Software Development Priorities:
Current Register

Last updated: December 24, 2020
Status: Finalized

TABLE OF CONTENTS

OVERVIEW

This document tracks possible feature enhancements for the open-source software applications used by SFU Archives. These are primarily Archivematica and Access-to-Memory (AtoM). Over time others may be included as the Archives incorporates them into its workflows (e.g. ePADD).

The main purpose of the list is to support planning when the Archives sponsors software development via contract. But the Archives can also draw on the list whenever it has an opportunity to provide input on the development roadmap of one of these software projects.

The Archives uses its GitHub repository to register features as GitHub issues (assigned a number, with brief description, tags, and staff comments). Use the present document to provide fuller feature descriptions, especially for high priority items.

Priorities should be periodically reviewed / reassigned, typically at the end of a development contract or beginning of planning for a new one. The Archives last reviewed and updated it in December 2020, as its' 2019-20 development contract with Artefactual was nearing completion and the Archives considered an extension for January 2021.

HIGH PRIORITY

Archivematica error handling [7]

Support better error handling. There are two main use cases we have in mind:

- We have several processing configs that are fully automated, i.e. archivist starts ingest, all decision points are automated, process completes. Later the archivist confirms completion by expanding microservices to make sure no errors occurred. It would preferable if Archivematica produced a report at the end that showed any errors. (anything else that could / should go into the report? e.g. option implemented at each decision point?)

- During ingest if file format identification fails with the preferred tool, Archivematica should try another. Could be an admin setting to state the order of preference.

Minor AtoM user interface enhancements [8]

Identify and implement a set of relatively minor user interface enhancements. Preference given here to "low-hanging fruit" type issues if they are not complex.

High priority:

- Add a "recently added" widget to the AtoM Home page to display new descriptions (similar to "Popular this week"); provide an admin interface to manage content that appears.

- Support a rotating set of images for the AtoM Home page; provide an admin interface to allow easy upload / deletion.

- Support creation of on-the-fly collections of digital objects (e.g. for temporary exhibits, outreach).

Medium priority:

- Expose the AtoM version number on the Home page; right now only an administrator can access this info (can be useful to know for liaising with AABC and MemoryBC).

- Make treeview sticky: when you save a record, treeview reverts to the top of the tree, losing view of current record's position in the hierarchy.

Low priority:

- Restructure *Related materials*, *Associated materials*, and *Sources* fields as repeatable sets (Identifier, Note, URL); for *Associated materials* the URL can point to holdings in other repositories.

AtoM authority records and access points [25]

Review issues relating to shared authority records across multiple repositories.

High priority:

- AtoM subject terms can be structured hierarchically, so we can set up multiple thesauri (e.g. BC Thesaurus plus taxonomies specific to Archives or Special Collections). But the interface for applying subject terms just gives a flat list of all terms. Idea would be to have an interface that supported faceted filtering to show just terms from a selected parent taxonomy.

Medium priority:

- When exporting EAD, specify which taxonomy terms to include in the export. Use case: when sending a description to MemoryBC, they only want subject access points drawn from BC Thesaurus terms. This feature would let us exclude from the export any subject access points that use terms from an SFU-specific taxonomy.

- Support controls on data entry for personal names, e.g. LastName, FirstName. Inconsistency in data entry leads to duplicate authority records (Braid, Kate and Kate Braid).

- Allow use of previous forms of a name as access points. E.g. Human Resources was formerly the Personnel Office. If a series was created during the time it was called Personnel Office, that should be the form of the name used as the access point; but the link should still point to the same authority record with its current name (Human Resources).

Other possible enhancements (low priority):

- Allow multiple admin histories / bio sketches for one authority (e.g. contributed by different institutions).

- Allow linking of authority records to URIs maintained by other agencies (e.g. link **Simon Fraser** authority record to entries in SNAC, VIAF, LC Name Authorities, Wikipedia etc.).

MEDIUM PRIORITY

Accessions module [21]

Create an AtoM module to implement the Canadian Archival Accessioning Information Standard (CAAIS).

Possible features:

- Support all CAAIS elements.

- Add multi-repository functionality.

- Allow linking AtoM archival descriptions at all levels to an accession – would allow create finding aids by accession (e.g. show all files – which may be distributed across multi series – by accession).

AIP repository manager [11]

Develop repository management functionality for managing objects in the AIP store.

- E.g. reporting, searching, checking checksums, bulk re-normalization based on changes to the FPR, backup and recovery.

First phase could be a research project to determine the best model. Options might include:

- Enhance the Archivematica Dashboard (e.g. make some of the analysis tools from the Appraisal tab also available on the Archival storage tab).

- Enhance Storage Service (e.g. add management functionality).

- Build on existing work with Binder (make it fully open source).

- Build tools to search / report across METS files in the AIP store.

- Build integration with an existing repository management application (e.g. D-Space, Islandora)

- Build the manager as a stand-alone module separate from Archivematica pipelines.

AtoM search enhancements [20]

Review existing search capability in AtoM, identify and implement improvements. Possible features:

- Add Level of description as sort option on search results

- Build admin interface to select fields for indexing (e.g. for us searching on restrictions would be very useful, add PREMIS rights fields to search index)

- Better ranking formula

- Support exclusion of records belonging to certain fonds / series from search results

CSV AtoM import utility [26]

The Archives has an Excel template to capture minimal file / item list data for use when processing records. It needs to be transformed into the standard AtoM csv template (with columns for every AtoM field). Currently RD does this with a FileMaker utility. Ideally we would have a Python-based drag-and-drop app that would do the transformation.

- Map Archives' template fields to AtoM template.

- Develop script.

- Create drag-and-drop interface.

DIP on demand [12]

Provide interface to generate new DIPs from existing AIPs. Currently you can download a stored DIP (generated at AIP creation) and upload to AtoM via a command line script or re-ingest the entire AIP to create a new DIP. Ideally you should be able to upload stored DIPs via the Archivematica Dashboard, target specific files in a given AIP and create DIPs for just those files, and support bulk upload to existing descriptions via csv import.

Disk images [4]

Review practices for ingest of disk images. When should disk images be used (vs copying directories / files only)? Should disk images be ingested or processed prior to ingest? Experiment with BitCurator forensics tools, other reporting tools (e.g. Brunnhilde), assess integration with Archivematica.

Email [13]

Review work to date on email relating to transfer, arrangement and description, and access to email records. Investigate integration of Archivematica, AtoM and ePADD. The project would involve research and policy as well as software development.

Possible elements:

- Review scripts that manage export of email from Outlook as maildir / convert to mbox.

- Extract descriptive metadata from mbox DIP, map to AtoM description fields.

- Integrate email viewer in AtoM.

- Configure IMAP as a transfer source in Storage Service.

- Manage ingest and normalization of email attachments.

ePADD enhancements [31]

Identify features in ePADD that could be improved or created to facilitate use by Archives.

The Archives has never sponsored development with ePADD. But work with ePADD in 2019-20 turned up several issues that were reported to the ePADD dev team:

- Timestamps changing on import.

- Line breaks and formatting sometimes lost on import.

- Original folder structure not retain in ePADD-generated mbox.

Rights management [18]

Review the existing PREMIS rights area in AtoM, identify improvements, especially support for different types of conditional access, each with their own notice screen.

Current functionality:

- You can set different access rules for different rights bases (statute, copyright etc.).

- There are 3 restriction-states for each basis (allow, disallow, conditional).

- A popup notice screen for exists for one use case, where basis = "copyright" and restriction-state = "conditional".

Possible development:

- Allow each rights basis to have its own conditional pop-up notice (rather than hard-coded to copyright).

- Integrate CC license icons

- Integrate standardized rights statements (from RightsStatements.org).

Note that recent AtoM development has created some problems for the existing implementation, e.g. user can now directly download an object via right-click, thus by-passing the rights notice.

Semi-active digital transfers environment [11]

Provide support for semi-active transfers of born-digital records that have long-term retention requirements but may not be archival. Create a separate Archivematica-AtoM environment for semi-active transfers. Build scripts to automate disposals (i.e. destructions and transfers of material from semi-active to archival environment).

LOW PRIORITY

AtoM SQL interface / support [30]

Provide an interface to allow AtoM administrators to interact with the SQL database, e.g. queries, bulk edits.

Bulk extractor enhancements [3]

BulkExtractor is currently available on the Appraisal tab (when transfer is set for _Examine contents_ = "Yes"); it identifies credit card and US social insurance numbers. Build on pattern-recognition capabilities to identify other potential pi.

Collections as data [22]

Provide support for users to work with descriptions and digital objects in AtoM as data.

- E.g. allow users to download all digital objects for a given level of description (e.g. from a fonds, all child objects) or set (clipboard).

- Use case: researchers want to run visualization or data analytics on set of objects.

Format Policy Registry (FPR) [5]

Provide more user-friendly interface to let administrators make changes to the FPR. Look into / test FPR2PAR application for sharing / querying FPR data.

Link AIPs to fonds / collections [1]

Support searching and reporting on AIPs by fonds.

- Currently the Archives relies on its FileMaker AIS database for this (via the Accession and AIP tables).

- Project would build better reporting into the Archivematica dashboard.

- We currently have ca 800+ AIPs, so need to consider how a solution could be applied retroactively.

Probably better to tackle in context of AIP repository manager [11].

Object metadata extraction [14]

Technical metadata embedded in digital objects (e.g. image files) is captured in the AIP, but not easily accessed for descriptive purposes. This project would develop tools to retrieve metadata from the METS file, map to AtoM / RAD fields, populate AtoM descriptive record on DIP upload.

Physical location module [19]

Identify requirements for and build out a physical locations module in AtoM to manage locations, boxes, and storage objects.

Selective de-indexing of digital objects [27]

Support ability to remove indexing from selected digital objects for privacy reasons.

- Use case: individual requests removal of name from public document (e.g. Convocation Book). Record would remain available in AtoM, but individual's name would not be picked up in searches.

- Currently have a manual work-around: manually redact DIP, reingest AIP.

Submission documentation [6]

Review practices for managing relationship between AIP and submission documentation (information about the transfer and its objects).

- What info would we like to capture?

- Where should it live (e.g. inside the AIP)?

- How is it related to the AIP?

- Can it incrementally change?

Transfer analysis and validation tools [2]

Build / integrate tools for better analysis, assessment, and validation of digital transfers. E.g. compare checksums against existing AIP store to identify duplicates, flag password-protected files.

User contributed content [24]

Provide support for AtoM public users to contribute descriptive content. The first phase would likely be a research project to identify and scope out requirements.

Web archiving [15]

Research and develop solutions for capturing, preserving, describing and providing access to archived websites. Investigate use of WebRecorder, ReplayWeb, integration of WARC view with AtoM.

IN PROGRESS / REVIEW AFTER COMPLETION

The following features were sponsored in the Archives' 2019-20 development contract. Artefactual will complete the work in January 2021. Review these features after completion / deployment and for each determine whether to close the issue or keep it open with any follow-up work as required.

Archivematica processing configuration [9]

You can create multiple processing configurations to support different workflows with different decision-point options pre-selected. As of Archivematica 1.9 you can download a custom processing configuration as an xml file, then include it in a transfer to use that configuration. This project would make it easier to use configurations by providing a selector in the interface to switch between processing configurations. Currently you can do this only for AIP re-ingest.

CAS integration [28]

Allow staff to log on to Archivematica and AtoM using SFU's Central Authentication Service (CAS).

Metadata-only DIP upload [10]

Improve Archivematica's metadata-only DIP to AtoM so that it is obvious to the AtoM user when a digital object exists, but is not currently displayed in AtoM (currently this is not clear). There should be a clear distinction between technical metadata for the original file vs metadata for the access copy.

Minor Archivematica user interface enhancements [8]

Enhance Archivematica by making improvements to various UI features. The main focus of the current work is to make the Backlog and Archival storage tabs more consistent with each other in terms of the information displayed by creating a column picker that allows users to select which data elements to display.