

Report on the Coalition for Networked Information Spring 2015 Meeting

13-14 April 2015

Jody L. DeRidder

The [Coalition for Networked Information \(CNI\)](#) seeks to transform digital information technology to advance scholarly communication and enrich intellectual productivity. Approximately 220 institutions are CNI members, each of which may send two representatives to the semi-annual membership meetings to discuss new and ongoing projects and plan for future initiatives. [Cliff Lynch](#) is the Executive Director, and the Associate Executive Director is [Joan Lippincott](#). This document reports on the sessions I attended at this meeting.

Table of Contents

Opening Plenary: “Providing Universal Access to Modern Materials – And living to tell the tale”	1
Visualization on the Big Screen: Hands-on Immersive Environments Designed for Student and Faculty Collaboration	2
Social Networks and Archival Context: From R&D to Cooperative Program	2
Software Curation as a Digital Preservation Service	3
Addressing Institutional Challenges to Providing Accessible Digital Content	4
Moving Ahead with Fedora 4	5
Networks of Expertise: A Model for Implementing and Sustaining new Information Services	7
Conclusions and Recommendations	10

Opening Plenary: “Providing Universal Access to Modern Materials – And living to tell the tale”

[Brewster Kahle, Internet Archive](#)

[The Internet Archive](#) is a non-profit, independent, \$12 million-a-year archive OF the internet. However, they want to become the archive ON the internet, not just hosting old web pages. We need to be trying to build libraries together, with distributed collection services and content. Brewster believes the archive should be free to all.

He reported on their progress so far, including not only web pages, but books, audio, TV, videos, and software (a [full report is on my blog](#)). Despite the warnings of lawyers, they have succeeded by being respectful, avoiding posting content that is for sale elsewhere, and by quickly taking down any content upon request. He believes that changes to the laws will follow the needs of the people. Universal access to all knowledge is a project involving everyone. Google books was stopped because it was a monopoly. He is calling for assistance from libraries and other institutions, to save our cultural history. If we cannot keep the originals, send them to him, and he will store them and digitize them. We need to work together, and do it collaboratively, and make the content free to the people.

Visualization on the Big Screen: Hands-on Immersive Environments **Designed for Student and Faculty Collaboration**

Bryan Sinclair, GA State: Jill Sexton, UNCCH, Joseph Hurley, GA State

Bryan Sinclair spoke first and said that what they are doing is not a video wall, it's a visualization wall. It's not preformatted. A visualization wall is an interactive experience, which enables users to interact in real time with content.

Virtual reality and Google Earth are becoming mainstream. See [Google Cardboard](#) for an interesting immersive technology that is available for only about 10 bucks.

The focus of big screen visualization is the experience of "being present in data." It is the social and collaborative aspect that draws people to the big screen.

Jill Sexton described the [research hub at UNC libraries](#), and the [Liquid Galaxy installation at UNC Chapel Hill](#). This is a low budget large screen visualization, limited in what it displays – but what it does, it does very well. It is very successful and highly trafficked.

She showed us some virtual field trips. [Open Street Map](#) enables contribution of data that is then visible online in a short time, open access. People have a high level of emotional investment in what they contribute.

Joe Hurley, interim director of CURVE ([Collaborative University Research and Visualization Environment](#)), then talked about GA State's work with "[photogrammetry](#)" -- generating 3D from 2D images which is an up and coming thing.

Social Networks and Archival Context: From R&D to Cooperative Program

Daniel Pitti (UVa) and Brian Tingle (CDL)

The [Social Networks and Archival Context \(SNAC\)](#) project seeks to build a vast social-document network connecting the past to the present to the future. They are funded by the [Mellon Foundation](#), [IMLS](#), [NEH](#) and the collaborators include UC Berkeley School

of Information, California Digital Library (CDL), [UVA Institute for Advanced Technology in the Humanities \(IATH\)](#), and the National Archives and Records Administration (NARA).

Their Research and Development objectives are two-fold:

- 1) to demonstrate that data describing people in existing archival description can be used to address the challenge of finding distributed historical resources and
- 2) to lay the foundation for an international cooperative for centrally maintaining the collectively created biographical data.

So far they have nearly 190k EAD Finding aids, primarily from the U.S, 300k British library authority records, NARA authority records, and 2.2M WorldCat MARC records. They extract the content they need for the [EAD-CPF](#) metadata, match the records against [VIAF](#) authority records (and add to it), and have created [a prototype historical resource and access system](#). The biggest challenges thus far are different names for the same person and different people with the same names. For about 40,000 names, they have Wikipedia images; by hitting "refresh", a new set of content is displayed. Right now they support an A-Z browse of 3 million records. Color coding is used for person, family, organization, and they offer a radial graph visualization (which they hope to improve), as well as a list of collection locations (which they hope to map). This will enable users to identify co-located archives for visiting.

Why is it best to do this as a cooperative? For archivists, one biographical statement for a person is better than thousands; there's economy in sharing descriptive work. Also, we can link the people descriptions to one another and to record descriptions, to build the social network. By doing so, we will also build an international, internet-based linked archival authority system.

The SNAC project is cooperating with [ISNI](#) (International Standard Name Identifier) and hopes to become part of their network someday. A fairly small number of enthusiastic institutions are involved, including NARA, the Library of Congress, and the Smithsonian. A cooperative pilot project will launch in July 2015. It is exceptionally complex to turn this into a program, socially more than intellectually or technologically. It requires international community building. They hope to contribute a significant component to the international humanities research infrastructure.

[Software Curation as a Digital Preservation Service](#)

Keith Webster, Dean of Libraries, Carnegie Mellon, and Euan Cochrane, Digital Preservation Manager, Yale.

There is a tremendous amount of software-dependent content, and even the software is dependent upon operating systems, which keep changing. We need older software to authentically render old content. How? We need to curate operating systems, software applications, fonts, scripts, plugins and other dependencies, whole desktop environments (to preserve the experience of interaction) and create pre-configured disk images with software already installed, for running emulated hardware. We create

an emulation software package that creates a virtual version of one computer within another. Virtualization is the same as emulation, but with compatible hardware. Virtualization bridges the gap between recently obsolete hardware, and the arrival of hardware powerful enough to emulate it. Some of the host machine's hardware is used directly by the virtualized computer.

We need persistent identifiers and catalogues or registries for software, disk images and virtual hard drives, and emulated/virtualized hardware configurations.

Normally emulated environments are accessed via a single dedicated desktop, but that doesn't make sense anymore. Modern users don't expect to come to the reading room to access digital content. We need emulation as an online service, providing remote access to pre-configured environments via a browser. Changes to the environments can be saved or discarded at the end of a session. Interactivity can be restricted where appropriate. There is a relatively simple way to provide custom online environments, demonstrated by [the EaaS \(Emulation As A Service\) project](#) from the University of Freiburg in Germany. Their methodology enables emulation to be provided remotely using disk image archives, or locally. Small derivative environments can be created from base environments, saving space. Standard environments can be reused and customized. Content is controlled, and cannot be copied without authorization. Additionally, users can cite the environments. This has the potential to transform scholarship.

Carnegie Mellon has started an archive for software, games and other executable content, called the [Olive Archive](#).

Addressing Institutional Challenges to Providing Accessible Digital Content

- Judy Ruttenberg, Program Director ARL
- Jonathan Lazar, Prof. of Computer and Info Sciences, Towson University
- Sheryl Burgstahler, Founder and Director, [DO-IT Center](#), U of Washington

[Note: a fuller description of this presentation is [on my blog](#).]

Sheryl of UW noted that in the past few years, 14 higher education institutions have been sued, for their IT not being accessible; this includes Harvard, MIT, Penn State, and U of Cincinnati. Legal basis:

- Section 504 of the Rehab Act
- The Americans with Disabilities Act and its 2008 amendments
- State and local laws

The definition of "accessible" in terms of IT means that we must provide the same information, same interactions, and same services in an equally effective and equally

integrated manner, with substantially equivalent eases of use. Disabled people must be able to obtain info as fully, equally, and independently as a person without a disability.

Lessons learned from legal cases include:

- Conduct accessibility audits and develop corrective action strategy
- Develop and disseminate accessible IT policy
- Set and disseminate IT Accessibility standards
- Provide training & education
- Develop procurement policies and procedures!!! Include language for vendors asking "what are your future plans for incorporating accessibility?" We are responsible for buying the product; the vendor is not at risk, so will not support accessibility without pressure.
- Develop and publicize a grievance procedure which is very public, not just a help link. You want patrons to have their grievance addressed before they take it to legal entities.
- Address developed, procured and used IT: websites, learning management systems, classroom technologies, banks and ATMs, purchased software

Settlement requirements are far beyond the requirements of the law, which encourages others to be ahead of the game. Always ask vendors for a [VPAT: a Voluntary Product Accessibility Template](#). Vendors lie.

Ohio state tests the interfaces and if it doesn't meet what the vendor promised, they make them sign an indemnification statement, so Ohio State can't be sued – the vendor would be.

Put vendors on notice: give them deadlines to make things accessible, or stop use. Pressure on vendors is very important. The Universities are the ones at risk.

One institution's approach is to agree only to one year conditional use and if not accessible by the end of the year, may not resubscribe.

Have a plan and a clear timeline, so you can show you're working on it, if a legal issue arises.

Moving Ahead with Fedora 4

- David Wilcox, product manager, Duraspace
- Robert Carolano, Associate VP for Digital Programs and Technology Services, Columbia U
- Robin Ruggaber, CTO, UVa
- Jon Dunn, Interim Assistant Dean for Library technologies, Indiana U

David Wilcox, Duraspace, spoke first.

There were 64 Fedora members in 2014, and membership is growing. Funding is growing also, exceeding the 2014 funding goal of \$500k, which enables them to take on new initiatives.

Newest developments include:

- Production release of [Fedora 4](#) in November 2014 (now on 4.1.1), built by 34 Fedora community developers with support for Hydra and Islandora, and semantic web. By this summer, Islandora will have a front end on Fedora 4.
- Tools for migration now exist: Generic + Islandora (migration-utils) and Hydra (fedora-migrate)
- Pilot projects include Columbia U, Nat'l Library of Wales, Simon Fraser U, U of New South Wales, and York U.

Fedora 4 uses the [Portland Common Data Model \(PCDM\)](#), which is flexible and extensible, defines collections, objects, and files, and encourages linked data best practices. PCDM is currently being adopted by [Hydra](#) and [Islandora](#).

There will be a number of one-day workshops in 2015 attached to conferences, and a 3-day camp planned for fall 2015. [All training materials will be available on the wiki](#). They expect to complete "upgration" pilots by end of April and release the next major 4.x by the start of [OR 2015 \(10th International Conference on Open Repositories\)](#).

[Jon Dunn, Indiana, spoke about "HydraDAM2 and Fedora4"](#).

[HydraDAM2](#) is an NEH-funded project to extend the current version of HydraDAM (a Fedora 3 based preservation repository and hierarchical storage management system). Partners include the WGBH Media Library and Archives in Boston, and Hydra. The intent is to develop a preservation-oriented digital asset management system for time-based media, using open source Hydra heads. ([Indiana is digitizing all their rare and unique assets across the university, a \\$15 million project](#)).

[Robert Cartolano, Columbia: "Columbia's Fedora 4 Pilot"](#)

They want to align with the emerging [Portland Content Data Model Specification](#). The upcoming Sufia (a component that adds self-deposit institutional repository features) version will support list browse of facet values, which is very important to their repository folks.

Ingest is the slowest part; Fedora 4 improves this some, as it allows for asynchronous events. Derivative generation takes time. Fedora 4 uses [Apache Camel](#) for asynchronous tasks like this on other servers. Camel is a message listener. It pushes RDF to a triple store or whatever, eliminating performance bottlenecks. It is much more distributed and event-based.

Fedora 4 is built on top of [modeshape](#), with [infinispan](#) clustering under it. So you can cluster for redundancy, and put a load balancer on top of that, to balance your read performance. When you update one, the others update in sequence. And you get failover.

They have one node for ingest, one for reading. They do not yet have clustering for ingest – it's not a use case supported by infinispan. So ingest is to one node and it updates to the cluster. On the storage side, infinispan has options they have not yet

explored. If you have all your stuff on the file system in a [pairtree](#), it will be slower than if it's in the database.

They are trying to maintain as little custom code as possible. The next version of Fedora may simply be a set of APIs.

IMLS is funding a 2.5-year collaborative development of "[Hydra-in-a-Box](#)" which should be a turnkey, cloud-ready Hydra solution which will be used to foster a greatly expanded network of open-access content-hosting "hubs" for the [Digital Public Library of America \(DPLA\)](#).

Networks of Expertise: A Model for Implementing and Sustaining new Information Services

- Jon E. Cawthorne, Dean of Libraries, West Virginia University
- John Culshaw, University Librarian, University of Iowa
- Geneva Henry, University Librarian and Vice Provost for Libraries, George Washington U
- Joy Kirchner, Associate University Librarian for Content & Collections, U of Minnesota

This group studied centers of excellence, expecting them to be good models for libraries to follow to provide the kinds of services we can't do alone. They were surprised by their findings.

They performed a preliminary investigation of more than 100 centers, narrowed it to 35 centers and 10 funders, and interviewed staff at 19 centers and 7 funders. They focused on cross-institutional centers. They presented preliminary results last April at CNI, and then held focus groups with library directors.

Reactions:

- Term "Center of Excellence" is often viewed with skepticism. There's an implication that excellence discourages innovation and collaboration, and doesn't value the people.
- There seems to be a focus on status instead of on value of effort.
- Science and engineering more comfortable with the term and fully embraced it.

Agreement on characteristics of a center:

- Engine of innovation
- Solving problems across institutional boundaries
- An evolving, adaptable, agile entity

Organization structure: most are part of a larger institution. Benefits include shared services (HR, financial, utilities), shared faculty appointments, access to students.

Partnerships are valued

- Peers, research and education, industry, supporters
- Some partnerships fail: 'Obstructive partners can really bring down a project'

Assessment:

- Motivation driven by funder requirements. Generally just part of reporting process, annual reports.
- Provides checks and balances to help with prioritization and staying on course.

When they asked about "sunset plans", everyone agreed they were important, and almost no one had one. Planning for closing is not a priority.

Funding sources:

- Grants, endowments, state funds, institutions, partners, members
- Core staff on institutional funds (e.g. director)
- Centers spend 30-50% of their time fundraising
- Funders willing to invest short term, but expect centers to find other funding long term.

Challenges:

It takes years to establish credibility and trust. Having a charismatic leader is very important. He or she must already have a presence in his/her profession. This is often a key criteria often for funders. But there are challenges – that person can go away. When leadership changes, that can destabilize things – and funder priorities shift. Also, funders have a priority to fund innovation, not provide ongoing support.

Networks of Expertise Conclusion and recommendations:

Nearly all centers appeared to be facing issues of purpose, stability, sustainability, assessment, leadership, succession planning among others.

Should libraries consider such a paradigm, or should we modify what we are doing in our existing orgs?

The group decided against "Centers of Excellence" as a model. But we have networks of experts – perhaps we need "expert networks" instead?

- A means to leverage institutional strengths and consider solutions beyond local environments.
- Allows active network of institutions to address issues
- Integration of new "expertise" back into local environment fostered from the collaborative network.

The way forward:

- Identify opportunities to test the notion of "networks of expertise"

- Create these networks with careful consideration of environments that foster collaboration and growth
- Develop mechanisms for ongoing and regular assessment and analysis of program deliverables and services
- Develop community-building strategy
- Create a pilot innovation team (e.g. tiger teams) to engage on an urgent issue affecting the library community
- Develop a taxonomy for collaborative activities for library and information services.

Expertise should not be siloed, but leveraged across the institution. When they are woven throughout the organization, it has the most success. The group thinks libraries are well positioned to set these up. They should be introduced as a natural outgrowth of sharing knowledge, and should help build a culture of experimentation and innovation.

“...libraries must find a way to ‘assert governance’ and become ‘less receiving and more influencing’” – quote from a center director.

- Do you see a network of expertise model as a transformative avenue to retool our workforce, develop new skill sets in libraries, and create library services that are more agile and responsive to the rapidly changing information landscape?
- If an agile network of expertise is the way for libraries to proceed, what should a management structure look like?
- What role, if any, should information schools have in these structures?
- Where shall we begin?

This is a preliminary funding grant from Mellon. It might be possible to do a test pilot.

There was lots of concern about how you keep it going, what happens to the work you take people away from – how do you sustain the resources? Sprints are doable. But it’s hard to keep efforts ongoing. Perhaps efforts should only be for a sprint, then we focus on the next thing.

A network of expertise is not a service model. There’s a very limited set of things they can do. You’re looking at more general needs. You can rally a small group to deliver it, and then they’re out of the pool. We need to separate developing an infrastructure from the short term delivery model. You still need someone to convene, someone to lead... the only collaboration that works has a phone number. Someone has to take the lead.

It has to be perceived that the goal is to the benefit of each organization. To continue over time, you have to have a sentiment. We would have to almost create an organization that would have to live somewhat permanently with resources going elsewhere, that creates good in areas where we want it to be. It might be an almost competitive entity, and it needs a programmatic vision. Digital Curation Centre in the UK might be a model.

What if the network was set up as a lab, and they bring it back to the institution? Provide exposure to experts beyond your institution, and then bring the knowledge back home. Seems like a hybrid between consultant and contractor. We might want to experiment on this within an institution, before trying to do this cross-institutionally. There is more risk and more cultural issues cross-institutionally unless the partners are bound in other ways. Next would be a logical consortium.

Conclusions and Recommendations

The CNI meetings are an excellent resource for both networking and learning about current efforts and challenges in our field. I highly recommend we have at least one regular attendee who provides reports back to UA Libraries faculty, staff, and administration.

Many attendees were impressed by the effectiveness Brewster Kahle has had in saving our cultural history, and the responsiveness of high-profile attendees leads me to believe there will be movement towards collaborations among libraries and other institutions to further expand on his effort and support his vision of a shared online archive. We should consider evaluating obsolete materials we have in Special Collections that are already beyond our ability to digitize, to determine whether we should send these to the Internet Archive for digitization, hosting, and preservation. Additionally, we should consider engaging scholars and archivists in creating collections in the Internet Archive, selecting valuable online content that should be preserved for long-term access, but which is outside our collection policy or rights to obtain or manage.

For scholars and researchers to make sense of large corpuses or large amounts of information, they need to be able to analyze large data sets and develop methods of visualizing the contents. More and more research institutions are developing research hubs which provide storage, high-powered desktops, Internet2 connections, access to cloud services, software and infrastructure for analysis and visualization, and often large screen visualization rooms. I recommend that we partner with other entities on campus and OIT to determine what services would be most useful for our researchers, and develop a research hub in Rodgers Science and Engineering Library.

The SNAC project is one to which we have already submitted finding aids; however, I don't believe ours included the level of metadata necessary to be of much use to their project. Names, occupations, geographical locations, and other entities must be encoded there in order to be extractable for use in the shared environment that will enable us to leverage this effort to build cross-linking and shared controlled authority lists. It would be to our benefit to revisit how entities are encoded in our finding aids and work with Metadata Librarians to ensure we are providing what is needed long-term for interoperability.

Carnegie Mellon's vision of an online registry for virtualized environments for emulation is innovative, striking in its clarity and viability, and rather desperately needed. Even if we do not provide such an online service, we should track their efforts, contribute images

of hardware/software environments, and ensure our users know about their service and how to use it, in order to access their older content and software.

The extent to which institutions are at risk of lawsuits for failure to provide accessibility to information technology is alarming. I have shared the information from this session with several entities across campus already, and am aware that Web Services is deeply involved with developing accessibility for library web pages. However, Acumen is not accessible by legal standards, and this needs to become a priority. Also, we need to provide online a very visible method to complain about accessibility, so that problems come to us for remediation before they go to a legal entity.

The discussion on "Networks of Expertise" highlights the need for greater collaboration between institutions to address issues that impact us all. We are attempting a bit of this with ASERL webinars, and we are beginning something like this with ADPNet, but perhaps we should expand beyond it and consider other collaborative efforts. Some that come to mind include the development of a Digital Public Library of America hub, cross-institutional working groups to develop support for online researchers, data management and digital humanities efforts, and workshops on accessioning born-digital content, and preparation of digital content for preservation.