

Collection Development in the Age of Networked Libraries and Digital Scholarship



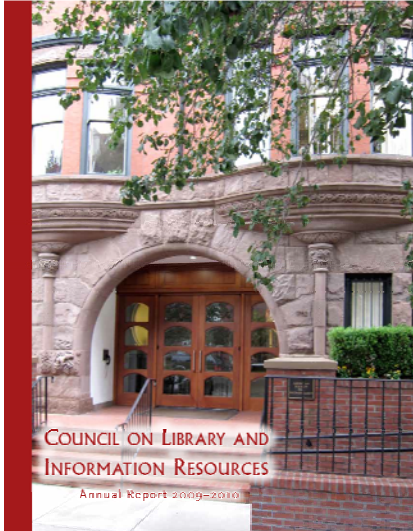
13th Fiesole Collection Development Retreat

May 12, 2011
Rachel L. Frick

Thank you for the opportunity to speak with you today about Collection Development, networked libraries and digital scholarship.

To start off the conversation, first, I would like to spend a little time discussing the concept of “Macro solutions”.

Macro Solutions:



“Common to these efforts will be developing strong coalitions that bring together diverse institutions within a national framework; federating shared resources and interests, including collections, technology, and expertise; and creating a genuine, volitional dependency on other participating institutions for the provision of what was once a locally owned and managed asset. We are calling these collaborative projects *macro solutions*.”

CLIR Annual Report, 2009-2010, p. 3

The President of CLIR, Charles Henry, has written much about macrosolutions, and I recommend his most recent article in the January issue of *educuase review*. Basically, macrosolutions are where institutions come together, share resources and create solutions that create convergence – or an integral dependency – to provide a service that was once locally owned.

Paul Courant and John Wilkin, of the University of Michigan refer to this as “above-campus” library services in an *Educause review* article from August 2010.

The concept of macrosolutions to shared challenges, can be characterized at group scale or web-sale

Dempsey: Taxonomy of Scale

- Institution Scale:
 - Within an Institution / Local Target
- Group Scale:
 - Supra-Institutional / corresponding groups
 - Macrosolution Example: ReCAP
- Web Scale
 - Networked level / potential for all users
 - Macrosolution Example: HathiTrust

<http://orweblog.oclc.org/archives/002058.html>

Lorcan Dempsey breaks this idea of scale very nicely into this taxonomy. We have local – which is institutional based and has a local target. Group – which is “supra-institutional” – or “above campus” – and the groups that come together tend to be organized by region or by a specialization or like characteristic. Group a good example is shared legacy print archives like ReCap -Research Collections Access & Preservation consortium or WEST. Web scale – is when a solution is applied at the network level and contributors and or beneficiaries can be potentially anyone connected to the “net. , a good example of library at web scale is the hathitrust



Research Library at Web Scale

8,660,577 total volumes

4,744,677 book titles

210,762 serial titles

3,031,201,950 pages

388 terabytes

103 miles

7,037 tons

*2,317,408 volumes public domain, 27% of the collection

How many are familiar with HathiTrust? HathiTrust is a partnership of major research institutions building an immense digital preservation repository. A majority of the content is from Google book scans, but other digital collections are represented.

This project has achieved.

Economies of scale for digital preservation and associated services

Grown the pool of digital preservation expertise by through real world experience

Trusted collaboration

As research libraries face financial pressures and weigh the relative value of print and digital volumes, this growing digital aggregation of research library content has the potential to support current local collection development decisions.

Understanding how trusted collaborations that enable macrosolutions at web scale, are key when thinking about 21st century collection development.

Here is why.

Cloud Sourcing Research Collections

Managing Print in the Mass Digitized Library Environment

Constance Malpas, 2011

Findings:

- 1/3 of U.S. ARL content duplicated in HathiTrust
- Space savings = 36,000 linear ft. / 45,000 sq.ft.
- Dollar savings = \$500,000 to \$2 million per library

<http://www.oclc.org/research/publications/library/2011/2011-01.pdf>



In a recent OCLC research report, the feasibility of outsourcing management of low-use print books held in academic libraries to shared service providers, including large-scale print like ReCAP and digital repositories like the HathiTrust is examined.

Based on a year-long study of data from the New York University's Bobst Library, HathiTrust, ReCAP, and WorldCat, they concluded that there is sufficient material in the the HathiTrust to duplicate a portion of virtually any academic library in the United States, and there is adequate duplication between Hathi and large-scale print storage facilities to enable a great number of academic libraries to reconsider their local print management operations.

As of June 2010, the median rate of duplication between titles held by university libraries in the U.S. Association of Research Libraries (ARL) and the HathiTrust Digital Library exceeds 30%; that is to say, nearly a third of the content purchased by research-intensive libraries in the United States has already been digitized and is preserved in a shared digital repository.

If the current growth trajectory of the HathiTrust Digital Library is sustained, it is projected that more than 60% of the retrospective print collections held in ARL libraries will be duplicated in HATHI by June 2014.

This growth rate far exceeds average annual acquisitions in ARL libraries, suggesting that the digital replication of legacy collections will outpace growth of new physical collections, enabling a transformation in traditional library operations, staffing and space requirements.

The median space fort an ARL library approximately 36,000 linear feet or the equivalent of more than 45,000 assignable square feet (conservative estimate).

The total annual cost avoidance possible available today would amount to \$500,000 to \$2 million per ARL library depending on the physical environment (e.g., open stacks on campus or high-density off-site storage) in which the titles would be managed locally. or (13,828,000 to 55,312,000 roubles),

There are some obstacles to achieving this vision of a cloud sourced library.

There is only 27% of Hathi in the public domain; it requires a network of shared print services – good news, a small number of print providers needed to achieve 70% collection duplication; and there needs to be a service to manage access to this print repository network.

Cloud Sourcing Research Collections

Managing Print in the Mass Digitized Library Environment

Constance Malpas, 2011

Recommendations:

- Advocate in favor of licensed access to mass-digitized resource
- Support HathiTrust's efforts to expand public access to the mass-digitized
- Communicate to stakeholders: Externalizing traditional functions improves ability to fulfill mission

<http://www.oclc.org/research/publications/library/2011/2011-01.pdf>

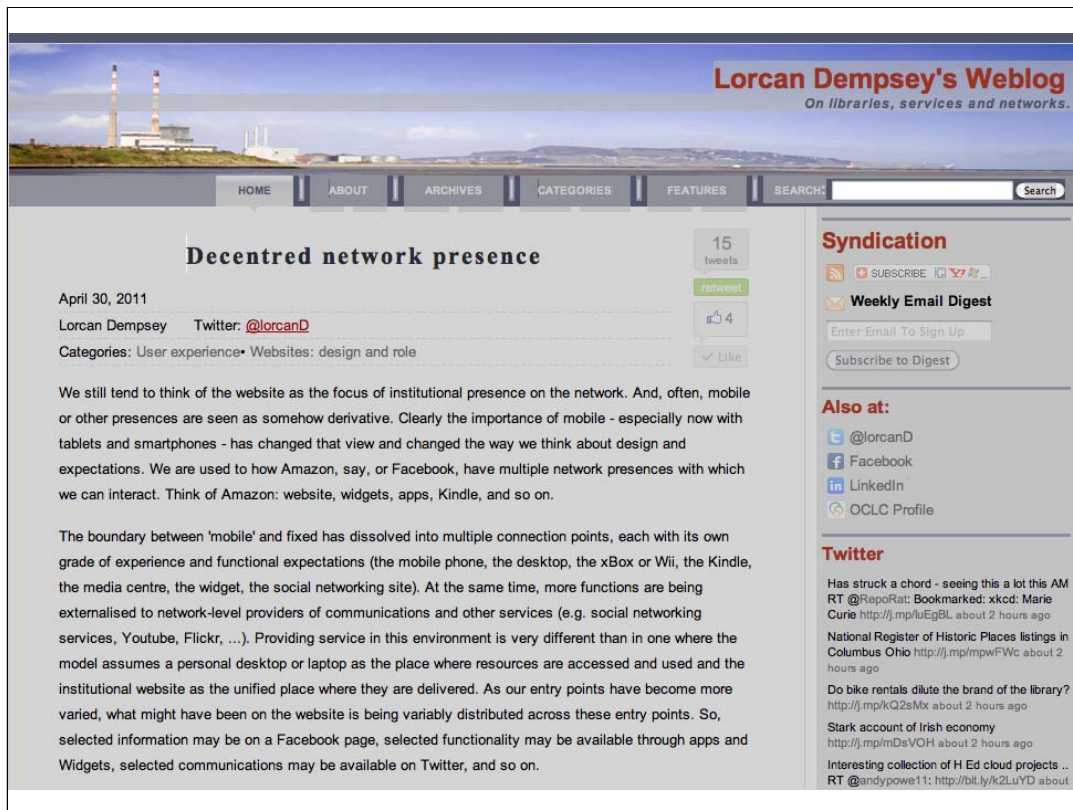


in order to achieve the promise of the cloud library potential – the report makes many recommendations for Libraries, print repository providers, library research organizations, Like OCLC Research and the Council on Library and Information Resources, and funders, like JISC or IMLS. Among the ones for libraries that I want to focus on today is the idea that we need to engage directly with faculty and academic officers to communicate a compelling strategy in which selective externalization of traditional functions improves the libraries' ability to fulfill an academic and research mission.

So what do I mean by externalizing traditional functions? You might be thinking – didn't we do that with collaborative collection development back in the 1990's? Didn't we do that when we outsourced some of our acquisitions and cataloging functions?

Externalizing functions is not new – but doing so in a highly networked environment, at webscale, is.

A good example of functions being externalized to the network level is seen in how mobile devices are currently affecting an organization's web presence.

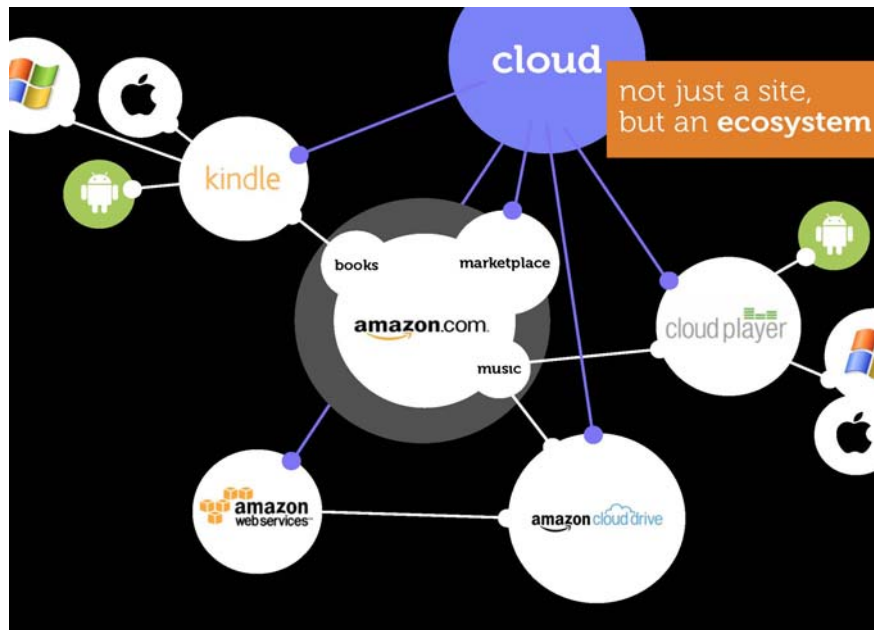


. Lorcan Dempsey discusses this in a blog post from April 30th, entitled “decentered network presence”

Tablets and smart phones are changing the way we think about web site design. Services like identity management, event registration, payments are externalized service streams – and organizational branded content is also streamed out to mobile devices, access through apps and widgets. Providing service in this environment is very different than in one where the mode assumes a personal desktop or laptop as the place where resources are accessed and used and the institutional website is the unified place where they are delivered.

Lorcan challenges in his web post how would libraries look in this model.

The De-centered Network presence



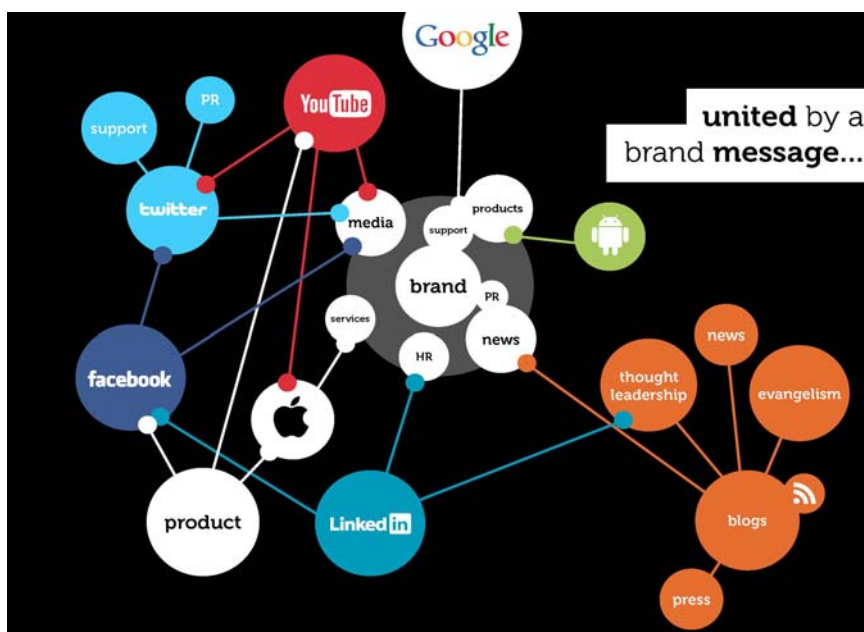
<http://www.slideshare.net/yiibu/beyond-themobilewebbyyiibu/128>

This is how Amazon is represented as a decentered model.

Replace Amazon with your library - and look at the cloud as “cloud library” as envisioned by the OCLC report.

By approaching library functions in this way – the library itself, ceases to be a stand-alone island, a world unto itself. It transcends the idea of place, and functions more like an ecosystem, enabling the freedom to experiment and respond proactively to user needs

Decentered Networked Library



In the Decentered Library model, Collections and services are united, not by a place or a website – but by the library brand and its message. By building distributed networked services and collections rather than “big giant all purpose institutional containers” libraries will begin to look less like this

<http://www.slideshare.net/yiibu/beyond-themobilewebbyiibu/125>

Less like this



<http://www.flickr.com/photos/ducatistaraul/4130178584/>

.....And more like this.

More like this



<http://www.flickr.com/photos/ppix/843626096/>

<http://www.flickr.com/photos/ppix/843626096/>

Light weight, agile, responsive.



In the ecosystem approach Libraries become a network of touchpoints, utilities, collections, and conversations.

Collection development becomes more about managing the relationships associated with the collections. Those with faculty, students, and our vendors– and less about managing the stuff.

Its about relationships in the network – the ecology of information services – both in the immediate service community context – and in the larger information network.

The ecosystem approach transissions collection development from an abstracted, passive back room function to one that is public and plugged in – actively supporting teaching, learning and research.

This is very important in today’s library environment, as more librariy directors’ view this activity of the library – that of supporting scholarship, teaching and learning - as their highest priority.



ITHAKA S+R

Insights from U.S. Academic Library Directors

Matthew P. Long
Roger C. Schonfeld

<http://www.ithaka.org/ithaka-s-r/research/ithaka-s-r-library-survey-2010/insights-from-us-academic-library-directors.pdf>

As identified in the most recent ITHAKA s+R report, Library directors envision research and teaching support and facilitation functions as their highest level strategic priority (over 90%)with a shift away, from traditional collections acquisitions and preservation functions.

By viewing and managing library services as stand alone functions, as research support separate from collection development, - ignores how the networked enabled library functions as an ecosystem. Taking the “monoculture” approach to libraries services – could have an overall detrimental effect to libraries, and their associated information providers.

“How I Fell in Love with a Fish”



A good analogy for this networked library environment is one from an ecological framework that can be found in the TED talk by Dan Barber: “How I Fell in Love with a Fish.” In this presentation, Barber discusses an alternative approach to aquaculture. The Veta La Palma fish farm in southern Spain is located on an island in an estuary 16 kilometers from the Atlantic Ocean. This farm is different for many reasons, but the most notable is that Veta La Palma measures its success by the health of its predators. Because the farm is also the largest private bird sanctuary in Europe, 20 percent of the fish and fish eggs are lost to birds each year. This is good, in the eyes of the farm’s biologist, Miguel Medialdea. “We farm extensively, not intensively. This is an ecological network. The flamingos eat the shrimp. The shrimp eat the phytoplankton. So the pinker the [flamingo] belly, the better the system.”⁸

Veta La Palma provides an alternative to the more common agribusiness model, which tends to be monoculture (farming only corn, only soybeans, or only salmon, for example) and resource-intensive, relying on high-touch intervention through chemistry and machines. In this story, success is measured in terms beyond the current best-practice realms; the model succeeds and thrives because old metrics have been discarded and newly aligned components have become the norm.

Its about relationships in the network – the ecology of information services – both in the immediate service community context – and in the larger information network.

Intensive vs. Extensive

- The Legacy of Tech Services outsourcing
- Patron Driven Acquisition
- Atomization of acquisition (least purchasable unit)
- Focus on unique, local, specialized
- Digital Collections / Curation

So how are we changing our approach to collection development to that takes into mindful consideration –our information ecology?

I say we take a page out of aqua culture approach, and shift our efforts to being more extensive, as opposed to being intensive.

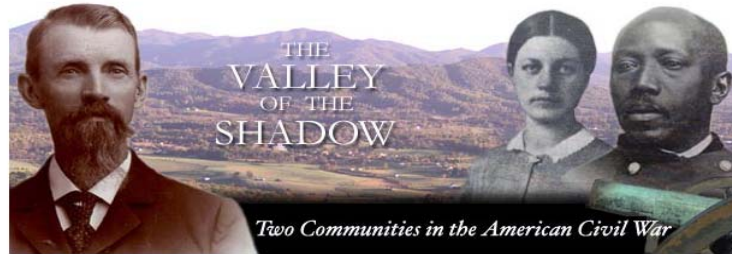
What do I mean by that – simply – collect with your relationships in mind.

The past 25 years, we have systematically rooted out professional discipline specific expertise out of our libraries, we have automated – or systemize collection decisions to approval plans, based on formulas, and purchase big deal subscription bundles. We reduced technical services staff dramatically. We were guided by the thought that if we outsourced 80% of what was duplicative, we could focus more on the 20% that made us unique. We could shift personnel to more “public” functions. We did this, thinking this was the best course of action to deal with the growing deluge of information choices. At it was, based on the information we had at the time. But, I challenge, have we really shifted our work, our professional focus, and realigned staff to the 20% that makes your library and its collections distinctive?

Have we hired more archivists, and special collections expertise to deal with our rare materials? Are we digitizing these collections, and doing so in a way that is informed and guided by our communities’ research and teaching strengths and needs? Is digitization integrated into core collections development functions?

I recently had a great conversation with a colleague who is an interim director at an academic library – he said – we should spend less time on collection development, and more time of finding out what our faculty need besides books and journals. We should acquire and digitize collections for better use in research and the class room, acquire special collections, that reflect our universities history and special scholarship areas, and help our faculty to acquired and preserve data. I looked at him and said, Well, that is “collection development”. Maybe you mean – less time on selecting, processing, and managing the 80% of things everybody else is going to have? He agreed

Collecting Digital Scholarship: UVA Case Study



The Valley Project details life in two American communities, one Northern and one Southern, from the time of John Brown's Raid through the era of Reconstruction. In this digital archive you may explore thousands of original letters and diaries, newspapers and speeches, census and church records, left by men and women in Augusta County, Virginia, and Franklin County, Pennsylvania. Giving voice to hundreds of individual people, the Valley Project tells forgotten stories of life during the era of the Civil War.

[Enter the Valley Archive](#)

Copyright 1993-2007, All Rights Reserved Edward L. Ayers

Talk about Ed and UVA

Local / Global

DataONE <http://www.dataone.org>

Data Conservancy <http://www.dataconservancy.org>

Text Grid <http://www.textgrid.de/en/>

Europeana <http://www.europeana.eu/portal/>

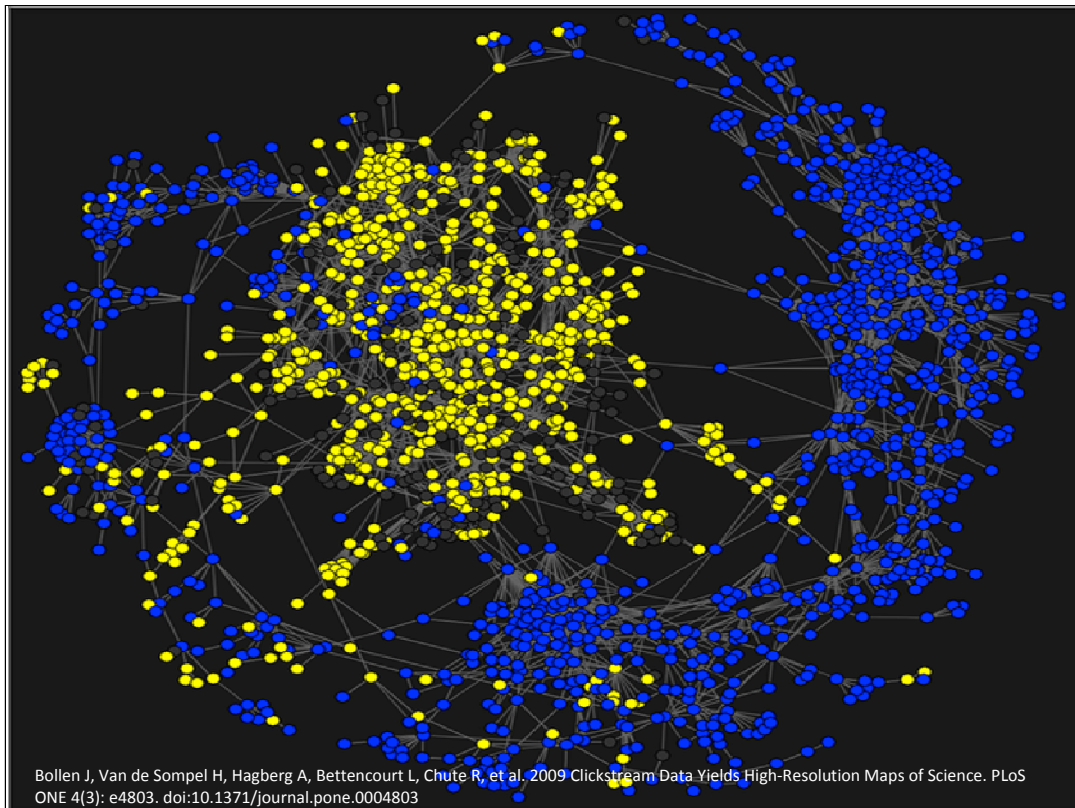
Digital Public Library of
America <http://cyber.law.harvard.edu/research/dpla>

Although the benefits of libraries at web scale are great, and the increasingly socially networked world demands a change in library operations, most of us hesitate or decline to work in a dynamic collaborative environment in which we rely on others for services. There is a tension between local priorities and contributions to the greater good; we lack metrics to measure the value of collaborative effort. Current metrics that measure service quality and library value encourage competition between organizations and undermine the network-scale effort. A competitive environment does not facilitate the development of trust networks and is often a causal condition of distrust in large, complex collaborations.⁵

But we know this is the direction we need to move in, as we see more and a more large scale collaborative efforts emerge, like HathiTrust, or the DPLA. In order to deal with increasing complexity of our collections in today's data-intensive, ubiquitously networked environment. Digital Collection development needs to be a lot more than buying e-books, and journals. It has to focus less on the content you bring into your library – and more on what you push out.

For being an active node in the information network, organizations have to contribute, not just pull down, from the services layer.

Being a strategic in our collection development strategy will have to include understanding how your organization can be both a beneficiary and contributor to solutions at web scale, like the ones listed here. By actively engaging, libraries and librarians, have the potential to actually transcend their local boundaries and roles, and be a part of a greater good. The healthier the overall information ecology – the healthier are all of the individual functions that play. Because....



As the process of creating new knowledge, pursuing academic inquiry, and teaching the skills and competencies to navigate a technology rich world becomes more complex and interdisciplinary - the need to collaborate becomes more evident and imperative. This goes beyond old school cooperative collection development, that we tried out when Faxing articles was cutting edge technology. Digitized collections, whether google books or hathi trust, resources, like Europeana, or the digital public library of America, are a game changer – when deciding how much local effort and resources is allocated to collection development, and specifically, in regards to locally produced and curated digital collections, that 20% that makes you unique. Digital collections that are interesting and powerful standing alone, can have a transformative affect when they are put into a larger context.

Working from a shared set of values, and creating the awareness and applying global values when considering the larger context of local collection development/ digitization decisions, prepares individual library organizations for the unique opportunities of the networked environment.

This map colors journals according to whether the AAT classifies them as either social sciences and humanities journals (yellow) vs. natural science journals (blue). Highly connected clusters corresponding to biology and psychology contain a mix of journals classified in either the social and natural sciences.

The resulting model was visualized as a journal network that outlines the relationships between various scientific domains and clarifies the connection of the social sciences and humanities to the natural sciences.

Katy Borner work – Atlas of Science 2010 MIT publication



- Thank you for your time and attention