High-Level Requirements for a Bit Preservation System
University of Maryland Libraries

Final revision date: October 4, 2013

Authors:
Babak Hamidzadeh, Associate Dean for Digital Systems and Stewardship (babak1@umd.edu)
Jennie Levine Knies, Manager, Digital Programs and Initiatives (levjen@umd.edu)
Ben Wallberg, Manager, Software Systems Development and Research (wallberg@umd.edu)

Mission

Provide digital content management services in all phases of content's lifecycle, including selection, creation, acquisition, and disposition.

Scope and Assumptions

- Unit of information is a file.
- Directories can organize files.
- Semantics of what is included in a file (content, virus, encryption, intellectual content, etc.), the file format, and how files are semantically interrelated are out of scope.
- Services include preservation and access to digital files.
- "Persistent" in the scope of this document means as long as a file or its name exists.
- The System will not take action to recover corrupted files without informing appropriate users.
- The System will define a maximum allowable file size.

Designated Community

- Internal: producers, collection managers, administrative managers
- External: producers

Requirements

1. Identity

File properties to ensure that a file can be traced and retrieved uniquely, consistently, and persistently over time.

a. The following characteristics will allow for identity management of files
    i. Pathnames
    ii. Filenames
    iii. Ownership
    iv. Access Control
b. The System will assign unique and persistent identifiers to all files
c. The System allows an owner of a file to change its filenames and pathnames

2. File Fixity

File fixity is the property of being constant, steady, and stable. Fixity checking is the process of verifying that a digital object has not been altered or corrupted.

a. Check file fixity on ingest if it is provided
b. Create fixity info if it was not provided at ingest
c. Check fixity of content at fixed intervals
d. Maintain logs of fixity info; supply audit on demand
e. Original filesystem properties for a file, such as relative pathname and last modification date should be faithfully preserved as metadata along with the file
f. File content should not be modified by compression or headers
g. Ability to detect corrupt data
h. Check fixity of all content in response to specific events or activities
i. Ability to replace/repair corrupted data

3. Versioning

Versioning recognizes that digital files may intentionally change over time.

a. The System will provide a versioning service
b. The System will provide a deleting service
c. Ability to version and delete is determined by use

4. Submission Interfaces

A submission interface is a tool that allows for deposit or ingest of files into the bit preservation system.

a. The submission interface will facilitate submission of a file or set of files into the bit-preservation system and provide an acknowledgment for the submission
b. The submission interface may be command line, GUI, web interface, or integrated into existing systems (i.e. Fedora/DSpace)

5. Access

Access refers to the process of retrieving or disseminating content from the System.

a. The System will allow for direct human user access (dependent upon use)
b. The System will allow for indirect programmatic access through applications or an interface

6. Authentication/Authorization

Authentication verifies a user's identity and authorization verifies what a user is allowed to do within the System. The combination of the two allow for security within the System.

a. The System will accommodate different levels of access
b. The System will require a registration profile prior to deposit

7. Disaster Recovery

Disaster recovery refers to the process, policies and procedures related to the recovery or continuation of the System in the case of system failure.

a. The System will be backed up regularly in such a way as to allow restoration of files

8. Reporting/Audit Trail

An audit trail provides a report of the System's status, as well as history of the actions taken, providing information related to fixity, submission, identity, etc.

a. The System will have the ability to generate reports on the status of its contents
b. The System will generate an audit trail indicating if files are damaged or lost