

0-effort Ingest

0-effort Ingest

Problem: Getting stuff recognized by our systems take way to long

Current state of most collections: Limbo

Files on disk

Disk is backed up

Files are checksummed

Metadata reside somewhere and referenced in the collection document

New Way

All known collections are created as Bitrepository collections right now.

- This should be easy, as the Digital Preservation Group have catalogued these and decided on preservation levels already.

Files are put in the Bitrepository, rather than Limbo

- This require a generic tool, but such a tool have already been developed for the Yousee workflow.

In addition to checksumming, they are analysed with Tika or Fits or similar.

- This should also be extremely easy, if we do not attempt to validate the files, just characterise them.

Create doms collection corresponding to bitrepository collection

- Create one object, should be easy

Attach metadata sources to doms collection

- We need to determine where to store the metadata. If it is something simple, store it in DOMS. If it is big, we should probably store it in the bit repository and reference it.

Create file object for each file in the collection

- Ie. create the simple datamodel. The datamodel is as follows. Only objects are file objects. These must have a content datastream and a datastream for the characterisation information. If we have metadata for each file separated (ie. separate xml files), it should be added the file objects.

End result

- Collection in Bitrepository
 - Can be monitored by Bitrepository GUIs
- Collection in DOMS
 - Can be edited by DOMS Gui
- Collection metadata stored
 - Easier time finding the metadata at a later day
- Collection part of our preservation infrastructure