PRESERVATION OF ELECTRONIC GOVERNMENT INFORMATION
CONFERENCE CALL, MARCH 8, 2017, 4:00PM CENTRAL TIME

Present: Martin Halbert, Robbie Sittel, Bernie Reilly, Marie Walsh, Shari Laster, James R. Jacobs, Scott Matheson, Marie Concannon (minute taker), Julia?

Agenda:

1. Review the PEGI proposal, its goals, and discuss initial thoughts about the project as a group initiative.
2. Discuss preparatory steps, especially things for us to do before the face-to-face meeting in May.
3. Discuss other aligned efforts that we are aware of, notably the nascent DataRefuge effort.

MARTIN recaps developments since our meeting in December. There is a lot of sudden and strong grassroots interest in preserving electronic government information. From CRL's Leviathan in 2013 to the more recent Data Refuge projects, this is a good time to coordinate and seek alignment.

Regarding grant funding for PEGI, UNT approached both NHPRC and IMLS. NHPRC cannot fund the proposal because PEGI's focus is government records, which is too close to NARA mission to avoid the conflict of interest. The IMLS proposal has been submitted, it is pending and we should hear back in a few month. Because the amount sought was small, CRL and UNT realized we could do this with internal money instead of the grant. UNT can commit two graduate researchers and their government information specialist staff time. UNT can also host functions such as the one to take place in May. Bernie can authorize Marie Walsh's time and some travel funds. To save travel money, we will meet in conjunction with events like ALA, and we'll each do the analysis at our home institutions.

MARTIN recaps the PEGI proposal. Our work would catalyze other projects already underway, such as Data Refuge. Data Refuge planners have excelled at generating interest up and marshaling energy in the community. The organization/coordination element was chasing to keep up with the surge in initial momentum. The organization/coordination work will help avoid duplication of effort between EOT crawl and other efforts to preserving government information by unrelated groups.

First, we will triage agencies, starting with those agencies identified by NARA as being most at risk. Second, there would also be a registry to account for all the various work to preserve government information, such as Data Refuge. Third would be a study of agency workflows. There would be interviews, analysis, and ultimately perhaps a plan for intervention.

JAMES notes that there are many organizations and entities attempting to save government information after the fact – GPO, NARA, individual libraries, etc. – but without regulations or guidance to agencies that directs them to handle this in the first place, it will be an endless game of catch-up.   James pointed toward work that could be done to get such regulations or guidance in place, notably, advocacy by library associations. ARL, ALA, AALL. Martin added SPARC.

BERNIE: The grassroots groups are rallying around the threat of data *disappearing*, but it may not be actually deleted, it may be more a matter of ineffective data management. Refuge groups are doing what NARA *ought* to be doing. Even if NARA is collecting data, it is not necessarily accessible.

GROUP DISCUSSION: NARA's function may be too passive of because they don't do enough to address findability. There is an electronic records archive (ERA) but it seems to function as something you can put stuff into, but you can't get anything back out. NARA in fact highlights their own frustrations with their limited authority to do the work we all agree needs to be done. They know that data and databases exist, but they are not sure if it can be called government record, and so they may not have the ability to put the pressure on. The electronic records archive was once considered something that would solve all the problems, but we now see that what it fails at is providing access. Plus, there is the issue of records that should be put into the ERA but never make it in. Bottom line – maybe we need to clearly identify what we want from NARA. What SHOULD they be able to do?

SHARI: NARA's mission does not involve optimizing public access. The focus is more on collecting whatever they can.

SCOTT: Data Refuge is trying to focusing on the information that is not "scheduled" as being among the material that NARA must collect. NARA cannot collect unscheduled materials.

BERNIE: wonders if the whole apparatus of scheduling records isn't completely obsolete. Maybe it was designed for a print environment. Everything about electronic information happens so much faster. Perhaps scheduling is a 19th century construct. Bernie suggests digging into architecture of the ERA (Electronic Records Archive) as a task for our group. Find out when it was designed and what the expectations were.

JAMES: The law gives executive agencies great leeway in what they schedule and how they schedule it – it is not necessarily technical issue involving speed of information creation, etc. The problem might be more a function of organizational culture and skewed institutional priorities at agencies.

MARTIN: Regarding the time line of this project.   We decided to move it up a year rather than wait for a cycle of funding. Everything can be moved up 12 months, starting with today's teleconference. We'll meet in Dallas in May. Regarding the meeting to coincide with ALA Annual in Chicago, Martin sees it as being preparatory, to talk about our game plan, more than telling the library community what we're doing. After that, we'd meet at the FDLP meeting in October. The fact that it is in Washington is important, we could get some synergy there because of the presence of Washington-based groups. Between all of these meetings, we could each be involved at our home institutions in data gathering.

Second year of project: analyzing data findings. Conduct educational awareness advocacy outreach. Meet at ALA Midwinter. One of our main outcomes will be the registry and an analysis of workflows of preservation entities. Each one of us could take one preservation entity to analyze. In June 2018 at ALA Annual, we'll do public programming, perhaps reporting final outcomes, discoveries and recommendations, including our view on further collaboration efforts. In Oct 2018 at the FDLP meeting, we'll do a final set of outreach presentations and a rollout of publications.

We are getting a lot of traction on this issue right now, the same energy that is fueling the Data Refuge is energy we can use.

SCOTT observes that the Data Refuge planners are just now beginning to talk about sustainability. So far they have been emphasizing outreach. Scott was on an organizing committee for Data Refuge and is in a debriefing phase right now. He described some of the nuts and bolts of the DR event, for example, they created guides for harvesting a new agency. Yale claimed the Department of Justice and got halfway

through their website in a day. He described the app that manages workflow. The Archivers.Space App is open to the public. They use the EOT harvest capture tool at DR events.

> The workflow app is http://www.archivers.space/
> The final public repository is https://www.datarefuge.org
> Our fork of the Penn documentation is: http://bit.ly/datarescuenhv2017

SHARI says that she along with James R and Jim A Jacobs are all going to be presenting for the HELP I'm an accidental Government Information Librarian webinar series regarding Data Refuge.

SHARI also notes that Digital Libraries Federation started a records management/transparencies group, very focused on FOIA and supporting the capacity of the government to save its own information. Digital Libraries Federation group name is "Records Transparency/Accountability Interest Group."   Shari and Robbie are in the group.

Martin says that he will fly his team to DC for more in-depth interviews with selected federal agencies. Robbie suggests bringing Meg Phillips back into the conversation, she is willing to make herself available. She talked about public/private partnership at last week's DLF meeting, and she was instrumental in the first Spring PEGI meeting.

SCOTT clarifies the connection between DR and http://librariesnetwork.org. The Univ of Penn people started talking with ARL, and the LibrariesNetwork.org site is ARL's way of supporting their idea. The presumption is that ARL libraries would volunteer to adopt agencies, like the Centers of Excellence idea. At ARL's May event, they will talk about how they can facilitate efforts going forward. ARL invited James and Jim Jacobs to write articles for their website. One article tells what libraries can do "A Long-Term Goal For Creating A Digital Government-Information Library Infrastructure," and another tells what agencies can do, "A rare opportunity to make a long-term difference." (Source)

SCOTT: ICPSR is doing Data Lumos now. It is a passive way for researchers to deposit data into OpenICPSR. ICPSR isn't really going out and getting data, but they are allowing people to upload data that they find.

BERNIE would like to dig more into existing federal apparatus for capturing and preserving records. The federal government has been in continual retrenchment regarding preserving records and making them available, #1 because money is not there, and #2 because there are legal issues with NARA's ability to compel agencies to take certain actions. Still, the federal government has a baseline responsibility to society. The tendency to generate new efforts to harvest is creating a crutch that won't hold up for the long term; we need something stronger and more permanent. He is interested to learn more about what the ERA is capable of, same with FDSys. Also the successor system.   Do we allow the responsibility that the government has always had to fade away while we shoulder it ourselves, or (worse) let the private sector pick it up? Martin suggests a system of checks and balances.

SCOTT says that although advocacy for ongoing programs such as American Community Survey may not fit the purview of this group, it should be footnoted somewhere that we do want such data collection to continue.

ACTION ITEM: We will start a website to tell what we are doing. CRL will host the webpage, perhaps linked off of their Leviathan page. Robbie will assist in providing content. Martin also wants to start an

official listserv for our group, perhaps using Google Groups, so our correspondence can be in a central place and referred back to when necessary.