Appendix H: Technical Development

1. Functional Requirements for Arrangement and Description

The functional requirements presented here were developed by the AIMS partners over several months of discussion and testing of various tools that can perform various activities within the born-digital archival workflow. The functional requirements are described in 13 overall sections. Within each there may be "Further Questions or Comments" — areas of discussion that were not decided on before the end of the grant period or that required some development work before they could be decided — and "User Stories" — examples of the proposed tool in use in a hypothetical situation. Although these requirements are unfinished and were only partially implemented in the Hypatia demo application (see *Appendix H.3*), the partners present them here so that they may fuel future work in this area.

Functional Requirements for
AIMS Hydra Head ("Hypatia")

A&D_00: Fundamentals

The arrangement and description tool must provide a mechanism which allows an archivist to do the following:

- Define an intellectual arrangement of transferred archival records that reflects the provenance and original order of the records. The original files and directory are not moved or modified in any way.
- Create and edit descriptive metadata for those records. It must also be possible for the archivist to add descriptive data to individual files in addition to adding descriptive data for any of the given levels of arrangement.

Levels of arrangement as defined within archival practice, and accordingly, this tool set includes collection, series, subseries, folder, and item. (see **A&D_12 Overview**)

Each archival collection will have its intellectual arrangement, that is the arrangement of the material in a hierarchical nature that intends to reflect its original creation or arrangement within a recordkeeping system. Over time additional material may be received and these accessions will be integrated into the collection and the intellectual arrangement will be updated. The arrangement is used to portray and distinguish critical elements of context. Software tools like Archivists' Toolkit and CALM allow archivists to create the intellectual arrangement with description based on content standards like DACS or ISAD(G).

Other tools might be used to create exhibitions but any organization of the material for this purpose should not be confused with the intellectual arrangement. A user is able to sort a collection into a particular order that suits them (e.g., by date) via the discovery and access tools.

AIMS partners can supply BagIt-based SIPs, either in directory or zip/tarball form. Rubymatica packages files with technical metadata from FITS/DROID into SIPs.

Further Questions or Comments

The terms used in this document are common within American practice. Arrangement terms used in the UK are collection, sub-fonds, series, sub-series, item [the unit of production; e.g., one file] and piece [pages within a volume or individual letters within a bundle etc]

A&D_01: Graphical User Interface

The arrangement and description tool must have graphical user interface (GUI) and implement and reflect best practices and conventions of user interface (UI) design. The application should operate within a web browser for best cross-platform compatibility. The tool set should be relatively easy to use and should likely reflect user interaction paradigms to which archivists are accustomed, such as those found in applications already in use by the AIMS partners (namely Archivists' Toolkit and CALM). Accordingly, in some cases these functional requirements may refer to other functional requirements, documentation, or specifications as applicable to demonstrate user interfaces requirements. Individual requirements within this document may also explicitly describe specific user interface requirements.

The original organization of the files and directories within an ingested accession and the archivist-defined intellectual arrangement have special status, and that status should be obvious in the UI and should be enforced by the UI. For example, it is essential that users authenticate as an archivist in order to modify the intellectual arrangement. (Keep in mind that a detailed description of collection permissions may not be covered by this document.)

When working on the intellectual arrangement, archivists will need ready access to technical metadata such as the original full path of a given file (see **A&D_02: Technical Metadata**). It may be useful to have a "show original" function within a contextual menu that would show the originally ingested file in the left pane.

Further Questions or Comments

It would be useful to be able to associate digital photographs of media with imported collection components. For example it would be useful, particularly for minimally processed collections, to be able to show images of the source media (floppy disks in particular) alongside the digital files it contains. These photographs must be distinguishable from actual content from the media, possibly via an explicit metadata folder or similar (this could also contain an original 'manifest' and/or web survey information if so desired).

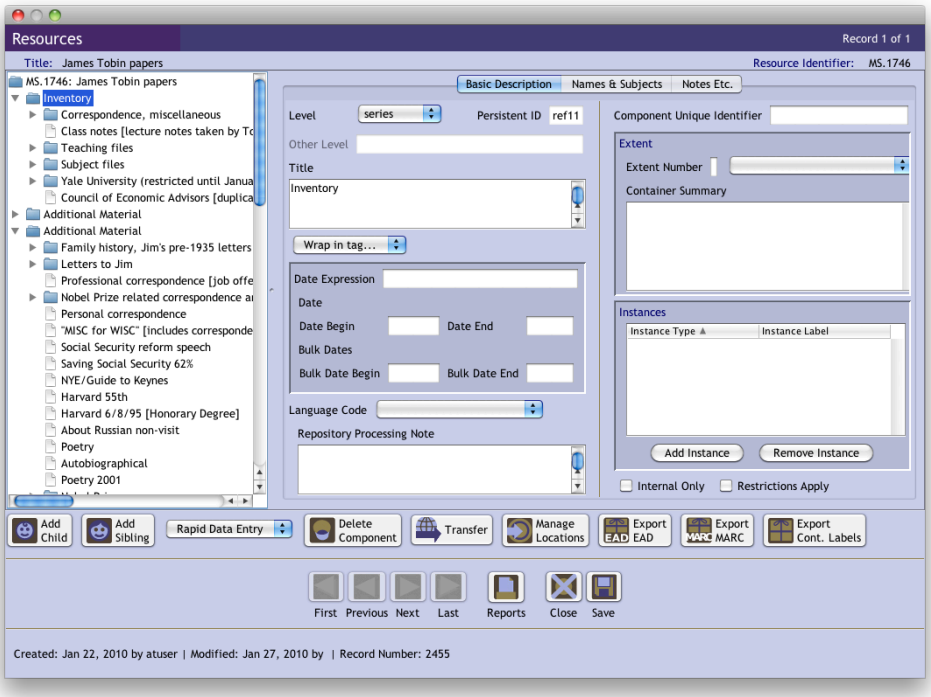
A&D_01.01: Representation and manipulation of hierarchy

The graphical user interface should allow users to view and interact with hierarchical structures representing the intellectual arrangement and the original arrangement of files and directories within ingested accessions. There should be distinct panes representing the structure of the intellectual arrangement and representing the accessions. For each component level in the intellectual arrangement, the user interface should present associated digital assets (see **A&D_01.02** and **A&D_12**) and an interface to view and edit descriptive metadata elements (see **A&D_03.02**).

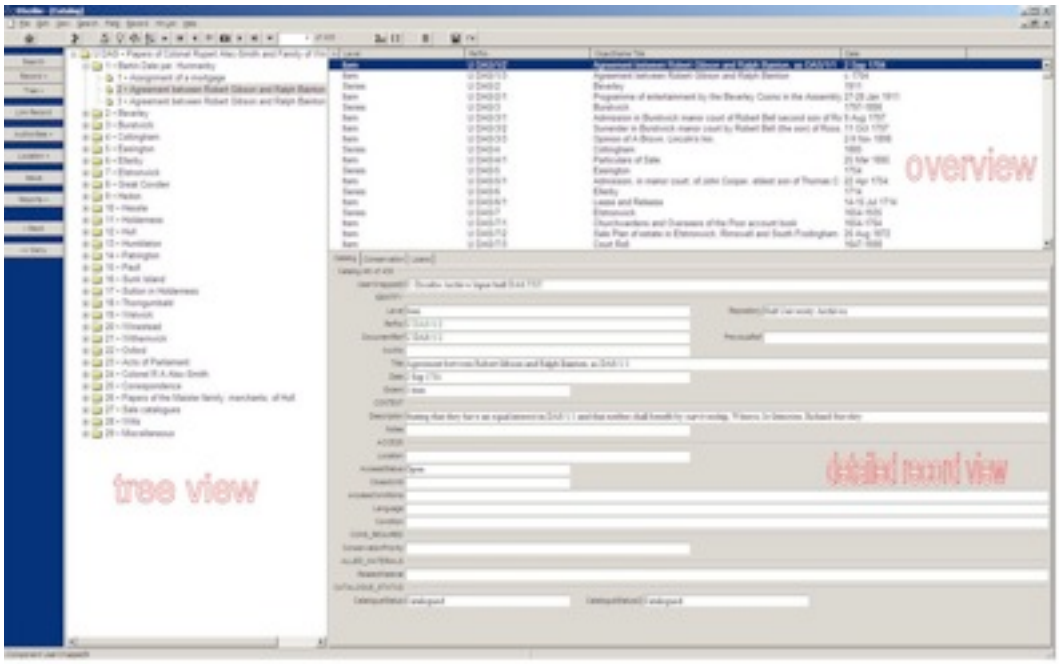
In addition, the tool should allow the following operations (applies to intellectual arrangement only unless otherwise specified):

- Collapse and expand record nodes for viewing (applies to both the original ingest and the intellectual arrangement)
- Add new child record (see **A&D_12**)
- Add new sibling record (see **A&D_12**)
- Copy all or part of an existing structure to the intellectual arrangement. Ideally, we could copy structure of the original ingest, or copy all or part of an intellectual arrangement.
- Delete a record in intellectual arrangement. This applies only to the intellectual arrangement. Recursive folder delete is a dangerous operation, and the UI must add special safe guards. We should be able to delete a record, only if it has no children in order to avoid orphan entries.

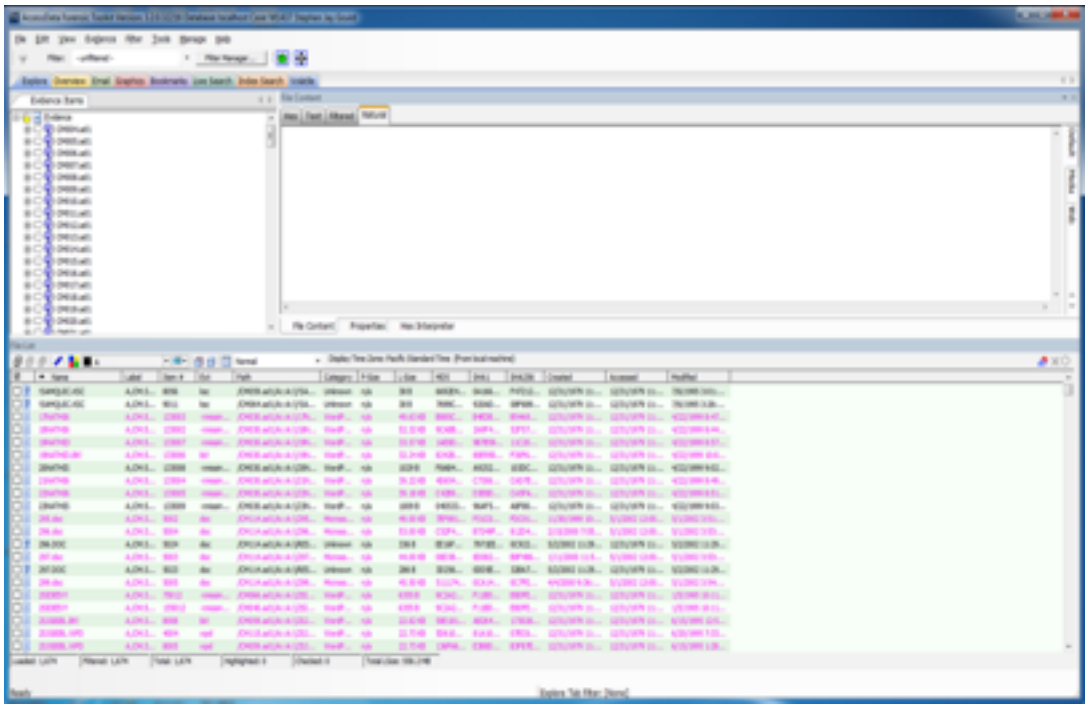
Sample Screenshots



Archivists' Toolkit



CALM



Forensic Toolkit

A&D_01.02: Drag and drop functionality

NOTE: This is heavily interrelated with **A&D_12**. *Please refer to functional requirements in detail below.*

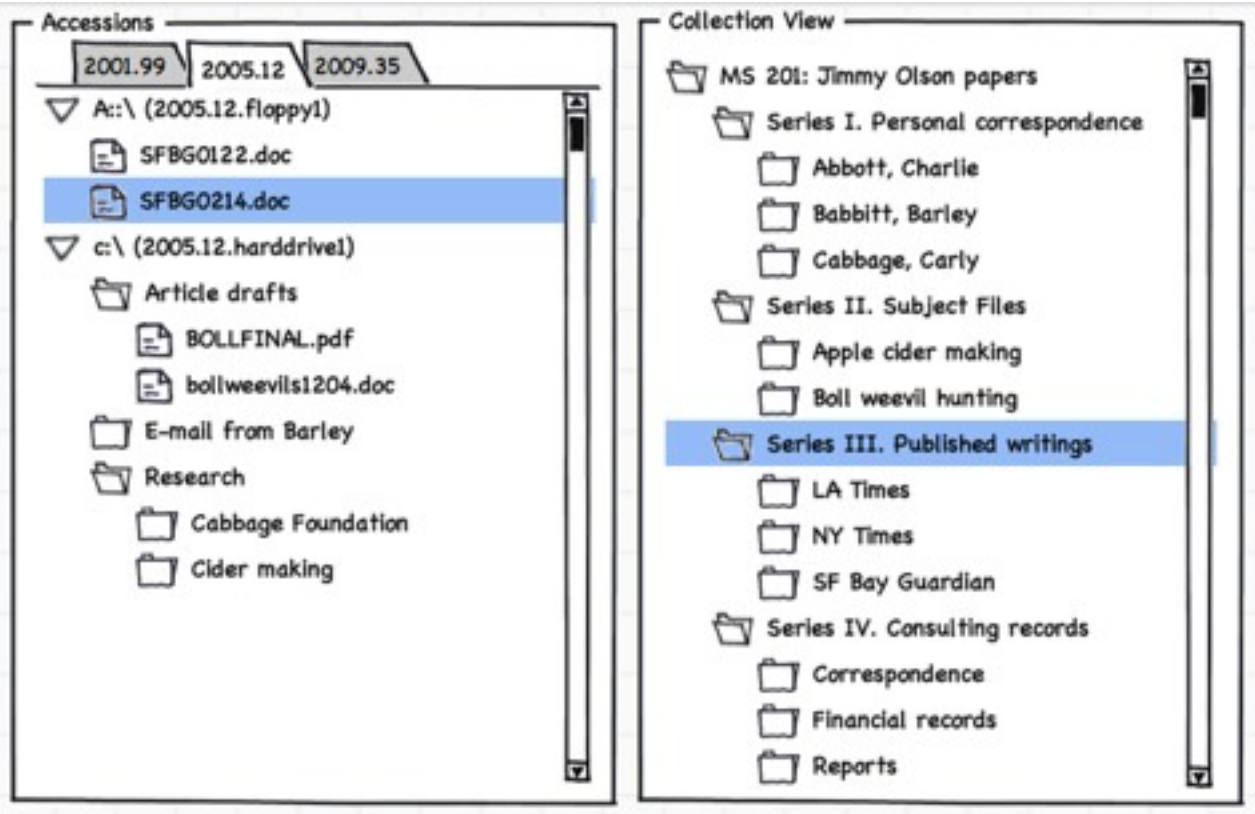
As noted in the Overview above, drag and drop is part of the UI necessary to create an intellectual arrangement for an accession. The original accession is read-only and cannot be modified (with the exception of appraisal actions; see below and **A&D_10**) and represents the original directory structures as they existed within an accession. Dragging a directory, file, or multiple of either to a component in the intellectual arrangement will establish a relationship between those directories and files and that component level.

Component levels also must be draggable to allow for ordering and changing the level of hierarchy. This includes changing the sequence of nodes, promotion and demotion nodes, and auto-renumbering sequences of intellectual units in accord with the modifications.

TomL (Feb 7): Are series ordered by number? Up to this point in the requirements, a programmer would assume "folders" are ordered by the usual rules: date or alphabetic. It is a special requirement that series folders have a numerical sequence.

The UI needs to make it clear which files and folders in the original ingest have or have not been assigned to a component within the intellectual arrangement. Deleting the relationship between a directory or file and the component to which it is assigned should update the status (to "unassigned") as appropriate.

Following the user interface conventions of desktop file managers, the original ingested accessions could be represented in a pane on the left side of the window, and the intellectual arrangement could be represented in a pane on the right side of the window. Files and folder can be dragged from left to right. The left side should be impossible to modify, with the exception of the ability to remove files during appraisal (see **A&D_10.01** below).



A&D_01.03: Sort records

Archivists will need the ability to sort and filter items within the list of ingested files. Both would apply to these fields: full path, base folder, file, time stamp, size, file type (PRONOM PUID). Ideally, we could apply more than one filter and allow filters to at least have and/or logic against other filters. We will probably need to group PUIDs by larger types: text files, word processing document, HTML, XML, various types of data, etc.

A&D_01.04: Copy and paste of hierarchical structure

NOTE: See also **A&D_12.05** *and* **A&D_12.06**

Archivists should be able to copy and paste intellectual arrangement from a number of sources. First, they should be able to copy directory structures from accessions to replicate them in the intellectual arrangement when the directories represent a clearly defined original order. They should also be able to copy existing intellectual arrangements that have been either imported into or created within the tool and paste them into the arrangement pane to duplicate structure as needed.

A&D_02: Technical Metadata (PC)

Overview

The decisions archivists make in terms of appraisal is partly reliant on technical metadata. Technical metadata should

be only viewable and not editable. Technical metadata may also be used to sort records (see A&D_01.03) or in the generation of reports (see A&D_08).

A&D_02.01: File-level technical metadata

The A&D tool should be able to import and provide access (and batch applicable) to the following technical metadata for a given file.

Filename.

Original full file path.

MD5 Hash. The MD5 (16 bytes) hash of the file

SHA-1 Hash. The SHA-1 (20 bytes) hash of the file

File Dates. Lists the Dates and Times of the following activities for that file on the imaged source:

- Created
- Last accessed
- Last modified

File Size.

File Format, as represented by PUID or MIME type. In addition, file format information ideally should have user recognizable names such as WordPerfect 4.2, Lotus 1-2-3 1.2, Word 6.0, etc. and be grouped into the following file categories:

- Archives. Archive files include Email archive files, Zip, Stuffit, Thumbs.db thumbnail graphics, and other archive formats.
- Databases. Database files such as those from MS Access, Lotus Notes NSF, and other database programs.
- Documents. Includes recognized word processing, HTML, WML, XML, TXT, or other document-type files.
- Email. Includes Email messages from Outlook, Outlook Express, AOL, Endoscope, Yahoo, Rethink, Udder, Hotmail, Lotus Notes, and MSN.
- Executables. Includes Win32 executables and DLLs, OS/2, Windows VxD, Windows NT, Java Script, and other executable formats.
- Graphics. Lists files having the standard recognized graphic formats such as .tif, .gif, .jpeg, and .bmp, etc.
- Internet/Chat Files. Lists Microsoft Internet Explorer cache and history indexes.
- Multimedia. Lists .aif, .wav, .asf, and other audio and video files.
- Presentations. Lists multimedia file types such as MS PowerPoint or Corel Presentation files.
- Spreadsheets. Lists spreadsheets from Lotus, Microsoft Excel, QuattroPro, etc.
- Unknown Types. Lists files whose types the A&D tool cannot recognize.

Further Questions or Comments

Categories might need to be configurable for individual institutions.

A&D_02.02: Directory-level technical metadata

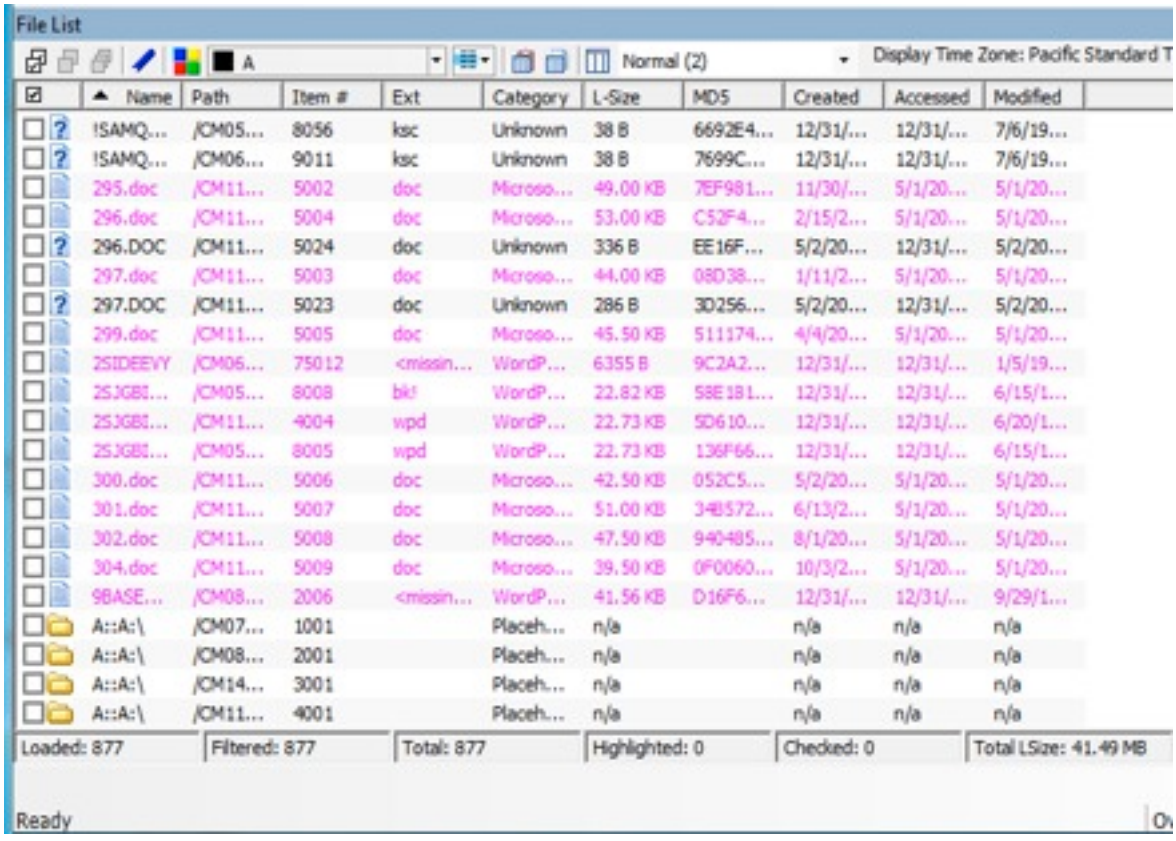
If possible, the tool should also provide the following technical metadata at the directory level:

- File Count. The total number of files within a directory.
- Size. Total size of all files in a directory, as expressed in kilobytes, megabytes, gigabytes, etc.
- Creation dates. A range of all files within the directory.

A&D_02.03: Presentation of technical metadata

Users should be able to view the technical metadata presented in a column format that presents the metadata as key/value pairs.

Sample Screenshots



Forensic Toolkit

A&D_03: Descriptive Metadata

Overview

It is essential that this tool is able to create, edit, import and export descriptive metadata about the collection (which might include paper archives not present or represented in Fedora in any way) for use in a third party collection management software (including, but not limited to Archivists Toolkit and CALM). The elements of the descriptive metadata should map to the descriptive elements of Encoded Archival Description (EAD). The tool does not need to store the metadata natively as EAD (e.g., it could store it as MODS), but the tool will need mappings to EAD for both import and export.

The sheer scale of born-digital files means that work is likely to be done over a prolonged period (i.e., over weeks/months). The solutions/workflow must be able to accommodate the flexibility of being able to save work whilst this sorting and processing is on-going.

Further Questions or Comments

It may also mean that more needs to be automated or will be done in less depth - e.g. automating inclusion of subjects / names via "entity extraction" or something like that **or** not doing detailed hierarchies.

A&D_03.01 Importing existing EAD

For hybrid, and multi-accession born-digital collections, there is a strong likelihood that the archival arrangement of the material will already have been undertaken and that the new material will need to be incorporated into the existing structure so that it can be presented as a single collection/finding aid. If material (especially born-digital) is added to a collection, then the existing intellectual arrangement and descriptive metadata must be imported into Hypatia. After importing the existing structure of this collection into Hypatia, the born digital material can be arranged into existing or new series / sub-series etc and then exported as an updated EAD (see *A&D_03.03*).

User Stories

Digital Archivist Carol has just received a deposit of born-digital material from an individual whose paper archives were deposited at the same institution ten years earlier. This additional material from the same depositor will form part of the same archival collection and so Carol would like to import the existing EAD structure into the tool to use as a guide for the arrangement of the born-digital material.

The intern Asok has conducted some initial processing of a new born-digital collection. After reviewing this with the digital archivist, he created the intellectual arrangement that is to be used for this material using AT/CALM. He then exports the entire EAD record which at this time may contain brief details about the accession (i.e., scope/content) and the proposed structure for the collection only - i.e., no descriptive data of the born-digital material.

This information will be used to create the groupings for the intellectual arrangement of the born-digital assets and the foundation for the EAD record. After further work adding descriptive data etc he then exports the updated EAD record so that he can overwrite the version originally created in AT/CALM.

A&D_03.02 Viewing/editing descriptive metadata

The tool will provide the ability to view and edit metadata using a form-based interface. The structure of the collection's intellectual arrangement should be viewable using a tree view (see A&D_01.01). The tool should allow for fields with controlled values (compare with screenshots in A&D_01.01) and allow for both short strings and full text notes for some values.

User Stories

Digital Archivist Tina has imported EAD for a particular collection and uses Hypatia to create an updated intellectual arrangement for the collection. As part of this process it is essential that she is also able to add descriptive data about the series of digital assets that will form the finding aid so needs a data-entry mechanism to add information including title, dates, extent etc. and ideally would like the system to suggest possible content for these fields based upon the items populating that set/folder/series (**see A&D_11**). Tina also needs the ability to assign rights and permissions (**see A&D_04**) at both a folder and individual file level depending on the nature and content of the digital assets.

Further Questions or Comments

Specific clarification is needed for the relationship between the PID for an asset held in Hypatia and its reference in EAD. This could be at least two places in EAD — either the unitid tag or the id attribute on the component levels for the PID associated with the set for a given component (e.g. the series). DAOs should contain references to the PIDs for the files themselves. In addition, the relationship between Hypatia and particular archival data management systems, such as Archivist's Toolkit and CALM, will be needed if users are going to be exporting EAD back and for the between the two system.

A&D_03.03 Creating new description and intellectual arrangement

For some collections the born-digital material will represent the first accession from that individual/ organization and we must offer the ability to start a completely new intellectual arrangement in the tool rather than force a user to create a skeleton record in AT/CALM and then import it into the tool (as per user story in **A&D_03.01**).

A&D_03.04 Exporting EAD data

Integrating born digital material into an existing arrangement requires that the updated description and arrangement can be successfully re-imported into software such as AT, CALM, and discovery platforms to enable further work or discovery.

Further Questions or Comments

Issues will exist if an institution uses an archival data management system like CALM and Archivist's tool kit. It would be technically very difficult to reconcile an EAD record edited outside of the Hypatia environment with one already ingested, especially if the differences relate to the arrangement of digitized assets. Resolution of these workflow issues are outside of the scope of the tool and will have to be resolved through local practice.

User Story

Digital Archivist Catbert has been working on a hybrid collection for a while and sucessfully imported the EAD (see *A&D_03.01*) for the paper material and used the Hypatia tool to integrate some born-digital archives. The revised EAD is then exported to CALM and made available to the public as part of the online catalogue.

Two years later a second accession of digital material has been deposited and Catbert then goes through the entire process again by importing the EAD.

Alice is working on a large ingest of born-digital material and having completed the work on one series of born digital assets she now wishes to export the descriptive data into their collection management software so that this material can be made accessible (see *A&D_04*) without having to wait until the entire collection has been processed. She exports the entire EAD back to AT/CALM so that the latest version is held there and can be made discoverable through other procedures. When Alice wants to continue processing the files from this ingest she can re-import the entire EAD from AT/CALM and continue.

A&D_03.05: Controlled vocabularies

Archivists will need to be able to use controlled vocabularies to assign access points at the collection level as well as component levels throughout the intellectual arrangement. The tool should either be able to import existing vocabularies (see *A&D_07.02*) or provide dynamic lookups against existing web services. Additionally, archivists will need to occasionally define new terms (e.g. authorized forms of names that don't currently exist in authority files).

A&D_04: Rights/Restrictions

Restrictions may affect the discovery, retrieval, or delivery of archival material, and will need to exist as controlled values that are machine actionable that have notes for human interpretation.

From the SAA Glossary of Archival Terminology: ***Access restrictions*** *may be defined by a period of time or by a class of individual allowed or denied access. …* ***Use restrictions*** *may limit what can be done with materials, or they may place qualifications on use. For example, an individual may be allowed access to materials but may not have permission or right to copy, quote, or publish those materials, or conditions may be imposed on such use.*

In terms of the implementation of this tool, access restrictions are the most critical. Archivists using this tool will need to set both date-based access restrictions and access restrictions based on class of user. They will also need the ability to add notes providing human-readable detail for both access restrictions and use restrictions.

Access restrictions should apply to a given component level and all the related files associated with that component level. Occasionally, related files may have more restrictions than their associated level.

A&D_04.01: Date-based access restrictions with automatic removal

Archivists will need to set date-based access restrictions that will be lifted automatically on a given date.

User Story

Miss Piggy, archivist at the Porcine Institute, is processing the Porky Pig papers. Mr. Pig is a well-known celebrity. The deed of gift for the collection states that two sets of digital records, subject files and correspondence on bacon

addiction, will be restricted only to archivists at the Porcine Institute until 2012. Miss Piggy needs to specify the date-bound access restriction to ensure no one except Porcine Institute archivists will have access to these records. However, Miss Piggy wants researchers to be able to discover these sets of records because they will be open for access soon. Miss Piggy also wants to ensure that the restricted material is available as soon as 2012 begins (i.e., on January 1, 2012).

A&D_04.02: Access restrictions to be removed manually at a later date

Archivists will need to add date-bound access restrictions that cannot be calculated automatically. These will need manual review and presumes that there will be a mechanism to report on restrictions for a given collection (cf. *A&D_08*).

User Story

Andrew, university archivist at Wilkes-Krier University, is describing the records of the Faculty Committee on Weasel Recovery. This committee discusses student academic issues, and folder titles identify students by name. For FERPA compliance, the records are restricted for the lifetime of the student plus 50 years, or 100 years after the date of creation. Since this restriction cannot be lifted automatically, Andrew wants to add a note describing the restriction as well.

A&D_04.03: Access restrictions for multiple classes of users and individual users

Archivists will need the ability to grant varying levels of access to archivist-defined groups of users and the occasional individual user.

User Story

Andrew (archivist from A&D_04.02 user story) needs to restrict these folder descriptions so only archivists can discover and view them. He needs to ensure that they are not discoverable or viewable by the public, but he may need to grant permission to current committee members or administrative staff at the University on a case by case basis.

A&D_04.04: Variable levels of discovery and access

Archivists will need to have variable levels of gated discovery and access. Levels of access should include "discover" (allowing items to be searched), "view" (allowing metadata to be viewed), "render" (allowing browser-renderable representations of an asset to be displayed), and "download" (allowing associated files to be downloaded).

User stories

Pepe, archivist at Feels Goodman College, needs to set access restrictions on a set of digital records so that they can only be viewed or downloaded from within the FGC Special Collections Reading Room. He wants them to be discoverable, however, and he wants to be able to give individual researchers permission to view them offsite from within their browser. He also needs to add a note describing the on-site use restriction since the records include proprietary software for which FGC has received a special license.

Frank N. Furter is an archivist at the National Organization of Hot-dog And Nitrite-laden Delicious Sausages (NOHANDS). To protect the intellectual property of NOHANDS, he wants to ensure that digital records made available through their discovery and access system are not downloadable. However, he needs researchers to be able to view the browser-renderable versions of the records when they use the system. He wants to set these permissions as he arranges and describes records.

Further Questions or Comments

More nuanced restriction setting may be needed in different situation in the future.

A&D_05: View Files / Representations

Archivists will need to view files or representations of those files to assist in the processes of arrangement and description. The file viewer should have a zooming function. There will obviously be some limitations in providing a viewer for some file types, so alternatives need to be available in some cases. Viewing files should not alter the technical metadata associated with the files, such as access and modification timestamps.

A&D_05.01: View original files

Whenever possible, users should be able to view the original files as rendered in the browser. At a minimum, this should include files that are easily rendered within web browsers (e.g., JPEG, GIF, PNG, text, HTML, PDF, XML, etc.). Ideally, the tool should provide a mechanism render common formats such as Microsoft Word and WordPerfect as well. The viewer should present original formatting whenever possible.

A&D_05.02: Extract and view text strings

For all files (particularly for file formats that are not easily renderable within a web browser) the tool should present a plain text representation of the data within a file by extracting strings.

A&D_05.03: Download files

For all files, archivists should be able to download the files to their local machine to allow them to view them with supplemental software. This can include native software (e.g. versions of WordPerfect) or software that can parse a number of file formats (e.g. Quick View Plus).

A&D_06: Export Metadata (excluding EAD)

Exported metadata formats required:

- METS for an entire collection
- MODS for a single object
- CSV export of all file objects, with associated PIDs/URLs, to be imported into an archival data management system like Archivist's Toolkit.

Questions

Technical metadata could also be exported. The ability to import technical metadata into CALM is a requested feature. However, this type of export is not seen as a priority by the AIMS partners at the moment.

A&D_07: Import Metadata (excluding EAD)

Archivists may want to import descriptive and arrangement metadata from another tool into the arrangement and description tool.

A&D_07.01: Import metadata from Forensic Toolkit

The A&D tool should be able to import the bookmarks, labels and flag "privilege" in collections. Bookmarks will be mapped to series, subseries, etc. Flagged "privilege" items will be mapped to "restricted" materials. Mapping of the "labels" will be decided later.

User Story

Peter, digital archivist at FRED Institute, used the bookmark, label, and flag "privilege" functions in AccessData FTK to assign intellectual arrangement to several collections. There are new accessions to the collections and the A&D tools is available, he wants to import the bookmarks, labels and flag "privilege" he assigned to those collections and process the new accessions using the A&D tools.

A&D_07.02: Import controlled vocabularies

The tool should be able to import controlled vocabulary terms for use within the tool. Sources of data could include Archivists' Toolkit, CALM, and web services such as id.loc.gov.

User Story

Peter, digital archivist at Present Institute, would like to use subject headings from Archivists' Toolkit to describe (series / subseries) of the collection he is working on.

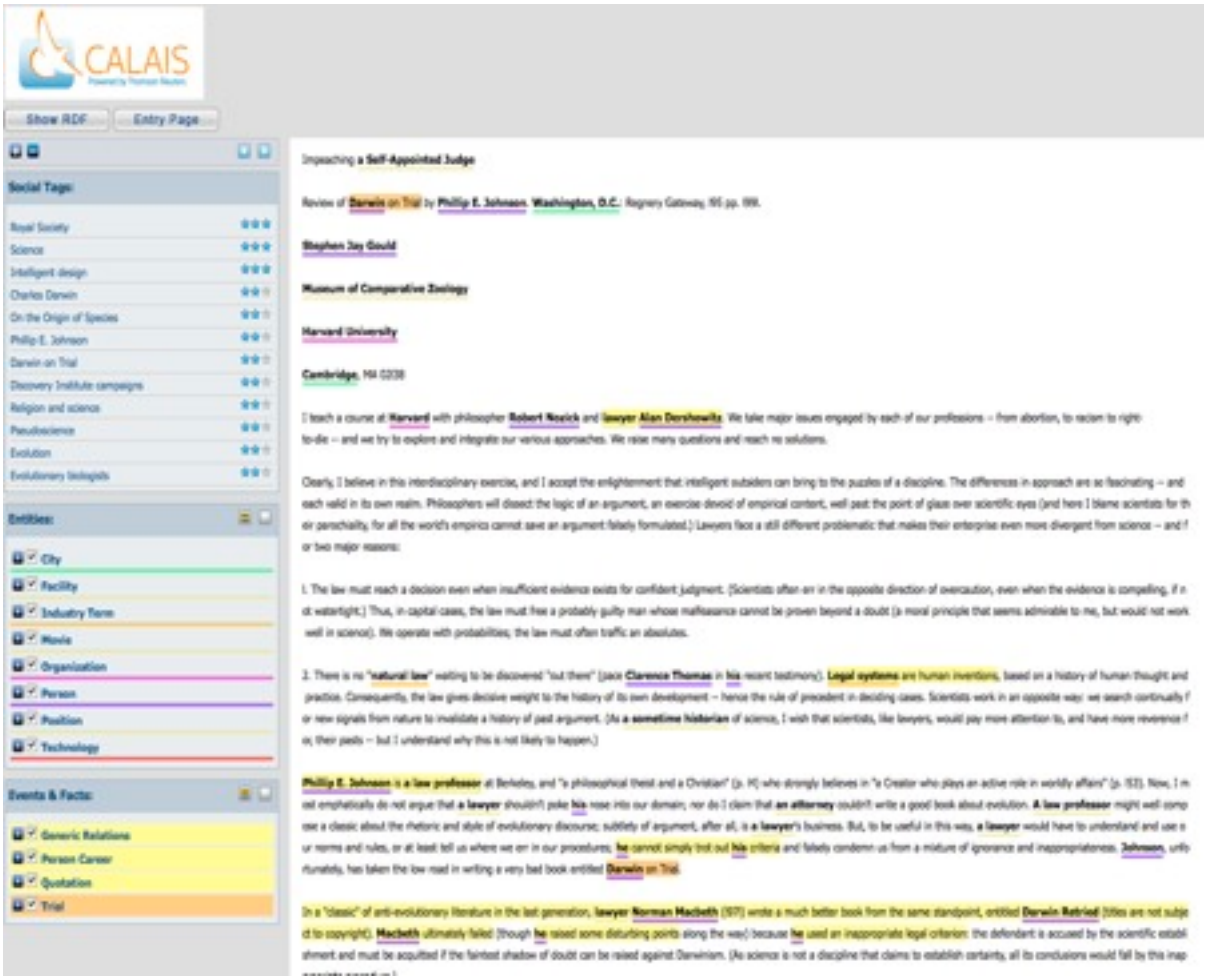
A&D_07.03: Import descriptive metadata using entity extraction software/service

The A&D tools should be able to produce entities (name, subject, place) using its an entity extraction engine, third party entity extraction web service, or third party entity extraction program and store the entities extracted in RDF format in Fedora. The entities will become the facets of the collection.

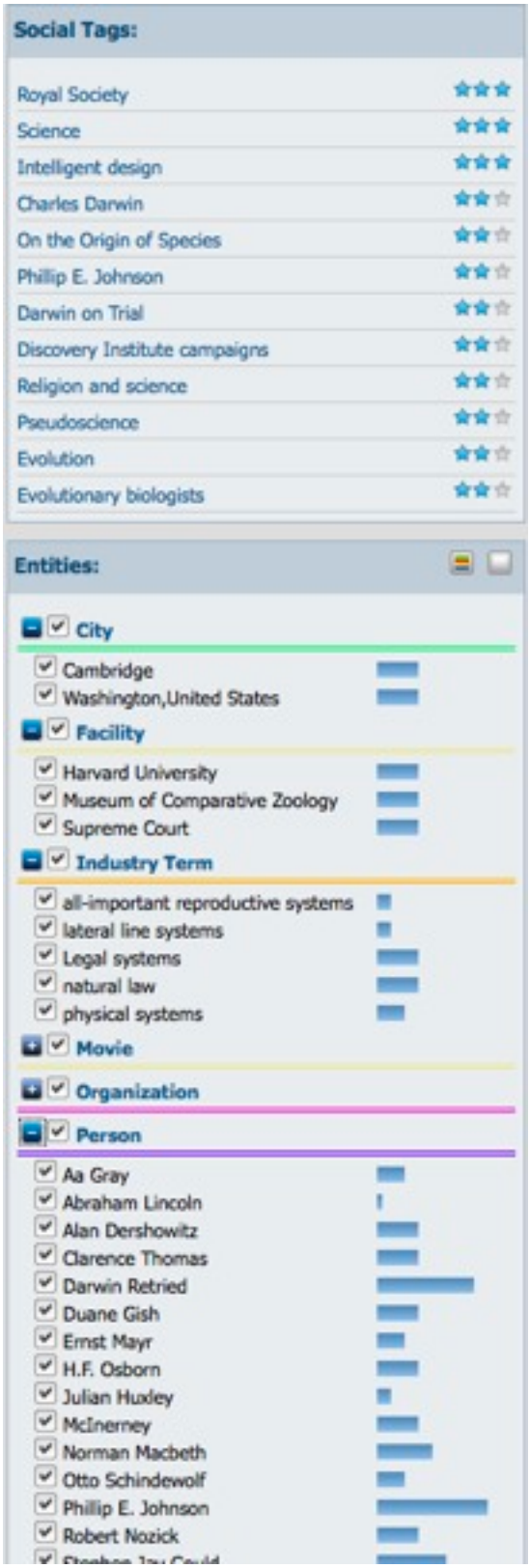
User Story

Peter, digital archivist, at Future Institute, is asked to process a collection with 5 million files. The files are not very organized. He was given 2 weeks to assign descriptive metadata to the files. He expects people using the collection would rely more on full text search and entities (people, places, etc.) browsing but not so much on EAD. He decided to prepare a EAD with very very high level arrangement of the collection and to publish entities extracted using OpenCalais, a very popular entity extraction service.

Screenshot



OpenCalais


Social Tags:
Royal Society
Science
Intelligent design
Charles Darwin
On the Origin of Species
Phillip E. Johnson
Darwin on Trial
Discovery Institute campaigns
Religion and science
Pseudoscience
Evolution
Evolutionary biologists
Entities:
City
Cambridge
Washington,United States
Facility
Harvard University
Museum of Comparative Zoology
Supreme Court
Industry Term
all-important reproductive systems
lateral line systems
Legal systems
natural law
physical systems
Movie
Organization
Person
Aa Gray
Abraham Lincoln
Alan Dershowitz
Clarence Thomas
Darwin Retried
Duane Gish
Ernst Mayr
H.F. Osborn
Julian Huxley
McInerney
Norman Macbeth
Otto Schindewolf
Phillip E. Johnson
Robert Nozick

OpenCalais

A&D_08: Reporting

Reporting in an arrangement and description toolset should allow for arbitrary queries. Reports generated from metadata about the records may inform external decision making processes or be used for the calculation of statistics. Reports should be produced in an output format such as CSV or XML that will allow simple post-processing.

A&D_08.01: Report on duplicate items

The tool needs to provide a reporting mechanism that will identify files that have an identical MD5 or SHA1 hash. Because the hash is independent of the filename, identical files may actually have different filenames. (See also: *A&D_01.04, A&D_04, A&D_10.02*)

User Story

Marmaduke, digital archivist at the Great Danish State Library, is processing the Scooby Doo papers. This collection was very disorganized when it arrived and processing the paper component lead to discover of lots of duplicate material that Marmaduke's supervisor wanted him to remove. Marmaduke wants to create a report of multiple files with identical checksums to help him identify records that can be removed.

A&D_08.02: Report on restricted components and collections

Archivists will need to generate reports listing all collections containing restricted material, as well as all component levels within a specific collection that are restricted.

User Story

Peter Peter, archivist at the Pumpkin Society, wants a report containing all the collections with electronic records that have restricted components. He just needs high-level information. Once he has this information, he discovers that collection MS150, the Gourdie Howe papers, has restrictions. He wants to create another report containing a detailed list of restricted components for MS150 as it is a heavily used collection.

A&D_08.03: Report on file formats

The tool should be able to provide a breakdown of file formats within a collection. This presumes and requires that the technical metadata already is associated with the files. The assumption on our part is that this information is included or generated during ingest.

User Story

Grimace is an archivist for the McDonaldland City Archives. He needs a report containing counts of all the different types of files in RG 12/4/2009, the Mayor McCheese records. He doesn't need to know where each file falls in the collection hierarchy. He also needs an approximate calculation of the total size of the record group. He needs to share this information with H.M. Burglar, the IT director for the City of McDonaldland.

A&D_08.04: Report on appraisal status

Archivists will need to create reports listing the various appraisal statuses as defined in *A&D_10*. There should be both a combined report and separate report for each of the individual statuses.

A&D_09: Email

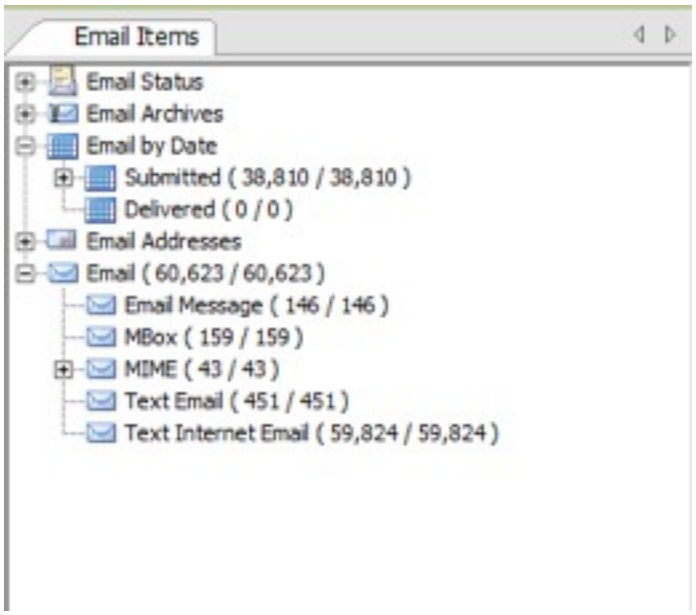
The tool should provide a set of tools to allow work with email messages that may be contained in accessions. The tool should be able to work with email created by different programs (Outlook, Outlook Express, AOL, Yahoo, Hotmail, Lotus Notes, and MSN, Eudora, etc.) and in different formats (mbox, mime, etc.).

A&D_09.01: Display emails by group

The tool should allow archivists to view groupings of emails as follows:

- Email Attachments (Contains only attachments to emails);
- Email Reply (Contains emails with replies);
- Forwarded Email (Contains only emails that have been forwarded);
- From Email (Contains everything derived from an email source, i.e. email related)
- Date (organized by Year, then by Month, then by date, for both Submitted and Delivered);
- Email Addresses (organized by Senders and Recipients, and subcategorized by Email Domain, Display Name, and Email Addresses).

Screenshots



Forensic Toolkit

A&D_09.02: Export/download email

The A&D tools should be able to export emails (cf. *A&D_05.03*) to work with other programs (e.g. network graph, etc.). Ideally, users would be able to select what fields and range of the value of the fields to be exported: to, from, date, cc, bcc, subject, email body.

A&D_10: Appraisal of Material

An archivist will appraise the material to ensure that only items wanted for long term preservation are retained. This is a key professional skill and the approach to this will vary from collection to collection. It may occur either pre or post ingest. Where it occurs after ingest there is a need to record the decision along the same lines as with duplicate files (see *A&D_10.02*). With paper archives we usually ask the depositor whether they want items that we do not wish to retain returned to them, recycled (for non-confidential material) or confidentially destroyed.

A&D_10.01 Marking files for deletion or other appraisal actions

It would be nice for the appraisal process to be able to flag the status of files and folders within the accession ingest as either "keep" "unsure" and "bin". This should be applicable at any level and inherited downwards but with the ability to change individual file(s) as needed - for example the vast bulk of a series of nested folders should not be kept but there are a few individual files that should be retained (or vice versa)

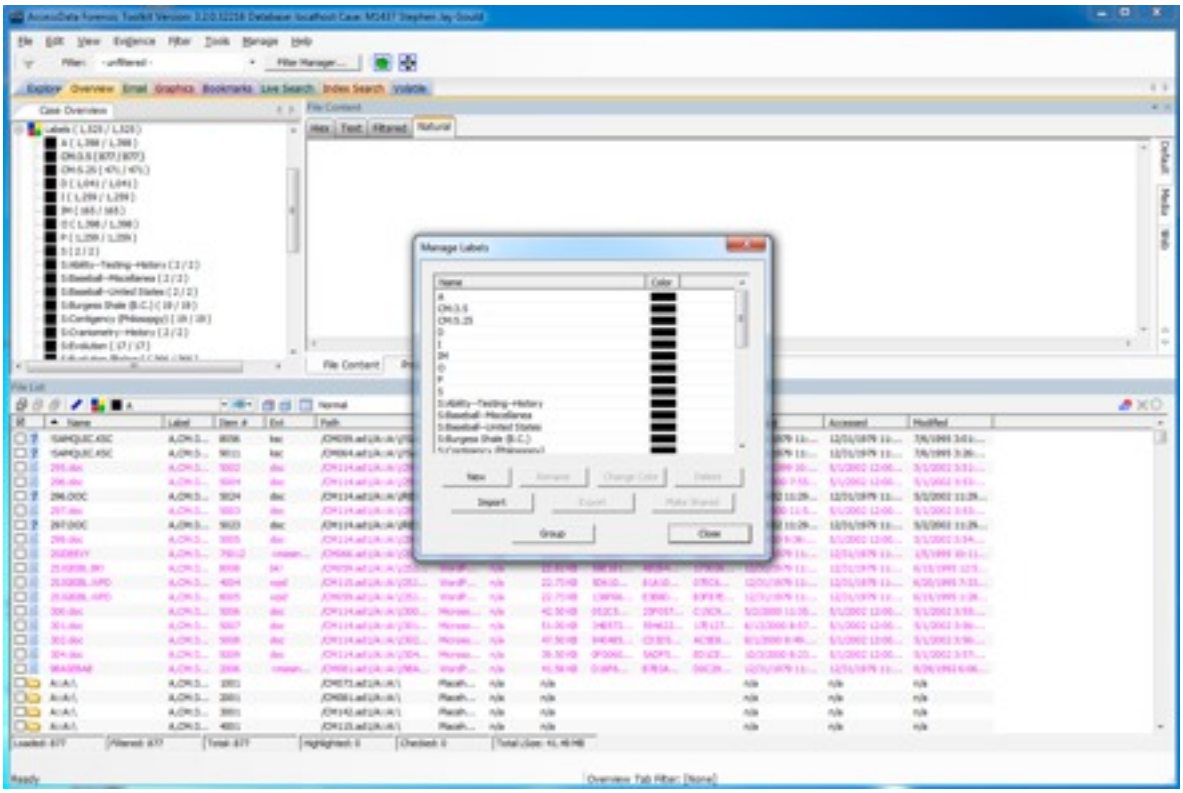
With large accessions it might be that similar material is held in different folders - so the ability to sort or filter the files (see *A&D_01.03*) in the accession ingest could offer the flexibility of looking at the material in alternative ways - but this should be a **temporary situation** and should not change or over-ride the arrangement of the folders/files at the point of ingest.

It is likely that appraisal will be conducted over time so the appraisal status flag would assist with recording progress through the material or allow additional staff to review particular sections of material (i.e., everything marked unsure).

In many situations it will not be possible to make an appraisal decision based purely on the technical metadata that is available so the archivist would need to open a file to make the decision about whether it is kept or not. It is essential that this appraisal process does not impact upon the technical metadata (especially last opened/accessed) date that we subsequently wish to present to researchers. Therefore, two options to achieve this might be to generate an access copy at the point of ingest or to ensure that "last modification time" for a file is the last mod time the file had when it was originally ingested and will not be overwritten if the file is opened.

For material to be deleted there should be a two-step process requiring confirmation etc. — one option could be to get Hypatia to generate a report listing all of the files to be deleted but I am not sure how useful/practical this would be if hundreds (or more) files are being deleted. We should consider making deletion of files from the ingest something that is restricted to particular user roles with appropriate permissions (ref to **A&D_04**). The return or destruction of the born-digital files may have been indicated during the deposit/transfer process. This work will be done outside Hypatia. A note regarding the removal of material as part of the appraisal process, such as a broad note like "third-party publications removed" should be possible any component level and at the collection level. It should correspond with the EAD note element <appraisal/>.

Screenshots



Forensic Tolokit (application of labels)

A&D_10.02: Duplicate Files

Either as part of the appraisal process or otherwise there is a requirement to be able to detect files that are exact duplicates of another file in the repository and to then be able to either to hide or delete the file. It is critical that all actions on the file are automatically record to provide a full audit trail.

User Stories

Digital Archivist Dilbert, based at the Scott Adams University, likes to keep a tidy ship and knows that this includes the digital repository and hates the thought of storing, preserving and providing access to multiple versions of the same digital file - whether this is because of a user accidentally misfiling a file into a specific folder or because the file has been transferred as part of multiple ingests over time. He does know that they are exactly the same because he has asked the processing archivist Wally to run a report (see A&D_08) using their checksum value.

Having run the report to detect duplicate files within a single accession, a single collection (i.e., multiple accessions) or across everything, Wally can look at the report data (this should

include ingest ref?, creation date, last viewed data, filepath and/or location of the matching file(s) and then has three options.

- Hide: This hides a file from view so it is not visible for the archival arrangement, discovery or access elements of the workflow
- Delete: This marks the file(s) as ready for deletion but suggest a further prompt to confirm that you want to delete the file from the system completely. This technique could also be applied to files/folders that are identified for deletion as part of an appraisal process and/or files that subsequently need to be removed (e.g., for copyright purposes) [should Wally be able to delete files?]
- Ignore: This says I know that the files are the same but do not wish to hide or delete it

For the purposes of creating the report and accessing the audit trail etc all hidden and deleted files need to have a datastream updated to reflect the change with possibly a default content - *this file was identified as a duplicate by XXperson on YYdate or deleted by XXperson for ZZ reasons.*

Screenshots

Accessed	Modified	Flagge...	Duplica...
12/31/...	7/6/19...	False	
12/31/...	7/6/19...	False	
12/31/...	4/22/1...	False	
12/31/...	4/22/1...	False	
12/31/...	4/22/1...	False	
12/31/...	4/22/1...	False	
12/31/...	4/22/1...	False	
12/31/...	4/22/1...	False	
12/31/...	4/22/1...	False	
12/31/...	4/22/1...	False	
5/1/20...	5/1/20...	False	
5/1/20...	5/1/20...	False	
12/31/...	5/2/20...	False	
5/1/20...	5/1/20...	False	
12/31/...	5/2/20...	False	
5/1/20...	5/1/20...	False	
12/31/...	1/5/19...	False	Primary
12/31/...	1/5/19...	False	Second...
12/31/...	6/15/1...	False	
12/31/...	6/20/1...	False	
12/31/...	6/15/1...	False	

Forensic Toolkit

Further Questions or Comments

Archivists should be able to decide which duplicate file should be the primary. Technical metadata could be used to determine the oldest duplicate and make that the primary.

When deleting files, it is unclear if Hypatia should delete the files itself or just "identify" them for deletion. If we choose the latter, this would mean that the deletion would happen outside of Hypatia.

When you delete a file (or series of files) should we delete the physical file including any derived versions etc. but leave a "shadow" or "tombstone" record that includes an audit trail and reason for deletion (i.e. duplicate or appraisal). This should be available as distinct report (see *A&D_08*). In addition, the issue of whether or not preservation copies of "hidden" files or files marked as "deleted" but not removed should be created. This may be out of scope for Hypatia, but local implementation should consider the issue.

A&D_10.03: Immediate (unstaged) deletion

The tool should provide an option to delete files immediately if needed. This must present a confirmation screen as files may not be recoverable. This functionality should still retain a "tombstone" record that includes the date of deletion.

A&D_11: Batch Application of metadata from files

The sheer volume of files means that we should try to automatically use the extractable metadata to form the proposed basis of the descriptive metadata. For example:

A&D_11.01: Apply filename to title field

A&D_11.02: Apply creation/modified date to "from" date field

A&D_11.03: Apply access/modified date to "to" date field

A&D_11.04: Apply creator to creator field

A&D_11.05: Apply file format to descriptive/technical note

A&D_11.06: Apply number of files to extent

A&D_11.07: Apply size of file(s) to extent

Further Questions or Comments

When multiple assets are being described by a single descriptive entry, the information could be derived from the first file it encountered (sorted by name or date etc) or from the directory level, see *A&D_02.02*.

The ability to define the date format to be used would be a "nice" feature given US vs UK local practices. However, the date will probably be stored in a machine readable format which will allow us to easily customize how it gets presented to the end user.

A&D_12: Intellectual arrangement

The contents of this overview have been adapted from section D2, "Resources in AT Description," in the functional requirements for the Archivists' Toolkit Description Module (http://archiviststoolkit.org/sites/default/files/description.pdf, pp. 4-9).

Record Types

A **collection** is 1) an item or aggregate of items generated or collected by an individual, family, or organization in the course of their activities and deemed to be of enduring value, 2) and is in the custody of an archival institution.

Collections may also be linked to **components** to form multi-level descriptions.

Hierarchical Levels

These two record types and their associated interfaces for descriptive metadata (see **A&D_03**) accommodate the twelve levels of description permitted in the Encoded Archival Description standard. In other words, a **collection** in the A&D tool may be represented by up to twelve hierarchical levels of records. A **collection** record may be the parent of a **component** record that is parent to a **component** record that is parent to a **component** record, and so on up to twelve levels deep. There may be an unlimited number of component records at each level, that is, there is no limit on the number of series records or file records. Records at the same level are referred to as **sibling records.**

EAD provides a standardized vocabulary of labels for the permitted hierarchical levels in an archival resource. These labels (**class, collection, file, fonds, item, otherlevel, record group, series, subfonds, subgroup, subseries**) each correspond, or map, to one or more of the collection or component records (See Table 1 in AT Description specification).

When the operator chooses to add a new component record to an existing collection or a component record, she or he **must choose** a level label for the component. The options given the operator are driven by a set of rules for acceptable children for a given level. For example, the parent of a subseries can be a series, but not a collection. (See Table 2 in AT Description specification)

Intellectual and Physical Order of Archival Resources

Intellectual hierarchy will be captured by **tracking the relationship** of the collection records and component records to each other. **Both parent/child record linkages and sibling record sequences must be captured and stored.**

A&D_12.01: Create new collections

Archivists must be able to create new records representing archival collections. Descriptive metadata should follow the collection-level elements available within Encoded Archival Description and **must include** creator, title, date ranges, identifiers, and call numbers.

User Story

Eugene is processing the Absurd Theater Records. The collection does not have an existing EAD finding aid or description in another system. He loads up the tool and logs in to his account. Once logged in, he selects "Create new collection." He enters the metadata about the collection and clicks save. Once he saves, he is redirected to a page for that collection.

A&D_12.02: Create new component levels

Archivists must be able to create new component levels that are children of collection records or siblings or children of other component levels. See **A&D_03** for description-related requirements.

User Story

Eugene then needs to create a new series of subject folders in the collection. He is logged into the system and is viewing the collection page. He selects "Create new component." In the "Level" field, he selects "Series." He fills out the metadata about the series and clicks save. Once he saves, he is redirected to page for that series.

A&D_12.03: Associate files and directories to component levels

Archivists must be able to associate files and directories from accessions with component levels. They may also need to remove or change the associations. Assigning a directory to a component **should not** create a new component within the intellectual arrangement.

The tool should allow for multiple associations during the arrangement process. **However, a file or directory must have relationships with no more than one component within a "finalized" intellectual arrangement.** This reflects constraints on arrangement as defined in archival practice.

A&D_12.04: Associate accession with collection

There will be cases where accessions will not include metadata that relates them to a specific collection, so the tool will need to provide the ability to allow archivists to associate accessions with collections.

User Story

Eugene wants to take files from an accession that have been ingested and assign them to this series. He associates the accession with this collection by selecting the appropriate accessions by number. He also can see a list of all unassociated accessions.

A&D_12.05: Replicate directory structure from accession into intellectual arrangement

Archivists may discover that an accession's directory structure demonstrates that a creator had a clear existing arrangement that should be maintained. Accordingly, archivists using the tool should be able to replicate some or all of the the directory structure from an accession into corresponding component levels. See also **A&D_01.04**, **A&D_02**, and **A&D_11**.

A&D_12.06: Duplicate components and structure in intellectual arrangement

Archivists are used to being able to copy component structure during the arrangement process to prototype various intellectual arrangements. Contents of the descriptive metadata should be duplicated automatically.

A&D_13: Searching within files

Ability to do pattern or keyword searches in order to discover files that should be restricted - credit card or social security information; passwords; student or medical files etc.

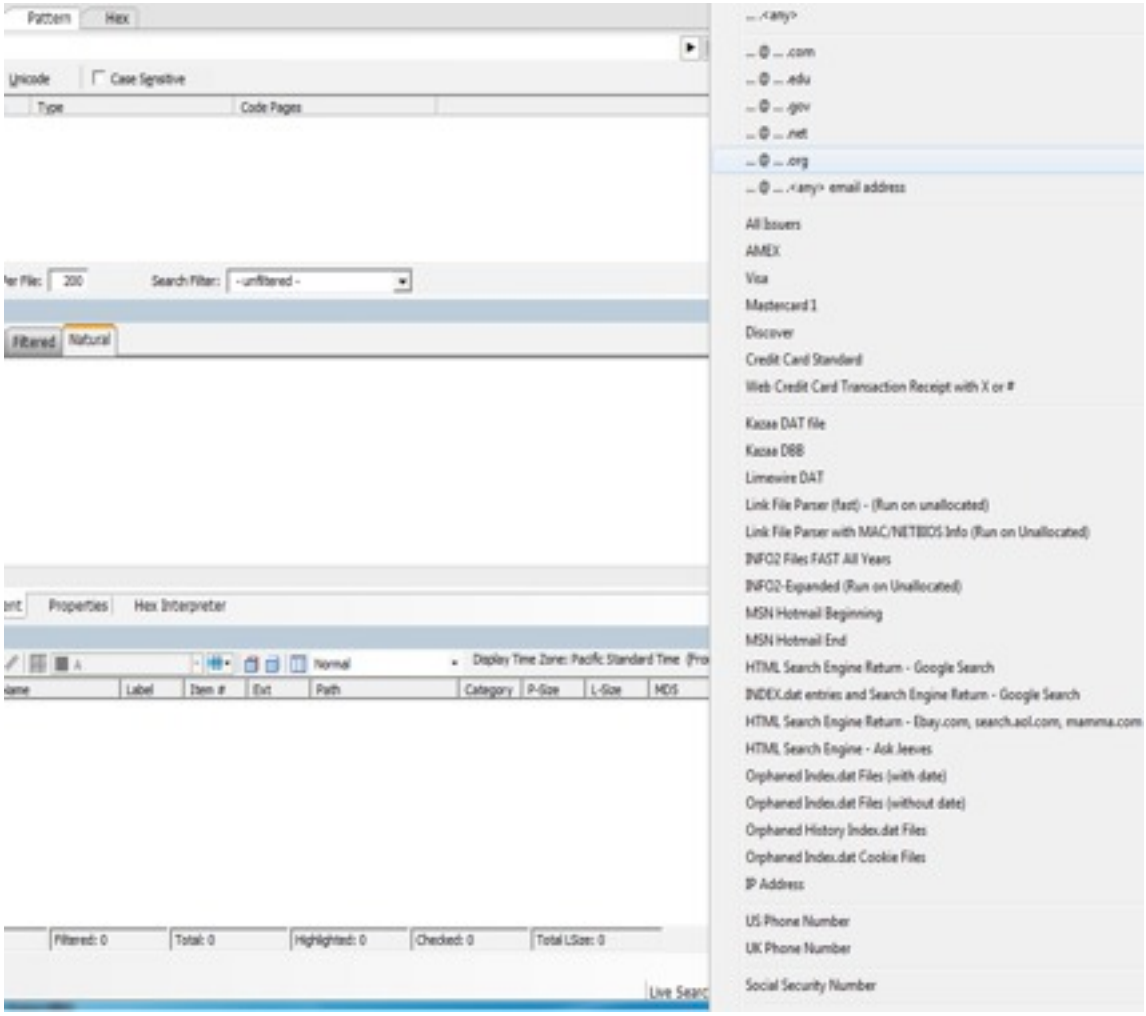
A&D_13.01: Pattern searching

For pattern search, it is desirable to allow users to define their own patterns as well as to include commonly used patterns such as social security number, phone no., credit card nos. etc.

User Story

Peter, digital archivist at FRED Institute, is processing a born digital collection. He is concern on the existence of social security numbers in the files. He would like to perform a search on the whole collection so that files containing texts with XXX-XX-XXXX (X- numeric) pattern will be grouped with the text XXX-XX-XXXX highlighted for him to review for restriction.

Screenshot



Forensic Toolkit (Pattern Search)

A&D_13.02: Full-text searching

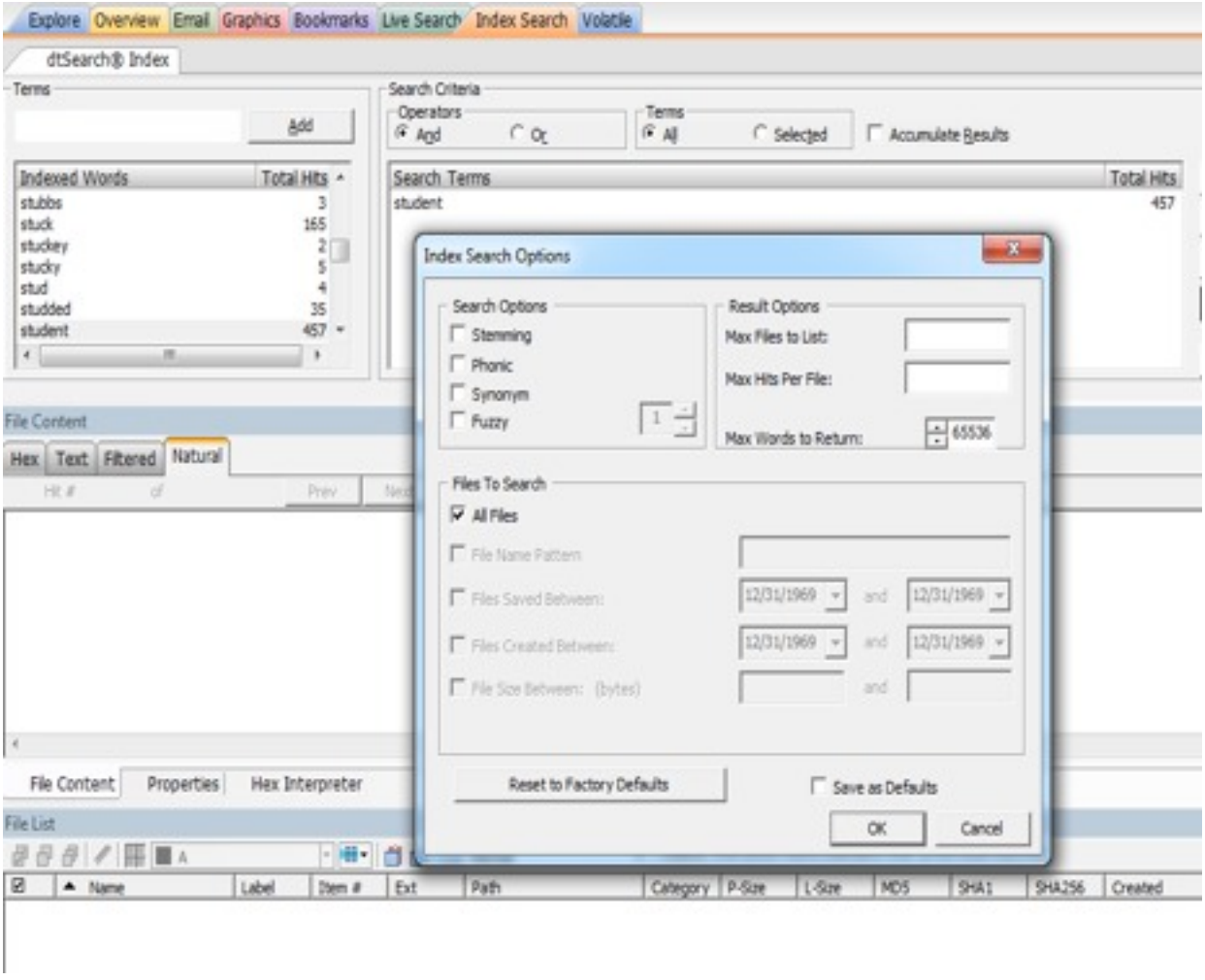
For full text search, it is desirable to have the following options:

- Stemming. Words that contain the same root, such as raise and raising.
- Phonic. Words that sound the same, such as raise and raze.
- Synonym. Words that have similar meanings, such as raise and lift.
- Fuzzy. Words that have similar spellings, such as raise and raize

User Story

Peter, digital archivist at FRED Institute, is processing a born digital collection. He was told by his supervisor that all files containing student grades should be restricted. He would like to perform search on the whole collection so that all files with "student", "students", "grade, "grades" will be grouped and the texts "student", "students", "grade, "grades" highlighted for him to review for restriction.

Screenshot



Forensic Toolkit (Full-text search)

Addendum: Functional Requirements Based on Yale Workflow Diagram (For Discussion)

Prepare for Arrangement

Use accession / acquisition records, existing surveys and descriptions, inventories, biographies, donor documents / correspondence to prepare for arrangement.

Component Tasks

- Select collection for processing
- Gather records and information
- Assign to processing archivist
- Restrict collection during processing

Survey Collection

Use appropriate tools to survey the records.

Component Tasks

- Assess / analyze type / condition of the media
- Assess / analyze file formats, sizes, dates
- Assess / analyze existing arrangement (e.g. folders)
- Assess / analyze content
- Assess / analyze context, functions

Arrange records intellectually

The intellectual arrangement of archives is a critical process in making the material, regardless of format, accessible to users. Wherever possible/practical the skills and terminology applied to paper materials should be applicable to born-digital materials. Archival collections are usually catalogued according to ISAD(G) or DACS cataloguing standards which are non-format specific.

Archival arrangement is a key professional skill, with a user likely to make a number of assumptions about the material depending upon its intellectual arrangement. Working with born-digital material is likely to be harder than working with paper records due to the increased practice of mixing both original and material from third-parties into a single filing system and the much greater volume of material concerned. For organizational records there is the increased complexity as a result of individual, team and institutional file stores.

Digital archives offers the potential to place a single digital asset in multiple locations with-in an archival arrangement and we should resist this temptation and retain the current practice of a single unique place with-in the intellectual arrangement and to include appropriate cross-references to aid users.

Component Tasks

- Review the material including its context and content and see if it is possible to identify or determine the "original order" for the material
- Create a logical order if the original order cannot be identified
- Identify the processing and cataloguing requirements for this particular collection [Note this could be part of the survey stage?]
- Place material of similar nature (e.g., all files relating to Book X) or function (e.g., all minutes of a specific committee) into series
- Create a hierarchy of the material into cascading series' from Collection at the top to Item (a single digital asset) at the bottom [the sheer volume of digital material means that the predominant practice is likely to be cataloguing at series level]
- If material is already held, the deposit of additional materials (whether paper or born-digital) will need to be integrated into the existing intellectual arrangement

Note: Heavily linked with issues of the GUI (**A&D_01**), import and export of EAD (**A&D_03**) and feeds into Descriptive Metadata

Further Questions or Comments

Consideration of access and permission issues (see **A&D_04**) should be done at the highest level first - e.g., apply the conditions to the collection and then modify specific series/items that vary from this position (e.g., a collection may be generally open but a specific series of records closed for xx years under Data Protection legislation)

Arrange records physically

With paper records an important element in their management is that relating to its location to enable easy retrieval from the store by the archives staff. For born-digital materials the original file will be ingested into the repository and preserved and an "access copy" version derived from the original created for individuals to access and use.

Whilst the "location" of the original file will remain in the Repository there is a need to create a link so that individuals with appropriate access and permissions can retrieve the access copy digital asset without further involvement by the archives staff. It is important that this link remains persistent for authenticity and citation purposes. It must also avoid revealing the true path of the Repository and so risk unauthorised access to other digital assets.

Component Tasks

- For retrospective cataloguing there will need to be a systematic process of identifying born-digital material within existing material
- Removing these file(s) for processing and subsequent ingest into the Repository
- associating ingested files/folders with Fedora sets to place it into its place in the hierarchy / intellectual arrangement

Create descriptive tools

Use information about content, context, physical characteristics, and archival management processes to create standardized or customized information products for various purposes. A description tool should be able to import (for existing or hybrid collections) and export EAD files. At a minimum, the tool should be able to export the structure of the container list. [Note: see A&D_03 for similar functions]

Component Tasks

- Describe biography/history
- Create scope notes and component description
- Create intellectual structure/box list
- Summarize preservation and appraisal actions
- Create subject and name access

Further Questions or Comments

Defining cataloguing standards for digital materials out is out of scope for the requirements, but the AIMS project at large may want to comment on this.

What level of descriptive activity is necessary for an AIMS-specific description tool?

Is assigning subject terms for different descriptive levels necessary?

What about biographical/historical notes?

In other words, is the AIMS-specific tool narrowly scoped, with the description then exported to something like the AT/CALM for further work?

Perform physical control

Assign archival records to containers and storage locations appropriate to their physical composition, technical characteristics, extent, and condition. Pack, label, and store materials so they can be retrieved and moved as needed. Assign identifiers to groups or containers of archival records. Storage assignments follow plans that reflect the archival institution's policies governing the placement of various materials within facilities.

Component Tasks

- Assign call numbers, locators, barcodes, and other identifiers
- Label boxes, folders, etc. [OUT OF SCOPE]
- Create and updates holdings and location records.
- Send materials to storage location [OUT OF SCOPE]

Further Questions or Comments

We won't really be "assigning" storage locations per se. True storage management is going to be out of scope. What will be needed is 1) basic workflow management that can represent a "commitment" action that work is

completed for a given collection or set of records, 2) a tool that may allow us to add mnemonic identifiers (e.g. based on call numbers), and 3) exposure of PID/URIs in the interface to allow linking back from descriptive tools. There is a larger integration with AT (and possibly CALM) bulk importing digital object locations that probably needs to be hashed out at some point.

Disseminate descriptive tools and records

Prepare records and descriptive tools for dissemination in access systems. Publication and indexing of descriptive tools. Digitization/transcription as appropriate. Creation of dissemination packages of records as appropriate.

Component tasks

- Publish descriptive tools
- Publish records
- Index descriptive tools
- Index records [LOW PRIORITY?]

Further Questions or Comments

This again is largely about workflow and "promoting" our descriptions etc. to a discovery and access tool. We will need a way to signal that these are ready to be discoverable. A way to build formally defined dissemination packages is not needed within the A&D tool, but discovery and access requirements suggest the need for ways of traversing the object relationships that would allow, for example, someone to retrieve all records from a given series, or perhaps all records in a collection. Indexing descriptive tools really falls under the domain of an access and discovery tool.

Complete processing

Signal that records are ready for access. Receive final approval from supervisors and document completion.

Component Tasks

- Remove processing restrictions [OUT OF SCOPE?]

Out of Scope

Requirements that were discussed but deemed out of scope are included here with reference to the section from where it was originally proposed in the document.

From A&D_03.03 Creating new description and intellectual arrangement

Can the information about the donor, deposit etc from the Donor Survey as the basis of an accession-type entry be accessed? Clearly there is a high chance that the potential material identified in the survey will not reflect the actual material subsequently transferred.

How to do this is an issue. Copy and paste would be easy to implement, but painfully slow. Drag and drop would be visually nice, but also painfully slow for more than a few fields. Various aspects of A&D would benefit from a

scripted approach instead of a visual UI. The "scripts" would be akin to macros. Historically, this type of functionality is reasonably easy to implement, and add a huge amount of power and flexiblity to a product.

We would also already have access to this information already in multiple ways. the EAD finding aid is the public view, but in AT/CALM which are collection management systems there is also a host of other data you need to record/manage but not divulge to the public within the accession/depositor tables of AT/CALM. e.g., depositor contact details, terms of deposit.

This information is out of scope for Hypatia unless it is relevant to how we arrange material or it's important enough to include in description. However, it should still be considered as it relates to information flow between Hypatia (for discovery, access, and management of digital objects) and AT/CALM for larger archival management. We need to illustrate it has not been forgotten/ignored especially as some of this information may be captured via the web survey

From A&D_03.02 Viewing/editing descriptive metadata

The "master" version of the EAD file should be dictated by local practice when using files created in AT/CALM and subsequently edited in Hypatia.

From A&D_04.04: Variable levels of discovery and access

Tools like FTK allow for restriction at the individual file level. Questions still remain as to whether it is good practice to allow mixing unrestricted files and restricted files into a specific level of arrangement, but I added this as an option above.

Functional requirements from Yale Workflow Diagram - Complete Processing

- Receive final approval [WORKFLOW]
- Document completion
- Announce availability of records and descriptive tools [OUT OF SCOPE]

If we have a project archivist working, we may want to have a senior archivist review work before we mark it as "done" but this is really a question of workflow. "Document completion" should be about generating documentation about what was done in processing, and clearly relates to A&D_08 above. It is unclear what form a processing report would take in this case, or whether it would be important enough to create.

2. Rubymatica

Rubymatica is an open source software project written in Ruby and adapts some of the convenient workflow provided by Archivematica. It is primarily an application programming interface (API) with the purpose of creating an arrangement of files for ingest. The project contains a simple demonstration web site which is open to the public on a by-request basis. Rubymatica adapts some aspects of the SIP to AIP transformation phase of Archivematica as a means to build SIPs ready for ingest.

There were several reasons to create a Ruby version of Archivematica. Rubymatica is written in Ruby so that it can easily be integrated into Hypatia. Ruby has become prevalent for developing web applications, and the University of Virginia (UVA) has standardized on Ruby and Java. At UVA, legacy Python, Perl, PHP web applications are being superseded or converted to Ruby. Writing the tool in Ruby also offered the opportunity to create some additional functionality than what is present in Archivematica.

Rubymatica, being a program to prepare files for Hypatia ingest, has somewhat different goals than the SIP ingest phase of Archivematica. Because of this, there is some different logging, and the creation of metadata databases that aren't necessary in Archivematica. The workflow and general architecture are similar to Archivematica. In both Rubymatica and Archivematica, many of the same external applications handle tasks such as unpacking archives, generating checksums, and checking files for malware. Ruby scripts do the bookkeeping, workflow, and data management. Each external application processes files without any knowledge of the overall workflow. Rubymatica has both command line and web interfaces.

This work only took a few days to complete and as part of the development process, members of the UVA Library software team did a code review of Rubymatica for both legibility and security issues. Rubymatica is on Github along with extensive, technical documentation.

https://github.com/twl8n/Rubymatica

Rubymatica processes each ingest as a single process, copying and transforming the ingest into a new directory tree containing a subdirectory with same structure as the original, plus metadata subdirectories. The ingest may be in the form of an archive file (ZIP, tar, or rar files) or a directory. Rubymatica has additional functions to create a BagIt bag, to integrate a Tufts TAPER submission agreement, and to integrate a donor survey. The current version also has a feature to create categories for PRONOM file identifications. PRONOM's DROID application is run via the FITS file identification suite.

Rubymatica runs several applications on every file in a logical copy of an ingest. The processing steps are:

1. Copy original files into a working directory tree
2. Recursively unpack any archive files
3. Cleanse (detox) file names of characters not supported by MS Windows, MacOS, and Linux
4. Check for malware, create checksums, identify file types via FITS and DROID, and write a METS file

Log files are maintained, and several very small databases are created to track metadata and status of the ingest. After this processing, the collection is in a form suitable for assessment and eventual ingest into a repository for further processing.

Archive files are unpacked into new, uniquely named directories in order to avoid directory name and file name conflicts. File name conflicts are also avoided during file name cleansing. Processing happens as a background process in order to prevent a timeout. The background process can (in theory) run as long as necessary to process an ingest.

3. Hypatia

Hypatia is an initiative to create a Hydra application (Fedora, Hydra, Solr, Blacklight) that supports the accessioning, arrangement / description, delivery and long-term preservation of born digital archival collections. Hypatia is being developed as part of the AIMS Project ("Born-Digital Collections: An Inter-Institutional Model for Stewardship"), funded by the Andrew W. Mellon Foundation.

Hypatia is a cross-institutional effort that includes University of Virginia (grant lead), University of Hull, Stanford (Hypatia development lead), Yale, and a third party software development company called MediaShelf.

Functional Requirements for Application Development

At the beginning of 2011 the AIMS digital archivists' created functional requirements for the application. These functional requirements are primarily focused on how an archivist would arrange and describe born digital collection materials in a browser based software application. The functional requirements were used to develop technical development tasks that have been translated into tickets in an Hypatia JIRA project (https://jira.duraspace.org/browse/HYPAT). At the end of the current development cycle only a partial set of these requirements will be supported by the Hypatia application. Complete implementation of these requirements will not be complete by the end of the current development effort.

Current Status of Hypatia development

The current phase of Hypatia development will be completed at the end of the October 31, 2011. Hypatia is being developed using an Agile methodology with weeklong iterations and weekly code submissions. The Hypatia application will not be completely functional at the end of this grant cycle and it is anticipated that the institutions supported by the current grant will seek additional funding to continue developing the application. By October 31st Hypatia will have the following functionality:

- A demonstration application hosted by Stanford that contains records for all of the AIMS collections
- A polished interface that allows for the discovery and display of AIMS born digital collections
- A small subset of the AIMS collections will also contain descriptive metadata and digital objects from these collections. All of the content loaded into the demonstration application will be viewable by the public.
- The ability to download disk images and file level assets.
- The ability to create groupings of digital objects (sets).
- The ability to edit descriptive and technical metadata for collections, sets and digital objects.
- Drag and drop functionality to assist archivists in the arranging and describing of born digital collection materials.

Additional information on the Hypatia application can be found at:

Hypatia Project Wiki:
https://wiki.duraspace.org/display/HYPAT/Home

JIRA project
https://jira.duraspace.org/browse/HYPAT

The Hypatia demonstration application is hosted at Stanford and publically available at:
http://hypatia-demo.stanford.edu/