• Agenda Day Two

  – Check In and Roll Call
  – Review Agenda & Protocols
  – Review additional questions and responses not covered in the initial e-survey that participants asked during Day One activities
  – Off-line activity: what can this group do together?
  – Next steps
1) Does your repository include born or reborn digital materials?

- **USU**: yes some (ETDs, posters in e-format, word docs, pdfs)
- **CSU**: yes
- **UMN**: yes
- **Alaska**: some, but not much born digital
- **OSU**: yes
- **Arizona**: yes, various repositories hold both
- **Florida**: Yes
- **UC Davis**: Yes for both repositories (Davis has two)
- **UC Riverside**: Our general depository for UC wide (e-Scholarship) includes e-dissertations and journal articles submitted digitally to publishers. The Water Resources Collections and Archives (WRCA) contains both born and reborn digital materials as does our citrus digital collection.
- **Purdue**: Yes, we are evenly balanced in looking at both
1) Does your repository include born or reborn digital materials? (cont.)

- **WSU**: Both (Born only)
- **UNR**: No, but could if needed
- **UWYO**: Yes, mostly student scholarship born digital right now; Herbaria collection is focus of “reborn” digital;
- **NAL**: Yes, repository is comprised mostly of reborn digital
- **Cornell**: Material in e-Commons (our D-Space installation) or Locale have both born and reborn digital. Our DLXS collections are reborn digital materials. Our materials in HathiTrust and the Internet Archive are reborn digital materials. (Also) for institutions that *want* to include reborn digital material an overview of digitation options might be discussed in terms of cost, robustness of metadata, “ownership” of images, and additional discovery mechanisms provided by the digitization organization. Secondly, a discussion on workflow for born digital and scanned materials might be fruitful.
- **Hawaii**: Both born digital and scanned
2) Is your institution using RDF or Linked Open Data?

- **USU**: No
- **CSU**: Not at this time, but looking at Linked Open Data
- **UMN**: Not yet, but working to do this
- **Alaska**: Not at this time, but excited about it
- **OSU**: Investigating and testing use of RDF
- **Arizona**: Planning to implement for “Global Rangelands” repository this coming year (in CALS); Library is not using at this time
- **Florida**: Not at this time
- **UC Davis**: No for both
- **UC Riverside**: Not at this time. We are considering RDF as a component of our Next Generation Technical Services (NTGS)
- **Purdue**: Purdue e-Pubs repository is not yet leveraging either RDF or Linked Open Data
2) Is your institution using RDF or Linked Open Data? (cont.)

- **WSU**: (Library response) Dublin Core, RDF, METS, and ORE (unknown – Karla Dolph)
- **UNR**: No, maybe
- **UWYO**: No, but output of many public applications are OAI harvestable, based on MySQL/PHP and can output in XML
- **NAL**: Not currently, but we are moving in that direction
- **Cornell**: DLXS and Locale definitely do not. I can't speak for eCommons, though I doubt it. (Also) This questions was asked in the context of the “Future Challenges based on the conference theme?” that was associated with each of the session areas. The challenge in this case is how (or if) existing repositories can expose their content as Linked Open Data/RDF. VIVO certainly uses LOD/RDF but it’s not really used as a repository. However, [given] the objective of the workshop, the ability to harvest information about the contributors to those repositories, essentially creating a network of scientists at land grant universities working the Agricultural domain, would be quite useful.
- **Hawaii**: We are planning on using RDF soon, but note that we are currently OAI compliant.
3) What system is your institution using?

- **USU**: Bepress Digital Commons platform for IR; CONTENTdm for digital collections created from special collections and archival material
- **CSU**: DigiTool
- **UMN**: DSpace; Islandora (Drupal & Fedora)
- **Alaska**: ContentDM; possibly Fedora or Dspace for future
- **OSU**: Dspace for IR; ContentDM for photos and media; looking at Islandora from ContentDM
- **Arizona**: Dspace; ContentDM; OJS; UAir (home grown; Drupal)
- **Florida**: SobekCM (integrated digital collections and IR)
- **UC Davis**: ANR = SQL + COLD FUSION; ESCHOLARSHIP = UC VERSION OF BEPRESS SOFTWARE SUITE
- **UC Riverside**: We use DSpace for some projects. In addition, our Water Resources Collection and Archives (WRCA) uses CONTENTdm, Web Archiving Service, and eScholarship.
- **Purdue**: Digital Commons, CONTENTdm, and HubZero
3) What system is your institution using? (cont.)

- **WSU**: Dspace (web-based online store for Extension publications)
- **UNR**: ContentDM
- **UWYO**: Colorado Alliance of Research Libraries (Denver); uses software: Drupal, Fedora, Fedora ingest content stored on DuraCloud (commercial); includes preservation
- **NAL**: Fedora (in development); moving from Dspace
- **Cornell**: Institutional repository, e-Commons, is a DSpace installation. Other digital collections are delivered via DLXS, Greenstone, Internet Archive, and HathiTrust. Also have the ESMIS (USDA statistics) system. (Also) CULAR and Harvest are both Fedora implementations. (Also) The ESMIS site is using a homegrown solution for managing documents. However, it’s designed so the storage layer is “pluggable” so theoretically a DSpace or Fedora installation could be used to store the documents but keep the user interface intact.
- **Hawaii**: DSpace
4) Is the content in your repository open access?

- **USU**: Of course
- **CSU**: Yes, but we do have some ETDs that have a one year embargo
- **UMN**: Yes for IR and Subject repository and most of our media repository
- **Alaska**: Majority is open access
- **OSU**: Yes
- **Arizona**: Yes, with a few exceptions for restricted items
- **Florida**: Yes
- **UC Davis**: ANR = Depends on user owner definitions; ESCHOLARSHIP = Yes
- **UC Riverside**: Yes
- **Purdue**: Yes, primarily
4) Is the content in your repository open access? (cont.)

- **WSU**: Yes, (I guess so, pdf’s are open to the public to view/print)
- **UNR**: The ag collection is open access; some other collections are not
- **UWYO**: Yes, for now everything is open, but some archival collections may be restricted when they come up
- **NAL**: Yes
- **Cornell**: Generally, I'd say yes, but it depends on what is meant here. (Also) If Open Access means that the contents of the collection are viewable to the general public, then DLXS and Locale qualify, and eCommons is *mostly* public. Of the three, DLXS is the only one that is designed to prevent users from downloading materials en masse. (Also) I don’t think that throttling downloads applies in the context of this question. Perhaps a better question would be “is most of your content open access and how do you deal with content that is restricted?”
- **Hawaii**: Yes
5) Do you have long-term digital preservation plans in place?

- **USU**: Yes, created this year, but continue to refine
- **CSU**: No, but we are working on pieces of it...
- **UMN**: Yes, but still working on some pieces
- **Alaska**: Started but waiting for digital projects position to be hired
- **OSU**: No
- **Arizona**: In progress, plan to develop this year
- **Florida**: Yes
- **UC Davis**: Yes for both
- **UC Riverside**: Plans are in development through a platform provided by the CDL and via HathiTrust (hopefully)
- **Purdue**: No, we are still exploring digital preservation options
5) Do you have long-term digital preservation plans in place? (cont.)

- **WSU**: No, in process (Not at this time)
- **UNR**: No
- **UWYO**: Yes, the Islandora/DRUPAL/Fedora/DuraCloud is only repository software that has long-term preservation as part of its makeup.
- **NAL**: We have preliminary plans in place and hope to improve and finalize them over the next year.
- **Cornell**: Under development as part of the creation of the preservation repository. Not done, yet, but [maybe by] the next year. (Also) CULAR (CUL Archival Repository) will be considered to be in production at the end of September. It is a Fedora-based archival repository that will NOT be publicly accessible.
- **Hawaii**: Yes for certain formats as part of the Dspace Community; also have weekly back ups to offsite locations
6) Are you participating in any Google scanning projects? If so, are ag collections included?

- **USU**: No
- **CSU**: No
- **UMN**: Yes, as part of the CIC agreement with Google Books, but not much ag material
- **Alaska**: No
- **OSU**: No, our Extension pubs are harvested by Google, but not scanned
- **Arizona**: No, but IR content is harvestable by Google
- **Florida**: No, all digital collection and IR material is harvestable
- **UC Davis**: ANR = YES TO BOTH, BUT NOT PART OF THE REPOSITORY; ESCHOLARSHIP = YES TO BOTH
- **UC Riverside**: Yes. We do not know, however, if specifically ag related materials have been included
6) Are you participating in any Google scanning projects? If so, are ag collections included? (cont.)

- **Purdue**: Yes, participating in Google’s gov docs scanning project, which may include some material related to ag
- **WSU**: No (No)
- **UNR**: No
- **UWYO**: No
- **NAL**: We are not currently
- **Cornell**: Yes. Our agriculture, life science, human ecology, and other related material were the first to be digitized. I would say 250,000+ ag related volumes were digitized from our collections.
- **Nevada**: No
- **Hawaii**: No
7) Are you using the Hathi Trust infrastructure for any of your digital collections?

- **USU**: Yes, contributed USU Press books and will probably do more in the future
- **CSU**: No
- **UMN**: Yes, only public domain content that is scanned by Google Books from our collections is given back to us and put in the Hathi Trust; also some local Minn. historical publications
- **Alaska**: No
- **OSU**: No
- **Arizona**: One project “TRAIL” is involved with Hathi Trust, but no others
- **Florida**: Not yet
- **UC Davis**: ANR = NO; ESCHOLARSHIP = YES BECAUSE OF A CALIFORNIA DIGITAL LIBRARY MANAGED GOOGLE SCAN PROJECT FOR THE REGIONAL STORAGE FACILITIES AND A COUPLE OF UC CAMPUS INITIATIVES
7) Are you using the Hathi Trust infrastructure for any of your digital collections? (cont.)

- **Purdue**: Not presently using Hathi Trust for repository program
- **WSU**: WSU is joining the Hathi Trust but not as a contributor (No)
- **UNR**: No
- **UWYO**: We are a Serials Solutions customer and have enabled/access capabilities to the Hathi Trust collections through Serials Solutions; but not contributing digital collections
- **NAL**: We are not
- **Cornell**: Currently only our Google images are in HathiTrust. This may change in the future.
- **UC Riverside**: Yes, for printed public domain monographs
- **Hawaii**: No
8) What metadata schema do you use?

- **USU**: Dublin Core
- **CSU**: XML, DC-XML, MARC-XML
- **UMN**: Dspace uses Dublin Core; Fedora uses MODS; all will use Islandora and MODS in coming year
- **Alaska**: Mainly Dublin Core; Star Archives proprietary format, but OAI-PMH compatible
- **OSU**: QDC (qualified Dublin Core)
- **Arizona**: Dublin Core-based across repositories; have also incorporated other schemas based on unique user needs
- **Florida**: METS/MODS, auto-transformed into MARCXML, and qualified Dublin Core as well
- **UC Davis**: FOR BOTH, THERE IS A MINIMUM SET OF FIELDS TO COMPLETE, SO THERE IS A STANDARD FORMAT FOR EACH REPOSITORY RECORD. ANR USES THE NATIONAL AGRICULTURAL LIBRARY’S THESAURUS FOR KEYWORDS, ESCHOLARSHIP ALLOWS DEPOSITORS TO USE OWN WORDS TO COMPLETE EACH FIELD
- **UC Riverside**: Will provide response later today
8) What metadata schema do you use? (cont.)

- **Purdue**: Content exposed in both qualified & unqualified D.C. streams though OAI gateway
- **WSU**: (Library response) records are exposed and are accessible in RDF, an essentially unused capability at this time, but the foundation supports semantic searching of the repository. HTML source includes semantic info, not just mark-up (No metadata used at this time – Karla Dolph, WSU Extension)
- **UNR**: Dublin Core
- **UWYO**: MARC, Dublin Core, Darwin Core; the Islandora/Drupal/Fedora stack is open source & can accommodate new metadata schema
- **NAL**: Dublin Core and MODS
- **Cornell**: For DLXS, we use TEI_Lite. For Locale and eCommons, Dublin Core. In all three cases, the schema are probably not used strictly according to standard.
- **Hawaii**: Dublin Core
9) Is your repository ADA compliant?

- **USU**: Use standard and widely used platforms and interfaces which we assume comply with ADA
- **CSU**: No
- **UMN**: I don’t know
- **Alaska**: Not sure; different levels of ADA compliance
- **OSU**: I don’t know
- **Arizona**: Not as yet, but it was identified as an issue for all repository projects this year
- **Florida**: Yes
- **UC Davis**: Yes for both (CDL has accessibility compliance goals for all web applications; check lists available from 508 guidelines; W3C, and XHTML – with W3C validation; testing from Cynthiasays and lynx
- **UC Riverside**: Not sure. Will check further into this.
9) Is your repository ADA compliant? (cont.)

- **Purdue**: Have not checked repository for ADA compliance
- **WSU**: No
- **UNR**: Not sure, think perhaps not
- **UWYO**: For the sight-impaired, the Islandora front-end has viewers that allow items to be magnified. FlexPaper and the Djatoka viewers are part of the software stack; colors are ADA compliant
- **NAL**: Much of it is, some is not. As a Federal government entity, we must comply, so we are working to that end
- **Cornell**: Not at present, but we are aware of this issue.
- **Hawaii**: Dspace is not ADA compliant
Summary Review Comments/Categories: Day One

– General comments
  • Diverse digital repository environment
  • Limited human and financial resources
  • Need to build on existing initiatives
  • Need to share in areas of strengths

– Content focus options
  • Historical
  • Current
  • Born digital
  • Print (reborn digital)
  • Topical (defining disciplinary boundaries)
  • Id, coordinate, focus on unique collections
  • AES & Extension materials
  • Multi-media
Summary Review Comments/Categories: Day One, cont.

- **Technical Issues**
  - Metadata schemas (open to Google and sharing)
  - Open access
  - RDF/Linked Open Data

- **Policy & Organizational Issues**
  - Federal policy environment, i.e. USDA or national
  - Google vs. subject-based discovery systems
  - Silos vs. “centralized” systems

- **Existing Initiatives**
  - Hathi Trust
  - AgNIC
  - VIVO
  - Google
  - Internet Archives
  - Regional Consortia (Colorado Alliance)

- **Scholarly Communication**
  - Motivating faculty to submit content
Strategic Questions and Comments from Day One:

- **UC Davis:** (1) Create a direction that leads sufficiently into the future so there's maximum flexibility: not building a single system but rather working on sharing existing metadata fields for harvesting and effective search capabilities; (2) Are there any lessons associated with what metadata schedules or practices used?; (3) Google scanning projects may be depositing those files in Hathi Trust for preservation; (4) Internet Archive does have “ag’ materials (including UC Ag Extension publications?)

- **CSU:** (1) Ensure that we identify what content and why this content is a priority; (2) analyze what we are digitizing institutionally; id what missing; share among institutional strengths

- **Cornell:** (1) are we focusing on historical or current; (2) how will this initiative link to AgNIC, VIVO, & others?; (3) determine if we are going to do both: share documents and share data and if so, how? Are they two different deposit systems? How will we consider linking to other related content? (4) how much [ag materials] have been scanned by Google?
Strategic Questions and Comments from Day One: (cont.)

- **Hawaii**: What are the boundaries of this process re: organization resources, inclusiveness, and relevance when there are several repositories in question
- **UNR**: Determining the strong point of collections; institutional content vs. faculty submissions
- **Oregon**: Should we focus on digitizing ag extension content, is this a priority?
- **USU**: The challenge of interdisciplinary content vs. Ag content in a repository; how to identify those relevant for harvesting
- **WSU**: What ag content is in Hathi Trust?
- **UWYO**: What are the digital preservation infrastructure plans, and collaboration plans between campus IT and library IT departments?
Comments and Chat Messages from Day Two:

**Question 2:**
- **Cornell:** If we have IRs across institutions, like in VIVO that allows to create linkages across institutions... we can harvest content and produce linked open data that is searchable from one location.
- **Arizona:** Is there something to drive adoption of RDF & Linked Open Data?
- **Hawaii:** Provides shared information, authority control, and easier collaboration

**Question 3:**
- **NAL:** We use Fedora with a SOLR index and Blacklight on the index as the discovery interface
- **Cornell:** Clarification on DuraCloud (storage); should look at SOLR indexing to search across institutions (NAL and VIVO are using this; Purdue has some experience with SOLR, too).
Comments and Chat Messages from Day Two:

**Question 4 – open access:**

- **NAL:** Is implementing embargos as with the NIH policy and federal research outcomes; USDA is not requiring open access, but White House is pushing for a national policy; NIFA is working on policy. Reports are required by not the data.
- **Alaska:** NSF requires data management plan to get projects funded.
- **UMN:** NIFA requires all publications in public domain to be submitted.

**Question 5 – long-term digital preservation:**

- **OSU:** involved in digital preservation as a member of the MetaArchive group that uses LOCKSS.
Comments and Chat Messages from Day Two:

**Question 6 – Google scanning projects**
- **Arizona:** About 5 participants are involved in Google scanning projects
- **OSU:** Extension pubs were scanned locally and not part of Google Books
- **NAL:** cannot participate because of requirement to limit search engine. U. of N. Texas has quite a bit, but not sure it went through Google project.
- **Cornell:** it is likely that some other universities such as Michigan and Harvard include some ag materials

**Question 7 – Hathi Trust:**
- **NAL:** Question about their preservation purpose; some documents are incomplete
- **Arizona:** North Texas dean of libraries is head of MetaArchive initiative
- **UMN:** NIFA requires all publications in public domain to be submitted
Comments and Chat Messages from Day Two:

**Question 8 – Metadata schema:**
- **NAL:** It makes sense that DC be common standard. However, WSU’s use of RDF is where we should all aim; OAI-PMH is good “sharing” method (harvesters)
- **Alaska:** Is reason for DC because of ContentDM?
- **Cornell:** Using RDF/LOD is not an alternative to a good schema...OAI-PMH supports DC and other schemas, i.e. METS. Dspace, Greenstone, ContentDM, and others can expose as OAI-PMH.

**Question 9 – ADA Compliance:**
- **Cornell:** Has worked with Djakota viewer. JAVA application that can integrate with Fedora to create tiled images. Advantage is that it supports jpg 2000 as well as Tiff. Good for providing magnified images
- **UC Davis:** CDL has accessibility compliance goals for all of its web applications including: US Fed Gov Section 508; W3C Web Content Accessibility Guidelines (priority 1); XHTML Transititonal 1.0 validation. Testing of eScholarship, they rely on checklists from these resources, W3C validation, testing results from CynthiaSays (http://www.cynthiasays.com) and testing with text browser such as lynx.
Other Comments from Day Two:

USU offered Blog quote re: importance of developing interfaces

(Karen Coyle) – So for every HathiTrust or DPLA that focuses on the resources we need a related project – equally well-funded – that focuses on users and access. Not just creating a traditional library-type catalog but providing a whole host of services that will help users find and explore the digital library. This interface needs to be part search engine, part individual work space, and part social networking. Users should be able to do their research, store their personal library (getting into Memex territory here), share their work with others, engage in conversations, and perhaps even manage complex research projects. It could be a combination of Zotero, VIVO, Zoho, Yahoo pipes, Dabble, and MIT’s OpenCourseWare.
Day Two Off-line Activity:

• Discuss what you think this group could do together and send in as email
  – UA compiled preliminary results and presented quick summary with categories: Purpose/Organizational; Collection Building (Standards); Discovery (user tools)

• Planning Team will send out revised list for prioritizing and determining next steps
  – Who will work on what; report & proposal for deans; report for participants