

# Global Format Registry Service Architecture

March 11, 2003 – MS

Following the emerging practice of identifying required system functionality in terms of “services” according to the Web services architecture paradigm, the following high-level services have been identified from the available use cases. Detailed service profiles and service models will be developed at a later stage of development of the registry, and taking into account more of the use case details. Some services were not explicitly identified in the use cases, but were implied to exist (e.g. a human review process for new submissions to the registry) and these have been added explicitly.

The following services fall into four categories: those required to build and maintain the registry itself (*management services*), those that retrieve information from the registry (*access services*), those that manage quality control of digital objects (*representation services*), and those that process objects (*brokerage services*). Not all of these services should necessarily be implemented by the registry, but could be (see discussion below). Terminology used here for services is just for discussion and not meant to be normative.

## *Management Services*

1. **Monitoring service** (Leeds)  
Perform environmental scans and other operations to detect new formats, new versions of formats, format obsolescence, etc.
2. **Approval service** (implied)  
Perform review and approval of newly submitted formats, changes to existing formats, or deletion of formats.
3. **Maintenance services**
  - a. **Addition service** (JSTOR, MIT, OCLC, PRO)  
Add representation information to the registry for new format following convention/approval process
  - b. **Update service** (OCLC, Leeds, Penn)  
Update registry with new format representation information or format services (local or external)
  - c. **Deletion service** (for completeness)  
Delete entry from registry (e.g. if superseded or duplicate)
4. **Notification service** (JSTOR)  
Notify registry clients of new/updated format or of format trigger events (e.g. obsolescence, new transformation service, etc.)
5. **Introspection service (from data model)**  
Determine local policies (scope, coverage, implemented services, etc.) of a given registry to identify appropriate registry to use

## *Access services*

1. **Description service** (OCLC)  
Registry returns representation information for single format on request
  - a. By supplied ID
  - b. By search query
  
2. **Export service** (JSTOR, OCLC)  
Registry exports its entire contents to a repository, or subset selected by various criteria (e.g. recent additions and updates, query match, etc.)
  - a. Periodic broadcast from registry to repository
  - b. Harvest of registry by repository

## *Representation Services*

1. **Identification service** (Harvard, OCLC)  
Determine format of a specific DO by comparing its attributes to the attribute profiles retrieved from the registry
  
2. **Validation service** (Harvard, OCLC, PRO)  
Verify format of a specific DO by comparing its attributes to the attribute profile retrieved from the registry for that format.

## *Brokerage Services*

1. **Rendering service** (MIT, OCLC, Penn)  
Identify current rendering conditions for supplied DO
  
2. **Transformation service** (OCLC, PRO)  
Convert DO from current (source) format to target format
  
3. **Metadata Extraction service** (PRO)  
Registry returns information supporting automated extraction of attribute metadata from a DO of a specific format

For each of the Access and Brokerage Services two distinct approaches have been described in the use cases: the registry returns data to support a process within the local repository (e.g. validation, transformation, metadata extraction), or the registry accepts a DO and performs the process itself, returning the result.

There seems to be an assumption by most participants that repositories will handle processes performed on DOs locally, and that the registry merely returns requested representation information to enable these. However it has been pointed out that if these processes are functions of the registry, or of other externally available service known to the registry, it would save many repositories the effort of developing them locally. It is probably the case that not all repositories would want to rely on the registry for the critical processes underlying their digital preservation service agreements. However, in this world of Web Services it bears considering that the registry could itself offer to perform processes, optionally, for common functions like format identification and validation, and even more complex rendering and transformation services.

Therefore, an initial scoping decision for the registry will be the extent to which it a) performs services directly on Digital Objects, and b) records the existence of external services for digital object access and preservation, or c) just records the representation information for each format and returns it on request.