

Fedora Migration Validation Tool

- [Fedora Migration Validation Tool Architecture](#)

Brief outline of specifications to validate a Fedora 3 to OCFL migration. The validation tool should begin with coarser tests, then progressively handle finer-grained tests.

TBD: specify the verifications to be performed as a runtime parameter (list)?

To be determined: validation of external and redirect datastreams?

Objects

Validate: number of objects (optional validation enabled by command line flag) ✓

Valid: number of objects in the OCFL repository is equal to the number of objects in the Fedora 3 repository

Validate: object IDs ✓

Valid: every object in the OCFL repository has the same ID as its corresponding object in the Fedora 3 repository.

Object Content

Validate: object metadata

Valid: The HEAD version of the OCFL `object.nt` metadata (HEAD determined from the object's top-level `inventory.json` manifest) matches the current version of the Fedora 3 object metadata

- `lastModifiedDate` ✓
- `createdDate` ✓
- `ownerId` ✓
- `label` ✓
- `state` ✓

Note that Fedora 3 content models will be verified as part of the examination of the RELS-EXT datastream (lower-level validation).

Validate: list of datastreams ✓

Valid: every datastream listed in the object's top-level (HEAD) `inventory.json` manifest matches the list of current (HEAD) version of the datastreams in the Fedora 3 repository.

Datastream Version Content

Validate: datastream version metadata matches corresponding OCFL version

Valid: The version of the OCFL `<DSID>.nt` metadata matches the same version of the Fedora 3 `<DSID>` metadata\

F3 Type Key = ("I" - Inline, "M" - managed "E" - external, "R" - redirect)

- `size` ✓ : M only
- `mimeType` ✗ (* hold off on this one)
- `state` ✓ (lower priority if it is not trivial)
- `title` ✗
- `identifier (DSID)` ✓
- `lastModified` ✓
- `created` ✓
- `externally referenced content (URI)` ✗ (Type E and R only)

Validate: datastream size

F3 managed (type "M") datastreams ONLY ((ie not inline XML):

Valid: the size of the version of the datastream in OCFL matches the size of the datastream **on disk** in the Fedora 3 repository ✓

Valid: the size of the version of the datastream in OCFL recorded in `<DSID>.nt` matches the size of the OCFL file **on disk** ✗

Validate: datastream checksum

F3 managed (type "M") datastreams ONLY ((ie not inline XML): 

Compare the F6 checksum with the corresponding checksum of the F3 datastream version. The latter will need to be generated using the same algo used for the F6 checksum.

Optional flag to allow users to skip this checksum validating step (as it may be resource/time intensive) : checksumming is turned off by default 

Versions

Validate: object versions (N/A)

To the best of my knowledge, Fedora 3 does not version changes to object metadata. Suggestions for validating changes to object metadata, as opposed to datastream versions, are welcome.

Validate: number of datastream versions

Valid: number of datastream versions for a datastream in an object in OCFL matches the number of versions of the same datastream in the same object in the Fedora 3 repository

Validate: datastream versions

Valid: each datastream version in the OCFL object has a corresponding version in the Fedora 3 repository, as determined by matching the datastream creationDate recorded in the OCFL version <DSID>.nt file to a datastream version creationDate in the Fedora 3 repository

Only validate the most recent version

Command line flag for disabling non-head version checks.

Other Requirements

Report

The results of a validation should provide a report that allows the user to understand an overall summary of the results, lists of error types by count, the list of objects by error type, summary of errors by object and detailed validation logs by object.

HTML 

TSV 

Validate objects in list

The user should be able to provide a list of object IDs to include in a validation routine.

Validate all objects

The user should be able to instruct the validator to validate all objects in the repository.

 Highest priority