Hydra Newspapers Interest Group Call: 2016-12-01

Time: 1:00 PM EST / 10:00 AM PST

Call-In Info: 712-775-7035 (Access Code: 960009)

Moderator: Eben English (Boston Public Library)

Notetaker: Nabeela Jaffer (Etherpad link: https://etherpad.wikimedia.org/p/Hydra_Newspapers_Interest_Group__2016-12-01)

Attendees:

- Eben English (BPL)
- Joshua Westgard (U. Maryland)
- Brian McBride (University of Utah)
- Gordon Leacock (University of Michigan)
- Michael Friscia (Yale University)
- Michael Kanning (University of Georgia)
- Jocelyn Triplett (University of Virginia)
- Peter Binkley (University of Alberta)
- Betsy Post (Boston College)
- Steven Anderson (BPL)
- Nabeela Jaffer(University of Michigan)
- Stefan Boddie (Veridian)
- Drew Heles (Johns Hopkins)
- Jenn Colt (Cornell)

Agenda:

1. Introductions (brief (no more than 2-3 minutes per institution) description of your current newspaper efforts/situation, why you're here, etc.)
    1. BPL: No repositories at the moment. Large content being digitized. Have a grant
    2. Boston College: Use same workflows for journals, 155,000 pages in 2 separate instance. Interested in whats happening with Newspapers
    3. Yale: Instance of Content DM, digitized from 1866 to 1999, interested in moving it from Content DM to Hydra, Holdback is article level..
    4. U. Maryland: National Newspaper Digital Program member, submitting to Library of Congress, ongoing project to digitize the student newspaper; going to go into Fedora instance with BlackLight. Working on content modeling for Newspaper
    5. U. of Virginia: Had a grant to work with public library, put it in a catalog and created a custom view. In general struggling with hierarchal newspaper
    6. Princeton: Digitized newspapers and periodicals in collaboration with local and historical libraries. All these are currently hosted in ...
    7. John Hopkins: A little over a 1000 issues of student newspaper, in the process of moving to Fedora based repo, looking at Sufia as an option, put the newsletters into Sufia, wasn't a good experience, looking for some other options
    8. Georgia: Scanned newspapers currently in legacy system, hoping to integrate into Blacklight app in the medium term. Also working on ChronAm instance
    9. Michigan: Soft launch of blacklight based app with IIIF for digitized students papers
    10. Cornell: Looking for options to move their digitized newspaper
    11. Alberta: Digitizing newspaper for several years, mostly METS/ALTO with article segmentation, will follow the example of reprocessing old Olive materials
    12. Utah: Recently switched from Content DM to custom system, starting to use Hydra as a RD project. Actively looking into a Hydra based solution with IIIF
    13. Veridian: Helps with digitization

2. BPL/Utah grant update
    1. scope, timeline, status: not received confirmation from IMLS, expected to hear back in a week. submitted the preliminary proposal in September, Jan 13th is the deadline for full proposal. If gets accepted, work will start in Summer 2017. Two year project. Aiming for article level segmentation. Will be based on current Fedora version
    2. 250,000 ask from IMLS to develop gems (pluggable admin and display gems). Something you can add to existing application. Although it can be used on its own too. Discovery and Display gems will be more like Blacklight IIIF dependency
    3. comments on preliminary proposal? (http://static.digitalcommonwealth.org/IMLS_Newspapers_Proposal.pdf)

3. Assessing the state of the HydraSphere with regards to newspaper content
    1. how to collect information? how to gather requirements?
        1. Start with what we have, may be in a matrix form
        2. Like DSPace interest group, collect use cases (Data Migration from DSpace to Hydra)
        3. What type of data do we want to get? When they want to migrate? What are they looking for?

4. What should the Hydra Newspapers Interest Group focus on?

1. Alberta: Looking forward to exposing the integration of IIIF, want to make sure the foundational layers are strong enough. Better to do it together than to do it alone. Fundamental layers are PCDM modeling, article level, Full OCR Text, everything tied together with well-thought-out URIs. Identify key components.
2. Princeton: Struck by the convergence of the community on Newspaper and IIIF, there is a growing demand of representing newspapers and periodicals. It will be a great opportunity to coordinate the activity. Involved in IIIF newspaper community. Full Text and Marked-Up text is stored in Solr poses challenges..
3. BPL: Come up with a set of recommendation about how to store data, full text OCR
4. A way to keep in touch with other related groups. Great to have standing agenda items to do consistent information sharing from other newspaper-related efforts, e.g., IIIF newspapers SIG just formed.
5. Michigan: Interested in best practices for working with various components used for newspapers in Hydra. The x-api extension architecture of Fedora. Interested in discussion on what are the legal and other barriers, and how people are dealing with them.
6. Yale: Interested in usability, user testing, user experience, and accessibility audit.
7. Utah+BPL: Help us with grant application, provide feedback on docs

8. activities, roles, deliverables, etc.
9. list of possibilities on Samvera Newspapers Interest Group page

5. Meeting schedule
   1. frequency? monthly time? 1pm EST, 1st Thursday of the month, Next meeting (Jan 5th, 2017. 1pm EST) medium?

6. PCDM profile for newspaper content
   1. review of some existing efforts: share ideas and best practices
      1. U. of Maryland: https://wiki.duraspace.org/download/attachments/77447979/maryland2016hyrdraconnect.pdf
      2. National Library of Wales: PCDM Mapping for Welsh Newspapers (NLW)
      3. PCDM Mapping for Welsh Newspapers (NLW)CDM Profile: How you can model different types of content. There is a template to create a PCDM profile.
7. Other business