Authority 2.0 and 3.0:
The Collision of Authority and Participation in Scholarly Communications

Michael Jensen

When scholarly and academic communications moved from a context of information scarcity (the era of physical movement of information goods) to information abundance (the era of digital movement of information goods), it radically transformed what metrics could be used to establish authority, significance, and even scholarly validity.

"Scholarly validity" itself is too much to take on in a short essay, but I'd like to address the changes I see taking place, and that I think are likely to take place, regarding authority and significance metrics in the scholarly arena of the humanities and social sciences.

Scholarly authority, in the pre-Web world, was driven as much by market forces as by pure academic merit, as I hope I'll demonstrate below; consequently I'll be looking at business and market pressures in the online arena for indicators of the drivers affecting scholarly communications and scholarly authority.

Like so much of the Web world, scholarly authority is being influenced by many of the features that have collectively been dubbed "Web 2.0" by Tim O'Reilly and others, and in keeping with the VeeDotting habit, I'll be using "Authority 2.0" and Authority 3.0" in order to elore more fully the shifts that seem likely in the near future. While the trends I'll discuss are enabled by digital technology, this isn't about technogy per se -- I learned years ago that technology is less the driver of change than is our cultural response to technology.

In Web 1.0, authoritative, quality information was still cherished, and "content [was] king." Content in all forms -- information, perspective, knowledge -- was presumed to be scarce, and intrinsically valuable. In general, the online publishing business models were variants on the standard "wholesaler" model -- duplicating the realities of a physical world. We manifested our expectations of the old world in our approaches to the new one, and garbed the new business and publishing models in 20th century clothes.

Web 2.0, however, takes more for granted: it presumes majority broadband, with unlimited, always-on access, and few barriers to participation – in fact, encourages participation and mechanisms for, in O'Reilly's words, "harvesting collectie intelligence." Its fundamental presumption is one of endless information abundance.

This changes greatly both the habits and business imperatives of the online environment. The lessons, derived from the successful enterprises, include a fracturing of markets into micromarkets, a general user impatience with any impediments, and changes in many entertainment, information, and education-gathering habits, across multiple demographics. Information itself is so cheap as to be free. "Always-on" information abundance leads to immediate price comparisons and fact-checking, which changes the

"authority market" substantially. The ability to participate in most online experiences, via comments, votes, or "ratings," is now presumed, and when it's not available, it's missed.

We see the rise of Google and micro-advertising, and its impact on digital publishing business models; we see the rise of volunteerism in the "information commons" leading to immense free resources; we see Wikipedia and its brethren rise up and slap the face of Brittanica. We also see increasing overlaps of information resources -- machine-to-machine communications (like RSS feeds), mash-ups (like Google Maps and real estate), and much more.

Web 2.0 is all about responding to abundance, rather than scarcity, which is a shift of profound significance. Much of what I'll be addressing has this "response to abundance" as a primary driver.

Abundance vs. Scarcity

Abundance creates expectations, just as scarcity does, but since we've moved into a world of information abundance, we'll need to consider both the scarcity-based models *and* the abundance-based models of scholarly publishing and scholarly communications, if we are to examine the new models of authority effectively. But the impact of information abundance needs some framing, in order to tease out the impacts.

Let's try framing it metaphorically. Imagine you're a member of a prehistoric hunter-gatherer tribe on the Serengeti. It's a dry, flat ecosystem with small pockets of richness distributed here and there. Food is available, but it requires active pursuit -- the running down of game, and long periodic hikes to where the various roots and vegetables grow.

The shaman knows the medicinal plants, and knows where they grow. This is part of how shamanic authority is retained: specialized knowledge of available resources, and the skill to pursue those resources and use them. Hunting and gathering is energetically expensive, and requires both skill and knowledge, to bring the meat and vegetables and roots back to the tribe. The members of the tribe that are admired, and have authority, are those who are best at gathering, returning, and providing what they could identify, for the benefit of the tribe.

(By the way, I'm not trying to imply that the work we were doing in the 20th century was primitive -- I'm trying to construct a useful analogy for the massive distance between scarcity and abundance.)

Tribes like the above all have admired authorities, and the admiration metrics pertain to the knowledge of locales of value, the strength to hunt down scarce resources, and the ability to return the resources for the good of the tribe. This is a scarcity-based authority model.

Contrast that with the food world now: for most of us, *acquiring* food is hardly the issue. We use food as fuel, mostly finding whatever is least objectionable to have for lunch, and coming home and making a quick dinner.

Some of us are better than just fuel-burners -- they are the ones who take the time to creatively combine flavors, textures, and colors to make the food more than just the raw

materials. These are the cooks, and if a cook suggest a spice to me, or a way to cook a chicken, I take their word as gospel, because they're more knowledgeable than I am.

Among cooks, the best are chefs, the most admired authorities on food around. Chefs simply couldn't exist in a world of physical friction and universal scarcity. Chefs know the provenance of combinations of flavor, texture, and color; they can access the works of the great chefs of yore and their magic; they know the history and scholarship of the art of preparing food in a low-friction world, and can make their own magic by acquiring and combining anything they want.

Chef Henri could not exist in a scarcity-driven food environment – food scarcity plainly limits what's possible. To be a chef requires constant availability of all food resources desired, whether spices or meat or vegetables or fruit.

I think we're speeding toward a time where scholarship will be affected by information abundance just as powerfully as food preparation has been.

But right now we're still living with the habits of information scarcity, because that's what we have had for hundreds of years. Information, knowledge, and perspective were historically rare, precious, and valuable resources. To read the ideas of a scholar from Oxford, we had to have a library that purchased the scholar's book or the journal containing a scholarly article. Very few works were everywhere – and those that were, were canonical.

Authority 1.0: Driven by Scarcity, Physical Limitations, and Cost
Scholarly communication before the Internet *required the intermediation of publishers*. The costliness of publishing became an invisible constraint that drove nearly all of our decisions. Because of the costliness of the physical world, it became the scholar's job to be a selector and interpreter of difficult-to-find primary and secondary sources; it was the scholarly publisher's job to identify the best scholars with the best perspective and the best access to scarce resources.

Before risking half a year's salary on a book, the publisher would go to great lengths to validate the scholarhip (with peer review), and to confirm the likely market for the publication. We evolved immensely complex, self-referential mechanisms to make the most of scarce, costly resources. Consequently, scholarly authority was conferred upon those works that were well published by a respected publisher.

Scholarly authority was also something that could be inferred by the scholar's institutional affiliation (Yale or Harvard, vs. Acme State University). My father got his Ph.D. from Yale, and had that implicit authority the rest of his professional life. Further, authority in conferred as well by the hurdles jumped by the scholar, as seen by degrees and tenure status.

Scholarly authority also could accrue over time, by the number of references made to all of a scholar's work by other authors, thinkers, and writers -- as well as the other authors, thinkers, and writers that a scholar referenced.

The above examples of authority metrics are all mutually constraining and confirming; that is, the systems that encourage selectivity also encourage the development of

other implicit, explicit, and de facto authority metrics, all of which were part of scholarly publishing for centuries. Fundamentally, scholarly authority was about *exclusivity* in a world of scarce resources.

Web 1.0 Authority: Old Habits

The Web, in version 1.0, mimicked these fundamental conceptions of content value and business models. It presumed that virtually any information had intrinsic value, or else it would not have been "published." The presumption was that information scarcity still ruled.

These old authority habits are most entrenched (in scholarly communications) in the humanities and social sciences, I think. In contrast, non-scholarly online arenas, new trends and approaches to authority have taken root.

As outlined above, the Web 2.0 presumptions include treating the Web as a dependable, always-on platform for communications and participation; a presumption that harnessing and harvesting the collective intelligence of hundreds of thousands of participants is of value in itself. It presumes openness and ease of access *and* ease of production, and thus ease of inclusivity.

This set of presumptions enables some very interesting experiments, projects trying to find strategies for dealing with abundance. Let's look at some examples available in the Web 2.0 category, as of November 2006, and then explore what that means for scholarly authority.

Web/Authority 2.0

One of the oldest and most "authority makers" in Web 2.0 world is of course Google. The magic of Google's PageRank system is fairly straightforward. From google.com/technology:

> PageRank Explained
>
> PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."
>
> Important, high-quality sites receive a higher PageRank, which Google remembers each time it conducts a search. Of course, important pages mean nothing to you if they don't match your query. So, Google combines PageRank with sophisticated text-matching techniques to find pages that are both important and relevant to your search. Google goes far beyond the number of times a term appears on a page and examines all aspects of the page's content (and the content of the pages linking to it) to determine if it's a good match for your query."

This model takes advantage of "conscious attention" by Webizens, and aims to monetize "the wisdom of crowds." This is authority conferred mostly by applause and popularity, and has its limits, but it also both confers and confirms authority, because people tend to point to authoritative sources to bolster their own work. For the current status of the Web, it continues to be a great way of finding "answers" from authoritative sources.

There are currently a wide variety of "group participation" news and trend-spotting sites: slashdot.org ("news for nerds. stuff that matters.") remixing and commenting and linking to other information web-wide; digg.com, enabling up-or-down votes on whether a news story is worth giving value to; del.icio.us, a bookmark-collection-and-tagging collector; dailykos.com, a semi-collective of left-leaning bloggers op-edding, commenting, diarying, linking to, and responding to news; and plenty of other large community sites, all dealing with the problem of how to prioritize postings, refresh and display appropriately, while providing the services *for free* to the user -- that is, without a budget for staff or much editorial intervention.

The challenge for all these sites pertains to abundance: to scale up to hundreds of thousands or millions of users, they need to devise means for "harvesting collective intelligence" productively. For digg.com, it's a simple binary yes/no vote from participants, and credit for being the first one to spot a story others find interesting; for del.icio.us, credit comes both from votes and tags and degree of participation. For slashdot.org, every posting or comment has a "score" which:

> "... includes things like moderations up and down, default posting bonus, and so forth. However, that same comment, when displayed in the context of a discussion, reflects the bonuses or penalties associated with any number of user preferences. These options are all configurable, and include settings like the small comment penalty, the long comment bonus, and any reason modifiers you may have defined." -- CmdrTaco, FAQ, slashdot.org

Registered users -- read, members of the community -- are able to rate postings *and comments* as they read: Normal, Offtopic, Flamebait, Troll, Redundant, Insightful, Interesting, Funny, Overrated, Underrated. This engaged participation helps filter out crap, and is a version of the "hive mind" authority, where increasing populations simply add more richness to a resource.

Slashdot in particular came to define a subculture, or "interest tribe"; we're now moving into sub-subcultures, I think, and are likely to see communities built around more specific topic areas. These, like usenet newsgroups in the 90s, will get so big that cross-participant ranking systems will be required, for the sites to remain valuable to the participants themselves.

Flickr, Youtube, and other media-collection sites tend to use a variant of "voted on by tag," as well as using the number of viewers as a metric of interestingness and value. The more votes-by-tag a picture has, the more likely it is to be found, and to be tagged some

more; the thumbnail version that gets lots of clicks to see the full version is likely to be given more attention by new viewers, and thus more attention.

This is the nature of socially-powered "User Generated Content," and its strength. It can also mean that valuable pictures -- or in Google's case, documents -- get skipped because they weren't famous. This is the problem of a fabulous actress who chooses to only perform in Tulsa, Oklahoma, to an adoring community. Her "Ophelia" may be the best ever performed, but we would never know it, because there are no reviews in the New York Times. Youtube uses a five-star rating system, as do other similar sites, which can mitigate the "celebrity" effect – being popular because you're popular.

MySpace, Friendster, Facebook, and other consciously "social networking sites" so far are mostly about self-expression, and the key metric is often "How many friends do you have? who pays attention to you? Who comments on your comments? Are you selective or not?" While these systems have not been framed to confer "authority," they certainly addresses many of the standard referents for "popularity," and especially as they devise means to deal with predators, scum, and weirdos wanting to be a "friend," they are likely to expand into "trust," or "value," or "vouching for my friend" metrics – close to "authority" – in the coming months.

Wikipedia is another group-participation engine, but focused on group construction of authority and validity. Anyone can modify any article, and all changes are tracked; the rules are few – stay factual, unbiased, cite your sources – and recently some "more authoritative" editors have been given authority to override whinging axe-grinders. But overall this is still an astonishing experiment in group participation. Interestingly, in Wikipedia, most seem to believe that the more edited an entry (that is, the more touched and changed by many different people), the more authority the entry has. This is a kind of democratization of authority -- or diffusion into a "hive-authority" -- which is nearly unique to Wikis, since it's not observation, but active participation in improvement, which is the authority metric. The more authors, the more likely to be accurate and authoritative.

Among the most venerable "authority makers" takes the other tack, by relying on the judgment of a few smart editors, but getting lots of recommendations from users/readers. In many respects boingboing.net is an old-school edited resource. It doesn't incorporate locally any feedback or comments, but rather is a publication constructed by five editor/writers. It has become a hub of what's interesting and unusual on the Web. Get noticed by these guys -- by having someone submit your site to them -- and you get a *lot* of traffic. They are conferrers of validity, and constructors of cool, for a great deal of the technophile community. In certain circles, boingboing is like NPR, or like Cronkite news was, back when there were only 3 networks. It has become a glue for its subculture: "did you see boingboing yesterday?"

Boingboing is an example of very savvy early adopters who claimed an information space. As the online environment matures, most social spaces in many disciplines will have their own "boingboings." Authority in this case derives from editorial judgment driving the site's popularity, which stems from the diversity of interests of the five. That they all are

fiercely smart and good writers doesn't hurt, while the external authority that derives from their other work is a constant reminder of value. If there was a "boingboing" that was paying attention to, say, 19th century material culture, with a handful of fiercely smart good writers, they could likely acquire similar significant authority in that interest tribe.

These are by no means fully representative of all "authority" mechanisms currently in place – I haven't addressed ebay.com buyer/seller ratings, or technorati.com's rating of "authority" of blogs, but these and the above are examples of different models for computed analysis of user-generated authority, many of which are based on algorithmic analysis of participatory engagement. These approaches are indicators for the future of scholarly communications and scholarly authority.

Authority 2.0: Many weeders make a tidy garden

Much of the above is still on the periphery, and the authority analysis is not transparent yet. But is still indicative of the ways in which the Web is being put to use in new ways. The emphasis is often not on *finding value*, but rather on *weeding abundance* -- that is, filtering out garbage to find gems.

So what will be the next step? What is Authority 3.0, or Web 3.0? Most of the technophile thinkers out there believe that Web 3.0 will be driven by "artificial intelligences" -- automated systems that can make reasonable decisions on their own, to pre-select, pre-cluster, and prepare material based on established metrics, while also attending very closely to the user's individual actions, desires, and historic interests, and adapting to them. It's very clear that mechanisms for "responding to abundance" -- filtering, homing-in, exploring available resource -- will increasingly be computer-assisted.

Making Sense of Abundance

"Responding to abundance" is the operative element in the Web 2.0 and 3.0 world, and the models of algorithmic filtration, of connecting of dots, and of "soft AI-based clustering" are just now beginning to show up. I've been involved with a few Web projects that may hint at some characteristics that are semi-"Web 3.0." One of them is the National Academies Press book-specific "Search Builder," and another is its "Reference Finder" – both of them tools that take algorithmically-extracted key phrases (pulled from a specific chapter or book) and put them to new purposes.

The Search Builder enables a researcher to select terms from a NAP chapter, generally "terms of art" around a topic, and make term-pairs for sending to Google, MSN, Yahoo, or back to the National Academies Press. This sort of re-use of language – essentially using language as a filter rather than as an identifier – holds promise for automated intelligences. The Reference Finder is a working prototype (of a type we will no doubt see more of, later) where a researcher can simply paste in the text of a rough draft, push a button, and get likely related NAP books returned, based on the algorithmically extracted and weighted key terms from the rough draft.

These are simultaneously about precision and about "fuzzy matching" – but most of all is about adding algorithmic, automated assistance to enable a human to make sense of abundance.

Connecting the Dots

Among the most exciting applications I've seen recently comes from Microsoft – its "Photosynth" product, which automatically links up multiple photographs of an area, enabling an almost three-dimensional exploratorium. While now in prototype stage, it shows the extent of how "connecting the dots" can become a meaning- and coherence-making mechanism on its own. Hundreds of photos of a site are algorithmically "knit" to enable a nearly three-dimensional exploratorium. The "dots" are the walls, corners, windows, towers, and other elements-in-common that the pictures have. Imagine how this same model could be applied to concepts, ideas, and themes derived from the extracted language of textual material, and one begins to see how a such a navigational tool could help make sense of immense content resources.

Photosynth in particular represents "swarm relatedness" – which could grow to become a multiple-intersection approach to scholarly interconnections and authority.

In the Web 3.0 world, I think we start seeing heavily computed reputation and authority metrics, based on many of the kinds of elements we see being used now, as well as elements that can *only* be computed in an information-rich, user-engaged environment.

Authority 3.0: Algorithmic Weights and Measures

Web 3.0, is about automatic intelligences – software which will "read," identify, interpret, cluster, and present material to individuals, based on multiple factors. Given the advances in technology, remarkable things are likely to happen. Authority 3.0 will both influence and drive these algorithmic models.

We are closing in on a time of utterly cheap and very fast processing cycles. In our hunter-gatherer/cook/chef analogy, it is as if the Chef Henri now has a matter constructor, instantly making anything, perfectly fresh, on command. In a world of unlimited computer processing, Authority 3.0 will likely probably include:

- prestige of the publisher (if any) as part of "contextual authority"
- prestige of peer pre-reviewers (if any)
- prestige of commenters and other participants
- proportion-quoted-by-others: out in Web (plagiarism as admiration)
- raw-links-to the document
- valued-links: in which the value of the linker and all her other links is also considered
- obvious-attention: discussions by blogspace, comments to the posts, reclarification of and continued discussion
- nature-of-attention language: positive, negative, lateral

- quality-of-context: what else is on the site that holds the document, and what's their authority status
- proportion of "discipline-valued phrases" in document
- author/PI institutional affiliation(s)
- reference-network (the significance rating of all the texts the author considers)
- significance-of-author's-other-work: the value of elders, or the prolific, also matters to generic authority
- author's participation in other valued projects, as commentor, or editor, or anything.
- how long a document has existed
- the document's inclusion in lists of "best of," in syllabi, in indexes, and other human-selected distillations
- the types of tags asserted to it, the terms used, the authority of the taggers

These are things that simply could not be computed reasonably by human beings, but which can be computed by "artificial intelligences of authority." They differ from current models mostly by their feasible computability in a digital environment where all elements can be weighted, measured, and whose digital interconnections provide computable context.

The preconditions for scholarly success in this arena include digital availability for indexing (but not necessarily individual access -- see Google for examples of journals that are indexed, but not otherwise available), digital full text availability, and some metadata of some kind that identify the document, categorize it, contextualize it, and summarize it -- and perhaps provides the key phrases of it, while also allowing others to enrich it with their comments, their tags, their own contextualizing elements. These data facilitate "computed authority by algorithm," and encourage the reuse of documents to "find more like it," and much more.

In the very near future, we're talking about a universe of tens of thousands, if not hundreds of thousands, of other documents that are *very similar* to any given document being "published." If you are writing a scholarly article about the trope of smallpox in Shakespearean drama, how do you ensure you're well read?

By competing in computability, by:

- encouraging your friends and colleagues to link to your online documents
- encouraging participation and commentary, and online back-and-forth with interested readers
- encouraging free access to much or all of your scholarly work
- recording and archiving all scholarly participation, for consideration in future tenure cycles
- recognizing others' works via links, quotes, and other tips of the hat online
- taking advantage of institutional repositories as well as open access publishers

- etc.

I hope it's clear that I'm not saying we're just around the corner from a revolutionary Web in which universities, scholarship, scholarly publishing, and even expertise are merely a function of swarm intelligences. I actually think that that's a long way off. Many of the values of scholarship are not well served by the Web, and the traditional models of authority will hold sway in the scholarly arena, I think, for ten to fifteen years, while we work out the ways in which scholarly engagement and significance can be measured in these participatory spaces.

But the new metrics of authority will be on the rise. And further, ten to fifteen years isn't so very long, in scholarly careers. Perhaps most important, if scholarly output is locked away behind firewalls, or locked away in print-only, that scholarship risks becoming invisible to the automated Web crawlers, indexers, and authority-interpreters that are being developed. Scholarly invisibility is rarely the path to scholarly authority.

The authority metrics discussed above are on the near horizon, and will be changing the ways in which we think about, use, and participate in scholarship, and will transform how our works will be found, explored, and given authority in the years ahead.

Sites referenced:
Slashdot: http://www.slashdot.org
Dailykos: http://www.dailykos.com
Digg: http://www.digg.com
Delicious: http://del.icio.us
Google: http://www.google.com
Myspace: http://www.myspace.com
Friendster: http://www.friendster.com
Facebook: http://www.facebook.com
Wikipedia: http://www.wikipedia.com
Tim O'Reilly: http://www.oreilly.com
Boingboing: http://www.boingboing.net
Flickr: http://www.flickr.com
YouTube: http://www.youtube.com
NAP Search Builder: http://www.nap.edu, see any book
NAP Reference Finder: http://www.nap.edu/nap-cgi/reffind5.cgi
Microsoft Photosynth: http://labs.live.com/photosynth