

Archive Processing Application: Use Cases

Users:

- theatre management
 - a member of the steering group
 - an ADC archivist
 - the application itself
 - a member of the public (will only use the front-end, not /really/ the same dev project)
 - an UL archivist (will only use Janus and their own tools, with EAD files we gave them)
-

(Front-end not processing:) As a member of the public, I want to:

- browse through lists of items and thumbnails of archive images, by clicking on tags or attributes, because I want to see what kinds of archives the ADC has.
- click on a thumbnail to view a larger version, because I want to be able to read the text on the page.
- view all tags and attributes associated with a specific item, because I'm interested in using it myself, or finding other items like it.
- suggest to the ADC more tags or attributes that could be applied to an image, because I know more about it myself.
- suggest to the ADC that two items are physically related, because I've looked at them both and they don't make sense separately.
- request that an item entry, its image, and/or certain tags or attributes of the entry, are removed, because they are inappropriate, sensitive, or under copyright.
- suggest to the ADC that an item's tags or attributes need more work, because I can tell that they're incorrect or inconsistent, or that there aren't enough for its existence to be useful.
- suggest to the ADC that a new digital scan is made of a specific item, because the existing scan is wrong or useless in some way.
- be prevented or discouraged from saving to disk for my own use images found on the archives website, because they are low-resolution copies and may be under copyright.

As a member of the steering group, I want to:

- remotely view some kind of visual overview of all the information currently held in the database, to gain an appreciation of how much information on how many items has been added, because it is important to know how advanced the project is, and how near completion of processing the backlog.

As an ADC archivist, I want to:

- be given an item to work on that is as-yet unprocessed, or one which is marked as incomplete and not being worked on by somebody else, because this is easier than picking an item to work on from a huge list, when I don't have any particular selection criteria.
- be guided carefully through adding tags and attributes in all applicable categories to a digital image, because I won't have received much training.
- view an explanation of each tag category when dealing with it, because it's important that I don't add tags in the wrong categories and they're not all intuitively obvious.
- view the TIFF image file relating to the item I'm working on, move through multiple TIFF pages, and be able to zoom in on bits of each, because that's where I'm getting my information from.

- create an entry for a new show that is mentioned, and tag the item I'm working on with that show, because this collects items together in a meaningful way.
- start typing a show name and view a list of matching shows that already exist in the database, with enough detail to reliably choose the right one to tag the item with, because there may have been many productions of 'Macbeth'.
- add one or more performance date ranges mentioned on the item, because these provide distinguishing information about the show they relate to.
- create an entry for a new performance venue that is mentioned in relation to a performance date range, and tag the date range with that venue, because this provides distinguishing information about the show.
- start typing a venue name and view a list of matching venues that already exist in the database, so I can choose the right one with which to tag a performance date range.
- create an entry for a new organisation that is mentioned, and tag the item I'm working on with that organisation, because this collects items together in a meaningful way.
- start typing an organisation name and view a list of matching organisations that already exist in the database, then be able to tag the item with the correct one, because this collects items together in a meaningful way.
- tag the item with how an organisation I've added is related to the production mentioned, because it may have produced, just funded, or been otherwise involved.
- begin or add to a list of cast and crew (including names and roles) involved with a show that the item is tagged with, because this adds valuable information that may interest the public.
- tag a cast or crew member as being directly a part of an organisation, either by selecting an existing one or by creating a new organisation, and describe how they are a part of it (what committee member, what period), because this adds valuable information that may interest the public.
- tag an item with a broad type such as "programme", "publicity material", "administrative document" from a predefined list that I cannot add to, because this will help the public search more effectively but I won't necessarily know whether it's appropriate to create a new category.
- modify three specific attributes of the TIFF file itself, namely 'Artist', 'DateTime' and 'Copyright' headers, with information about the creator of the item, the date of creation of the item, and the copyright status of the item respectively.
- mark an item as needing a new digital scan, because the existing scan is wrong or useless in some way.
- mark an item as not being appropriate to list publicly at all, because the entry, tags, or attributes are sensitive or under copyright.
- mark an item as having an image that should not be shown publicly, because the image contains sensitive or copyrighted information, but the basic text entry and tags are not inappropriate.
- mark an item as not being appropriate to list outside the CUDN, because the entry, tags, or attributes are sensitive (but freely available within the CUDN) or under copyright (but permitting non-commercial research use).
- mark an item as having an image that should not be shown outside the CUDN, because the image contains sensitive information (that is freely available within the CUDN), or under copyright (but permitting non-commercial research use) but the basic text entry and tags are not inappropriate at all.
- mark an item as having very little or no information on it, because the ADC can bring such items to the attention of a historian or other researcher for further work.
- mark pairs of items I've seen as physically related, because they don't make sense separately.

- refresh my view and have a pair of related items now appear contiguous, exactly as if they were a single digital file, because then I can apply tags and attributes together and save time.
- save my progress easily at any point, because I might need to serve a customer or respond to a problem away from the computer.
- return quickly to exactly where I left off, with visual indicators as to what to work on next, because I might only have a few minutes before needing to move away again.
- be able to mark the item I'm currently working on as incomplete when I leave the system for an extended time, so another user can finish it off if using the system before I do.
- create 'task lists' of existing items that satisfy specific database criteria (e.g. "broad item type is 'other'" or "item type is 'programme' but has no cast/crew roles associated") which I can then work through by clicking on each item in turn to add/edit information associated with it.
- crop an item's image to remove large blank areas, by drawing a rectangle around the area I want to keep and clicking a button; this is because reducing lots of images' area by say 10-20% each would have a big saving in terms of data storage capacity.

As the application itself, I want to

- interpret dates and date ranges in an intelligent fashion by: guessing the year based on day of week and date of month; understand a date range as the sum of a set of individual dates where appropriate; assume that a person "involved" at multiple distinct points in time was involved from the earliest to the latest point, as a date range; ... because this will save archival staff time, reduce tedium and increase efficiency.
- automatically set default values to three specific headers of the TIF file (Artist, DateTime, Copyright) when the file is first processed in the database application, because in many cases we won't know the "correct" values and it is better to resort to general informative or worst-case values than empty or incorrectly guessed ones.
- understand that the image filenames will be a ten-digit number with the '.tif' extension, and that some filenames will be preceded by either a 'd' (indicating that the item is digital-only) or an 'e' (indicating that the item is part of an external collection); 'd', 'e' and 'blank' collections each start numbering from zero, so it is possible to have items 0123456789.tif, d0123456789.tif, and e0123456789.tif; this is because it is the sanest long-term method of distinguishing but integrating archives from different epochs and sources via the same processing application.
- be happy for two or more ADC archivists to be simultaneously using me, provided they work on two separate items, and to prevent them from ever working on the same item in the same time-frame, because this will increase efficiency considerably.
- automatically add into any EAD export a field in a suitable location that can be subtly used to indicate that the EAD was generated by me, rather than by any other method.

As theatre management, I want to

- add, remove or edit the list of broad item types (such as "programme", "publicity material", "administrative document") that an archivist can tag items with, and have the tags that have already been applied be updated sensibly, because we may find the useful definitions and boundaries shift as we uncover more material.
- review what items were added or edited in the recent past, because it's useful to keep a track on how casual staff are using/abusing the application.
- be able to easily undo particular edits or groups of edits, based on time-of-edit or workstation, because some ADC archivists may not get the hang of it and it's better to spend double time than have poor-quality information.
- view some kind of visual overview of all the information currently held in the database, to gain an appreciation of how much information on how many items has been added, because

it is important to know how advanced the project is, and how near completion of processing the backlog.

- define a set of rules for creating one or more **EAD** files from the information in the database, where rules determine what tag categories the files are split up by, which tag categories the items in each file are sorted by, in which order, and which other pieces of information are included in the item description; static text may also be defined to be added into the file at specific places, and the rules may exclude certain subsets of the information from being included in the file at all; this is because we may not understand the optimal **EAD** layout when the project is still in earlier stages.
- export all information currently held in the database into **EAD** format according to currently defined rules, because these can then be imported into the **UL**'s own archive catalogue: it is important to ensure the **UL** remains 'in touch' with our collections and that University researchers are able to find our archives in as many ways as possible.
- batch-import digital-only archive files into the database, adding a set of tags to the whole collection at once, but leaving the items marked as incomplete so that **ADC** archivists will go through and add more tags to individual images as appropriate; this is because we anticipate receiving sets of digital-only files such as a set of production photos, that inherently have a lot of information in common.

Total: 57 use-cases (17/Aug/09 – SJC)