

Electronic Records

AXAEM has a significant feature set when it comes to electronic records, but there is also a significant amount of features yet to be added. Here is a summary of what the system currently does:

1. Electronic records can be ingested either from the desktop over the network, or directly from an AXAEM server location, which is useful when the accession is measured in terabytes.
 - a. The user selects a base directory, then the application displays a directory tree of that location, allowing the user to include whole folders or omit certain records. Optionally, the user can indicate the intention of uploading just a single record.
 - b. The ingest process offers the user several options:
 - i. Tie the batch to a bib record (required)
 - ii. Tie the batch to a transfer record
 - iii. Tie the batch to a digitization workorder where metadata is already recorded
 - iv. Run a virus scan
 - v. Run and validate any BagIt bags it finds
 - vi. Extract files from container such as .zip, and allow the user to indicate the disposition of the container file.
 - vii. Run a variety of metadata extractors against the upload
 1. FITS
 2. JHOVE2
 3. Droid
 4. MediaInfo (used to capture data from audio/video files)
 5. Can also run the New Zealand metadata extractor, though option is not selectable in the ingest window currently. The NZ extractor is also bundled with FITS, so running it standalone may be superfluous.
 - viii. Extract descriptive metadata from source files
 1. FGDC metadata as used with geospatial records
 2. Dublin Core as exported from CONTENTdm systems
 - ix. Metadata extractor data is then attached to the electronic object, and metadata fields within AXAEM are updated to contain extractor data.
 - c. Each ingest performs an MD5 and SHA-2 checksum on each file, which is recorded in the database.
 - d. After ingest is complete, a report is auto-generated that displays all the new records, any extractions done, and any error messages received.
 - e. The ingest process creates several new records in the database: one for the physical file, one for the electronic object, which is where the descriptive metadata is contained, and one for the object group, which records folder structures. The electronic object is complex, in that it covers descriptive

metadata for multiple representations of a record. One representation could be one physical file or many that are used as a unit. Multiple representations per electronic object are common, especially as original files are migrated to alternative formats. To use PREMIS terminology, the electronic object is the lowest-level intellectual entity contained in a transfer batch. Object groups contain similar metadata fields as electronic object, and they also allow subgroups to be created, allowing the archivist to describe records at any level regardless of the depth of the file structure.

- f. A batch record is recorded in the database, and audit trail data assigned to it, such as when it was uploaded and by whom. It also records when deleted and by whom, should a batch need to be removed from the server. Batch deletions remove all traces of the files, including metadata in the AXAEM database as well as the file on the server, and the folder created to contain the file on the server.
2. Format types and their associated PUIDs are recorded in AXAEM, and migration rules are associated with them. The format interface allows the user to record the status of the format, links to specifications, and various other notes. To date, migration rules have only been set on a limited number of formats.
3. Transformation tools integrated with AXAEM include Open Office (via a converter utility), and ImageMagic. More will be added as time permits and tools become available. The user can now direct an ingested file to be transformed to some other format, and that new file is then placed in a directory on the server, and associated metadata records added to AXAEM and associated with the different representations of a given electronic object.
4. Checksum compliance can be determined at any time, either file by file by clicking a button, or through queries that run a utility against a set of records and produce a report. Checksum audits can be done in the background if desired.
5. The BagIt tool has been integrated with AXAEM. It can interact with a desktop installation of that tool, providing the user with a graphical interface to create or validate bags, as well as on the server side, where bags are validated as they are ingested. Validation failure results in the contained files being rejected from the ingest process.
6. Records that fail the virus check are also rejected from the ingest process.
7. Electronic record metadata is indexed by the Solr search engine, provided the user has checked the field to “publish”. The search engine then allows the records to be downloaded if an access copy exists. Note that currently, the ingested file or its transformed version are the only copies on the server, and located by default in the place set by the user in parameter files. The system does not yet have the capability to send a copy to some other server, although such capability would be easy enough to add.
8. Among the metadata recorded in the electronic object representations is an indication of whether the representation should be considered the original, the preservation copy, the access copy, or the “record copy”, and whether it is redacted. If access restrictions are set within the bibliographic record, the electronic object may not override that. If the bib record indicates that some material may be open for access, then the electronic object can place its own

- access restrictions at the item level. Unless access restrictions indicate that the record is open or redacted, the “publish” checkbox mentioned above is to remain uneditable.
9. AXAEM offers a place to record the hosted locations of copies, which identifies the institution and its address. Metadata inside the electronic object can be added to link to these locations.
 10. Media types are recorded in the system. Metadata includes manufacturer, model, refresh timetables or conservation actions needed, dimensions, and other notes.
 11. Individual media can be linked to media types, which contain dates specific to that media instance: when manufactured, when purchased, when written, when accessioned; plus media barcode information, disposition information, and indicators of errors from testing.
 12. Metadata can be exported from the electronic object file, edited in a spreadsheet, and then re-imported to the electronic object. This also helps the user reuse metadata for other workflows, such as the creation of online digital collections.

The metadata contained in the Electronic Records portion of the system begin with Dublin Core, but also expand beyond that as more granular information is obtained from metadata extractors and descriptive metadata that comes with some records. Certain file types also produce unique metadata, which have their own fields to contain this information. For instance, when email is ingested into AXAEM as either an .mbox file or .eml, the application will grab header information such as sender, recipient, date sent, and subject line, and record that directly into the database. The other content is stored as an .eml file. Future development will extract attachments from these .eml files and save them as their own electronic object with a relationship to the email record. Other record types have similar unique properties, such as geographic coordinates and projections found in geospatial records, or image details found in scanned records. More fields are added to accept this unique data as we come across them.

Here is a sample list of the metadata recorded:

1. Title
2. Contributors
3. Creators
4. Dates
5. Extent
6. Identifiers
7. Publishers
8. Abstract
9. Coverage
10. Description
11. File Name
12. Folder Paths
13. Format Types
14. Genre Types
15. Genre-Specific Metadata

- a. Email
 - b. Geospatial
 - c. Images
 - d. PDF
 - e. Audio/Video
16. Geographic names
 17. Subjects
 18. Table of Contents
 19. Transcripts
 20. Access Classifications
 21. Hardware
 22. Language
 23. Media
 24. Rights
 25. Software
 26. JHOVE data
 27. Location of copies
 28. NZ extractor data
 29. Droid data
 30. Records transfer documentation links
 31. URL/File paths
 32. Attached extractor metadata files
 33. Related Objects
 34. Sources
 35. Thumbnail Images
 36. Alternate Titles
 37. Checksum
 38. Checksum log
 39. Migration history

Some features have been added to facilitate use of the interface. For instance, to view an electronic record within the client, click a button and a copy of the file will download to your desktop, then auto-launch software that is associated with that file type. If you click this button while viewing an object group, the entire group of files will be placed in a .zip file and downloaded to you. Files that can be viewed in a browser can also be viewed directly in the client, without requiring the user to have additional software.

An additional feature that has not yet been started but is anticipated as a necessary component of the electronic records feature is the ability to wrap ingested records in a METS file and export them somewhere. AXAEM can already convert data from/to base64, and XML is not difficult to generate, so putting this in place is just a matter of time.