

Global Digital Format Registry (GDFR) Implementation Plan

V1.0 2006-06-05

Scope

- The Global Digital Format Registry (GDFR) will provide sustainable distributed services to store, discover, and deliver representation information about digit formats.
- “Format” is defined expansively as a serialized digital encoding of an abstract information model.
- The 2 year project will be conducted with maximum transparency and will actively solicit review and comment by the international stakeholder community.

Deliverables

- Implementation plan
- Abstract data and service models
- Inter-nodal protocol
- Functional requirements
- Technical specifications
- Reference implementation and documentation for GDFR node
- Editorial process for technical review
- Technical mechanism for editorial process
- Inter-nodal and integration testing
- Policy statement
- Succession plan
- Initial representation information population

Oversight

- Ongoing policy review and oversight will be provided by a Steering Committee (SC) of international library and archive administrators recruited for this purpose.
- Technical review and oversight of the design and implementation phases of the project will be provided by a Technical Working Group (TWG) of international preservation experts recruited for this purpose.
- The operational editorial process for technical review will be staffed by volunteers recruited from among international technical experts.
- Direct oversight responsibility for all project work is provided by the Harvard University Library project manager.

Community

- A project web site will be established at gdfp.info (or formatregistry.org?).
- All relevant documents concerning the design, implementation, and operation of the GDFR will be publicly available on the project web site.
- All documents will be classified into:

- Working draft – an internal working document of a GDFR committee or working group that is subject to change at any time.
- Final draft – a final form document released by the appropriate committee or working group for public review and comment.
- Recommendation – a final form document approved by the appropriate committee or working group.
- Dated Final draft and Recommendation documents will never be changed.
- All public documents will display the following information (based on W3C practice):
 - Title
 - Date
 - Status
 - URI
 - URI of current version
 - URI of previous version (optional)
 - Change control?
- The project web site will conform to XHTML 1.0, CSS 2.1, and the WCAG 2.0 working draft.
- A public community forum will be provided (listserv, wiki, blog?).

Development goals

- To the extent possible without compromising other project goals and schedule milestones, development efforts should first be expended to create a generic registry system that is then specialized to the GDFR application.
- The reference node implementation will exhibit good service-oriented architecture design.
- The underlying storage and network transport technologies will be easily swappable, i.e. the design should include appropriate abstraction layers to isolate technological dependencies.
- The full information content of the GDFR will be expressible in XML form.
- The full information content of the GDFR will be re-instantiatable from its external XML expression.
- All technical components should be design with extensibility in mind.

Development standards

- All code newly produced for the GDFR project will be open source under the GNU LGPL license.
- Third-party code incorporated into the GDFR code base (excluding pre-existing OCLC software) will be open source, preferably also under the GNU LGPL license.
- The implementation language is Java 1.5.
- The project will establish coding standards for object-oriented modeling granularity; class, method, and field naming; and source code Javadoc.

Data model

- The data model should be consistent with the PRONOM 4.0 data model.
- A complete instance of the GDFR data model for a particular format is referred to as a record.
- A record is uniquely identified by an “info” scheme URI
- There are three (four?) primary classes of records:

- Local – no further distribution (or perhaps to distribution up, but full distribution down: this supports the case, for example, of domain-specific “gateway” nodes, e.g. ocean science)
- Non-vetted – submitted to the root node for unmediated global distribution
- Vetted – submitted to the editorial process for verification
- The data elements of a record fall into the following categories:
 - Identifiers – canonical and aliases
 - Statement of responsibility – authorship, ownership, maintenance
 - Classification – format taxonomy
 - Specifications – bibliographic description, identifiers, actual contents for public domain, actionable pointers; have facility for recording a local shelfmark for locally acquired copies of documents
 - Signatures
 - Grammar – fully-typed to permit parallel grammatical notations
 - Relationships – taxonomy including subtype, version, encapsulation, affinity
 - Tools
 - Assessment – fully-typed to permit parallel methodological expressions
- Complete record-level change history to support the recovery of arbitrary states in time.
- Descriptive element values will be drawn from controlled vocabularies, which are themselves updatable and distributable via the inter-nodal protocol.

Service model

- A GDFR node will provide discovery and delivery interfaces for humans and machines.
- The human interfaces will be HTML, conforming to XHTML 1.0, CSS 2.1, and the WCAG 2.0 working draft.
- The machine interfaces will be SOAP, REST, and Java APIs.

Inter-nodal protocol

- A presumptive model for the inter-nodal protocol is OAI-PMH.
- Each child node will be assigned a primary parent node from which it will routinely synchronize its holdings.

Node reference implementation

- The underlying storage technology is ? [RDBMS, XML DB, RDF DB, LDAP?]
- The design and implementation will incorporate community best practices for security, thread safety, and performance.

Editorial process

- A presumptive model for the editorial process is the IETF RFC process: initial public comment period, iterative review by area editors, and final approval followed by instructions to the root node to release the new information for distribution.

Policy

- Participation in the GDFR network requires an agreement between the new child node and an existing parent node.
- The specific terms of this agreement are a matter of local policy of the parent node.