Unified Digital Format Registry (UDFR)
Community Meeting: Understanding the UDFR System and Service
Washington, DC
May 4, 2012

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. The UDFR builds upon and "unifies" the function and holdings of two existing registry solutions: PRONOM, from the UK National Archives; and GDFR (Global Digital Format Registry), from Harvard University. While these services rely on older relational database technology, the UDFR uses a semantic database in which all information is represented in RDF form and exposed as Linked Data. 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) at the California Digital Library (CDL) with funding from the Library of Congress as part of its National Digital Information Infrastructure Preservation Program (NDIIPP). Since the UDFR is an important resource available to the international preservation community, the members of that community should take collective responsibility to participate actively in the long-term administration, operation, maintenance, and enhancement of the service. Towards that end, UC3 and the Library of Congress are sponsoring two open community meetings focusing on (1) an in-depth technical review of the UDFR system and service, in order to expand the community of knowledgeable developers; and (2) building a sustainable administrative, operational, and governance structure.

The first meeting, providing a thorough review of the technical aspects of the service, will be held in Washington, DC, on May 4, 2012, in conjunction with the International Internet Preservation Coalition (IIPC) General Assembly meeting on April 30-May 4, 2012 <http://www.netpreserve.org/events/2012ga.php>. The meeting is open to all interested members of the preservation community; space is limited, however, and prior registration is required. The meeting will include technical presentations on the UDFR architecture and code walkthroughs of the major components of its open source

technology stack, including OntoWiki, RDFAuthor, Virtuoso, Zend, PHP, Apache, and Noid; code walkthroughs; and a review of the four main ontological models: OntoWiki system configuration, UDFR user profiles, UDFR class and property ontology, and UDFR data.

The tentative meeting agenda is:

Time	Topic
09:00 – 09:10	Introduction and review of outcomes and agenda
09:10 – 09:30	Background on the UDFR project
09:30 – 10:00	Demonstration of main features of the UDFR service
10:00 – 10:30	Technology stack and architecture
10:30 – 10:45	*Break*
10:45 – 11:45	Code walk-through
11:45 – 12:00	Questions and discussion
12:00 – 13:00	*Lunch*
13:00 – 13:45	Ontological models
13:45 – 14:15	Administrative procedures, account management, bulk data loading
14:15 – 14:45	Community building
14:45 – 15:00	Questions and discussion

This community meeting is generously sponsored by the Library of Congress.