, edit, and delete representation information for UDFR *resources*, such as File Format, Agent, etc. (Resources correspond to RDF *classes*. See Appendix C, *UDFR class structure*, for a comprehensive list of resources.)

5.1 Create a resource

Use the Navigation pane ontology browser to find the desired resource and then select the "Create Instance" option in the resource's contextual menu (see Figure 31).

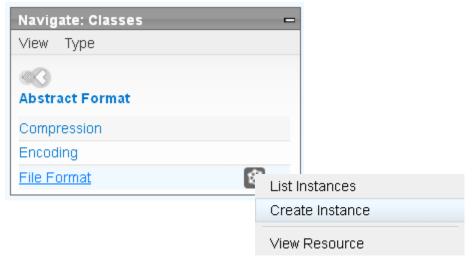


Figure 31 – Create instance menu option

This will display a popup window with input fields for all properties defined for the resource, either directly or by inheritance from ontological super-class relationships (see Figure 32).

UDFR properties are of three types:

- Text strings.
- Closed enumerations.
- Open enumerations.

Text properties can be entered directly in the form text box. Closed enumerations have a fixed set of values defined in the UDFR ontology. Clicking on the "Select Predefined Value" link for property, for example, the "byteOrderType" shown in Figure 33, will

UDFR User's Guide Page 26 of 48







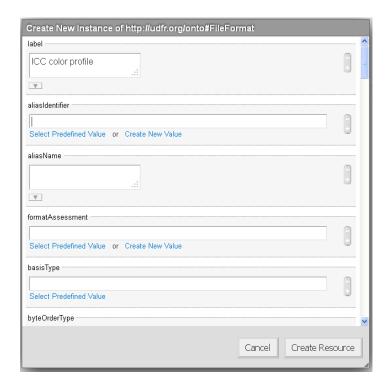


Figure 32 – Create instance popup window

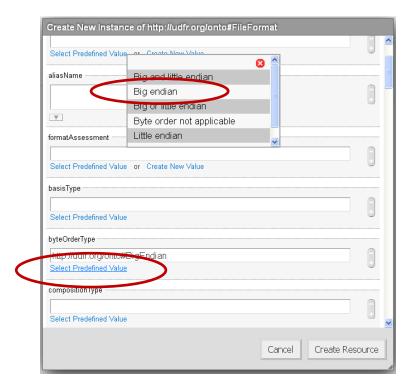


Figure 34 – Byte order type closed enumeration

UDFR User's Guide Page 27 of 48







display a popup window with all of the enumeration values. Once a value is selected, its RDF URI will be automatically displayed in the form box.

Open enumerations can be selected from a pre-defined list in similar fashion, or if the appropriate value is not available, entered directly in to the form box (see Figure 35) and clicking the "Create New Value" link (Figures 36 and 37).

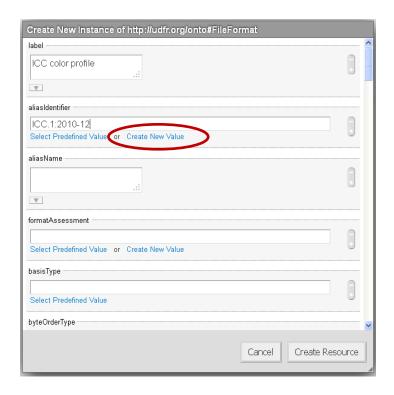


Figure 35 – Open enumeration, step 1



Figure 36 – Open enumeration, step 2

UDFR User's Guide Page 28 of 48







Figure 37 – Open enumeration, step 3

Additional instances of repeatable properties can be defined by clicking the "+" icon to the right of the property form box (see Figures 38 and 39).



Figure 38 – Add property icon



Figure 39 – Added property

5.2 Editing a resource

Existing resources can be edited by selected the "Edit" button on the resource display (see Figure 40). The behavior of the input boxes is same as described in Section § 5.1, *Creating a resource* (see Figure 41).

Editing can also be initiated via the "Edit Resource" option on the contextual menu for a resource in a results list (see Figure 42).

UDFR User's Guide Page 29 of 48







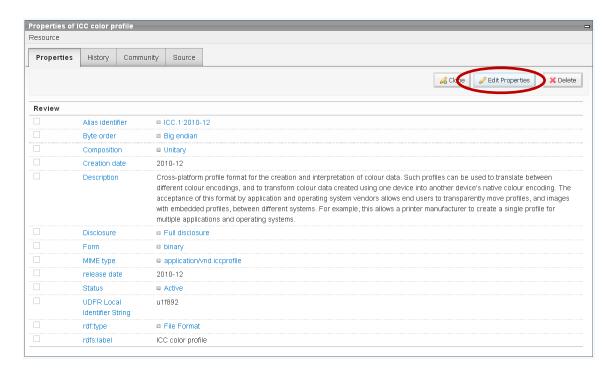


Figure 40 - Resource edit button

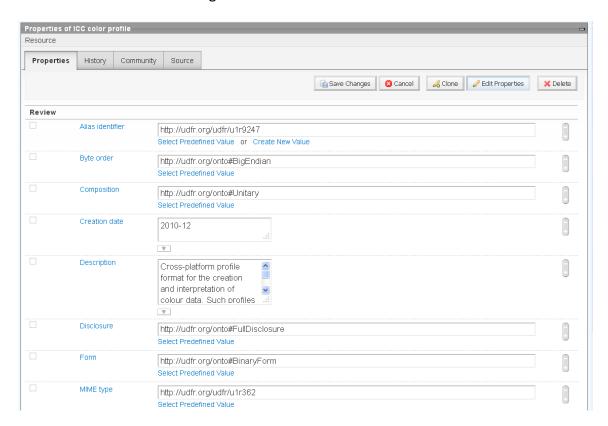


Figure 41 – Resource edit display

UDFR User's Guide Page 30 of 48







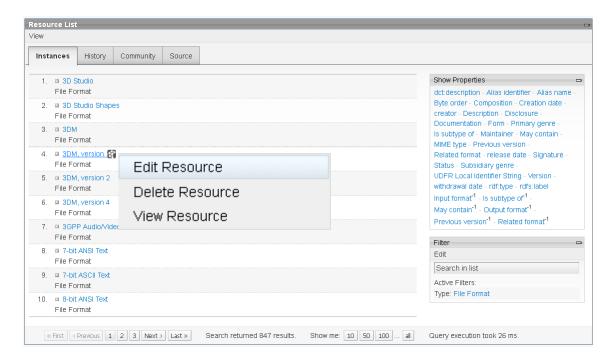


Figure 42 - Edit resource menu option

5.3 Deleting a resource

Existing resources can be deleted by selected the "Delete" button on the resource display (see Figure 43). Deletion can also be initiated via the "Delete Resource" option on the contextual menu for a resource in a results list (see Figure 44).

5.4 Deleting a property

Individual properties can be deleted using the "—" icon to the right of the property form box in the edit resource display (see Figures 45 and 46).

UDFR User's Guide Page 31 of 48







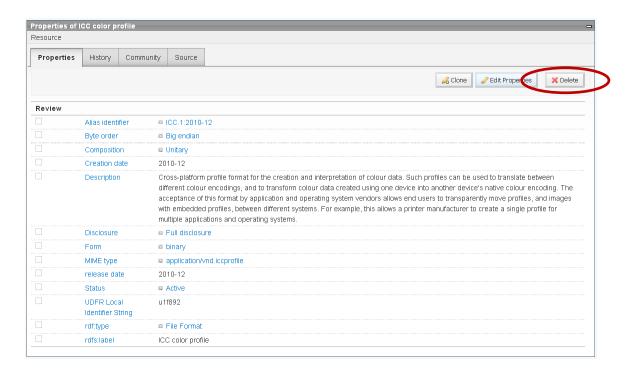


Figure 43 - Resource delete button

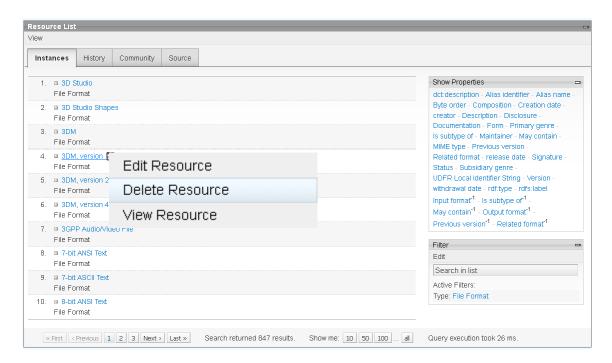


Figure 44 – Delete resource menu option

UDFR User's Guide Page 32 of 48







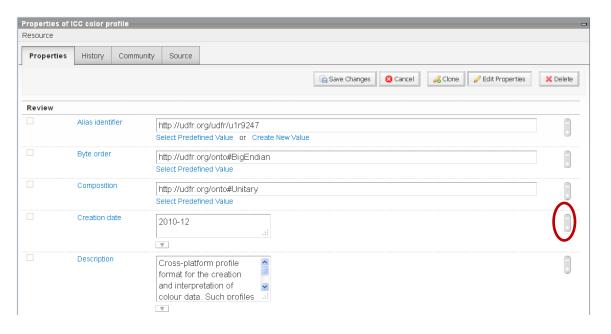


Figure 45 – Property delete icon

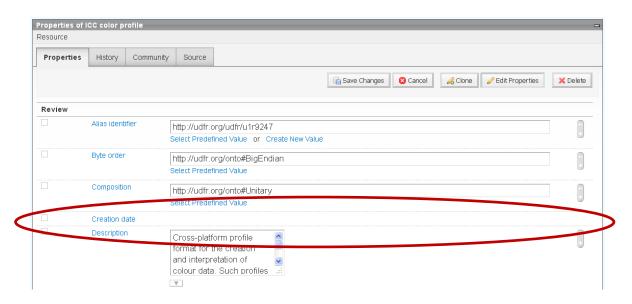


Figure 46 - Deleted property

UDFR User's Guide Page 33 of 48







6 Advanced features

This section describes a number of advanced UDFR features.

- File upload.
- SPARQL queries [16].

6.1 File upload

The OntoWiki pane "Extras / File Manager" menu option enables the upload of reference files that can be associated with particular formats (see Figure 23).

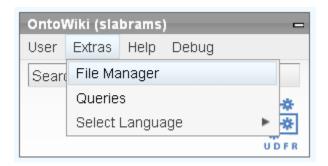


Figure 23 - OntoWiki pane file manager option

The file manager display shows all uploaded files by resource URI and MIME type (see Figure 24).



Figure 24 – File manager

Clicking the "Upload" button displays the upload form (see Figure 25). The "Resource URI" field should *not* be modified; UDFR will automatically assign the appropriate URI to the uploaded file.

The file can be associated with a format by selecting the "Edit Resource" contextual menu option for the format resource, clicking the "Select Predefined Value" link for the "File" property, and selecting the file from the pop-up window (see Figures 26 and 27).

UDFR User's Guide Page 34 of 48









Figure 25 - File upload form

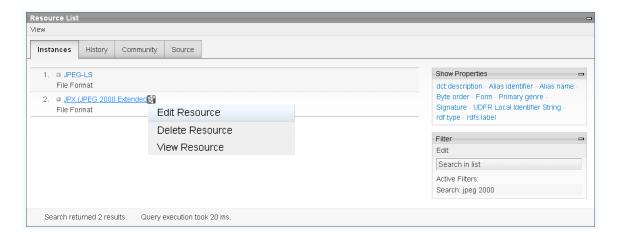


Figure 26 – Edit resource contextual menu

6.2 SPARQL queries

Direct SPARQL queries against the underlying RDF in the Virtuoso quadstore [14] can be performed via two mechanisms.

- OntoWiki query form, for manual queries by human user agents.
- SPARQL endpoint, for scripted queries by automated agents.

An explanation of the SPARQL query language is beyond the scope of this document. Many comprehensive online tutorials are available.

6.2.1 Query form

The OntoWiki pane "Extras / Queries" menu option enables an Working pane form for submitting SPARQL queries (see Figures 28 and 29).

UDFR User's Guide Page 35 of 48







Figure 27 – Associating a file with a format



Figure 28 – OntoWiki pane query option

UDFR User's Guide Page 36 of 48







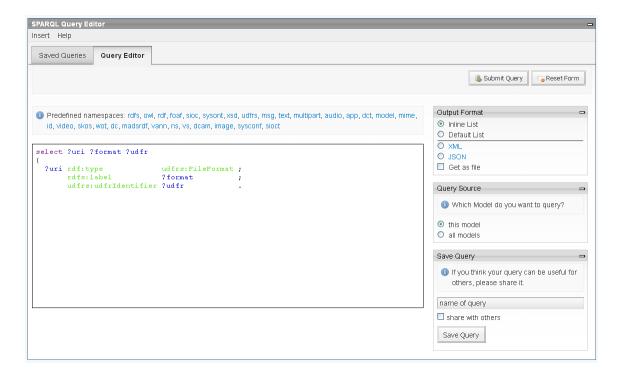


Figure 29 – SPARQL query form

Inline results are displayed at the bottom of the Workspace pane (see Figure 30).

When formulating the SPARQL query, note that a number of fundamental namespaces are pre-defined, including:

rdf : http://www.w3.org/1999/02/22-rdf-syntax-ns#

rdfs : http://www.w3.org/2000/01/rdf-schema#

udfrs: http://udfr.org/onto#

The delivery mode and format of the query results can be specified via the radio buttons in the upper right hand corner of the display (see Figures 31 - 33).

- Inline (default) or downloadable file.
- Plain text (default), XML, or JSON.

6.2.2 SPARQL endpoint

A SPARQL endpoint is a web service that conforms to the SPARQL protocol [16]. UDFR presents a read-only endpoint at http://udfr.org/ontowiki/sparql/.

UDFR User's Guide Page 37 of 48







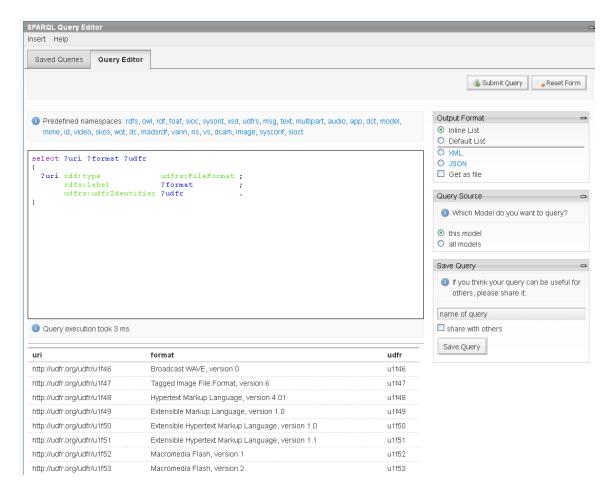


Figure 30 - Query results

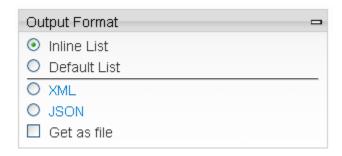


Figure 31 – Query results output format

UDFR User's Guide Page 38 of 48







```
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
 <head>
      <variable name="uri" />
     <variable name="format" />
     <variable name="udfr" />
   </head>
  <results>
            <result>
              <binding name="uri">
                  <uri>http://udfr.org/udfr/u1f46</uri>
                </hinding>
              <binding name="format">
                  <!iteral</pre>
          >Broadcast WAVE, version O</literal>
                </binding>
              <binding name="udfr">
                  Kliteral
          >u1f46</literal>
                </binding>
            </result>
```

Figure 32 – Query results in XML

```
{ "head": { "link": [], "vars": ["uri", "format", "udfr"] },
 "results": { "distinct": false, "ordered": true, "bindings": [
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f46" } , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f47" } , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f48" } , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f49" }
                                                                      , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f50" }
                                                                      , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f51" }
                                                                      , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f52" }
                                                                     , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f53" } , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f54" } , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f55" }
                                                                      , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f56" } , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f57" } , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/ulf58" } , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f59" }
                                                                      , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f60" }
                                                                      , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f61" }
                                                                      , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f62" }
                                                                      , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f63" }
                                                                       , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f64" }
                                                                     , "format": { "type": "l:
   { "uri": { "type": "uri", "value": "http://udfr.org/udfr/u1f65" }
                                                                     , "format": { "type": "l:
```

Figure 33 – Query results in JSON

UDFR User's Guide Page 39 of 48







Note that unlike the use of the SPARQL query form, in which the most important namespaces are defined, as described in Section § 6.2.1, *all* namespaces must be explicitly defined in the queries sent to the endpoint.

Assuming the SPARQL query:

in the file "example-query.sparql", the query can be sent to the UDFR endpoint via the "curl" command:

```
% curl --data-urlencode query@example-query.sparql \
       http://udfr.org/ontowiki/sparql/
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
  <head>
      <variable name="uri" />
      <variable name="format" />
      <variable name="udfr" />
    </head>
  <results>
            <result>
              <binding name="uri">
                  <uri>http://udfr.org/udfr/u1f46</uri>
                </binding>
              <binding name="format">
                  teral
          >Broadcast WAVE, version 0</literal>
                </binding>
              <binding name="udfr">
                  teral
          >u1f46</literal>
                </binding>
```

UDFR User's Guide Page 40 of 48





UDFR User's Guide Page 41 of 48





Appendix A RDF and the Linked Data

All format representation information in the UDFR is represented in RDF (Resource Description Framework) form [23]. The fundamental RDF data structure is the *triple*, an assertion of the form:

```
subject predicate object
```

where subjects and predicates are represented by URIs (Uniform Resource Identifiers) and objects are represented by URIs or typed literals. For example:

```
udfrs:u1f46 rdfs:type udfrs:FileFormat .
udfrs:u1f46 udfrs:udfrIdentifier "u1f46" .
udfrs:u1f46 rdfs:label "Broadcast WAVE, version 0" .
```

which collectively assert that the RDF resource represented by "udfrs:u1f46" is a UDFR file format, that it has the UDFR identifier "u1f46", and that it represents the format, Broadcast WAVE, version 0.

Note that RDF has a number of different serialization formats. The example above is given in Turtle (Terse RDF Triple Language). Other equivalent forms are possible.

Linked Data is "a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF" [11].

UDFR User's Guide Page 42 of 48







Appendix B UDFR implementation

The UDFR is implemented by a stack of open source products, primarily the OntoWiki semantic wiki, http://ontowiki.net/Projects/OntoWiki, developed by the Agile Knowledge Engineering and Semantic Web (AKSW) research group at the University of Leipzig, http://aksw.org/. The other main components of the technology stack include:

- RDFauthor. RDFa JavaScript library, http://aksw.org/Projects/RDFauthor [20].
- Erfurt. RDF API, http://aksw.org/Projects/Erfurt [18].
- Zend. PHP application framework, http://framework.zend.com/ [24].
- PHP. Server-side scripting language, http://www.php.net/ [15].
- Apache web server (httpd). http://projects.apache.org/projects/http-server.html
 [1].
- Virtuoso. RDF quadstore, http://www.openlinksw.com/dataspace/dav/wiki/Main/VOSRDFWP [14].
- Noid (Nice opaque identifiers). https://wiki.ucop.edu/display/Curation/NOID [4].

UDFR User's Guide Page 43 of 48







Appendix C UDFR class structure

The two main class hierarchies of the http://udfr.org/onto# (udfrs) ontology are rooted at "Abstract Base" and "Controlled Vocabularies". The "Abstract Base" hierarchy defines the main properties used to express UDFR format representation information (see Table 3).

```
Abstract Base >
      Abstract Product >
             Abstract Format >
                    Compression
                    Encoding
                    File Format
             Document
             Hardware
             Media
             Software
      Abstract Signature >
             External Signature
             Internal Signature
      Agent
      Assessment
      Byte Sequence
      Digest
      File
      Grammar
      Holding
      IPR
      Identifier
      MIME
      Process
```

Table 3 – Abstract Base class hierarchy

The "Controlled Vocabulary" hierarchy defines controlled enumerations used as the values of various representation information properties (see Table 4).

UDFR User's Guide Page 44 of 48







Controlled Vocabulary >

Administrative Status Type

Agent Type

Assessment Type

Availability Type

Basis Facet Type

Byte Order Type

Composition Facet Type

Compression Lossiness Type

Constraint Facet Type

Country Code

Digest Type

Disclosure Type

Document Intent Type

Document Type

Domain Facet Type

File Purpose Type

File Type

Form Facet Type

Format Role Type

Genre Facet Type

Grammar Type

Hardware Type

IPR Type

Identifier Namespace Type

Individual Agent Role Type

Language Code Type

Legal Jurisdiction Code Type

Media Access Type

Media Type

Media Write Type

Process Type

Programming Language Type

Role Facet Type

Signature Obligation Type

Signature Position Type

Signature Type

Software Type

Transform Facet Type

Table 4 – Controlled Vocabulary class hierarchy

UDFR User's Guide Page 45 of 48





References

- [1] Apache Software Foundation, Apache HTTP Server http://projects.apache.org/projects/http server.html>.
- [2] Beckett, David, and Tim Berners-Lee, *Turtle Terse RDF Triple Language*, W3C Team Submission, March 28, 2011 http://www.w3.org/TeamSubmission/turtle/>.
- [3] California Digital Library, California Digital Library (CDL) < http://www.cdlib.org/>
- [4] California Digital Library, *Noid: Nice Opaque Identifier (Minter and Name Resolver)* https://wiki.ucop.edu/display/Curation/NOID>.
- [5] California Digital Library, *Unified Digital Format Registry (UDFR)* < http://udfr.org/>.
- [6] California Digital Library, *University of California Curation Center (UC3)* http://www.cdlib.org/uc3>.
- [7] Google, Welcome to Google App Engine < http://www.appspot.com/>.
- [8] Harvard University Library, Global Digital Format Registry (GDFR) http://gdfr.info/>.
- [9] Internet Assigned Numbers Authority, *MIME Media Types* http://www.iana.org/assignments/media-types/index.html>.
- [10] Library of Congress, Digital Preservation < http://www.digitalpreservation.gov/>.
- [11] Linked Data Connect Distributed Data across the Web http://linkeddata.org/>.
- [12] *Media-types* < http://mediatypes.appspot.com/>.
- [13] National Archives [UK], *PRONOM* http://nationalarchives.gov.uk/PRONOM/Default.aspx>.
- [14] Open Link Software, *Virtuoso RDF*http://www.openlinksw.com/dataspace/dav/wiki/Main/VOSRDF
- [15] PHP Group, PHP Hypertext Preprocessor < http://www.php.net/>.
- [16] Prud'hommeaux, Eric, and Andy Seaborne, eds., SPARQL Query Language for RDF, W3C Recommendation, January 15, 2008 < http://www.w3.org/TR/rdf-sparql-query/>.
- [17] University of Leipzig, Agile Knowledge Engineering and Semantic Web (AKSW)

UDFR User's Guide Page 46 of 48







<http://aksw.org/About>.

- [18] University of Leipzig, *Erfurt* < http://aksw.org/Projects/Erfurt>.
- [19] University of Leipzig, OntoWiki http://ontowiki.net/>.
- [20] University of Leipzig, RDFauthor < https://github.com/AKSW/RDFauthor>.
- [21] Wikipedia, Linked data < http://en.wikipedia.org/wiki/Linked data>.
- [22] World Wide Web Consortium, *OWL 2 Web Ontology Language: Document Overview* http://www.w3.org/TR/owl2-overview/>.
- [23] World Wide Web Consortium, Resource Description Framework (RDF) http://www.w3.org/RDF/>.
- [24] Zend Technologies, Ltd., Zend Framework < http://framework.zend.com/>.

UDFR User's Guide Page 47 of 48