Sprint Beyond the Book: Hidden Knowledge in Information Overload

Carol Anne Meyer

June 2, 2016

Six-word memoirs

It's here somewhere hidden on purpose Lack of curiosity leaves information hidden I think I know everything. Not Thirty years experience trapped inside

More verbose

Jenni Rankin of Annual Reviews said in the small data session today that Annual Reviews was founded in the 1930s to combat the problem of information overload confronting researchers. The volume of scholarly knowledge continues to increase, and despite any number of discovery and digestion tools invented, discarded, reinvented, and reinvented since then (and before), scholars today are in no better situation. There is too much to read and not enough time to read it all. Carol Tenopir and Don King have done good work in studying the habits of researchers. More recently, Simon Inger has published a study on how researchers discover services. (There was a whole session about that here at the conference that I missed. The report is available for purchase.)

Graduate students are taught (I think they are still taught this) the importance of doing a literature review before beginning their own research. This is a good thing. But if the discovery tools they are using are selective rather than comprehensive, they risk missing important information hidden as a result of the shear volume of information.

Traditional bibliographic discovery tools have some limitations, including the variability of the metadata fields and indexing terms. (Studies have shown that human indexers are inconsistent; even the same person may use different terms at different times of day. As one of our group members said about undergrad social science volunteers, perhaps they were hungry at one of the times.) Machine indexing in combination with human review can help, but nothing is yet perfect.

Text indexing has become a way of searching for hidden data within the full text of articles that previously could only be discovered through bibliographic searches. Many linguistic studies were available after the launch of JSTOR's full-text journal backfile services
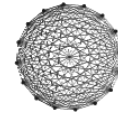
that had not been practical before, although such analysis was not a goal of JSTOR, at least in the beginning.

Newer text-mining techniques allow for analysis that was previously extremely time consuming. Word frequency is one simple example.

So discovery tools have appeared that make it easier to deal with the glut of scholarly information. Their effectiveness depends on whether the information seeker has access to the source material and the tool as well as the discipline, curiosity, and information literacy of the information seeker.

But what of the person who does not have basic research skills or who has basic research skills that he or she does not turn to the problem at hand? I'm particularly thinking of the political divide between researchers who are firm proponents of green open access and the publishers who have experiential (and documented) knowledge of the resources required to publish.

At a recent Boston-area SSP panel discussion on institutional repositories, the librarians and scholarly communications officers representing those well-funded organizations Massachusetts Institute of Technology and Harvard talked about the difficulty finding the resources to implement those institutions' open access faculty mandates. The hidden content of publishing expertise is slowly transferring to library publishing programs and institutional repositories.

Disruption in the scholarly communications ecosystem is a given and has been for 20–30 years or possibly longer. Some of this disruption has been caused/made possible by technological innovations, some by researcher pain, some by damaged economic models. There is room for new models. My postulate is that openness to both historical knowledge and innovation will create a new and more productive equilibrium than a stand-off between parties who have dug in as if the problem were a binary one with winners and losers.

Building a quality scholarly communications system cannot happen without publishing resources. Some of those resources may be commercial, which is not to say that they are evil.

Improving scholarly communication can include creative ways to make content accessible to more people. Some of these methods may be based on green open access, which is not to say that they are naive.

Even in this essay, I am polarizing the groups, but, in fact, there is a continuum. Red and yellow and pink and blue (to quote my mother's favorite song) and green and gold can be woven together in a stronger and more beautiful system. We just need to stop digging in, misinterpreting information that doesn't fit our world view, embrace change, be prepared to say "no" to things that don't work, and

move forward.