So, What Does a Chief Technology Strategist Actually Do?

Dean B. Krafft
January 13, 2010
Helps the Library manage the disruptive transition to digital tools and resources

I want to speak broadly on the challenges that the Library is facing and specifically on some of the projects I am working on that relate to these challenges
Why is this hard?
In the physical world, the Library was at the center of the cycle of scholarly activity. The digital transition challenges this position.
“The library as an institution, engendered over the past few millennium [sic] to provide gatekeeping to scarce physical resources, is reaching a terminus in its life cycle.”

Peter Brantley

Reality Dreams (for Libraries)

12/29/09
Highly complex library technology environment, (click) and one that is constantly changing.
Highly complex library technology environment, (click) and one that is constantly changing.
Need for a CTS

• Need someone to understand and track the explosive changes in CS, IS, and IT

• Need an informed, high-level advocate for necessary change within the Library

• Need a credible partner to work with central IT, external partners, and funding organizations to enable change

• I need to be out there synthesizing all these pieces to create new opportunities for CUL
Research Focus - 33%
• Lead and partner on strategic research (e.g. VIVO, Datanet, Fedora)

IT Focus - 10%
• Library representative to Cornell-wide IT groups (e.g., IT Managers Council, Faculty Advisory Board on IT, Acad Tech planning)
• Liaison between CUL and campus on IT policy and security issues

Library Change Focus - 57%
• Member of Library committees (e.g., LibExec, LOL, DataExec)
• Library representative to outside organizations and partnerships (e.g., 2CUL, ARTstor, CNI)

What are my hats?
Challenges

• Unbundling, disintermediation, commoditization, and consumerization

• IT, academic technologies, and the library

• Preserving digital scholarship and culture for Cornell and the world

• Leveraging our skills, strengths and trust to create services for the scholars of 2020
Definitions

• Unbundling: The disruption of vertically integrated business models: goodbye classified advertising

• Disintermediation: The elimination of intermediaries: goodbye Tower Records

• Commoditization: Price is the only significant factor in a product: hello Amazon and iTunes

• Consumerization: Technology solutions for the masses: hello Flickr, Wikipedia and Google Maps
Unbundling

Newspapers: dead tree printing and classified advertising; Libraries: processing physical books/journals – what’s strategic?
Research Library Competition

- Research libraries used to compete over physical holdings
- Could compete over subscriptions (but scholarly communication is changing)
- Continuing to compete over special collections
- We will compete over the quality of unique scholarly services

Identify our critical, strategic services and unbundle ourselves, before it gets done to us – or before Brantley & others dismiss us
Research Library Collaboration

- Must collaborate over physical and digital holdings
- Must collaborate over open access services and publications
- Could collaborate on web curation and data curation
With digital surrogates available for browsing and discovery, we can deliver physical from off-site storage
as well as digital holdings
Create collaborative solutions rather than doing it all ourselves
ARTstor Shared Shelf

- ARTstor implementation with SAH, Harvard, Yale, Cornell, NYU, Univ. of Miami, UIUC, Colby College
- Don’t have to create or buy a new metadata/workflow system to replace Luna
- Allows faculty to use common externally hosted system for all their teaching images
- Leverages a lot of expertise in creating the “right” system (lots of eyeballs)

Hosted solution for visual resource management and delivery
Collaboration as Competition

- For universities and research libraries, collaboration can be a competitive advantage (BorrowDirect)

- 2CUL: Cornell & Columbia Library partnership looking at opportunities for deep integration

- Collaboration is an opportunity to unbundle services

Collaboration is really an unbundling opportunity
Within the Technology Infrastructure part of 2CUL, here are our current focus areas

- Common Fedora-based preservation repository
- Web curation services
- Electronic records management and archiving
- Common discovery tools
Move away from mediating things that are being provided directly, and providing things available as commodities.
That may include many books
It’s already happened for many journals that we no longer hold directly
In our special collections, we aren’t the brokers, and we can’t easily be disintermediated
Outside use may be much greater than inside use – but our choice is to solve the “shared open access” problem, or just send all our money to Elsevier
We can build unique services as well as holding unique collections
Consumerization and Concentration

Integrate with and take advantage of consumer-level services and the concentrated attention that goes with them; and meet consumer expectations
LOL has piloted moving our collections to Flickr and taking advantage of the eyeballs that come with it
How do we ensure Cornell’s information resources are discoverable and accessible?

To our own faculty and students and to everyone – they expect the same level of discovery and access they get from consumer services.
“There’s a whole Web’s worth of context out there.”

John Mark Ockerbloom
Content and context in concept-oriented catalogs
1/7/10

We need to relate our resources to the larger context of the web
We have several problems with discoverability – many legacy projects, and growing hosted solutions
It’s not just what we hold, it’s also what we pay for
Many licensed databases of journals, images, books, reference works, datasets, and much more
Currently, we use a really bad federated search product to search a subset of these databases.
Do we just count on all our licensed information providers to give access to Google?
Lots and lots of citations and full text of scholarly materials, which the Dartmouth Library owns or provides access to, are in Summon now, including:

- 4,800+ publishers' content included, such as journals and books
- 50,000+ journal, newspaper and magazine titles
- 400+ million items indexed so far, from journal articles to maps to videos
- Dartmouth College Library records

For a list of participating journal and database publishers see [http://www.serialssolutions.com/summon/summon-content-participants.html](http://www.serialssolutions.com/summon/summon-content-participants.html)

Is one (or more) of the commercial providers going to win and provide universal access?
How do we fully index the resources that our users have access to (and ideally allow us to add value)?
Challenges

• Unbundling, disintermediation, commoditization, and consumerization

• IT, academic technologies, and the library

• Preserving digital scholarship and culture for Cornell and the world

• Leveraging our skills, strengths and trust to create services for the scholars of 2020
Academic Technology Providers

- Central IT - IT, infrastructure focus
- Library - scholarly, enduring access focus
- Research Centers - narrow, deep services
- College-level support - balanced between central/edge and general/specialized
- Department-level support - specialized, edge services
Library’s piece is the middle to the bottom right
Who does what?

- Structures and support evolve over time and are different at every institution.
- Allocation driven by history and personality.
- Central IT pushed by major business/administrative challenges (Peoplesoft).
- Fully merging IT and Library has been rife with difficulties - different missions and cultures (IT is from Mars, Library is from Venus).

However, there are clear opportunities to sort out what happens in the middle – the boundary between IT and scholarship.
Cornell ATS Report

• IT strategic planning committee charged Academic Technology Subcommittee to look at provision of AT at Cornell

• Chaired by Dean Krafft and Nancy VanOrman - Library members were Oya Rieger and Jon Corson-Rikert

• Delivered preliminary report to ITGC on November 17, 2009
ATS Recommendations

- Create an AT Support Council, consisting of representatives of all the local and central AT support providers, an AT governance group reporting to the ITGC, and SIGs as needed.

- Provide a single Helpdesk service that coordinates back-end support among all the support providers for AT, admin computing, and IT infrastructure.

- Provide a central service and support web portal that provides clear entry points and user views of academic technology services for all faculty, students, and staff at the university.
ATS Report Summary

- CIT and Library put equal overall FTE effort into supporting Cornell-wide AT
- There are currently overlaps in services and opportunities for much better coordination and service delivery
- In any scenario, both organizations will continue to have some level of AT responsibility

Next steps from the Provost: Library should work with CIT and ITGC to address provision of AT at Cornell
What the Library Must Provide

• Disciplinary understanding and expertise

• Full awareness and engagement with modern IT tools and services

• Our raison d’être: ensuring access to and long-term preservation of the cultural and scholarly record

• Close collaboration with IT experts
Challenges

• Unbundling, disintermediation, commoditization, and consumerization

• IT, academic technologies, and the library

• Preserving digital scholarship and culture for Cornell and the world

• Leveraging our skills, strengths and trust to create services for the scholars of 2020
Preservation Challenges

- Dealing with diffuse information sources
- Curating the Web
- Living with (and in) the Cloud

Figuring out how to preserve the bits may be the least of our challenges
It used to be easy – subscribe to magazines & newspapers in print, on microfilm, or through a subscription database
These are bad times for print-based newspapers
People are going elsewhere for both news
and culture. There are millions of blogs out there on every possible subject.
Newspaper Armageddon

• How do research libraries preserve news, journalism and culture when it is fragmented and scattered in a million pieces across the web?

• What do we do when we can’t just subscribe to a nice simple set of newspapers, whether paper or electronic?

• How do we avoid facing future scholars with a gaping hole in the cultural record?
Web curation

• Discipline-based selection and archiving of web-based materials

• Initial pilots: Human rights websites at Columbia; Women’s blogs at Harvard


Who should be responsible for ensuring that scholars 50 years from now can find what they need?
There are many new challenges for preservation on the web: Google/privacy, DRM, IP nukes
Libraries must live in (or with) the cloud

• Cloud storage, computing, and applications will have lower costs, better service, and a smaller carbon footprint

• Scholars will use these services - how should the library help preserve that work?

• How do we balance core library needs and values against business values and pressures?
Lost in the Cloud
Jonathan Zittrain, NYTimes Op-Ed 7/19/09

• If you entrust your data to others, they can let you down or outright betray you

• Data stored online has less privacy protection both in practice and under the law

• Cloud platforms may restrict the ability to write and run arbitrary code

• No reason to turn back, but legal changes may be needed to ensure critical rights
We rely heavily on the Internet Archive for snapshots of the historical web – will it outlive Brewster Kahle?
Google may not always be the same company with the same values
Here today - gone (or edited) tomorrow?

Amazon deleted copies of Orwell’s 1984 and Animal Farm off owner’s Kindles. It did refund their money, and promised not to do it again. Eastasia/Eurasia?
Here today - gone (or edited) tomorrow?

Amazon deleted copies of Orwell’s 1984 and Animal Farm off owner’s Kindles. It did refund their money, and promised not to do it again. Eastasia/Eurasia?
Organizational Lifetimes

- Google: 11 years
- Amazon: 13 years
- Internet Archive: 13 years
- IBM: 98 years
- Cornell University: 141 years
- Harvard University: 373 years
- Oxford University: 842 years

Even if disruption is coming to research libraries, which of these would you bet will still be around in 100 years?
Challenges

- Unbundling, disintermediation, commoditization, and consumerization
- IT, academic technologies, and the library
- Preserving digital scholarship and culture for Cornell and the world
- Leveraging our skills, strengths and trust to create services for the scholars of 2020
Our strengths

• Disciplinary expertise
• Training in selection and curation
• Understanding how to describe and organize information: FRBR, metadata, ontologies
• Ability to recognize the unique needs of researchers, scholars and students and create new tools, services and solutions
• A long and trusted organizational history
“From a campus wide perspective, fundamental components of campus library and technology organizations need to be surgically extracted from their current hosts and combined in a new data systems support unit [...] with the goal of enhanced information access, handling, and interactivity.”

Peter Brantley

Reality Dreams (for Libraries)

12/29/09

It is not obvious to everyone that the Library should play a role in the digital transition
NSF Datanet Challenge

“Integrate library and archival sciences, cyberinfrastructure, computer and information sciences, and domain science expertise to provide reliable digital preservation, access, integration, and analysis capabilities for science and/or engineering over a decades-long timeline.”

National Science Foundation Solicitation 07-601
Science must combine data from many disciplines to address grand challenge problems like climate change
Current state of describing and combining cross-disciplinary research datasets
The Data Conservancy

• Collaboration of Johns Hopkins, Cornell, NCAR, UCLA, UIUC, and others

• Carl Lagoze, Steve Kelling, Jim Cordes, Dean Krafft from Cornell

• $4 million/year for five years with possible five year extension

• Both DataStaR and arXiv have deliverables
Deliverables

• September 2010: Create SWORD deposit of research datasets into Data Conservancy

• March 2011: Pilot projects deposit datasets into DC - extend metadata/ontologies

• September 2011: Broaden DataStaR to support additional disciplines. Improve DataStaR integration with Semantic Web
Supporting eResearch

- Preserving primary digital data tied to intermediate work products and publication
- Ensuring fixity, security and recoverability
- Developing cross-disciplinary ontologies and metadata mappings
- Pulling order and information from data, descriptions, and observations across multiple disciplines
We are also bringing the Library’s strengths to respond to other unfilled needs
NIH VIVO grant

- NIH 2-year grant of $12.2 million to Univ. of Florida, Cornell, Indiana U. and 4 implementation sites including WCMC

- Builds on work by Jon Corson-Rikert and others since 2003

- Based on Semantic Web technologies, and will integrate with other Linked Open Data efforts

- Cornell has $3.96 million to create new ontologies and software to enable national, multi-institutional network of scientists. Team of 15 people led by PI Dean Krafft, Co-PI Medha Devare, and Development Lead Jon Corson-Rikert
VIVOweb

• What is it? Information regarding scientists’ interests, activities, and accomplishments. Scientists using software tools based on this information to support new and existing collaborations and teams.

• Why? Support the acceleration of team science. Improve visibility (accuracy, currency, breadth) of information about science. Improve competitiveness.

• How? Institutionally hosted information provided and maintained by institutions and scientists. Facilitation and support through the libraries.
In my hiring talk I made several specific recommendations to the Library. Here’s an update on them, 18 months later.
Put the Library in our users’ workflow

• Provide the right academic technology support

• Ensure that all the materials we hold and pay for are discoverable and accessible

• Take some risks to match library services to new technologies
Mediate Cornell’s unique assets to everyone

• Enable research data with efforts like DataStar and the Data Conservancy

• Fully integrate Cornell’s digital assets with the Web

• Leverage 2CUL and other consortial efforts to share the costs and build the best solutions
Enable creating unique context and organization

• The Cornell community is not enough; we must link with global contexts

• VIVOweb is an opportunity to build a web-scale system for research and scholarship

• Must create and leverage opportunities like arXiv that have global reach and impact: “[arXiv is] the cornerstone and the future of the Cornell library (perhaps the only piece that will be around 20 years from now from the current structure)” - anonymous physicist
Be a trusted broker

• We have a unique ability and obligation to preserve the scholarly and cultural record

• As a broker, we can ensure open and protected access to controversial knowledge

• We need to be a partner in efforts like ARTstor, 2CUL, and HathiTrust
Applamazoogle is putting a lot of information intermediaries out of business. Let’s not become extinct.

We must make unique contributions.
CUL in 2020

• The growth in Library resources and services will have been mostly digital - and the physical materials will primarily be accessed as digital surrogates

• Librarians will be the bridge between disciplines and the unique digital tools and services that they need

• The Library will be more specialized, providing some global services locally and drawing on consortial partners and outside services for others

• We’ll just be completing the upgrade to Voyager 9.1

The last bullet may sound flippant, but we will still have major ongoing production responsibilities – even in 2020
Help scholars and students navigate and use the digital cloud of information, tools, and services.

Figure out the strategic services that the library should be providing in support of scholarship.