

American West Partners Meeting Minutes

Digital Library Federation
New Orleans

Sunday, April 18, 2004

Attending: Dan Greenstein, (CDL), Robin Chandler (CDL), Steve Toub (CDL), Nancy Allen (Univ of Denver/Interim Exec Dir, CDP) Brenda Bailey Hainer (CO State Library), Curtis Fornadley (UCLA), David Loy (CDL), Roy Tennant (CDL), John Dunn (Indiana U), Perry Willet (Michigan), Christie Stephenson (Michigan), Dan Greenstein (CDL), Leslie Johnston (U Virginia), Laine Farley (CDL), [Mike McKenna (CDL), David Greenbaum (UCB IU)], Anne Lally (UW)

1. Dan Greenstein – welcome and thoughts about the American West project

Background: CDL is transitioning from a digital library focusing on end users to one serving on UC libraries who serve user directly. CDL builds services, and provides collections that other libraries can integrate into their environment. CDL fades into background, becomes a utility, “Intel Inside”. CDL asks what do we need to give to our libraries to enable them to offer high quality services? Libraries have best understanding of local holdings strengths, user needs, and how best to leverage that understanding with tools we can supply.

The American West partners were chosen because of personal connections, knowledge of relevant collections and services, institutional readiness, ability to contribute and willingness to innovate. For example:

Michigan: early developer, massive collection, layered services view, innovator.

CO: pool of content, experience of interoperability, needs assessment experience, pushing toward community idea—shared understanding and enthusiasm of distributed members.

UVa: source of rich content, pursuing model of separating content and services, how to store and manage content so that it can be delivered through various services, sought community to validate local effort.

Indiana: exploring same issues, making different specific architecture choices but similar approach, experience with other formats (streaming data).

CDL: OAC as a community bigger in scale, smaller in ambition, co-investment as a value. Collaborations work best when everyone is doing something different.

LC: content, strong interest in having services built on top (can't do a lot of interpretation themselves), have to deliver in a service neutral way.

So to summarize:

- Need for a Community Driven Model for Shared Digital Libraries
- Concept: layered service model – can't be proven only at UC
- How to sustain? Intersections of self interest
- Precedents: Colorado Digitization Program & OAC consortia
 - Becoming an enabler – leveraging your's institution's expertise to help other institutions

Dan's vision of American West's core goals:

- Pool our content
- Build a suite of tools for the 21st century “special collections librarian” to select virtual collections of interest / integrate with local content
- Investigate means to build configured views or instructional content from the virtual content
- Investigate means to Integrate instructional products into the repository
- Question: How can this project benefit your institution?
 - We'll present at this meeting areas that CDL has been working on for this project (OAI Harvesting, Metasearching, User Assessment, Layered Service tools – such as Scholar's Box, D-Space) and begin to surface what work your institutions are doing in these areas and begin to explore our shared agenda in these areas

Partner's Discussion / Response to Dan's thoughts:

- Collaboration needs a shared agenda / specific work to do
- How do we Manage or synch our “self-interest”
- Using American West as a means of Identifying the shared agenda and consider how best to organize to accomplish the work; noting that synching our self-interest will be a larger issue that we should keep in the back-of our minds as we move forward

2. Metasearching / OAI Harvesting Presentation (Roy Tennant and David Loy)

CDL is building a prototype harvesting service. Thus far we've harvested 100,000 parter records -- samples from contributors using the VaTech Perl havester, simple Swish-e indexer and Simple Perl to provide access

Discussion:

- Data providers can't anticipate needs of service providers. Shouldn't service providers declare mappings they want. Can service providers declare sets, subsets? Can sets be dynamic, e.g., searches?
- Others who have dealt with these issues: RLG Ricki Erway, Cultural Materials: have been trying to do this from service provider perspective.

- Columbia: tools for content analysis. (Judy?) Diane Hillman has done a lot of analysis; OAlster hasn't done a lot of analysis
- Collection description: what do service providers need to know to interpret the data; what tools can help content providers provide this information Other work on collection profiles may help (SRW, NISO metasearch, etc.)
 - Develop list of what service provider would like to get (related to user needs).
 - May have to add two different types of services to test flexibility of data provided. Different audiences with distinct needs.
 - Simple DC is too simple; need to get qualified if there is richer data. What do we need to know about higher level of data; what should have been in a profile to help in advance.
 - DC is the low threshold; Need to surface richer sets of metadata (MARC, MODS, qualified DC)
 - Expressed a desire for dynamic searches of content and review by a subject specialist to identify relevant sets
 - Need to have a better understanding about how and why harvestable collections were assembled to understand decisions about how the metadata was mapped

Action List::

- CDL will do more harvesting (its still preliminary) and analysis so that we can identify what we have and then ask can we get better richer sets? CDL will work on a "list of concerns with metadata"
- Establish a data profile for collections and sets: what additional information do we need to identify, i.e. what do you wish was available to understand the collections to make decisions as a service provider
- Envision multiple user scenarios - and possible profiles for users
- Metadata issues – we need a clearer vision of responsibilities of service providers and data providers -- Future Meetings: DLF Metadata service providers discussion (Tim Cole, Kat Hagadorn, Diane Hillman, etc.) look closely at CDP Dublin core builder for scalability
- Look in depth at CDP STD & crosswalks
<http://www.cdpheritage.org/resource/metadata/index.html>

Monday, April 19, 2004

Attending: Attending: Dan Greenstein, (CDL), Robin Chandler (CDL), Steve Toub (CDL), Nancy Allen (Univ of Denver/Interim Exec Dir, CDP) Brenda Bailey Hainer (CO State Library), Curtis Fornadley (UCLA), David Loy (CDL), Roy Tennant (CDL), Christie Stephenson (Michigan), Dan Greenstein (CDL), Leslie Johnston (U Virginia), Laine Farley (CDL), Ann Lally (UW), Stephen Davison (UCLA), Mike McKenna (CDL), David Greenbaum (UCB Interactive University)

1. Recap of yesterday: Reiterate need to develop a shared agenda; synch our self interests, manage the process in order to sustain it.

2. Project calendar/Task List reviewed.

Continue to meet at DLF (Baltimore – October is next)
Internal CDL team plus other related project teams (e.g., Metasearch)
CDL Staff: User assessment consultant hired; metadata coordinator, project manager positions are in recruitment; web crawler/OAI harvester consultant has been hired.

Held design team meeting with Brenda – Colorado Digitization Program -- to develop process for user assessment.

Next goal is to develop functional requirements for tool suite based on user assessment and conduct a technology review of similar products (commercial, open source) to Scholar's Box. Project tool development starting in October 2004.

General questions to keep in mind:

- Where does/can AW serve your institutional self interests?
- How would/could it be conducted accordingly
- How do we shape (longer term) a shared agenda and what should be on it?

3. Shared Agenda with partners – possible areas to work together in addition to Metadata harvesting

a. Text interoperability

- Develop a vision of services: what is the library tool set and what is the scholar's tool set:
 - 1) search across texts: to do deeper discovery and
 - 2) gathering texts to do work – textual analysis: suites of tools needed by users for texts
 - 1. 3) how can texts encoded by scholars be submitted to a repository
- Investigate the MICH, Cornell, and Gottenburg Math text interoperability project and SRW to determine if there are ideas that can be applied to furthering text interoperability with American West partners: Mich, UVA, Indiana, CDL
- Other models to investigate: Oxford Text archives?

b. A new area of interest: DC / EAD convergence: how can CDP's DC operate together with the University of New Mexico's EAD consortia

- Examine UIUC implications of harvesting of EAD

c. Interoperability of audio tapes / oral histories CDP project seeking funds to generate a useful transcript from the audio file – looking at the Michigan model

d. Group: User assessment

e. Instructional technology

f. Making EADs more useful in a instructional environment; what would need to happen to help a student use an EAD;

4. Partner Content Discussion

[see handout "Collection Descriptions"]

UVa notes that if their integration project is delayed, they could make their content harvestable.

Any new collections?

LC has identified a few; UW has more including the Western Waters Digital Library. CDP: working on historic newspapers, but not harvestable (in Olive); many other smaller collections

Other institutions: which ones, when might they be approached? After we have a better sense of data provider/service provider roles. (Utah, UT Austin)?

Scope question:

Tension between desire (expressed by librarians) to know "what's in it", "why would I send users there" versus a new model of special collections librarian who might craft a collection to meet their local needs. What can be delivered as an initial virtual collection versus a longer term collection that might be less well defined. Not mutually exclusive—could present examples of packaged collection with tools to create others.

Virtual collection search dimensions could be based on time, geography, content (topics or disciplines)

Are there other frameworks we could use to target content?

**another collection development strategy—see what this aggregation brings forth and then identify gaps for adding other collections.

There has been curation on the micro level with each collection.

- CDP did precoordinated searches to reveal what was common across them.
- CO Dept of Education organizes by content area with lesson plans and CDP links. User generated interpretation of collections.
- Curated exhibit view: pull from multiple collections.

Tag line: encyclopedia – as a starting point, not an ending point. But does this imply editorial process, comprehensiveness, etc.?

Curator needs to know roughly what is in collection – topic mapping?
Need to develop a few curated views and document the process.

We need to frame the collection to begin with. How does UVA identify what should be included out of their encyclopedic text collections.

One proposal: American West includes all states west of Mississippi River, all time periods, all aspects? And anything related to westward movement. Exploration and discovery.

5. UCLA Dspace (Curtis Fornadley)

Next version (1.2) will export to METS, support indexing of full text. Uses DC; supports OAI as data provider.

For AmWest, plan to import content (learning object) from Scholar's Box into Dspace. Can include different formats (SCORM, ppt most popular).

UCLA D-Space project seeks to identify incentives for faculty to submit objects. Must be simple (esp. metadata). UMich starting to experiment with librarians who partner with faculty to add metadata. eScholarship uses a model of having administrative support from research units to assist in preparing materials for deposit. What infrastructure could be used for this project?

6. User Assessment (Laine Farley)

Alex Wright consultant has held sessions with librarians (two groups of four) and faculty (two groups of 2; one alone); graduate students this week (three groups of five); community college faculty, and K-12 teachers later in May. His report will be posted on the American West website.

7. Scholar's Box Tools (Steve Toub)

During the User Assessment we hope to get ideas on key functionality required for scholar's box tools for research / learning and librarian curator tools; and on

platform requirements. We are getting feedback on wireframes for tools as part of the assessment.

8. Wrap Up

Next meeting: similar format before DLF in Baltimore, October 24/25