HubbuB, August 2019

In this HubbuB...

- Exporting from Archive systems to the Hub

- The EAD Editor - latest changes

- Repository descriptions on the Hub

- 'Themed Collections' are to become 'Online Resources'

- Google Analytics Statistics

Exporting from Archives systems to the Hub

1 - Exporting from Archival Management systems is a continual work in progress.

We've updated you on the exporting epic for some years now - it is always a work in progress, as we think about how data can be used, respond to contributor requirements, and try to improve on creating interoperable data.

Calm

Pipelines (processing routines) have been set up for a substantial number of Calm users, and they are working well. We have a new Calm export template, June2019 which includes:

- Export of Calm notes field

- Improved export of digital archival objects (taken from the 'URL' field in Calm)

- Addition of Calm 'Use Restrict' field to export

- Additions to the language export

- Addition of the Calm 'Access Status' field to export

- Checking the inclusion of mandatory fields to help with ensuring validation at the point of export

- Export of fields for photographs and books (i.e. Calm record types)

- Export of Calm field names within the EAD code (which could be useful for future developments and improves provenance information)

- Re-formatting of 'Physdesc' export to simplify the XML

- Export of all Calm creator names as index terms

- Removal of EHFD Notes and Revisions field, as these were causing validation errors

If you currently use the Calm template for Nov2018 or earlier, you might like to update to the newer June 2019 template. It is available from https://archiveshub.jisc.ac.uk/calm/

AdLib

We are working hard to progress the AdLib export. We are currently working on getting additional information exported from the persons and organisations database - namely (1) whether the name is for a person or organisation (although this information is not usually provided by cataloguers, so we won't usually get it even if we can export the right field content), (2) life dates of people and (3) titles.

We have already added digital objects in the export, enabled more than one creator to export, and added 'conditions governing reproduction', 'related material' and 'bibliography'. We hope this will be ready to test by the end of August.

AtoM

We have successfully harvested from some AtoM users, so we are confident we can continue to take AtoM exports. A current issue is that the AtoM harvesting endpoint doesn't recognise updates and revisions, so we can't sent a request such as 'give me everything that has been updated since the last time we grabbed your data'. Artefactual need to make the changes required for us to be able to harvest updates. Otherwise, we can only harvest the entire content.
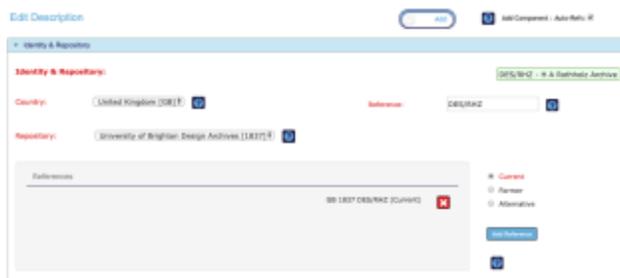
We are also working with some other systems, including KEmu and also AI+, although it is early days at the moment.

For more information see: https://archiveshub.jisc.ac.uk/archivesystemexport/

We have now run 3 exporting workshops, and these have gone very well. We are planning on an exporting workshop that focussed on AdLib in the Autumn.

Make sure you are on our Archives Hub contributors' list to receive announcements about workshops and other events and developments. Email contributors.hub@jisc.ac.uk

The EAD Editor - latest changes

The Editor now allows for adding and inserting new levels - there is a toggle button you can use to switch between adding components to a section, or inserting components above or below any chosen point.

Adding a component will create a new entry below the current set of references, e.g. if you have RHZ/1/1-12, the new component will be RHZ/1/13. Inserting a component allows you to create a new blank reference component directly after RHZ/1/1 and before RHZ/1/2, for example.

When adding components, you can choose to retain the automatically generated references, to ensure consistency and uniqueness, or switch this function off.

We also have a new lookup that allows you to search Geonames for place names. This searches the Geonames API directly, so sometimes it can be a little sticky, but it does ensure that you add a place name correctly, with a Geonames Id.

2 - Look up a name in Geonames.

Repository Descriptions

3 - Repository descriptions are useful content in their own right, and we are looking at ways of adding more value for end users.

We have started work to bring the creation and revision of repository descriptions in-house, as part of the Editor. This will make things easier, and allow us to think about how we can maximise the benefits of the repository descriptions that we have.

We are looking at providing links to the new Library Hub: Discover service and we want to explore other ways to use the repository descriptions as a way for end users to connect to other sources.

Themed Collections are to become Online Resources

We are changing 'Themed Collections' to become 'Online Resources'. This is partly in response to the feedback we have had from you. People have assumed Themed Collections means a set of collection

descriptions, maybe based around a theme. Whereas they are actually online resources that feature archives, such as project-based websites resulting from digitisation work, or interpretive websites celebrating a centenary. We plan to promote them quite a bit more over the next year, and encourage you to add to them, so that end-users can find an increasing range of archival resources from the UK that are available online.

We aim to make this change around the end of August.

Google Analytics Statistics

Google unfortunately experienced some problems a while back which means your quarterly reports were largely empty. This was not something we could fix - it was a problem at their end. But it now seems to be resolved, and you should all be receiving stats showing access. These are stats created by Google, so no-one other that folks at Google know exactly how they are compiled. We have found that they do not tally with our own logs, which is something we have never entirely got to the bottom of. But if anything this give a more conservative figure for access as Google Analytics seems to be very strict about filtering out anything that might be a robot (which is not always an easy thing to identify).

We have looked at providing reports from our own logs, but that would require significant additional work, as the stats system is not set up to send our reports in this way. In truth, stats are a minefield, and the more you look into it, the more complicated it becomes. For any open service, you have both human and machine access, and the value of machine access is hard to quantify - clear we want Google and other search engines to index the Archives Hub, but other more dubious robots sometimes hammer the service, and this can mean hundreds of thousands of hits in any one day.