

UNIFIED DIGITAL FORMATS REGISTRY (UDFR)

Proposal and roadmap

March, 2009

Proposal. There are two major efforts underway to create a format registry with complimentary strengths and weaknesses. PRONOM, created by The National Archives (TNA) in the UK, has a strong technological base, and has been building a database of original information about various digital formats. PRONOM at this point however is owned and maintained by a single organization, making it vulnerable to changes in that institution. The Global Digital Formats Registry (GDFR) effort, hosted by Harvard University, has developed a model for a registry based on shared governance, cooperative data contribution, and distributed data hosting. However, GDFR is technically less far along in development, and has not yet begun database building.

Given the paucity of resources in the digital preservation community it would be highly unfortunate if these efforts were to compete for resources. Therefore a group of involved and interested institutions have agreed to join together to create a single shared formats registry drawing on the individual strengths of the two existing efforts. The initiative would:

- * be technically based on the existing PRONOM system and database;
- * create a community governance model for the registry involving all institutions willing to contribute to its development;
- * develop a mechanism for the distribution of the registry data in such a way as to support local extensions and additions to the database;
- * develop both technical and organizational support for distributed input to the registry, including some form of quality vetting of contributed data.

Process. An interim governing body for UDFR would be established, composed of a small number of institutions already involved in registry planning. This body would be responsible for:

- * drafting a proposal for on-going registry governance and funding
- * drafting a plan for on-going registry operation
- * drafting a proposal for editorial oversight of contributed registry data
- * overseeing the creation of 2 or 3 mirrors of the existing PRONOM 6.2 system and database
- * investigate and set registry policy for intellectual property related to both the proposed registry software and data
- * identifying the scope of the future registry service and thus the scope and nature of any necessary developments or changes to PRONOM to create UDFR
- * identifying how development of PRONOM into UDFR in the directions specified by the Initiative would be funded
- * overseeing the further development of UDFR to support replication, distributed input, and editorial oversight.

Proposals for governance and for editorial oversight would be vetted by the larger digital preservation community. Over the coming 18 months the governing body would work with TNA to develop and change PRONOM into the UDFR system to support the multi-institution model proposed above.

Timeline.

April, 2009

- * Formation of an interim governing body.
- * Establishment of a mirror of the existing PRONOM 6.2 system and database at NARA.
- * Interim governing body initiates investigation of alternative technologies (1) for distributed data input and editing, and (2) for local extensions to file format signatures.
- * Interim governing body issues invitation seeking one or two libraries to establish PRONOM 6.2 mirrors.

May, 2009

- * Interim governing body solicits participation in the process of developing the functional requirements for the first phase of UDFR based on PRONOM

June, 2009

- * Establishment of one or two additional PRONOM 6.2 mirrors in libraries worldwide.

June-September, 2009

- * Interim governing body examines the current and proposed capabilities of the PRONOM system (versions 6.2, 7 and 8), GDFR vision and outputs, and wider requirements to determine which elements of PRONOM should be retained, which deprecated, and which new elements need to be developed for the purpose of the registry initiative and the creation of the UDFR system.
- * Interim governing body develops plan for the first phase of the UDFR system including staff, process and funding means.
- * Interim governing body develops the functional requirements for the first phase of UDFR based on PRONOM.
- * Interim governing body develops governance charter proposal for on-going governance, funding, and operation of the registry.
- * Interim governing body develops proposal for the registry policy for intellectual property related to both the registry software and data

September-October, 2009

- * Interim governing body makes (1) the proposed governance charter, (2) the proposed registry policy for intellectual property, (3) the functional requirements, and (4) the plan for the first phase of UDFR system development publicly available and issues invitation for public comment.

October, 2009

- * Interim governing body initiates development of technical specification for first phase of UDFR system development.
- * Interim governing body begins work on proposal for editorial process for registry contributions

November 2009

- * Permanent registry organization (including new governance and funding models) implemented – takes over any existing work being undertaken by interim body
- * Permanent registry organization reviews the remaining schedule and revises as necessary.
- * UDFR system development 1 is initiated based on technical specification overseen by interim governing body (having obtained funds from permanent organization) - to remove any unwanted existing functionality and support data replication for the purpose of the registry initiative.

December 2009

- * Permanent registry organization issues proposal for editorial process for registry contributions (based on work already started by Interim governing body)
- * Permanent registry organization develops functional requirements for UDFR system development 2

February 2010

- * Permanent registry organization initiates development of technical specification for UDFR system development 2.
- * UDFR system development 1 ready for testing.

March 2010

- * UDFR system development 1 deployed.
- * UDFR system development 2 initiated to support distributed data input and editing and any other identified additional functionality.

June 2010

- * UDFR system development 2 ready for testing.

July 2010

- * UDFR system development 2 deployed.
- * Governed editorial process for registry contributions implemented.