Archivematica User Forum
Bi-monthly Call:
2024/08/01 @ 9:00 am Pacific / 12:00pm Eastern
/ 6:00pm Central European

Call-in info:

Previous meeting minutes and agendas can be accessed via Google Drive:
https://drive.google.com/drive/folders/0B4hO2MJSeCa5cmtrclVMOVg0X00

NOTE: NEW ZOOM LINK! (SEE BELOW)

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

https://sydarkivera-se.zoom.us/j/83569274593?pwd=hQh1dnxvwY1fVzH36xfBNpa5rEjFBc.1

- To join by phone find the number here: https://sydarkivera-se.zoom.us/u/kcEzUveQvC
- Meeting ID: 835 6927 4593
- Passcode: e5R4KG

Participants

Please add your name and/or institution below:

1. Karin Bredenberg, Kommunalförbundet Sydarkivera, Sweden
2. Ya'qub Ebrahim, Qatar National Library
3. Tracy Popp, University of Illinois Urbana-Champaign
4. Chris Banuelos, Rice University, Houston, TX
5. Luke Menzies, University of Alabama at Birmingham (UAB)
6. Grayson Murphy, University of Alabama at Birmingham (UAB)
7. Kim Hoffman, Hamilton College, Clinton, NY, USA
8. Clayton Ruminski, United States Department of Agriculture, National Agricultural Library
9. Susan Borda, University of Michigan, Ann Arbor, MI, USA
10. Karla Roig Blay, University of Texas at Austin, TX, USA
11. Ima Oduok, Rockefeller Archive Center, Sleepy Hollow, NY
12. Deb Schiff, The College of New Jersey, Ewing, NJ
13. John Moran, University of British Columbia, BC, Canada
14. Alex Merrill, Washington State University Libraries, Pullman WA, USA
15. Amy Allen, University of Arkansas, Fayetteville, AR, USA

Archivematica User Forum
Bi-monthly Call:
2024/08/01 @ 9:00 am Pacific / 12:00pm Eastern / 6:00pm Central European

Agenda Items and meeting notes

- Convener is Karin Bredenberg
- Hellos and assign note-taker (convener) - *5 min*
  - *Notetaker is:* **Luke Menzies**
- Welcome - intros for new folks & roll call: name, institution - *5 min*
- Logistics:
  - DATES IN 2024 (next meeting in bold):

    - February 1
    - April 4
    - June 6
    - August 1
    - **October 3**
    - December 5

- Topic(s) of the day:
  - Archivematica intro videos
    - [A series of 5 videos on what Archivematica is and how it works by Artefactual](#)
    - These videos are good for newcomers, getting an idea of what Archivematica is
    - Sandbox is currently down; hopefully upgraded tomorrow morning
  - Experiences of generating reports from Archivematica
    - Option to download CSV reports
    - Q. What experiences have people had with generating these reports?
      - ?
    - Audit trail reports would be useful; Can Archivematica reports be used as audit trails?
    - Artefactual Labs/ AIPscan
      - https://github.com/artefactual-labs/AIPscan
    - Artefactual Labs/ Auditmatica
      - https://github.com/artefactual-labs/auditmatica
    - Hosted (or paid) instances of Archivematica can get support from Artefactual

- e.g. They have to set up the Fixity plugin
- If you're not hosted by Artefactual, you have to install plugins or additional tools yourself
  - Can download the CSV from the Archival Storage tab
  - Q. Has anyone experience with:
    - Archivematica API? Is there any API?
      - https://www.archivematica.org/en/docs/archivematica-1.16/dev-manual/api/api-overview/
      - *Can use the Fixity API to create tools to generate fixity reports:
      - https://www.archivematica.org/en/docs/storage-service-0.22/fixity/
    - Querying the DB's and generating reports from MySQL directly?
- Use of different workflows for ingest
  - Different workflows depending on how well we know the material
    - e.g. At Kommunalförbundet Sydarkivera lots of materials come from municipalities, and from old systems, homegrown or custom built systems, not sure what we're going to get
  - Different workflows depending on the types of materials
  - Lots of **pre-processing**, before materials get to Archivematica
    - e.g. To capture certain types of info. from certain collections
      - Processing configs
    - Might run packages/ files through DROID before Archivematica
  - Workflows that involve **purging** information that is no longer used
    - Esp. important for running env's where space is limited or potentially costly
    - Can use different tools for finding duplicates, long file paths, and cleaning up the storage drive
  - Migration from DuraCloud to Coretex (sp?), running through Archivematica to utilize some of the automations that it does
  - Converting Word to PDF is not available in Archivematica Normalization options
    - Can do manual normalization or format migration, outside of Archivematica
- Large packages
  - Large can be large number of files, or large size overall
  - Q. Experiences with managing large packages?

Archivematica User Forum
Bi-monthly Call:
2024/08/01 @ 9:00 am Pacific / 12:00pm Eastern / 6:00pm Central European

- At UMichigan, one package had 44k files!
  - Ran for 3 Days
  - Found out that a second file ID process/ microservice was running, which slowed things down
  - Was able to stop the ingest mid-stream
- Large Size - you can mitigate this with compute power, distribution of the workload;
- Large Number of Files - this is harder to mitigate with compute power;
  - Solution to tar the files up as a single file and run them as one large file
  - 10k is an upper limit
- At UT-Austin, some packages with hundreds of thousands of files!
  - Bagging at the collection level, with millions of files in the collection
  - Largest AIP - 6TB
  - Automation Tools - getting stuck w/ 150GB package
  - AIC - Archival Information Collection
    - https://www.archivematica.org/en/docs/archivematica-1.16/user-manual/archival-storage/aic/
    - AIC Search - only works if you have this column enabled
- Preingest actions to make transfer and submission easier
  - Python Scripts
    - MediaConc - run this to examine the materials, so we know if they can be ingested/ if they pass or fail the policy, improperly formed bag,
  - Normalization actions - e.g. migrate formats before ingest
  - *In many cases, it works better to run scripts and tools before ingest, rather than try to integrate these tools into Archivematica or get Archivematica to do these tasks
- **TO DO:** *Be Sure to submit topics for future discussions and meetings, especially those who are just starting out, submit questions so that some of the more seasoned users can answer those questions
- Verify Convener for next call on **October 3, 2024 12 PM Eastern / 9 AM Pacific** - *5 min*
  - Next convener will be: **Karin Bredenberg**
  - Link to email reminder text and suggested listservs: [Meeting reminder to be sent via listserv announcement](#)

Archivematica User Forum
Bi-monthly Call:
2024/08/01 @ 9:00 am Pacific / 12:00pm Eastern / 6:00pm Central European

- Potential topics for October 3, 2024 call
  - **Suggested Topics**: see today's topics if something remains and also add new ones in this list!
    - Continuation with reports and examples of Reports and Audit Trails
    - Workflows?
    - How to delete Pipelines => you can't remove the Pipeline if it has associated Locations; If you manage to remove or disassociate the Pipeline, how do you delete the Locations
    - Working with Duplicates (and near duplicates) - automations related to this