ASIS&T 2007: Keynote -- The Impact of Web 2.0

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Submitted at 10/22/2007 12:35:28 PM

Anthea Stratigos is CEO/Co-founder of Outsell.

Impact of web 2.0 on publishers, libraries, information providers, etc. We're in business of marketing experiences -- not information. Web 2.0 is happening because of a convergence of individual traits, social and technological forces. Cycle of disruptive technologies: online databases, CD-ROM (1982), web (1991), xml, Web services, RSS (2001), AJAX, Ruby on Rails, REST (2007).

Showed YouTube video: "Did you know? Shift Happens." Web 2.0 is about being global, being "flat". Shows famous "nobody knows you're a dog" cartoon -- web 1.0. Web was static. Not interactive. Now, web 2.0, everybody knows you're a dog -- and your likes, your activities, etc. Web 2.0 is interactive -- anything is a consumable.

Web 2.0 manifests itself as social networks, mashups, user-generated content, community/sharing, networking, crowdsourcing. Communities for any and every slice of life.

Web 2.0 enterprise

Google is class example of this; other enterprises are catching up. Quick, agile, global. "Open"-minded (i.e., IBM & Linden Labs' new avatar standard to enable avatars to move from world to world). Content without containers, play well with others, service-oriented, conversationalist.

Marketing -- new tools enable new research (Facebook, Second Life, panels using cell phones). Notion of physical focus groups diminishing; observing live interactions is rising. Lego is doing product development with power users by showing designs on web and redesigning in response.

Users

31% still struggling with information retrieval -- 31% doesn't generate information users want. Users don't want to pay for stuff. Want free content (60% of time); either free or fee if it serves my needs (36%).

Users wantto receive content by email alerts (85%), blogs (47%), intranet posting 41), podcasts (23%), RSS feeds (21%), videocasting (16%)

(Source: Outsell's information markets & users database).

Users are pulling together networks (MySpace, Facebook, LinkedIn, and enterprise networks). Enterprise networks: behind-the-firewall social networks. Visible Path is one such company.

Publshing & Information Provider

Information industry of yesteryear is flat or no growth. Google, Yahoo, Microsoft, AOL are exploding. Leads to new revenue models. Have choices: bundling, licensing, subscription, pay per view, advertising, syndication. Agility is rising -- publishers are reacting faster. Also lots of online ad possibilities (for additional revenue). Publishers facing great pressures.

Innovation areas:

• Pay per Answer (Gerso Lehrman, Nature Publishing Group, Sermo, Innocentive, Complinet, Corporate Executive Board).
• Pay per View: O'Reilly (buy a chpater, buy a page, etc.). ScienceDirect Info. Scitopia.
• Pay for Software and Tools: McGraw Hill Construction, Soucient, Visual Files. Mixing content and software and a particular user set to create a workflow solution.
• "Freemium" -- free basic services for all; premium paid services for those who want to buy them.

Library Environment

Library technology adoption is not keeping pace with real world. Libraries are slower to react. Denver Public library has a teen space "Zwinky". It's an environment to reach teen-agers where they are. PennTags -- users get to interact with content. Putting libraries in a mall -- for example, Camden Public Library. in NJ. Libraries with spaces in SecondLife. Libraries of things, not books -- library for designers of various kinds (Material ConneXion). Anyone can use; for pay, you get more access.

What Does It Mean

Some think next web will be more like SecondLife - 3D.

Quotes Yogi Berra: "the future ain't what it used to be."

Our industry is going through what other industries have. A new technology appears, it's disruptive.

Established industry must shift. Price pressure, ubiquity, accountability are results. Prices are pushed down, ubiquity increases, accountability -- people expect better results from old system to match new. Result is commoditization. A permanent shift in customer habits.

Odd behaviors occur. LIke products are available. Two customer types emerge (lagger and leading edge). Customer focus emerges (industry pays attention to different kinds of users). Partners become competitors. Competitors become partners. Segments and business models fall apart. Old business models fail, new ones arise.

Move is from product-centric to market-centric. Compete on market needs and differentiation. Google and Yahoo are our Wal-Mart and Target. Information as enterntainment, entertainment as information: Richard Saul Wurman.

Essential Actions

Become agile. Stay on top of trends, making sure you differentiate your service from "competitors". If we're in business of providing information, we need to be digital marketeers delivering digital experience.

Trendwatching: what is happening in the world. Follow the money -- where consumers spend that's where enterprises go. iPhone, green technologies, consumer spending habits. 2-3 year lag time between consumer web and information web. Think globally.

Q&A

Q: How do we reinstill or earn trust in products

A: Users are somewhat trusting automatically, but are highly aware of potential threats. User sophistication is rising. Increasingly jaded view of authority; but it's going to come full circle.

Q: How have people changed their information seeking?

A: Time spent with information is going up ,but time spent finding it is too. Users starting to recognize value of their time in finding information and are looking for more efficient ways to find. Turning to portals, expert communities, etc., not open web. This should move ratio toward more time spent using information, from where it is now.

Q: What does "semantic web" hold for us?

A: Semantic web is coming. And coming quickly. As are other developments in the web; things will look radically different in a few years.

Q: Talk more about 3D world. Where is this happening?

A: Look at virtual worlds. SecondLife is prime example -- it's a platform for duplicating Earth.

Q: In world where simple technologies (IM, del.icio.us, facebook) are booming, how to complex technologies fit in?

A: 3D will become simpler -- it's the next thing. But whatever it is, it must be simple, it must be viral.

Q: Our generation is developing the current tools baased on our models. They're successful. But what will be designed by the upcoming generation who think and interact with information and each other so differently?

A: It will be fascinating, whatever it is. But can't predict.

Q: Have people really changed that much? Card catalog represents a lot of research and a fit with society.

A: Yes and no. Can't throw out the old or reject the new -- but the old informs our reaction and implementation of the new. History doesn't dictate, but guides and informs.

Q: People want to be paid. What is upcoming for ways to pay each other on the 'net?

A: Copyright is a mess -- payment for use is failing. Technology to monitor content isn't keeping up. We'll see more digital fingerprinting -- where you can track content as it moves around. Google is creating transactional technology -- and they're in a position to provide content and payent mechanisms.

Q: Can you comment on shift taking place in power dynamics between engineers who created technologies and individual users' expression on these technologies.

A: Web empowers people; it's a platform for conversation. Web empowers more people; everyone has a voice. Everyone decides which voice(s) to listen to. Users will need to decide how to structure their online life; we've been empowered, but responsibility roles and mores aren't clear yet.

Quick Survey: Where Do You Read RSS4Lib

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Submitted at 11/8/2007 4:15:00 AM

I'm working on an article about blog readership via RSS and would like your help. If you have a few minutes, please fill out this three-question brief survey. Thanks!

ASIS&T 2007: Plenary Session: Clifford Lynch

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Will talk about issues Lynch has been thinking about -- role of universities and cultural memory institutions in a networked world. How is idea of collection changing in this world? When confronted with a confusing situation -- like today's information world -- in which economics, services have become dysfunctional, it's useful to go back to first principals. Refers to Ithaca, a Mellon spin-off. Has a research arm They've been looking at university publishing in the digital world. What is future of university presses? (It's ugly.) Their approach -- how can we fix the press -- not quite right.

Correct question is, what were we trying to do when we created university press, and is the press the right structure for that today. Or, are there different opportunities to achieve those goals?

Presses' purpose was to disseminate scholarship. Not to be house organs, but to publish for a circle of universities, provide some breadth, arms-length discipline. If that's the goal, then transactional, book-based model may not fit. We have lots of kinds of scholarship to work with.

Two notes: 1) History of university presses shows (Lynch thinks) that origins are complicated and less noble than you might think -- rationale includes procuring reasonably-priced printing services, for example; 2) is communicating scholarship part of fundamental mission of universities? In Netherlands, they have affirmed the latter point firmly; but not clear that's the case everywhere in U.S. Some institutions feel strongly yes -- that role of institution is to disseminate faculty's work (especially. publicly-funded state universities); others, not so much. itunes U and YouTube broadcasts of classes -- a follow-on to support of history of public broadcasting at these state universities.

Others feel that "publishing" belongs in technology transfer office. (Open Source movement in computer science departments conflicts directly with tech transfer, incidentally.) Libraries in universities are taking on "press-like" functions -- dissemination functions.

A big challenge for universities: do universities have fundamental role in stewardship of intellectual research? This is a fundamental role of research library -- but without funding (at federal/cultural level). There's a squeeze; technology increases. Libraries underwrite cost of data storage and preservation, run repositories, etc. Other entities in university do this, too: archives, museums also do this work.

Another problem in terms of resources for stewardship: Broad move to create digital surrogates of rare, unique/inaccessible material. Mostly non-book materials here. Museum tradition is "preserving authentic stuff" are in an interesting position. tension between preserving the real thing and creating surrogates. Ability to create surrogates is getting very good; Lynch says we can create surrogates that are good enough to satisfy a broad cross-section of scholarly, educational, and recreational interests. Mediated viewing allows, for example, 3D views of sculptures (such as Michelangelo's David) from viewpoints you can't have as a museum-goer.

You can, of course, duplicate surrogates endlessly and cheaply. Part of good stewardship should involve making those surrogates available broadly -- to protect against natural or man-made disasters. So if original, or original surrogate, is lost -- record isn't gone. This is counter to culture of collecting -- but world isn't the same as it was.

Another thing: for art that is repatriated, it should be thoroughly documented and "surrogated". After all, these works are "centuries out of copyright". National patrimony -- a way to have national digitized record of cultural elements that remain in the private sector at a level that's good enough for most purposes.

Heads of major research libraries are in a tough place: increasing expenditures for resources for researchers; budgets not kept up. At same time, need huge investments in digitization and -- in long term -- data curation. There are sources of money for this, but they aren't plentiful. NSF Datanet, private funding, start-up funding. This is all research and capability-building, not long-term. Lynch says funding will come out of traditional stewardship organizations. To change gears. Now talking about changing nature of scholarly publication and communication environment. There's an explosion of rethinking of scholarly work -- monographs, journal articles, data are all changing, evolving, becoming more complex. Data curation will be a big issue, not just in sciences but in social sciences and humanities, too. These challenges reach down into small science -- in fact, this is where the real challenge is. Big projects generally have good data collection and storage mechanisms. Small projects -- especially individual researchers, with no grant money -- don't have those resources (money or staff). The right support structures simply do not exist in most universities. Sometimes there's a bit in campus IT, sometimes in library, sometimes in departmental informatics groups... But scattershot and rare.

Growth of interest in "virtual organizations." Fundamental idea is that of "collaboratory." Researchers and students who want to work on a problem using the same data, the same instrument -- want ad hoc groups independent of institutional borders to get together, work, and go apart. Short-term or long-term, as needed. How do we support and curate data from this sort of project, when there's no there there? Proliferation of NGOs is similar -- often virtual organizations with similar demands and requirements. We are crossing threshold where people are authoring not just for people but for machines. Not just for indexing purposes, but for understanding, at some level, of research. Data needs to be available in forms that can be synthesized.

What does this mean? Lots of tagging and microformats for specific data types. Roles of publishers and authors in supplying this markup are unclear. How to attach structured data to article (and by whom?).

Overwhelming issues

1) Entire journal delivery system is not designed to allow text mining -- in fact, publishers stop this when they notice. Often contractually prohibited or limited. Some open access sites are text-mining friendly -- even zipping entire corpus and making it available. License and delivery mechanisms need updating.

2) Intellectual property issues vastly challenging. Definition (legally) or a derivative work is complex. Does an algorithm generate a derivative work? Legally not, probably. Output of a text summary tool may be a derivative work. Are your PubMed summaries derivative works? We're running up against a set of new challenges with very high stakes in copyright area.

Google is scanning everything, but in -copyright material is only provided as "snippets." Fundamental argument is that Google not doing economic damage by providing snippets. Google internally has a comprehensive database of literature which it can computer upon. We cannot know what they're doing with the results of computing on this database. This is a unique strategic asset. If they can develop text mining tools -- what can they do with it? It's a training set for a range of interesting purposes. Lexical analysis, AI systems... and more. We don't currently understand how to even talk about these questions.

Summing Up

We see an enormous amount of material produced outside traditional media. And mashups of things in and out of traditional channels. Pools of interesting content in Flickr, YouTube, hosted blogging services, The public don't really understand these as dissemination mechanisms; they see them as preservation mechanisms. These services are not preservation-oriented. Who fills that role? Who knows.

Problems of doing research are particularly acute in academia: human subjects, institutional review boards, etc. -- important roles, but get in way of rapid research. Corporate (Google, Microsoft, etc.) very concerned about individual privacy. Corporate researchers say they couldn't do their research in academe -- could not get through IRBs. Models of how we do research in academe need to be reviewed and updated. This is becoming a serious problem.

Interaction -- where will it lead us? Interaction is core of Web 2.0. We tend to trivialize this interaction. Where we need to go... Two sets of things around social tagging. One is language and vocabulary, how people want to describe things is in conflict with traditional stewardship organizations' methods. Users are often after different things. Other side of tagging is about assigning imprimatur -- things a person found interesting. Becomes a rating, of sorts. These are still simple interactions. Key point is that we're opening up our systems to the public in ways that have never been done before. Depth of description is potentially infinite; actual description often scant ("500 pictures of street life in Manhattan, 1951"). Enables a much wider conversation between cultural items and the audience. We don't know how to manage it. But the stakes are high: it's about building collective narrative and history. Revising and revisiting history. We are noticing that, if we do a good job curating what we have, they want to give more. How do we structure these collections across organizations? We can build virtual collections regardless of what makes sense geographically or organizationally. How do we structure resources (biographies, timelines) to be integratable into other tools.

Copyright remains a huge problem; most of the content that people will interact with was developed in living memory -- and therefore in copyright. How do we deal with that?

Validation of authority -- a library's opinion is seen as well-measured and accurate. How do you mediate disagreements between taggers or participants in these interactive worlds? It's very different from the challenges we're familiar with in annotating records the way libraries always have.

Q&A

Q: Google's document (code, documents, etc.) sharing work well; who owns stuff and what can happen to it while Google has custody of it?

A: General purpose tools to support scholars are important. We need to think more about what those tools should look like. Typically, when you use Google, etc., there's a license agreement you clicked through. You don't generally give away your copyright, but are giving limited rights to do things with your content.

Q: What about rights to digital reproductions of cultural works? Current practice gives those rights to the body that owns the physical work.

A: Museums don't own right to pre-1920 items (disclaimer from Lynch: I'm not a lawyer). They control access -- museum sets rules by which an image can be made (tripods, flash, etc.). On a policy basis -- we need to start talking about whether museums, as tax-exempt entities holding public cultural items, have right or obligation to distribute these items digitally.

Q: Are there records that should "gray out" after a while? Is there a "statute of limitations" on things like bankruptcies -- which vanish after so many years?

A: Sorting this sort of thing out is a huge social problem. Reportage should not be rewritten -- there's a slippery slope. There are public records and public public records (things that exist, but are hard to get; things that are truly public). When legal public records go public on the internet, there's a conflict. This needs to be sorted out, too -- as a social issue. Another question about how much you should be able to revise your own personal history. Facebook, Myspace, and their ilk open up these questions to a tremendous degree. Where's privacy boundary here?

Q: How can cultural heritage

ASIS&T 2007: Next-Generation Catalog: Prototypes and Prospects

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

OCLC
Chip Nilges
Nilges is VP of Business Development at OCLC. Currently working on WorldCat Local.
People view libraries favorably as source of great information (from Perceptions report). Report identifies a problem: where do you start your search? 84% say search engine; 2% started at a library site. There is a huge gap there.
How do libraries deliver value (collections, services, and community) to the user, on the network, at the point of need? This is what OCLC is trying to solve.
OCLC strategy to weave libraries into the web. Open WorldCat, WorldCat.org, WorldCat local came out of this strategic goal.
Open WorldCat a syndication project. Puts OCLC catalog records into Google, Yahoo, etc. Get data where it's being searched. Predictable URLs, machine interfaces. Hooked in to Google Scholar, for example. WorldCat.org -- a way to search the catalog. "Give away" worldcat data. Launched about a year ago; use of WorldCat overall has tripled in 3 years.
Things under development recently:
Personal profiles, citations (in various standard forms);
List creation/management/sharing, expanded metadata coverage to better expose collections of interest to users; Personalization -- features being developed now.
OCLC wants to get into job of citation management -- moving in that direction.
OCLC measuring traffic. in 2006/7, and 129.4 million referrals from partner sites to Open WorldCat landing page. 7.6 million clickthroughs from Open WorldCat to library services -- this is huge.
WorldCat Local: Not in original plan to release a next-generation catalog. But from library demand, it came about. OCLC "doesn't do portals" -- it's just a search box. Service is centrally-hosted, customized view and search algorithm. A library gets a search box and a custom URL. Standard search algorithm is 'tweaked' to present local items first. Local holdings displayed in record.
OCLC learning it's a different thing to design for librarians than for customers. Learning a lot about customers.
What's searched in WCL? WorldCat, metadata of 33 million articles, local repositories as indexed in WorldCat. Object is to bring in good enough data from OCLC sources that libraries can replace their federated search engine. Also indexing local repositories.
WorldCat Local fulfillment requirements: interoperate with local management systems and with local

delivery services. Pilot partners: University of Washington, Peninsula Library System, State of Illinois libraries, Ohio State University (12/2007), University of California System Melvyl pilot (spring 2008).
Upcoming features:
Institution search
Identities integration (http://orlabs.oclc.org/identities)
Big challenge for OCLC -- balancing local needs with global needs; local record vs. master record. User wants continuity, systems don't provide it. There may be an OpenURL resolver on the way; some clients are asking for it.
Q: Is inclusion of Open Access journals considered?
A: Yes -- open access books, archival materials, ejournals. Lots coming over next two years.

NGC: Next Generation Catalog
Andrew Pace
Our patrons are already "next generation"; it's our systems that aren't. Quick demo of Endeca -- faceted browsing, shelf browsing, etc. Why do Endeca? Unresponsive vendors; early experiments in NGC; casual conversation with Endeca; formal conversation with Endeca (2/2005-6/2005; fast implementation (7/2005-1/2006).
What's the big picture? Improve quality of catalog, exploit data already in the catalog. Build a more flexible catalog tool that can be integrated with future tools not yet invented.
Why do Endeca? Facets were a nice byproduct, but relevance ranking was the target. There's little in the literature about relevance ranking for bibliographic surrogates. Improved response time enhanced natural language searching, and true browsing. Automatic word stemming (for certain words).
Sits on top of library catalog system. Daily data load from catalog. Used to improve the discovery process.
Data and analysis
From July 06 to Jan 07... 67% of users do search. 20% do browse. 8% do pure navigation (through LCSH headings).
26% of navigation is by subject topics -- people are refining their searches by subject.
See Lown & Hemminger (2007) for a detailed transaction log display.
The "revolutionary war" problem. A search in catalog gives you LCSH subject headings. U.S. revolution gets 10 pages of subject records. In Endeca, working on this. Do you get the top n subjects in browse?
Expanding scope to 10million records in the Research Triangle libraries.
Emily Lynema and Tito Sierra -- a web service on Endeca that allows access to the catalog. Yields RSS new book feeds. Enables mobile device searching. New books wall w/jacket images. Resource lists for embedding in other web pages with web services.

Q&A
Q: Students when faced with too many options don't learn the best way to do something.
A: It's more important that they get what they want at the destination; entry path not so important.
Q: Endeca is "next generation OPAC"; what about next-generation catalog -- describing information?
A: NCSU hasn't done anything yet to change its cataloging practices; what they've done is exposed all that work so that it is accessible to users.

eXtensible Catalog
Judy Briden
The eXtensible Catalog (XC) is a project to design and build a system that provides libraries an alternative way to reveal library collection. Integrate library content into other systems. It will be open source and collaborative. Customizable locally. XC will have a UI with faceted browsing. Locally customizable without significant programming skills. Interface customizable. Multiple metadata schemas (MARC, DC, etc.). Informed by user research. Two phases to project.
1) One-year grant to write a plan. Completed in summer 2007. Proof of concept prototype, C4, that displays the basic UI that will be bundled with XC. Uses Lucene as search engine. Interesting feature.. from articles search, clicking a link (generated from MetaLib), rather than getting the OpenURL screen, user is directed straight to the full text.
2) Just funded -- starting the project.

XC can be used as a new interface to an existing single repository -- or integrate multiple repositories (at the interface level).
XC will address the needs of many libraries and be flexible, extensible -- anyone can contribute.
Q&A
Q: What open source license will XC be released under?
A: GPL.

Next Generation Catalog: the Minnesota Report
Janet Arth
In March 2006, Ex Libris demoed Primo prototype to UMn and others. They were looking for development partners. UMn became one of those partners. Bibliographic data are extracted from catalog and put into Primo.
Usability was in the contract between UMn and Ex Libris. Minnesota did studies. They have access to an amazing usability lab at Minnesota. Three usability rounds.
• First used proof-of-concept version (completely canned search results).
• Second used demo site with live, but anonymized, data.
• Third used live test site.

Most users actually use drop-down boxes to narrow their search (item type, with/without keywords, location) -- very few typed word and hit search without narrowing it.
In usability debriefing, asked about tags (a part of Primo). Users saw tags as way that future users could see what past users had thought. None thought they would use tags. Few in study actually used tags. Useful as a discovery tool -- way to expand search. But not strong support for tagging. Almost universally viewed as something others would use, not selves.

Q&A
Q: Are you happy with Primo?
A (Arth): Mostly yes; but realistically, we didn't have money to explore other tools the same way.
Q: Has University of Washington looked at how many people are using WorldCat Local vs. the native catalog?
A (Nilges): Not sure what 'take rate' is.
Q: Is there a web service interface to WorldCat local?
A (Nilges): ISBN, yes -- but not extensive yet. Coming soon.
Q: Preference for WorldCat local vs. native catalog?
A: In academic libraries, tendency toward WorldCat Local. In publics, the other way. Perhaps this reflects a difference between what's generally a system of libraries (academic) vs. a single library (public)?
Q: To what extent have we "bridged the gap" with these projects? Are we doing enough to get people to start their search at the library, or is this not even a goal?
A (Briden): Our content needs to be where students are doing their work; we can't change their behaviors. Library fits in their thinking, it's just not the first thing. It should be *one* of the first things, though.
A (Nilges): Ditto. Need to build interfaces that allow your services to be everywhere.
A (Pace): We need to avoid self-fulfilling prophecy. We need to make our catalogs useful, entertaining, helpful -- so when people do get there, they like the experience and find it of benefit. Make catalog "sticky".
Q: Does the underlying catalog data need to change to continue making improvements?
A (Arth): We have good data. Challenges lie in merging it.
A (Nilges): Separating inventory management and finding; pulling other data in with the cataloging. Not clear to what extent the data need to be unified; perhaps only connected.
A (Briden): Opportunity to bring tags into collaboration with subject headings; use tags synergistically. Catalogers have opportunity to work with user-generated data. Pull it together in ways that will make more sense.

ASIS&T 2007: Opening Science to All: Implications of Blogs and Wikis for Social and Scholarly Scientific Communication

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Bora Zivkovic
What are sci/tech bloggers doing?
Fun stuff... Changing policy. Scientists are not humorless automatons. A way for "fun" to appear within scientific literature. Science and art, history of science. Blogging from the field -- talking about field research.
Serious stuff... Snippets of research too "small" to be published, but valuable. Sometimes hypotheses and data -- open notebook science (in a later talk). Blog carnivals -- ad hoc popular journalism. One editor collects posts sent in by others, posts link list in a single place. Editorship rotates among group.
Popular magazine editors; some have blogs. Serious publishers do, too.
Blogs are starting to be locus of open access publishing and review -- reviewers don't comment on quality of paper, per se; rather, on value of information being added -- is it worth publishing? Trackbacks can allow one to see who else in the community is commenting on a paper. Scientists who are bloggers write comments in a few lines: short, blunt. Non-blogging scientists write paragraphs with references; very polite and subtle. A clash of cultures.
Impact of open discussion on research will be immense.

UsefulChem: An Open Notebook Science Project
Jean-Claude Bradley
Jean-Claude coined phrase "Open Notebook Science".
Speaker runs a chem lab at Drexel; manages student researchers. Talk is about how they share their research. Talk
There is a continuum from closed to open in how science is reported:
• Closed research: Model is the traditional lab notebook -- unpublished, fundamentally personal. Failed experiments are never seen by anyone.
• Traditional journal article: Mostly open; but you need a subscription to journal. Not as convenient.
• Open Access Journal: Available to anyone online. Some journals require authors to pay to be published.
• Open Notebook Science: full transparency. Everything that's done is recorded and available.

Where is science headed? we are between human-human communications and human-computer communication. Research is moving in direction where computers start to manage research -- plan experiments. It will be a self-organizing redundant projects. Critical factor: being able to read and write (publish) with zero cost. Publication of all aspects of the scientific process: open notebook science. Total transparency.
If machines "do" science, how do they know what's important? Ask humans. In other words, search texts for things like "next steps", "what's next" and answer those questions. Malaria is a good venue for this: big problem, no big money for drug companies.
Started out blogging things... Moved to wiki because wikis are better at organizing things. Wiki enabled broad discussion. The successes, and importantly the failures. Also, blogs don't have record of changes. Wiki enables the history to be preserved. Result is UsefulChem.
Things are indexed in Google, time-stamped, findable. History of editing is available to all.
How do people find experiments? Free tool, site meter, shows how people are finding the wiki. Some via RSS, some via searches (mostly Google). Molecules are tagged in wiki using InChI. Google handles these pretty well -- so a good tool for researchers to use. And of course, raw data are available for every experiment.
They are still using a blog, but using it do point to things in the wiki, define problems. Blog is targeted toward other chemists, not public. Open Science lets you connect with people at other institutions and collaborate -- you find each other in the course of your individual research. Interestingly, mailing list is still tool for intra-group collaboration than either wiki or blog. Also using Second Life to hold meetings.
Q&A
Q: How do you achieve institutional buy-in for open science? Many scientists/researchers/academics are not good at sharing
A: Need to find people who share the vision and lead by example. Growth of open notebook science is going to be slow. Impact will be big, though, over time.
Q: How easily are graphics handled in wiki software?
A: There's a free Java viewer for images -- to do "zooming", etc. -- so there's no burden on user. It's just there, part of the open source movement.

Social and Scientific Implications of Science Blogging
Janet Stemwedel
Interested in philosophy of science and ethics of science. Blogs at Adventures in Ethics and Science. Talk
Scientific communication is essential to scientific practice: to share results (with public, with each other), to articulate theories, to train new scientists.
Traditional channels of communication are peer-reviewed literature (this is how "score is kept"). Tenure, promotion, existence as a researcher all tied up in peer-reviewed process. Peer-reviewed literature is a back-and-forth between scientists over a long time scale. Research tends to be secretive until [eventually] published. Peer reviewers are necessarily your "competitors" -- experts in your narrow field.
Also conferences -- shorter timescale. Informal conversations and discussions. These tend to be ephemeral; thoughts vanish after being uttered, and those not at the conference don't take part.
Press releases, popular publications, etc. -- these tend to be one way, from scientist to public. Science journalists end up being gatekeepers.
Problem is the knowledge-building requires good communication. Only way to get to objective knowledge is by having many people comparing results and interpretations. Interdisciplinary tools and approaches are key. Challenge is avoid duplication and avoid already-discovered dead ends.
So what's wrong with traditional channels of communication? Most communication comes at end of project, not in midst. Not much collaboration or input. What's reported reflects author, reviewers, and journal editor. Not broad community. Vast amount of information is not reported, especially things that don't work.
Blogs hold promise to improve this. Offer back-and-forth on short timescale. Less ephemeral. Potential to expand audience broadly across geography, disciplines, backgrounds. Blogs may be free of existing pitfalls of peer-review (inherent conservatism in process). Quality control is interesting; posts are viewed and commented on more broadly. Through discussions on blogs, we get a window into science as process, not result. This is important to scientists, as well as to public.
How does community of science function? Blogs can open up this community a bit to scientists. Scientists are loathe to discuss process by which they communicate. The community is opaque from the outside. (And from the inside.) Blogs can help expose this to those thinking of entering the field. You can have a virtual community in place of the real one that may not exist where a person is. Opportunity to change mode of community conversation.
Your audience becomes the audience of the willing. Do you blog as yourself or anonymously? If yourself, there's risk; if anonymously, people don't know who you are.
Can blogs shift the culture of science? Now, see things as competition for scarce resources. Blogs could help make mentoring be taken more seriously. Expand audience to the non-scientists. Ongoing discussions will review that science is a process, not a result.

Q&A
Q: What are risks to intellectual property in open science?
A: Large -- if you're interested in a patent or IP, open science isn't right for you.
Q: How will wikis change university?
A: When people who have tenure feel the current process does not work anymore. It will be slow and evolutionary.
Q: What is key research question that you think is important to investigate (in terms of how to use blogs/wikis to support science)?
A (Stemwedel): How do scientists learn to be good scientists? How is that changing?
A (Bradley): Study how science gets done through Open Notebooks -- see how people change minds, react to data, etc. Interesting to see how other scientists "do" science.
A (Zivkovic): Blog is software, not way of thinking. What you do with it is what is important. Publication of paper is not end; it has a life after publication, and that life is now public and observable. A second stage of peer review.
Q: How do electronic lab notebooks (aimed to decrease "cheating" in science) interact with open science.
A (Bradley): Having a wiki enables me to mentor students, via wiki, several times a day. Also opens mentoring to anyone.
A (Stemwedel): Electronic notebooks are scary because disks can get destroyed -- centralized online storage is safer in the long term.
Q: How do we view authority in "science 2.0"?
A (Zivkovic): Nothing new; authority is built over time. Some blogs will be "citable". We will figure this out. Comments on Public Library of Science get DOIs -- the comments are citable. Idea of "citable unit" will change.
A (Bradley): Blog posts can go to Nature Proceedings; no peer review, but editorial review. And there's a DOI, too. Be sure to keep copyright if you want to do this.
A (Stemwedel): People are using authority of reviewer as a substitute for quality of reviewer.
Q: How do you know with whom to collaborate?
A (Bradley): I'll work with anyone with something to contribute. Can't rely on traditional authority; rely on actions.
A (Stemwedel): Interactions within scientific community, not narrow research. Blogging can be a powerful support tool for researchers.
A (Zivkovic): Open access science is critical to globalization of science. Helps reduce data privilege, especially outside developed world.

ASIS&T 2007: Understanding Information Work in Large Scale Social Content Creation Systems

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Submitted at 10/24/2007 1:58:31 PM

Wikipedia: Distributed Editorial Processes
Phoebe Ayers
Who is wikipedia? It's thousands of people behind the site. Lots of groups joined by shared values of openness and shared values: free content; open to all; key editorial policies (Neutral Point of View, no original research, verifiability).

How do tens of thousands of people with no top-down control write the world's largest encyclopedia? Wikipedia is governed by non-profit foundation. Has several sister projects -- we're only talking about wikipedia. No one is in charge of editorial decisions. Wikipedia has a modest goal: giving every person full access to the sum of all human knowledge. There are lots of self-organized tools -- for cleaning up articles, for defining NPOV, for style. Information works in wikipedia as a sum of distributed social processes and the technical structure of the wiki and culture of openness.

Technology, Theory, Community, and Quality: A Talk in Two Acts
Dan CosleyAct I:
Matching people with tasks they're likely to do motivates contributions The problem is that some articles need help in some way (items are tagged). These articles are listed on a community page. If you want to fix something, hard to find a page you want to fix. Built a recommender engine so that people are given pages to edit based on things that they are likely to be interested in fixing. This worked well in MovieLens. Translated to Wikipedia. Wrote SuggestBot -- it goes through list of articles tagged as needing help; finds items that are similar to items that person has edited, written, etc., before.
Through wikipedia, you can see if someone edited article. Four times as

many articles get edited through recommendation engine -- it works. Other communities should take this approach to editing/moderating. Or match a new user in a community to an older member who talks about similar things.
Theoretical basis for this: collective effort model says lower effort = great reward. Therefore we should build interfaces and algorithms that help people find work to do.

Act II:
Understanding community is huge for improving information quality Knowing system (wikipedia content in this case), knowing users, and knowing habits all help inform the recommended engine. A failure: an automated welcome to the community to new users (people with their first edit in December 2005, about 28,000 people). Looked for people who had "welcome" on their home page. People with "welcome" messages edited more entries. However, wikipedia culture was that only good members got welcomes (bad members got warnings). But there still seems to be an effect -- people with a welcome message went on to be a bit more active in wikipedia. But this is not strong.

Information Quality Work Organization in Wikipedia
Besiki Stvilia
Why do work organization models matter? To design effective, sound, robust models for different contexts/domains inexpensively through knowledge reuse. To establish benchmarks for analyzing and evaluating existing models.
Questions studied: How does the community understand quality? What processes exist? What are motivations of editors? What are dynamics of information objects? Why do people contribute? What IQ intervention strategies are used?
Percent of pages in wikipedia devoted to articles has decreased from 53% to 28% since 2005 -- more effort is going in to talk, discussion, and so

forth pages, less on articles themselves. More emphasis on community building by its users. IQ processes: content evaluation, editor evaluation, building and maintaining work infrastructure. Differences between wikipedia and other systems. First, user feedback and information creation are the same process in wikipedia, unlike other systems. Quality control and author of data are separate, for example, in library catalog. End user and editor roles are merged. Product creation and delivery environments are the same. Work coordination is informal and ad hoc.
Wikipedia controls quality through content and editor evaluation. Some parts of process are formal, others are informal. Because there's little built-in mediation, disagreeing parties must come to their own agreement (or else endlessly erase the other's contribution). Community experiments with different intervention processes when there are conflicts -- trying to find the best approach at any moment.

Wikipedia Reference Desk: Processes and Outcomes
Pnina Shachaf
A study to evaluate the quality of processes and outcomes at wikipedia reference desk. There is a reference desk at wikipedia. It uses a wiki to process reference transactions. Users leave questions; wikipedia volunteers help users find the info they need. Organized under seven categories: computing, entertainment, humanities, language, mathematics, miscellaneous, science.
Not a lot of work in social aspects of Wikipedia community. In particular, opportunity to learn from wikipedia reference desk as a way of improving service in traditional reference desks. What is quality of answers at reference desk? Looked at 210 transactions and 434 messages (in April 2007). In this month, there were 2000+ transactions and 11,000+ messages. Most were in science and

miscellaneous categories. Most responses per question in mathematics. (Entertainment and Miscellaneous had the fewest.) 170 users (122 expert, 48 novice); 34 participated in multiple reference desks. Experts are more active at reference desk. Novices submit more questions (44 vs. 33). Novices are more likely to ask questions (70% of novices, 29% expert); experts answer more questions. By profession -- computer/IT professional are plurality.
Most questions (96%) got an answer; 92% got a complete or partial answer; average time to first response is 4 hours and to last response 72 hours. Accuracy level is about 55%. Response completeness 63%. There is question negotiation; 28% of time there's a follow-up post from requester. There are elaborations -- improved answers -- 67% of the transactions. Additional resources, different point of view, different solutions, etc.
Wikipedia reference desk quality is "not too bad; can be improved probably". Collaborative effort yields interesting results. Future study will try to compare with small groups of librarians who use a collaborative process.

Q&A
Q: How did you determine accuracy of response?
A (Shachaf): Involved qualitative analysis of answers (reading them); results presented are preliminary, one-reader reviews. Final research will involve multiple reviewers.
Q: What are views on copyright of materials in Wikipedia? Is there analysis of plagiarized in Wikipedia?
A (Ayers): There should not be anything in Wikipedia that's under copyright. In practice, hard to deal with this.
A (Cosley): There's a tag in Wikipedia for identifying possible copyright violations.

Bloggers for Peer-Reviewed Research Reporting

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Submitted at 11/2/2007 8:33:18 AM

There is a movement afoot to encourage and support "serious" blogging in science. Bloggers for Peer-Reviewed Research Reporting [BPR3] is a group of scientists who have made a step in this direction by releasing a set of icons that scientists are invited to include in their blog posts when "they're making a serious post about peer-reviewed research." BPR3 is an initial effort to encourage scientists to identify their commentary on peer-reviewed

research articles -- whether the article is online or in print -- with an icon. The next step, according to BPR3's web site, is "to use bpr3.org to aggregate all the posts discussing peer-reviewed research from across the disciplines." If this effort succeeds it could well open up new doors to scholarly debate and discussion.

Why does this matter? Well, as was first brought to my attention at the ASIS&T panel discussion on Opening Science to All: Implications of Blogs and Wikis for Social and Scholarly Scientific Communication,

there is a great deal of communication, debate, and discussion of scientific research within the blogosphere. However, unlike letters to the editor in peer-reviewed journals, there is no standard method to capture, collect, or forum for evaluating the opinions of blogging scientists. To the extent that research -- and discussion of research -- moves into the public sphere, there is a great opportunity for the scientific community to add to and discuss research as it happens. [Via The Medium is the Message.]

ASIS&T 2007: Research Directions in Social Network Websites

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Submitted at 10/23/2007 9:03:26 AM

Where's My Fieldsite?
Danah Boyd
Looked at high school students out of school hours. Make sense of what teenagers are doing by looking at snips of their lives. Answer questions, what are the publics in which we live?
Public and private are different for teenagers than for adults. Children have geographically constrained lives. Culture of fear -- you might be

ASIS&T

hurt outside of home. No social spaces outside of home. Commercial spaces are increasingly constrained. So what do teenagers do? They go online. Cause and effect are reversed from popular conception: children don't hang out online because they want to, necessarily; they do because it's the only option.

Networked publics -- spaces or collections of people that exist within and through mediating tools that network people. Has 4 properties:
1. Persistence -- things stick around.
2. Searchability -- you can find things -- including your kids. Everyone is searchable. Problem is that you don't want to be searchable by anyone; you don't want to be found the wrong person.
3. Replicability -- conversations can move from forum to forum. You can edit things and repost. What's original?
4. Invisible audiences. You don't get feedback from those with whom you're addressing. In real world, speaker knows to whom she is speaking. We address our talk to that context. Not so in networked public. What are social norms online? They are different, and evolving.

ONline concept of friends -- putting audience into being. Defining to who you are speaking when you post. "Public by default, private when necessary."

Teens' idea of privacy is that they can control the audience, or have semblance of control They do this 3 ways:
1) structural walls -- they put up info that hides them.
2) social demand -- create a space that's mine, not yours.
3) playing ostrich -- if I don't see you, you don't exist.

Public life is changing. Mediated and offline are growing together. Conversations have fluidity -- they occur across media. Public life is incorporating all of this -- online and offline -- into something new.

Information diffusion and users' behavior in Fotologs
Raquel Recuero
Based on Fotolog users in Brazil. A two-year study. 20% of Brazil's population has online access; social networking sites are very popular (more profiles than online people). Fotolog is a simple site. People make fotologs about tons of topics. It's been extended by its users.
Identity appropriation -- create an identity. People select images and text carefully -- lots of thought goes into it. Pictures are carefully photoshopped; perception of self is important.
Social interaction appropriation -- most important thing in Fotolog. Comments are critical -- interaction with Fotolog is important to users. Unique fotolog nickname is important. Groups emerge and conversations take place across groups.
Fotolog is an information tool. Decide what to publish based on perceived gain of doing so. Value is related to social capital. Users think carefully about what info they will put on fotolog -- value based on interaction.
Information that creates social interaction spreads within a group before it spreads across the network. Spreads among people who are closely bound. Perceived value is to make people closer to you.
Perceived value of information is what defines what information will be

disseminated.

Activism and Social Network Sites
Alla Zollers
Activism: an intentional action to bring about change. Emphasis on change.
May Day protest 2006; students used MySpace to organize walk-outs.
Social network sites consist mainly of weak ties.
Studied 100 Facebook groups (Politics and Beliefs & Causes) and 100 MySpace groups (government and politics). Content analysis.
Does participation in online groups lead to offline action? Unclear. But there is discussion. Does the architecture of the site effect activist activities? Do people interested in activism go to a site because of the site, or because their friends are already there?
Analysis of Online Social Networks
Fred Stutzman
Research focus on: 1) privacy; 2) dynamics (how systems grow, how friend patterns change); 3) context (how networks answer situationally relevant needs); 4) affordances (what social networks offer to friend-seekers).
Analyzed network characteristics, connections in the network, status in service, privacy, consent, terms of service.
What to think about when doing this sort of large-scale data collection (in Facebook, in particular)? In Facebook, an out-of-network person has different ability to see others' information than an in-network person. Faculty see less than students. Anonymizing profiles to protect student privacy. What about consent? IRBs do not have a good way to deal with getting consent from users.

Dealing with terms of service of the site. Facebook granted exceptions until 2006.
Built a Facebook application, "Your True Self". Analyzes your friends' profiles, shows friends who share similar taste. A way to gather information about users via the Facebook Platform.
Question: what does a "friend" represent? In real world, "friend" is on a continuum; in Facebook, it's binary. But hard to know what it means.
Q&A
Q: What governs parental access to MySpace?
A (Boyd): Lots of things. Some kids want parents there, others don't. Privacy rules by service make a difference. Differences in privacy concepts based on race and class, as well; different concepts of privacy and of utility of tool.
Q: How does online community effect the decline of "belonging" that we see in F2F world?
A (Zollers): THere is interaction and debate in online world; this might translate into further, real-world, action.
A (Boyd): Lack of agency means lack of political engagement. Teenagers don't have access to meaningful public spaces; so they feel withdrawn and excluded, so don't participate.

Q: Are there any qualitative research methods to use?
A (Boyd): It depends on the question you're asking.
Q: Did Stutzman's analysis take into account kinds of schools?
A (Stutzman): Yes; it covered a wide range of schools.

ASIS&T 2007: Wrap-Up and Thoughts

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

I had a great time at the ASIS&T 2007 conference, Joining Research and Practice: Social Computing and Information Science, in Milwaukee. I blogged most of the sessions I attended -- see the list at ASIS&T Sessions. A few thoughts about particular sessions or things I picked up.

I experienced one of those so-simple-it's-genius moments during the session on "Live Usability Labs" by Paul Marty. The technique Paul employed -- running a usability test with two people, each in different roles for the event -- worked stunningly well. It completely avoided the awkwardness of one person thinking aloud -- hardly a natural state for most of us -- while explaining the actions being taken on screen. By play-acting, two people in the roles of graduate student and faculty member, or two colleagues, elicited great feedback from each other about what was on the screen,

how it worked, and how each person expected it to work. It seemed a particularly effective technique for teaching usability to others, but I'd bet it's very effective in a more traditional usability testing situation, too.

The session on "Opening Science to All: Implications of Blogs and Wikis for Social and Scholarly Scientific Communication" was one of my favorites because it showed both some empirical research as well as effect. Jean-Claude Bradley's presentation of UsefulChem as a place where scientists can record their experiments, successes, and -- this is the key point -- dead ends gave me another "Aha!" moment about the impact of blogs and wikis on science, education, and society. Janet Stemwedel's talk on the societal implications of blogging -- particularly within the scientific community -- was also very interesting. The divide in the sciences between those who embrace the openness two-point-oh technologies engender is even starker than in the social sciences and humanities,

domains in which I'm more comfortable. At the same time, the potential short-term benefits to the general population are even greater in the sciences than in the humanities.

Clifford Lynch's keynote address on open access was also informative and engaging. He asked the audience to consider what it was that academia wanted to achieve when it created the institution of the academic press and whether that role is currently being met. He says that one of the biggest challenges for universities is to decide if they still have fundamental role in stewardship of intellectual research. While this is a fundamental role of research libraries, their parent organizations expect them to accomplish it without the depth of funding or support that is necessary. If libraries, or universities, are the stewards of intellectual research, they must make great strides in technologies to ensure that today's research is fully usable in the future. Lynch left far more questions unanswered than he answered -- it was truly a thought provoking and stimulating talk.

On a very much related note, I was struck by the fact that numerous academic researchers made comments in the course of their presentations about how information -- reports, documents, data, etc. -- are all available on Google and so not much attention needs to be paid to stewardship. I fear that too many people, in and beyond the academy, view Google as the universal library. This is far from the truth. Perhaps Google is the universal card catalog, but even that is a stretch. Google's business model is very different from that of a library. Google is all about access (local copies for indexing aside); libraries are all about preservation and stewardship of information. (Saturday's Unshelved comic strip makes this point more humorously and succinctly.) As a librarian, I grow concerned when academics -- the primary user population I support -- so blatantly misunderstand the role of the library.

ASIS&T 2007: Social Computing, Folksonomies, and Image Tagging: Reports from the Research Front

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Submitted at 10/23/2007 3:48:30 PM

User Supplied Image Category Labels
Hemalata Iyer
Study's goals were to identify underlying structure of image tags. Analyzed 105 participants' labeling of 100 images. Images tagged and organized into groups. Identify a prototype image in each group. Identify significant feature of prototype image.
Example of hierarchy: furniture (superordinate), chair (basic level), kitchen chair (subordinate). The basic level has more distinctive properties than superordinate, but isn't too specific.
Out of the 899 category labels applied, ~58% were superordinate, ~38% were basic level, and ~4% were subordinate. Interesting -- it was thought that basic level would be most common.
A group of people displaying emotional behavior was grouped as "emotions"; facial behavior was prototype. Categories can be built around prototypes; for any category there is likely to be a single prototype. Familiarity, culture, environment effect selection of prototype.
Superordinate terms and significant features of prototype image are important in indexing. Retrieval and browsing: grouping facilitates browsing.
Social tagging: group labels tend to be superordinate. Individual images in that group tend to be tagged non-

hierarchic related terms. Associations, not hierarchy. There is not much structure (does this matter? unclear). First tagger influences subequent taggers. Perhaps first tag should be done by an expert, to subtly guide future taggers.

PhotojournalsmAndUADs geotagged:ASSISST2007MilwukeWi topresent
Diane Neal
Yes, title is intentional.
Needs of photojournalists are different from other photographers in terms of tagging.
Photojournalists select what to photograph and to store their photos in their publication's photo archives. Photo editors pick photos to go with stories. Also worked with photo librarians.
Where is the locus of control -- internal it's something you can control; external -- blame on something outside, beyond, you. We like to have control over our pictures (they're something we save in a disaster, we like to have them). Photojournalists and editors were studied.
People found named objects, specific events, browsing, user-assigned descriptors (UAD), metadata as the most important. Descriptors, in general, were most important kinds of labels. Started with a keyword, moved to browsing. Like metadata-based searching.
Problems with people doing tagging -- inaccuracy, errors, typos, lack of time. Need to formalize rules for tagging (somehow). tag guidelines

(ie., no plurals, no compound words, etc.).
Presentation
Abebe Rorissa
In classic info retrieval, a document representation (surrogate for document) is matched with a user query (surrogate for information need). In new world... We have huge multimedia digital librareis; not single items, but collections. Many things are not text, they are multimedia. Retrieval systems more complex to match queries and document representations. Now we're looking at slices of information space, not documents.
User is creator, annotator, indexer, searcher, and consumer of content - all roles formerly done by authros and professional indexers. Users have their own language, not the controlled vocabulary. Rise of tags and folksonomies, not controlled vocabs.
Challenges
Users' roles change, often in mid-research. They have simultaneous multiple roles. We have to react to individuals and groups of users. MNeed a more complex information retrieval model. We have "a million typing monkeys". We have to deal with free and uncontrolled sers' langauge and vocabulary.
Opportunities
The million typing monkeys are also an opportunity. Users are wiling to contribute descriptions of ocntent. Rich data to study tagging behavior (great for researchers). Need to find ways to let user tagging inform our retrieval systems.
What Next?

Probably no single model will capture whole information environment. Browsing is important feature of IR. Revise Ranganathan's second law: Every user his/her overview of the document collection". Still need way to get to single document.
Two tools to look at:
FlamencoPhotoMesa
How do you provide access? People tag at a high level -- broad terms. Best entry level in a browsing interface should be the basic level; where people search. Depth of hierarchy is a problem. Hard to display breadth of terms in a functional way.
Social tagging is an opportunity, not a challenge.

Semantics of User-Supplied Tags
JungWon Yoon
Wide gap between terms used by taggers and terms used by professional indexers. There is not a thesaurus to get from one to the other -- at least, none now.
Generic terms are most frequently used terms. 75% of generic terms are in formal index (LC TGM). Studied occurrence of colors as tags in Flickr and in LC TGM.
What are relationships that are most useful for users?
Tags of specific location were frequenlty used in Flicr. TGM doesn't include specific geographic locations. But related tags don't follow regular patterns.

ASIS&T 2007: Live Usability Labs: Open Access Archives and Digital Repositories

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Submitted at 10/22/2007 12:30:38 PM

A series of live usability tests, with volunteer repositories and testers from the audience.
1: dLIST (University of Arizona)
Not a single-institutional repository; it's a cross-disciplinary cross-library repository. Typical user of dLIST looks for information. We'll test 1) an author search; 2) browse for neural networks; 3) can the tester find usage stats for a specific article.
• Author search -- no problem. Found works by specific author. Tester had a hard time finding the "search" button -- it was screens down the advanced search page.
• Neural network. Had a hard time finding how to do phrase search. Search fields don't specify a phrase, just keywords in any order or any keywords.
• Found the article and

abstract/download stats handily.

Users tend to be "browsers" or "searchers" in Paul's experience. Search box says "search titles, abstract, keywords -- but doesn't search authors. They aren't a keyword (other searches showed that indexing sometimes includes authors, but not consistently. Also, on advanced search page -- Paul Marty says "if you need a cancel search button, don't make it bigger than the search button."
Home page is very detailed.
Q: How much prompting of the user in a session?
A: It depends; in a purely exploratory session, you give none; in other cases, if you're less concerned with how a task is completed than if it's completed, you can give more.
2: Illionis Digital Environment for Access to Learning and Scholarship (IDEALS) (University of Illinois at Urbana-Champagne)

IR at UIUC. Concentration is on scholarly research and output at the university. Mostly 'gray literature' and content from departments that are publishing technical reports. Most people find IDEALS content through Google, etc. -- roughly 10 times more access of full-text materials than through the IDEALS search interface. Task: 1) Upload an article to the IR. Then a page of legalese. Two volunteers -- a "faculty member" and a "graduate student". Submit an Item -- not called "upload". Then it asks for "choose collection"? What's that? Collections don't match expectations. With two people, the give and take was very rich -- since people don't think aloud, having a situation in which conversation is natural helps elicit conversation. Technique is called "constructive interactionism".

3: Minds at UW (University of Wisconsin)
It's a consortial collection -- all 26

libraries in the Wisconsin system. This implementation is almost purely "out of the box" -- DSpace is moving to a new platform soonish. Most uers come from Google direclty to an item page.
Tasks:
• You Googled your way to a particular work. You want to find other items by this author.
• Look for other examples of Urdu poetry.
• Look for other contributions from the same school (UW-Whitewater)

1) An author search (full-DSpace) pulls up many false hits. Browsed by author to get to him, found his works.
2) Look at record, look for subject heading -- but no clickable links.
3) Went to communities list, then found UW-Whitewater, then searched.

ASIS&T 2007: Social Computing as Co-Created Experience

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Social Computing as Co-created Experience
Karine Barzilai-Nahon
Gatekeeping: information control. Lots of concepts throughout literature, concept dating back to the 1940s. Somewhat fragmented. Ways gatekeeping is defined -- what's the rationale for gatekeeping: protection, preservation of culture/social, linking, facilitator, editoiral, disseminator, change agent, access.
Barzilai-Nahon's definition is Network Gatekeeping. Information control (not in negative sense) but in sense of channeling, facilitating, editing, adding, deleting information. Focus not on the gatekeeper, but on the "gated" -- those whom gatekeepers act on. Four attributes of gated:
P Political power
I Information production
R Relationship -- frequency, duration
A Alternatives -- information society creates more autonomy because we have more alternatives.
Dynamism between gated and gatekeepr.
Research asked two questions: What kinds of message do we delete and why? They found about 8 reasons for deleting messages in forums. Main reason -- if someone hurt the community. Then spam. Then off-topic messages.
Gatekeeping self-regulation mechanisms. For example -- censorship, editorial, channeling, localization mechanisms. There are designated gatekeepers and informal gatekeepers. Designated -- managers in a formal role -- and informal -- community members. All gatekeepers tried to keep homogeneity.
90% of "guest" users were, according to IP addresses, regular registered members who were entering anonymously. People often used guest account to make critical comments they didn't want made under their real persona. These comments often got deleted.
The Four Attributes
Do gatekeepers have politial power in any context, online or otherwise?
Two Takes on Virtual Design: The Construction of Expertise and Embodied Design in Second Life
Deisgn Teams
Kalpana Shankar
Collaborative Virtual Environments [CVE] (There, Active Worlds, Second Life). These are defined by collaborative design, mied reality, ecoerce, education, and enterprise. Not stricly games.
What's in a metaverse? Builds -- there's nothing there when the first user goes in. Users build the world around them. Users can occupy space at the same time or at different times. Live chat and leaving a message. Live interaction via avatar.
Research questions
1. How does virtual collaboration affect and influence deisgn activities in Second Life?
2. How does the designer's experience of embodiment shape emergent design practices?
3. How is design executed?
4. How does SL design become integrated in real world design? Methodology. Learned about SL through interviews and SL's "sandbox" -- a place to learn the space. Then recruited two design teams to observe and interview. Interviewed in SL, via chat.
Then, once understood what they wanted to observe, did ethnographic observations -- watched teams work, gathered chat logs, conducted follow-up interviews.
Embodiment: The bodily aspects of human subjectivity: the human body's physical presence. There's also the "experience of physicality" -- how users see themselves and present themselves to others.
Presence: Each user has a unique graphical representation. There are rules -- you can't walk through walls, you can't be invisible. Artifacts are similar. They can be given and received. Like in physical world.
Awareness: Understanding viewpoints and attention of team members is crtical in collaborative design activities. Gestures to point at things, verbal "over there"
Location: Space vs. place. SL users create spaces conducive to the activity they're doing. Even though there's no "need" for it in SL, people create elaborate spaces in SL.
View Manipulation: You can see youself on the screen, but can also look in other places. Avatar is not eyes.

Conclusions
Technical infrastructure and the notion of presence. People create a space in wich to work, and then build team identity. This requires management and knowledge of whom you are working with. Lots of uncertainty because you don't know about the avatars the way you would abe real-world people.

Q&A
Q: How does perspective (first-person vs. over-the-shoulder) change interaction?
A: Some research done, but not much. But perspective needs a lot more work.
Q: What are benefits of social capital in online communities?
A: Social capital serves the individual -- you get listened to more. For example, eBay's new ad campaign ("Shope Victoriously") -- idea that it's better to compete than to cooperate. Virtual social capital is not the same as real-world. But it does aid connectedness.
Q: What studies have there been to compare avatars with real world person, and why they choose the avatar?
A: Not really. Avatars are fairly limited -- out-of-the-box you can't change things a lot (with programming, you can).

RSS4Lib Survey Results

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Thanks to the 137 of you who have taken the quick subscriber survey that I posted on November 8. Based on my best-guess estimate of my readership on this date, I had 1513 feed subscribers on November 7, the day before I launched the poll. This represents a respectable 9% response rate.
I asked three questions in the survey:•
Do you subscribe to the RSS4Lib RSS feed?
• What tool were you using when you saw the post about this survey?
• Where did you first see the link to this survey?
Of people who took the survey in the first two weeks (7:15 AM EST November 8 - 7:15 AM EST November 22), 96.2% (127 of 132) respondents were subscribers. Interestingly, though not necessarily significant, two of the non-subscribers ho took the survey in week 1; the other three did so in the week 2. Part way through week 3, all five of the additional surveys submitted have been by subscribers.
Of the five non-subscribers who took the survey in the first two weeks, three were at RSS4Lib when they saw the survey and two saw it linked in another blog.
Web-based aggregators are the clear favorite among respondents. Bloglines has a 43.9% share of the first two weeks' respondents (including eight users of Bloglines Beta). Next is Google Reader, with 42 users (31.8%). The numbers then dwindle dramatically, with five people reporting they use Sage Firefox extension and three or fewer using a variety of other tools.
Finally, I asked respondents where they saw the link to the survey. An overwhelming number of respondents (115, 87.1%) saw the survey link in RSS4Lib's RSS feed. Of the remaining 17 respondents, five noticed it on the RSS4Lib site, five at unspecified "other" or "don't know," and four others in various other blogs. From reviewing the referer logs and respondent comments, I note that two of the four came from the University of Michigan library's "superfeed of library and librarian blogs and two others came from blogrolls at other sites.
I also captured the user agent (the way the web browser or application identifies itself to the web server). Firefox is the browser of choice for two thirds (88 of 132) respondents, followed at 30.3% (40 of 132 respondents) by a mix of tools that don't identify themselves, followed finally by Safari (2 respondents), Internet Explorer (1), and Vienna (1). Mac users, by the way, account for 17.3% of respondents whose user agent identified itself, with Windows users making up the remaining 82.7% (only two user agent's identified themselves as being Vista).
I've been interested to see the 'long tail' of survey respondents. More than half -- 58.3% -- of respondents took the survey on the day I posted it (November 8). Responses have dwindled to fewer than 10 on all days after that, but even now, more than two weeks after its appearance, one or two subscribers are still taking it a day.

EBSCO and RSS Alerts

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

Paul Pival learned something interesting about EBSCO, EZProxy, and RSS feeds: if you don't edit your EZProxy configuration just so, RSS alerts for saved searches in EBSCO databases get rewritten to pass through your library's proxy server. And if that happens, off-campus users (including aggregators like Bloglines or Google Reader) can't get to the RSS feed. (The link to each new item from the database in the feed is still routed through your proxy server so your patrons can still get to the full text at a single click.)
See the Distant Librarian for instructions. In short, all that's needed is adding a single line to Useful Utilites' standard EBSCO EZProxy configuration:
NeverProxy rss.ebscohost.com
I suppose other databases require similar proxy configuration changes. If you have examples, leave them in the comments.

Waiter, There's a Diacritic in My Feed

By rss4lib@gmail.com (Ken Varnum) (RSS4Lib)

How to best encode characters with diacritics in your RSS feed? That was the question posed in a thread on Web4Lib recently, started by a librarian working putting his new books list into an RSS feed. (A great idea in itself, of course.) Since many books, especially in an academic library, but also in other libraries with user communities speaking diverse languages, this is important -- you want to be able to show the title properly, especially to speakers of the language for whom you've bought the book.

There are, of course, several ways to encode many diacritic marks (HTML character entities and Unicode, for example); finding the best one for RSS engendered some discussion. The consensus in the discussion was that using character references is the best solution, particularly for the item title, which is arguably the most important part of an RSS item to get right (OK, the URL is also critical). If users cannot understand the title, why would they click to the full text? Character references take the form The numbers refer to the actual character; "402" is a ƒ or "florin"; "247" is a ÷ or division symbol. And so on... For a list of common character references, see this entities table.

ASIS&T

institutions improve training to reflect the issues you've brought up?
A: There should be more convergence in education programs -- among libraries, archives, museums. Museums, in particular, are often isolated from libraries and archives.
Q: Are there constraints on horizon to funding for these activities -- funding for collection digitization has been relatively good until now.

A: There should be more -- it's OK now, but could be better. Demand is still huge. Challenge is to think about priorities for applying the money. Humanities and social sciences should get together and decide on collective priorities for digitization. Should be discipline-driven, not opportunistic.