

Archivematica User Forum

Bi-monthly Call: 04/01/2021 12:00 EDT

Call info:

Meeting agenda and previous minutes can be accessed via Google Drive:
<https://drive.google.com/drive/folders/0B4hO2MJSeCa5cmtrclVMOVg0X00>

Join from PC, Mac, Linux, iOS or Android:

<https://ucla.zoom.us/j/338056800>

- Dial (U.S.): +1 669 900 6833 or +1 646 558 8656
Meeting ID: 338 056 800
- International numbers:
<https://ucla.zoom.us/join?m=nNXYLXsHF56wwy8HGCvOVBITbupMmKBH>

Participants

Please add your name and/or institution below:

1. Jamen McGranahan, Vanderbilt University
2. Tom Boone, WRLC
3. Kathryn Talbot, K-State University Libraries
4. Holly McIntyre, NASA Archives
5. Shira Peltzman, UCLA Library
6. Seth Kaufman, CollectiveAccess (seth@collectiveaccess.org)
7. Lauren Work, UVA Library
8. Sean Buckner, Texas A&M University Libraries
9. Stephen Klein, CUNY Graduate Center
10. Dallas Pillen, Wayne State University
11. Satya Miller, University of Ottawa Library
12. Sarah LeClaire, NASA
13. Katrina Cohen-Palacios, York University
14. Julie Bell, K-State University Libraries
15. Sarah Lake, Concordia University Library
16. Don Gourley, WRLC
17. John Richan, Concordia University

Archivematica User Forum

Bi-monthly Call: 04/01/2021 12:00 EDT

Agenda Items and meeting notes

- Convener is Wendy Hagenmaier from Georgia Tech (new to the Archivematica User Forum)
- Assign note-taker (Convener) - 5 min
 - Shira - thanks
- Welcome - intros for new folks & roll call: name, institution - 5 min
- Topic(s) of the day
 - Fixity tool in the Archivematica Storage Service - 15 minutes, topic idea from Holly McIntyre, NASA
 - From Holly- I'm most interested in hearing if anyone has experience using Fixity through the Storage Service across holdings that are in an AWS S3 bucket.
 - <https://www.archivematica.org/en/docs/storage-service-0.17/fixity/#fixity-application>
 - How are folks using the tool? Anyone using it with AWS?
 - Stephen: can you clarify? Archivematica looks at the checksums
 - Holly: my understanding is that after checksums are assigned through Archivematica and AIPS have been ingested to storage, you can batch-check fixity across your holdings. Have anyone performed this against content in the cloud?
 - Stephen: I set the checksums on my submission packages previous to transfer outside of Archivematica and my understanding and verified through rigorous testing is that Archivematica exams assets against checksums during transfer and ingerst and then Duracloud continually monitors checksums.
 - Dallas: for clarity, we're talking about this:
<https://github.com/artefactual/fixity>
 - Stephen: notice, on the github page that Archivematica's Fixity does not assign checksums, but rather 'checks fixity for AIPs stored in the Archivematica Storage Service.'
 - Holly: yes, and one of the things we're interested in is the price to 'touch' those materials once they're in a Glacier S3 bucket.
 - Holly: the cost part is interesting because if you do it in Glacier, for instance, you pay per 'touch' on the files, but it costs less than downloading the file

Archivematica User Forum

Bi-monthly Call: 04/01/2021 12:00 EDT

- Seth: I have some experience outside of Archivematica and the takeaway is that it can be financially damaging to do this without a very clear understanding of what your Service Level Agreement is.
- Shira: how often did you check your checksums? Where did you land?
- Seth: there are so many variables that go into a calculation of how often is often enough. One question is how large is the data, one question is how nervous are your clients. Once a week seems like the most frequent that makes sense to me, once a month is reasonable.
- Shira: it's also a waste of resources (ie, electricity)
- Seth: and wear and tear on disks.
- Lauren: AP Trust runs its own fixity service on a schedule (every 90 days) to calculate, and to avoid egress runs it on S3 servers in the same geographical zone
- John: At Concordia, our Fixity check runs once a month - to chime in on frequency.
- Seth: we were looking for a magic cloud storage service that just gives you an Md5 of what they have at a free or reduced cost price point but we haven't found that yet.
- Lauren: Folks may be interested in the NDSA Infrastructure call we had recently on fixity in the cloud - it touches on some of the topics here (see February notes):
<https://docs.google.com/document/d/1bNI1IZQ5RTuyOminiUmJ5hQyLiYn7a21EoN7P6CqZDU/edit#heading=h.pc6fia6hz282>
- Lauren: Andrew Diamond's (AP Trust developer, which is the UVA consortial distributed digital preservation repo) has on cloud fixity notes:
https://docs.google.com/document/d/1kSUNNNk7cSwZ93N_XM3z7Y82QKmp1xhoSc957Y23DiE/edit
- Julie:
<https://docs.aws.amazon.com/solutions/latest/serverless-fixity-for-digital-preservation-compliance/appendix-a.html>
- Dallas: the Digital Curation Google Group also had a discussion on this topic recently that i learned a lot from:
<https://groups.google.com/g/digital-curation/c/wSEvQScJkOM>
- Shira: there is evidence that this kind of bit rot (with AIPs at rest) may not be that common. Is there value in making the risk management calculation up front, putting effort into the protocol for repair when fixity doesn't match and running fixity less often
- Holly: we're thinking in this direction and trying to write our policy around this. Two things come to mind, one, is it risky to check fixity too often?

Archivemata User Forum

Bi-monthly Call: 04/01/2021 12:00 EDT

And two, I'm inclined to say we won't be checking fixity once per month, probably more like once a year. If we did that, would that mean that our backup would have to be from the year before? Amazon is big on saying that it can guarantee the safety of backups, but I need more clarity on backup protocols.

- Julie: you can turn versioning on and establish the timeline and then keep those for as long as you want. So if you find that there is corruption and they fixed it and you're not happy with that fix, you can go back to the first file upload using versioning and replace it with that version. I recommend talking with your Amazon representative. We don't purchase a support plan but they'll happily give you info. Also there's a cost estimator on AWS and that will give you a pretty close estimate of what it's going to cost you.
- Holly: we've also used that calculator and it's helpful but the tip off on versioning is super helpful.
- Stephen: Thanks, Julie. Does anyone know how versioning works with DuraCloud which is on top of Amazon?
- Lauren: I also think it's interesting like Shira and Dallas highlighted, to think about how/where we are thinking of risk. People thinking about restoration instead of fixity (in a targeted way) might be better able to highlight risk/target known format concerns in terms of actual access and rendering of content - which is a lot more contextual re: software needed, contextual documentation, etc.
- Shira: It would be interesting to do case study-type presentations on what the storage ecosystem is like at each institution. It would be personally helpful to me to do a deep dive.
- Julie: This is a great idea - especially from an institution that dealt with a fire recently.
- Demo from Stephen Klein - TABLED for next meeting
- Virtual Archivemata Camp Montreal May 17 & 19-21: John Richan / Sarah Lake - Concordia University (5 minutes)
 - John: it is sold out!!! But there's a wait list, so please add your name if you're interested. The camp is taking place over 4 days, there's some info on how Archivemata fits into your digi pres environment. Artefactual will spin up some virtual machines with Archivemata on it so that folks can practice some exercises and get hands-on experience.
 - Wendy: is there demand for more? Asking for a friend who's on the waitlist who is also me.
 - John: previously they were doing 2 per year so I bet they'd be happy to talk about doing more virtual events so give them a call.

Archivematica User Forum

Bi-monthly Call: 04/01/2021 12:00 EDT

- Katrina: thanks for organizing the camp, John and Sarah!
 - [add your topic ideas here, along with your name, institution, and a suggested length for the discussion]
- Verify Convener for next call on Thursday, June 3, 2021 12 PM Eastern / 9 AM Pacific - *5 min*
 - Seth Kaufman - thank you!
- Potential topics for June 2021 Call
 - Grant Hurley: File format identification tool choice (do you use Siegfried or Fido or something else?) and output handling in Archivematica for cases where there is no canonical PUID
 - How Siegfried vs. FIDO output shows in the AM METS file
 - Archivematica User Forum logistics (Shira Peltzman/Stephen Klein, 15 mins)
 - Case studies on storage ecosystem - what does this look like at your institution?