

ArchiveGrid and Related Work
OCLC Research TAI CHI Webinar Chat Transcript

23 May 2013

#archivegrid

from OCLC Research:

Welcome! Thank you for joining us for today's webinar, ArchiveGrid and Related Work. Please submit your questions or comments anytime via chat to All Attendees or tweet them with #archivegrid. Thank you.

from Tarez Graban:

Great overview -- thank you, Ellen.

from Merrilee Proffitt:

Remember if you have any questions or comments, I'm watching both here and on Twitter

from Merrilee Proffitt:

The Twitter hash is #archivegrid

from Amalin Ferguson:

We are in the process of converting to EOS.Web which has an API for WorldCat in the OPAC module. Is there going to be a separate API available for ArchivesGrid? FYI, I am the National Park Service (NPS) Library Program Manager and am in a position to market ArchiveGrid among NPS archivists. They do not use a MARC based metadata management system, and we have very few professionally trained librarians, so NPS MARC records for archival finding aids will not be high in number; ditto re: finding aids in EAD, as many parks don't have archivists and those that are in parks may not have the skill to do EAD -- most archival finding aids that have been digitized and posted on NPS.gov re in pdf and HTML format so far as I'm aware. Do you have a high-level contact in the NPS Museum Program yet?

from Melissa Renspie:

ArchiveGrid Blog <http://beta.worldcat.org/archivegrid/blog/>

from Melissa Renspie:

More details about the survey <http://beta.worldcat.org/archivegrid/blog/?p=631>

from Amalin Ferguson:

We are in the process of converting to EOS.Web which has an API for WorldCat in the OPAC module. Is there going to be a separate API available for ArchivesGrid? FYI, I am the National Park Service (NPS) Library Program Manager

from Lauren Bradley:

Do you have any plans to implement EAD-CPF?

from Amalin Ferguson:

just did -- first part anyway

from Merrilee Proffitt:

Response to a question from a participant about an API for ArchiveGrid

from Merrilee Proffitt:

Great question but no we are not. Since 90% comes from WorldCat, we'd encourage use of WorldCat APIs. Would like to find a way to get the remaining 10% of data into WorldCat.

from Amalin Ferguson:

When Bruce is talking about clustering, is he basically talking about the Linked Data model (i.e., triples)?

from Amalin Ferguson:

Thanks, Merilee, but will the OCLC WorldCat API ultimately include a limit to restrict search to ArchiveGrid subset of records then? (wink wink nudge nudge)

from Merrilee Proffitt:

It's a good question, and I don't know if we'll think about slicing WorldCat into different subsets. An intriguing idea.

from Amalin Ferguson:

Ability to limit to ArchiveGrid both in the public OCLC WorldCat interface and the API would be a huge value-added feature

from Merrilee Proffitt:

clustering is not like linked data. we're talking about getting like to huddle with like.

from Richard Urban:

Are there plans for a ArchiveGrid API?

from Merrilee Proffitt:

No, if you look upward in the chat I explained why no ArchiveGrid API

from Ellen Eckert:

Amalin - we are happy to have Yellowstone

from Francesca Pitaro:

We create html finding aids using Archivists Toolkit but do not yet make them available on our website. Would we be able to send you these finding aids for inclusion in archivegrid?

from Jane Stevenson:

In the longer term, is being comprehensive a priority for ArchiveGrid? Across the US...or internationally?

from Bruce Washburn:

OCLC Research has an interest in creating an ArchiveGrid aggregation that serves our Research purposes and does not duplicate or replicate work done elsewhere. For that reason we've tended to concentrate on North American materials to extend the aggregation, as other regional and national resources have well-represented their collections. To the extent that WorldCat's archival materials representation continues to become more comprehensive, ArchiveGrid will benefit from that as well.

from Jane Stevenson:

Fair enough. We can talk sometime about the Archives Hub and ArchiveGrid - if you are interested in more UK representation.

from Peter Goodman:

NUCMC started cataloging MARC records in RLIN in the mid 1980s; all of those are supposed to be in OCLC now; records before the mid 1980s are in ArchivesUSA

from Peter Goodman:

I think all NUCMC records in OCLC are in ArchiveGrid.

from Bruce Washburn:

We've included NUCMC records in ArchiveGrid, based on the NUCMC records in WorldCat. But we do apply our usual WorldCat filter to those records as well, so if for some reason there were NUCMC records that, for example, represented published materials, they would be in WorldCat but not in ArchiveGrid.

from Amalin Ferguson:

What's your relationship to DPLA?

from Ellen Eckert:

Francesca - we can harvest HTML finding aids if their source files are online, on a webpage you host so we can instruct our crawler to pick them up and put in our index. If you want me to work with you more on this, email me at eckerte@oclc.org. Thanks! :)

from Claire Lobdell:

Are there simple ways I can format my PDF finding aids so that you can ingest them more easily? I just started as archivist at a small historical society that hasn't had an archivist or digitized finding aids before, so I have the flexibility to format them in a way that would fit multiple needs (including yours).

from Bruce Washburn:

It will be helpful if you can provide a directory or web page that lists links to just the PDF files. And we'd highly recommend taking advantage of the PDF document properties to store metadata (a title and a description) that we can use to generate a brief display in ArchiveGrid search results.

from Ilhan Citak:

In ArchiveGrid, is it possible to search within a repository's holdings?

from Bruce Washburn:

You can search in ArchiveGrid for all of the descriptions contributed by a particular repository, and narrow that search further by keywords and other facets.

from Tracy Jackson:

our finding aids have just recently gone in to the grid, and when I searched our institution I found some results in the overview that don't seem to be ours (pulled from worldcat) - can that be fixed?

from Bruce Washburn:

We would like to get more details about that search result. There can be a few different causes of a problem like that, and we'd like to get that resolved.

from Barbara Aikens:

Is harvesting of XML files impacted by robots.txt?

from Bruce Washburn:

No it isn't. We're given a specific starting point for harvesting by each contributor, so given that understanding we've not made additional use of robots.txt to guide the harvesting process.

from Ellen Eckert:

Claire - You are getting your question answered now but email me at eckerte@oclc.org and we can work with you more on this.

from Tracy Jackson: check - I'll email you!

from Peter Goodman:

Yes, this has been an issue with WorldCat for NUCMC

from Melissa Renspie:

Bruce Washburn washburn@oclc.org

Ellen Eckert eckerte@oclc.org

Merrilee Proffitt proffitm@oclc.org

from Peter Goodman:

you can also contact NUCMC at nucmc@loc.gov

from Barbara Aikens:

If we disallow the directory

from Barbara Aikens: OK, thank you!

from Bonita Weddle:

If ArchiveGrid pulls data from both WorldCat MARC record and archival finding aids, can it identify and eliminate duplicate information?

from Bruce Washburn:

There isn't a way for us to reliably and consistently match MARC records and finding aids for all contributors, as there hasn't been a consistent and universally adopted method used by contributors to link to the two different forms of the collection description. We've tended to include both, to maximize recall for those searching for these collections (as access points will appear in one description that aren't in the other), and we've heard during usability studies that the apparent duplication isn't perceived as a problem. But if you'd prefer us to favor one form over the other, we can do that.

from Amalin Ferguson:

re: DPLA, so will they be enabled to harvest records from ArchiveGrid that do have links to digital resources?

from Merrilee Proffitt:

There isn't a way to harvest records from ArchiveGrid. OCLC is working with DPLA, but the ArchiveGrid team isn't directly involved.

from Blynn Olivieri:

How about harvesting from ArchivesSpace hosting site, like Lyris? Would you need the URL in the MARC 856 field?

from Amalin Ferguson:

but that's still a lot of collections (10% of a huge quantity of MARC records is significant)

from Ellen Eckert:

Send me your questions: eckerte@oclc.org. We would like to look at URL's of where your finding aids are so we can get better idea of how we could include them in ArchiveGrid. Thanks!

from Marly Helm:

After the presentation will the answers to the chat questions be sent to us via email?

from OCLC Research: We will post the chat transcript on our website and e-mail the link to all webinar registrants (along with the links to the recording and slides). Thank you all for joining us today!