

Sustaining the UDFR
DigitalPreservation 2012 Session
July 25, 2012
1:45 – 2:45 pm

Presenters: Stephen Abrams, UC Curation Center, California Digital Library

Attendees: 16

Highlights

Background

Uniform Digital Format Registry (UDFR)—“a semantic registry for digital preservation”

- Purpose was to unify the function and holdings of PRONOM and GDFR and make it available in an open source platform
 - PRONOM: <http://www.nationalarchives.gov.uk/PRONOM>
 - JHOVE: <http://hul.harvard.edu/jhove>
 - GDFR (Global Digital Format Registry): <http://gdfr.info>
 - UDFR: <http://udfr.org>
- We’re trying to manage representation information about various formats
 - What you need to know about something in order to exploit that thing meaningfully
 - Information that lets you answer important preservation questions

UDFR Specs

- Semantic platform lets anyone say anything about anything
 - Understandable to both people and machines
 - The web will soon be semantic
- Facilitates self-configuration and easy extension
- Open contribution: self-registration but no further barriers
- Complete change history at the assertion level
- Technical review
- Four Roles
 - Consumer: anonymous read
 - Contributor: read and write (self-registration)
 - Reviewer: read, write and review (administratively granted)
 - Administrator: read, write, review, and administer

The main part of the technology stack is OntoWiki (<http://ontowiki.net>)

UDFR Schema

All ontologies and code managed in public repositories at GitHub

(<https://github.com/UDFR>)

Initial Data Loads

- Complete export of PRONOM database in February 2012
- MIME types from Appspot

PRONOM data contributed under UK Open Government License. Other submissions contributed under Creative Commons Attribution License

User Interface

- UI layout: paned organization
- Contextual menus available by hovering over links and clicking on the “snowflake/graph”
- User’s guide available

Discussion

Q: Is it synchronizing in one direction? Is PRONOM interested in digesting info from UDFR? Parent-child or sibling relationship?

A: We don’t know yet. UDFR is open contribution. PRONOM is a very controlled contribution. On an interim basis, definitely PRONOM will continue.

Q: Is the assignment of identifiers in PRONOM more granular?

A: We have 10 identifiers for 10 versions of Word, but they have 10 identifiers for the .doc extension.

General Comment: The long-term sustainability of these tools is questionable because the funding ends soon. Enormous issue.

Action Items

- CDL will continue to host the UDFR for one year while a more permanent hosting strategy can be identified
- The “admin” role is held by CDL staff. Find ways to share this responsibility
 - Adding user privileges
 - Modifying the ontologies
 - Bulk imports
- Find ways to share the “committer” responsibility
- Technical development
 - Synchronization with PRONOM and other external sources of bulk imports
 - UI enhancements to provide low-barrier learning curve
 - RESTful API in addition to SPARQL endpoint
 - Replication to mirror sites
 - Others?
- Bring under the OPF code repository/issue tracking umbrella
- Import additional data resources
 - LC Sustainability of Digital Formats
 - IT History Society hardware database

- NIST NSRL (National Software Reference Library)
- Stanford CPUdb
- TOTEM (Trustworthy Online Technical Environment Metadata) database
- Other candidates?
- Encourage adoption and use
 - Identify an evangelist
 - Marketing/outreach
 - Cf. Chris Rusbridge's blog posing the question, "What was the problem" that UDFR was trying to solve?
<http://unsustainableideas.wordpress.com/2012/07/04/the-solution-is-42-what-was-the-problem/>
- Enable the reviewer function
 - Who will review?
 - What are the criteria?
- Sustainable community governance
 - Who will make the decisions?
- To provide feedback or subscribe to the listserv, email Leslie Johnston at LC