

ESDN Metadata Working Group

Capital District Library Council

28 Essex Street, Albany, NY

July 9, 2014

10:30 – 2pm

Present: Susan D'Entremont (CDLC), Claire Enkosky (CLRC), Jen Palmentiero (SENYLRC), Anne Karle-Zenith (METRO), Heather Bolander-Smith (NYS Archives)

ESDN Staff: Chris Stanton, Kerri Willette

Meeting Minutes

- Project background
 - The overview timeline includes two phases:
 - Phase 1: October 2013 – April 2015. Hire staff, convene advisory & working groups, develop infrastructure, establish workflows, gather data from NY 3Rs hosted collections.
 - Phase 2: May 2015 – April 2016. Create and distribute guidelines for contribution, outreach to LAMS statewide, grow content base.
 - Kerri was hired in March. Chris started this week. A technologist, John, starts next week.
 - Still developing infrastructure and workflows.
 - Initial partners (“aggregation guinea pigs”) will include New York Heritage, Hudson River Valley Heritage, METRO Digital Collections (in Islandora), and the New York State Archive (all represented in this group). This will include many different scenarios and data types for testing the workflows and infrastructure.
- ESDN technology & workflows
 - Currently using Repox. Open source aggregation tool developed in Europe and used for Europeana. It’s used by the North Carolina hub and in Missouri. Illinois might begin using it this summer (CARLI and the state library).
- Aggregation
 - Provider metadata -> Aggregation Software / Transformations -> DPLA
 - For example, the state library is one provider with many collections, as is New York Heritage, that will each have one stream of metadata.
 - Preference is for OAI-PMH (easily available in CONTENTdm). In Islandora, MODS data can be extracted. Dublin Core can be taken from CONTENTdm.
 - ESDN will take the metadata and apply “transformations” (which is in no way “cleaning up” or changing the content). “Transformations” merely means normalization so that it will all leave ESDN and enter DPLA in one format.
 - ESDN has chosen MODS as their format because it’s easiest to map to.
- Initial Ingest Timeline

- ESDN will contribute a large initial ingest to DPLA in Q4 of 2014. ESDN will contribute all collected data in one big “splash” in late fall.
- **First steps**
 - Chris will be the primary contact for data conversations about ingestion
 - **Rule 1:** Every contributed record must include a **title field** and a **rights statement**. Type and language fields would be fantastic, too.
 - **Rule 2:** No newspaper collections yet (a single issue of a newspaper is not problematic)
 - **Rule 3:** No EAD finding aids yet
 - DPLA Rights Working Group will be meeting in October about progress on standardizing rights statements. Greg Cram of NYPL is a member of the Working Group and will provide an update on any progress at a METRO event immediately following the DPLA meeting. Anne will pass on any relevant information to the Metadata WG.
- Things to keep in mind
 - DPLA is a pointer to your full records.
 - Sharable data is different from standards compliant data. Some things don't need to be shared.
 - Don't let metadata shame get you down. Don't not share just because a collection hasn't been cleaned up yet. Harvesting is done periodically (it's not done only once).
 - Un-mapping fields can improve your shared data.
- **Prioritizing Collections**
 - Chris and Kerri will be asking for initial collections in the coming weeks. Refer to the guidelines laid out in the Prioritizing Collections for ESDN Criteria and Considerations list.
 - Kerri would like at least 50,000 items for the first splash.
 - There will certainly be a variety of data scenarios, so look for ideal collections to start.
 - Start with a top five.
 - What ESDN can't fix, often, ESDN can at least get into a state that DLPA can fix as they need.
- **Communication with collection contributors**
 - Sample letter included in packet.
 - Initially, a more legal contract was expected. The ESDN Advisory Group concluded this was not feasible.
 - Put the template letter on collection contributor letterhead and reword as desired. This is necessary per contributor (not per collection). Send them to Kerri.
 - This agreement is not necessary before ESDN begins working on data.
- Next steps
 - Chris will walk contributors through the spreadsheet for ingestion. It is Google Docs based, so a google account is necessary. Each contributor will share the document with Chris so that comments can be made as needed. This will be the main tool in identifying collections we want.
- Metadata discussion from Chris

- This process is expected to be a one-to-one working relationship.
- Chris Stanton: cstanton@metro.org, 212-228-2320 x115
- One spreadsheet (tab) per CONTENTdm collection.
- We will fill in Source Collection field with a content sample & our mapped DUBLIN CORE field.
- Write “no map” if not used.
- “Use” fields come straight from DPLA map.
- Mapping consistently within a collection is more important than mapping incorrectly.
- List other schemas.
- Chris will send out an expanded list of resources after the meeting
- Chris will set up the spreadsheets tomorrow morning.

<http://cdm16694.contentdm.oclc.org/oai/oai.php?verb=ListSets>

This is the list of collections available for OAI.

http://cdm16694.contentdm.oclc.org/oai/oai.php?verb=ListRecords&metadataPrefix=oai_dc

This is (a 200 limited) list of records as they would be harvested by ESDN.

- Committee business
 - The ESDN site will have the meeting minutes, meeting calendar, and purpose statement posted.
 - The group likes the purpose statement as written. It’s concise and broad.
 - Meeting in person will be determined ad hoc. Ongoing group communication could be done through a google group.
 - Kerri would like to have everyone’s five collection sample by October.
 - HRVH will be aim for mid-September
 - NYH will aim for next week
 - METRO will aim for September
 - NYSArchives will need to configure OAI