DISC Digital Preservation Working Group minutes 6/18/19

Attendees: Lee Dotson, chair (UCF), Kelley Rowan (FIU), Dean DeBolt (UWF), David Russell (GCSC), Courtenay McLeland (UNF), Krystal Thomas (FSU), Chris Levesque (PSC)

Agenda:

- Review the individual institution’s preservation needs that were emailed to the listserv. (The text of the emails has been added to the end of the minutes)
- Prioritize the group’s goals

Action/Agenda Items for Next Meeting:

- The next focus of the group’s work will be the Working Group’s first goal: “To produce a guide with specific criteria required in a digital preservation platform, reviewed on an annual or biannual basis”.
- Members were asked to think about this goal and to jot down ideas. The group’s internal Google spreadsheet now includes a tab for this purpose: https://docs.google.com/spreadsheets/d/1K3CdPoj9xWHLSE8MrwJoZe7puE2rfsm9sqyimFsixG4/edit#gid=1989084886

Discussion of preservation needs (where you are now)

- Lee: UCF has no dedicated IT in the library, and is working with campus IT on a discussion of digital preservation vs. storage.
- Krystal commented that all members of the group appear to be in different stages of “not doing it” (digital preservation).
- Krystal shared how the system selection process worked for FSU:
  - FSU was paying for MetaArchive and LOCKSS but they were not a good fit for workflow
  - The took a step back and wrote a preservation policy and framework (https://docs.google.com/document/d/1PsHc2L6kAGo-Sa1rYwZWJq6-POcJ6YVmx8FP1k_amta/edit)
  - For dark storage they are using Amazon Glacier
  - Last summer they implemented a test instance of Archivematica
  - FSU has an IT staff member in the library and work with systems people and are lucky to have their own IT.
- Lee: UCF used to have library IT
- Courtenay: UNF IT doesn’t charge them for service or space
- Kelley: FIU has been looking at Portico. There are seven staff in digital collections. Jamie (Rogers) and their Dean met with Portico, and Portico is interested in moving in their direction. Institutional data is in Archive-It, and FIU has backups. They’ve shown Portico the FDA packages but haven’t received feedback yet. FIU would like a system that works with what they have.
- Lydia: Reminded the group about the upcoming Preservica demo.
- Lee: UCF viewed a CLOCKSS demo recently and found it interesting to hear about the various challenges.
- David: GCSC has a couple of hard drives of data and has inherited a couple of digital preservation projects. They want something more standardized.
• Chris: PSC is even more at the beginning stages. They’re at the beginning stages of working with FL-Islandora and are scanning special collections.
• Krystal: how FSU deals with Amazon Glacier: Amazon negotiates the contract directly with Library IT, and Library IT budgets library materials into the Glacier contract. Amazon Glacier is a stop-gap measure for FSU. They wanted storage that wasn’t easily accessible so that data would be fairly safe from unintended changes.
• Lee: Everyone would benefit from preservation analysis such as UWF will be performing.
• How did FIU proceed with their work? They’re at the very beginning stages.
• Three main areas of concern about digital preservation systems were identified:
  o Preservation workflow. This would include how users assemble and submit materials and how disseminations are requested and delivered and the format of disseminated packages. There was some agreement that ease of submitting materials for archiving is important.
  o Technical aspects of a system’s preservation actions and preservation strategies, and are they adequate for an institution’s needs.
  o Cost: pricing model, base costs vs. additional costs. The necessity of understanding all aspects of cost of a preservation system.

**Member statements about current preservation needs:**

**UWF, Dean DeBolt:**

In the University Archives and West Florida History Center, we are generating digital files rapidly. For example, since late February, we've added 122,000 digital photographs of University events and history, and in the last month, 22,000 digital files of materials loaned to us for addition to collections (manuscripts, WWII letters, photographs, and the like). These are loaded as jpegs, pdfs, and tiffs; then there are working files, editing files and the like.

Also on the library servers is the FDA data that was returned to us. Archives staff do not have access to the servers, only the drives indicated on our department computers. Our digital files (born digital, donations, scanning products) are stored on a variety of drives (Drive L, M, N, P, S, etc.). Most of the drives have 2T capacities, so based on that, my best guess is something in the neighborhood of 10-25T at present -- and that may be low.

The drives are subdivisions of library servers and our sole library IT person assures me that the drives are backed up daily to other library servers and also remotely to a library server at our Fort Walton Beach campus. At one time the Library had an Amazon cloud account for storage but that was apparently discontinued at some point unknown to me.

We have submitted a grant to CCAHA for a digital preservation analysis. My hope is getting an external evaluator or reviewer to help guide the library administration and the archives in how we should be handing digital data.

**FIU, Kelley Rowan:**

Here at FIU we currently have about 20 TB of data needing preservation! During this interim period we are using our local network for storage which provides backups, offsite storage, and versioning – which we pay for. We have two people using Archiv-it for institutional data storage and to crawl university
websites (special collections and our IR manager). It was paid for with a tech fee grant. We are currently exploring Portico and have met with them. Portico is very interested in trying a different approach to preservation with institutions and after reviewing our needs they seem very confident about being able to provide the service we need. Workflow is going to be an important aspect of picking a preservation system for us. It would be nice if whatever we end up with will take our METS files so that we don’t have to be creating an extra set of XMLs for everything as well.

**UNF, Courtenay McLeland:**

Like UCF we were very dependent on the FDA. We began looking into alternatives at the beginning of this year and currently have the returned material from the FDA on a backed-up campus network storage space designated for digital preservation purposes. We also use the supplemental files feature in Digital Commons heavily and upload our archival versions (though generally hidden) there. We currently have about 725 GB that will need to be submitted for digital preservation. That is a combination of FDA returned material and newer items. We also receive born-digital material through whatever means a donor chooses.

**FSU, Krystal Thomas:**

FSU Libraries ceases using the FDA a couple of years back when we started our own internal search for what would be the next digital preservation system. We were also members of LOCKSS and MetaArchive at the time and could not sustain the cost of participating in that service anymore. After a year-long review of all the potential options, as well as completing a Digital Preservation Policy and Framework for the Libraries, we selected Archivematica using Amazon Glacier as our new system. Due to various reasons, we’ve not actually set up a working workflow for moving items from Islandora into Archivematica at this time (it’s one of my big summer projects). We too are working on a standard workflow for ingest and processing of born-digital materials (right now, it’s sort of however the donor gets it to us – Google Drive, OneDrive, emails, thumb drives etc.). We also use Archive-It for web archiving needs. Currently, we rely on that service for long-term preservation of the WARC files but we are exploring how we may extract and store those in Archivematica over time as well to keep our data costs done in Archive-It itself. So, we’ve selected our tools – we just haven’t quite worked out how to best use them yet.

**UCF, Lee Dotson:**

UCF relied heavily on the Florida Digital Archive since its inception. We now have approximately 20TB of packages that are in need of digital preservation. Our two biggest needs at this moment are education and a low-cost digital preservation solution. Our biggest challenges are time, money, and staffing to implement new digital preservation procedures. We are in the process of transferring and verifying our exported FDA packages. Given our budget and IT constraints, our temporary storage solution (not digital preservation) is to use Azure for cloud storage and maintain two copies of the files on local hard drives.

**UF, Fletcher Durant:**

UF Libraries currently manages over 14 million pages in the UF Digital Collections (UFDC), with an estimated 280 TB of content. Since migrating from FDA, we are utilizing two different back-up systems: 1. UF brown-storage on spinning disk and 2. UF tape-storage with off-site redundancy. Our current needs for storage are to implement a system to enact fixity checks to
ensure the stability of our storage set ups. We are also looking to set-up a workflow for ingesting born-digital files from our archival collections to migrate CDs, hard-drives, diskettes, etc. with a forensic work station and make the files available through UFDC or local access depending on rights.

**Gulf Coast State College, David Russell:**

Gulf Coast State College uses Islandora for online hosting. We don’t have a lot of born-digital materials, so there is not currently a workflow set up for ingest. Current needs are to establish stable, long-term preservation storage solutions and workflows for completing digitization projects and potentially creating an institutional repository to support our Affordable Instructional Material Initiative.

**Pensacola State College, Chris Levesque:**

At Pensacola State College, we’re really just getting started with FL-Islandora. At the moment we are in the process of developing the site and digitizing our initial objects. We are not yet ready to focus on big “P” preservation of materials.