

Archivematica User Forum

Bi-monthly Call: 11/07/2019

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=nNXYLXsHF56wwy8HGCvOVBI](https://ucla.zoom.us/join?m=nNXYLXsHF56wwy8HGCvOVBI&pwd=1234567890)
[TbupMmKBH](https://ucla.zoom.us/join?m=nNXYLXsHF56wwy8HGCvOVBI&pwd=1234567890)

Convener: Grant Hurley

Participants

Please add your name and/or institution below:

Guita Lamsechi, Carleton University
Cindy McLellan, BCIT Archives
Shira Peltzman, UCLA Library
Travis Kinnear, Carleton University
Shannon Hodge, Carleton University
Holly McIntyre, NASA Goddard Archives
Mira Basara, Cornell University
Dallas Pillen, University of Michigan
Ed Warga, Texas A&M-Corpus Christi - first time attending!
Don Gourley, WRLC
Dina Sokolova, Columbia University
Stephen I. Klein, CUNY Graduate Center
Joe Carrano, MIT Libraries
Don Smith, Cornell University Library
Michelle Paolillo, Cornell University
Russell, William, William, Divya from CRKN (Canadian Research Knowledge Network)

Kari Smith, MIT
Grant Hurley

Agenda Items

Please add agenda items below; include your name and/or institution and a suggested length:

- I. Assign note-taker (Convener) - 5 min
 - Shira will take notes
- II. Welcome - intros for new folks & roll call - 5 min
- III. Topic(s) of the day - 30 min
 - Update on [Archivematica Con](#)
 - Submissions received and being reviewed
 - Re: Evelyn McLellan's request for feedback: "[Should we change Archivematica's default format normalization rules?](#)"
 - Bentley: we've gotten more comfortable turning off some of those normalization rules over time. Primarily turning off JPG/TIFF → TIFF normalization. Photo CDs were going in as tens of MBs and coming out as 10s of GBs, which seemed unnecessary.
 - Ed Warga: I haven't used the normalization processes yet.
 - Dina: we turn this off usually for AV material, TIFs and JPEGs since we already have preservation/access derivatives for these. Also we don't use PDF/As since these have gotten corrupted in the past
 - Shira: we don't turn on/off normalization rules at all — we use it out of the box, more or less
 - Cindy: I currently don't normalize for video files.
 - Grant: in some cases, if you have a TIF, Archivematica will normalize to TIF (!) — the logic is that is normalizing to an uncompressed TIF so if you have a compressed TIF, this will happen. BUT, if your source is uncompressed to begin with, Archivematica will just create a copy
 - Nick: Lossless compressed TIFs is 10:1 (efficient lossless compression) so file size balloons if you're expanding to uncompressed
 - Grant: is anyone transferring videos?
 - Ed W.: I have videos to add but I haven't transferred them yet
 - Nick: we don't do any preservation normalization from Archivematica. We don't find much value in these files. If we need to do this, we'll do it as necessary. With video transcoding, we want more specificity in our ffmpeg recipe than Archivematica can give us. For our born-digital, we can't apply a good rule for this since there are so many variables and we can't know what the best option is.
 - Grant: the current way that normalization happens isn't of great value for NYPL

- Nick: yes, and this makes it hard to sell to admin. Our general strategy is: keep the original and make the access derivative later when the need arises.
- Grant: it doesn't handle the obscure file formats (like Word Perfect files) — just the most common ones
- Kari: in terms of an overall idea of having sets of normalization rules that could be applied, what would be great would be to be able to say our Organization's digital preservation policy says we will preserve XX formats, so "for these 20 files types, we want to normalize; for everything else, don't normalize." For USA - these align with preservation formats, for Canada, align with xx standards, etc.
- Shira: it would be nice if you could select which files you'd like to normalize
- Grant: like an opt-in approach rather than an opt-out approach.
- Dina: if you go to the normalization settings it is very hard to figure out what you want to normalize or not because the list is so long (so many diff kinds of TIFs, for instance)
- Stephen Klein: the danger of it being an opt-in approach is that the normalization rules would be ignored, and this is one of Archivemata's strengths
- Kari: but isn't monitoring and setting local preservation strategies part of the role of digital preservation officer?
- Stephen: but it's better for one individual not to have to make these decisions
- Nick: but they way it's designed right now, Artefactual has made these decisions and it's hard to change them. It takes time how they have it built now, and it's hard to share rules with people.
- Shira: the [PAR project](#) is a mechanism for sharing these rules but that is technically separate from Archivemata (even tho Artefactual is involved)
- Dina: if you have different workflows for different collections or material types, the opt-out approach also makes it more difficult and time-consuming
- Kari: if you can set your processing configuration rules, could you also drop in your mcp that specifies your normalization rules? That would take some time to get going but would be great.,
- Dina: yes, having this MCP (for the FPR settings) file would be great and would save a lot of time
- Grant: there are a few key points to pull out of this discussion:
 - a) Making the settings more granular
 - b) The FPR interface is difficult to set and understand what actions are being run (relates to the opt-out approach)
 - c) The various rules and workflows out there for making normalization decisions in the first place
- Grant: does anyone reprocess archival packages if their approach to normalization changes over time?

- Shira: I wish I could make my AIPs into more perfect versions of themselves but it's not realistic :(
- Grant: on the point that Kari made earlier about having institutional format policies, in theory this is how it's supposed to work, but how many folks out there have those policies in place? Or are they relying on Archivemata as the policy stack?
- Kari: that's the crux of the matter. It is really tough to get those policies in place, so everyone's going bottom-up rather than top-down. That's why people are relying on these out of the box policies in Archivemata — because most orgs don't have this in place.
- Cindy: I have a policy. :-) Well it's just me. I try to work from the idea that if someone was to step into this job tomorrow they would be able to pick up where I left off. And I'm starting with just the digitized things. I have had a production version of Archivemata since August but I have been thinking about and working on documentation since the spring. So I have an evolving document that records my practices and decisions - which processing configurations I use for which formats. *BCIT Archives: Archivemata user guidelines* is very much an inhouse document that includes names of people to contact about certain future projects etc. I have filled out this simple table of Processing configuration options for my different processes for the different file formats. (* I cut and paste example at bottom of document)
- Stephen: if we're all subscribing to the same policies aren't we assured to be following best practices from the bottom-up?
- Kari: best practices according to what? Is this different from country to country? (in the context of what decisions are taken for normalization being on or off for any given file format)
- Nick: the way that the FPR works is PRONOM-based, so if there are flaws with PRONOM then we're stuck with those flaws/limitations.
- Kari: given that Archivemata is a community resource, who here has recs for a spec that would encompass what we're discussing so that it meets the community needs?
- Stephen: but what are we asking for? An opt-in approach, or something more granular? Also let's back up: if we're not choosing the best normalization rules then it's easier to ignore them
- Dina: but diff institutions use diff normalization rules
- Kari: yes, and this is why the approach to choose/modify your own was taken by Artefactual (I think). I don't think that they would ignore these rules if it was opt-in.
- Dina: I agree.
- Grant: there is a general interest in making normalization rules smarter. Right now they are ham-fisted
- Dina: Or even just giving the institution the choice to opt-in rather than opt-out
- Grant: is there a way of sharing what the actual practices are when it comes to normalization? Maybe there is an opportunity there.

- Shira: Would not want an all or none selection for Opt in or Opt out. But Opt In should have the ability to choose which formats to select .. or you'd do Opt In always so that you catch the things you do want to opt-in.
- Select next convener & create agenda doc for next meeting (Convener) - 5 min
- Would anyone volunteer to convene the next call?
 - Work entails 15 minutes total of prep time
 - Next convener:
 - a) Need to create a new Google doc for the agenda/Minutes and send a couple of emails. Not burdensome.
 - b) Suggestion to rotate alphabetically
 - Listserv announcement template:
 - https://docs.google.com/document/d/17Xsusz3ORJI9ag_GK8IOcUMDzKziwSv3kskJ6TDt4II/edit
- Next call is January 9, 2020 at 12 PM Eastern / 9 AM Pacific
 - First Thursday in Jan in the 2nd, so let's do Jan 9
 - Convener will be: Kari Smith smithkr@mit.edu
- Agenda topics:
 - Continue discussion about normalization - Re: Evelyn McLellan's request for feedback: "[Should we change Archivemata's default format normalization rules?](#)"

Excerpt from BCIT Archives: Archivemata user guidelines

Processing configurations

Note: Processing requirements can be downloaded and included in a SIP as more types of records are ingested this process might become more common. For now changing the settings in 'default' is fine.

The following table applies to these file formats for in house digitized materials and student thesis received from the BCIT IR: JPG, TIFF, PDF, PFD/A

Microservice	Setting	Note
Assign UUIDs to directories	No	Not required

Send transfer to quarantine	No	Not required.
Remove from quarantine after (days)	28	Leave at default. Not using quarantine therefore, not relevant.
Generate transfer structure report	Yes	Directory tree structure to be included in the AIP METS file
Perform file format identification (Transfer)	No	Not necessary for in-house scans
Extract packages	No	Not relevant
Delete packages after extraction	No	Not relevant
Perform policy checks on originals	No	Only relevant if using Policies on MediaConch. So, not relevant
Examine contents	No	Bulkextractor: forensics tool that can recognize credit card numbers, social security numbers, and other patterns in data. Not relevant.
Create SIP(s)	Create SIP and continue processing	
Perform file format	No, use existing data	

identification (Ingest)		
Normalize	Do not normalize	No need to normalize these (uncompressed and fixity checked files)