Audit

Audit

Name:	Audit
Version:	1.0 (2013-02-04)
Status:	Production
Specification:	Merritt Audit Service
Download:	Not available
More information:	Curation home page

UC3 staff presented a webinar on the Merritt Audit service on Thursday, March 7, 2011. The presentation slides and a recording of the voice/web stream are available.

The webinar situated the Audit service in the context of a comprehensive program for pro-active preservation management, such as is provided by the Merritt repository. Merritt currently provides robust solutions for persistent identification and storage, fixity, replication, and access. UC3 staff are actively working on additional user-facing services for content characterization and enhanced discovery and delivery, which will be provided by integration with the CDL Publishing Group's open source XTF platform (http://www.cdlib.org/services/publishing/tools/xtf/). UC3 staff are also moving ahead with plans for transformation and annotation services. All of these activities will be the focus of future webinars.

The webinar presented information on version 2 of the Audit service, which was then under active development. (All content currently in Merritt was subject to fixity verification using an earlier version of the service. Version 2 built upon the experience we have gained through this process.)

The Audit service is intended to provide a high level of confidence in the authenticity of managed digital resources. In other words, it is concerned with verifying the a given unit of digital content conforms to a known, and trusted, state. This is an important consideration in view of the multitudinous threats and risks that digital content is subject to, including media degradation, software or hardware failure, natural disasters, and inadvertent or malicious human behavior.

One of the key assumptions behind the design of the Audit service is that the content that may be subject to periodic fixity verification may be managed in a variety of services and systems, including, but not limited to, Merritt. So an important design decision was to represent the unit of verification (what we refer to as the "*item*") by a URL, which can point to an arbitrary web-accessible location. (The service also accepts "file" scheme URLs that reference content on a physically-attached file system.) Each item is associated with a known size and message digest value. Supported digest types include: Adler-32, CRC-32, MD2, MD5, SHA-1, SHA-256, SHA-384, SHA-512,. (UC3 recommends the use of SHA-256, which provides a reasonable balance between computational efficiency and cryptographic security.)

The status of a given item can be reported as:

- *Unverified* (verification has not yet been attempted)
- *Verified*
- *Size mismatch*
- *Digest mismatch*
- *Unavailable* (the item cannot be retrieved via its URL)

At the completion of a full iteration of item verification, UC3 staff will receive a summary of an fixity errors that have been detected. The summary fixity status of all Merritt collections will also be available on the collection landing page in the Merritt UI.

Although the service will be implemented to take advantage of multi-threading, the time to process all items will increase as the number of items increases. Since a complete verification of all items will be a long-running process, service can be suspended and restarted gracefully, to accommodate production backup and maintenance activities.

The campus UC3 partners participating in the webinar engaged in a discussion of the service following the formal presentation. Among the topics discussed were:

- Since items are represented by URLs, the service provides a means for URL link checking as a side effect of fixity verification.
- A comparison of the service with the UMIACS ACE tool being investigated by the UCSD/SDSC Chronopolis service (https://wiki.umiacs.umd.edu/adapt/index.php/Main_Page); The scope of the Merritt service is intentionally simpler than ACE. One of the things that ACE does provide, and that should be added to the Merritt service definition, is a means to verify the stored item information (size and digest values) to guard against verification spoofing in (the unlikely) case of a security breach on UC3 servers.

Other areas in which UC3 is eagerly seeking feedback from the UC3 community are:

- What is the appropriate periodicity for fixity verification?
- What types of reporting is desirable by content owners and curators?
- Are there real use cases for placing content external to Merritt under fixity control?

- Is there an interest in deploying the Fixity service locally on campuses?