

Reporting Functionality

Event Types

Library of Congress preservation events vocabulary: <http://id.loc.gov/vocabulary/preservationEvents.html>

- capture
- compression
- creation
- deaccession
- decompression
- decryption
- deletion
- digital signature validation
- fixity check
- ingestion
- message digest calculation
- migration
- normalization
- replication
- validation
- virus check

Some of these might typically occur in a single action (e.g., ingestion and message digest calculation), and some of these events might not be applicable to F4 (e.g., compression, decompression, and decryption). There may also be other events that users may want to track, such as adding and updating datastreams and modifying properties.

Usage Statistics

Because caching can result in users being served without any repository access occurring, it's difficult to track usage at the repository layer. So usage reporting is probably best handled by commonly-available analytics packages that track client requests in the browser. Front-end webserver logs can also be used.

That said, there is a lot of interest in usage statistics. Typical reports track usage over time, broken down by collection, format, and contributor. This is compared to storage usage. Reports also commonly list the most popular objects.

Repository Statistics

The basic statistics of repository contents are the number of collections, the number of objects, the size of data stored, the number of user sessions and the number of user queries performed.

In addition to reporting the total for the entire repository, it is also desirable to break these down by collection, contributor, format, etc.

Other Queries

Another type of report that would be useful is the ability to perform format-specific queries and perform statistics on the results. Examples are summing the total hours of video, or counting the number of video files broken down by codec.