2014

The Importance of Data, Information, and Knowledge in Scholarly Communication

Carol Tenopir
University of Tennessee - Knoxville, ctenopir@utk.edu

Recommended Citation

Tenopir, C. (2014). The importance of data, information, and knowledge in scholarly communication. In C. Bitso, & R. Raju (Eds.), LIS Education and Research in a Dynamic Information Landscape: Proceedings of the Library and Information Studies Centre 75 years Commemorative Conference (pp. 54-62). Cape Town, South Africa: University of Cape Town Libraries.

Chapter Seven
The Importance of Data, Information and Knowledge in Scholarly Communication

Carol Tenopir
University of Tennessee, School of Information Sciences
Email: ctenopir@utk.edu

Review status: Peer reviewed

Background

Research in scholarly communication, including the role and importance of data and publications, can reveal important insights into how knowledge is formed and transmitted. These insights can be interesting in and of themselves, as answers to fundamental research questions such as how formal communication helps science progress, and they also can help librarians and publishers and researchers create better information systems. This paper is about the intersection of these concepts in the study of scholarly communication with examples from my own research and the work of others that inspired me in formal scholarly publication.

I cannot begin a paper with a title like this without the obligatory reference to T.S. Eliot's "The Rock" and the data, information, knowledge, and wisdom pyramid derived from it (Eliot, 1934.) Eliot famously posed the questions: “Where is the wisdom we have lost in knowledge?”; “where is the knowledge we have lost in the information?” The pyramid, derived from Eliot, shows data as the bottom foundation or biggest slice of a pyramid. By adding value such as contextualisation and categorisation to data, information is created and becomes the next, smaller but more refined step in the pyramid. Value added by comparison or connections to information create knowledge, while adding value to knowledge such as possible actions or decisions can create wisdom, which is application of knowledge (Taylor, 1986.)

Research on how experts communicate is, of course, not new. The formal study of scholarly publication goes back to work in the early 1960s by Derek DeSolla Price (1963) and seminal studies by Garvey and Griffith (1967, 1972) and many others over the last six decades. Garvey and Griffith’s discussion of “communication means” includes dozens of formal and informal communication venues that researchers use throughout the research process to disseminate their ideas. These means and venues range from personal informal conversations to formal peer-reviewed publications and everything in between. The detailed view of communications means depicted in Figure 1 can be summarised as shown in Figure 2.

Figure 2 is a concise view of Garvey and Griffith’s more detailed analysis, where communication can be oral or written OR informal or formal. All communication means contribute to scholarship in sequential and iterative ways and are part of understanding and creating new knowledge.

As complex or as summarised as these standard diagrams are, however, they are incomplete when we think about the complete picture of how scientists or other scholars create knowledge. They leave out important intermediate steps or sources such as observations recorded in laboratory or field notebooks, videos, audio files and data sets. Figure 3 depicts how some of these sources contribute to the research work of a scientist.

Scientists do not just rely on oral communication or written reports to build knowledge and to create publications or new information products.

Figure 1: Communication Means from the work of Garvey and Griffith (reproduced from Tenopir and King, 2004.)

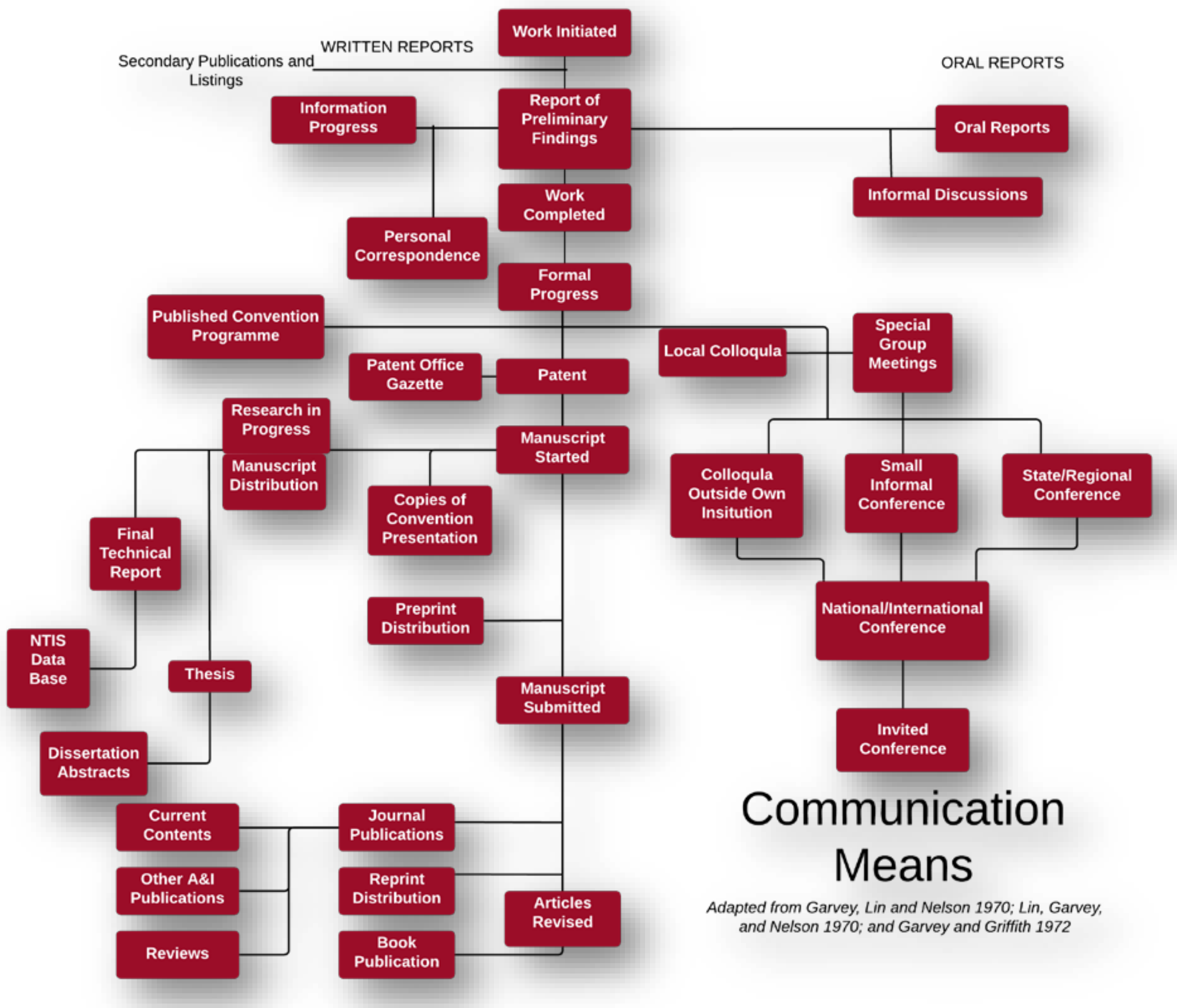
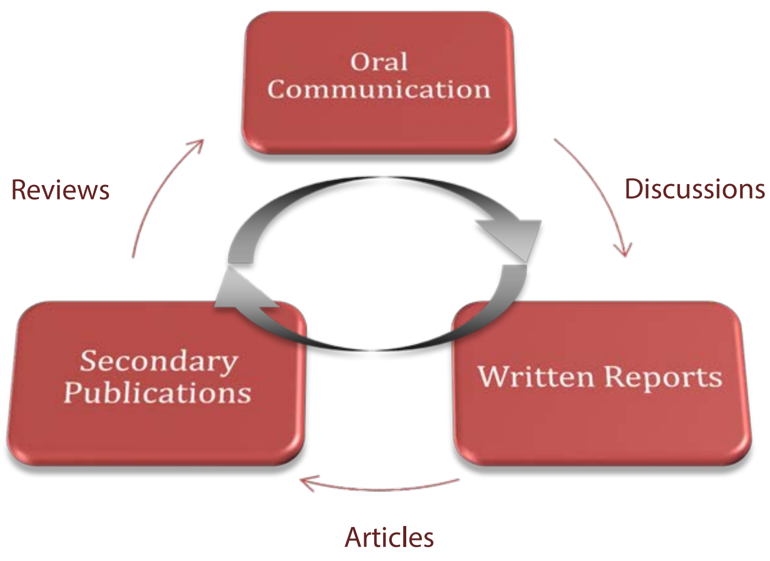


Figure 2: Summary of Communication Means



To understand the entire process, we need to focus more on workflows or the process of how work and knowledge creation go together. Data, specimens, sounds, images, and the like, together with analysis and narrative, all feed into the process of creating new knowledge and communicating science. In an electronic environment, multiple ways of communicating information can feed into enhanced e-publications, as depicted in Figure 4.

Traditionally, not all of these steps or resources have been reflected in the final or more formal products of scholarship such as books and articles. In an e-environment, however, all important artifacts of science and workflow can also be included as parts of publications to form a more complete representation of scholarly communication.

Putting everything in boxes as in the previous diagrams implies a separation and neatness that is not necessarily there. In science, for example, images can both be a transformation of data, that is, a way to understand underlying data collection by visualizing it, OR images can be the data themselves such as photos of animals, or x-rays.

Images by themselves, however, normally do not convey as much meaning as is conveyed when combined with explanatory text. Text, plus images, plus data convey the most meaning and allow the fullest picture (or lead to more knowledge being construed).

And, perhaps, scientific wisdom can be obtained by combining multiple sources of data, with text and visuals, and adding models and computations, which convey an additional level of meaning. This image from the DataONE (Observation Network for Earth) shows an example of how combining data on bird observations, land cover, and weather data that examines climate change, may affect bird migration. (Figure 5) (dataone.org.)

By adding analysis and narrative to the other components, electronic publications have the power to convey multiple levels of meaning to provide various levels or stages of information. Multi-leveled e-publications allow the consumer to select the level that is needed at the time.

One way to think of this is to think in terms of granularity, or parts, and how to deconstruct, combine and recombine them. Granularity can be defined as 'divisible, or made up of conveniently small and independent parts' (Business Intelligence Dictionary) Granular publications can be combined, divided, and recombined as desired. In the terms of the journal world this means we can go beyond the traditional idea of granularity in terms of just journals, issues, or even articles.

Again, this is not a new concept, the idea of granularity in publishing has long been talked about and is possible. This is not to say that traditional aspects should be eliminated—granularity can mean more parts (or grains), in addition to fewer.

In the traditional, formal, written communication process, especially from the viewpoint of libraries, granularity for writing and reading is often at the journal level. Libraries traditionally selected and purchased journals by journal title, rather than paying for parts of a journal. For distri bution and shelving, print journals in turn were broken into issues. In the current e-access and search engine world, granularity from the reader's viewpoint (and certainly the author's) is most often at the article level. Libraries still make purchases at the journal title level or, often, at an even higher level of granularity in bundles of titles grouped by subject or publisher.

On the other hand, granularity of searching and using today may be broken down even further to help users get the level of granularity they need to construct meaning for their purposes. This granularity may be by allowing separation of sections, paragraphs, graphs, tables, photographs, and other components. In addition, if all component parts of a bigger whole are labeled and linked together, an article or a table or a paragraph can be a starting point to go bigger—to the data sets behind a table, for example. This type of flexible thinking shows the value of an ar-

Figure 3: Sources Used by Scientists in Research

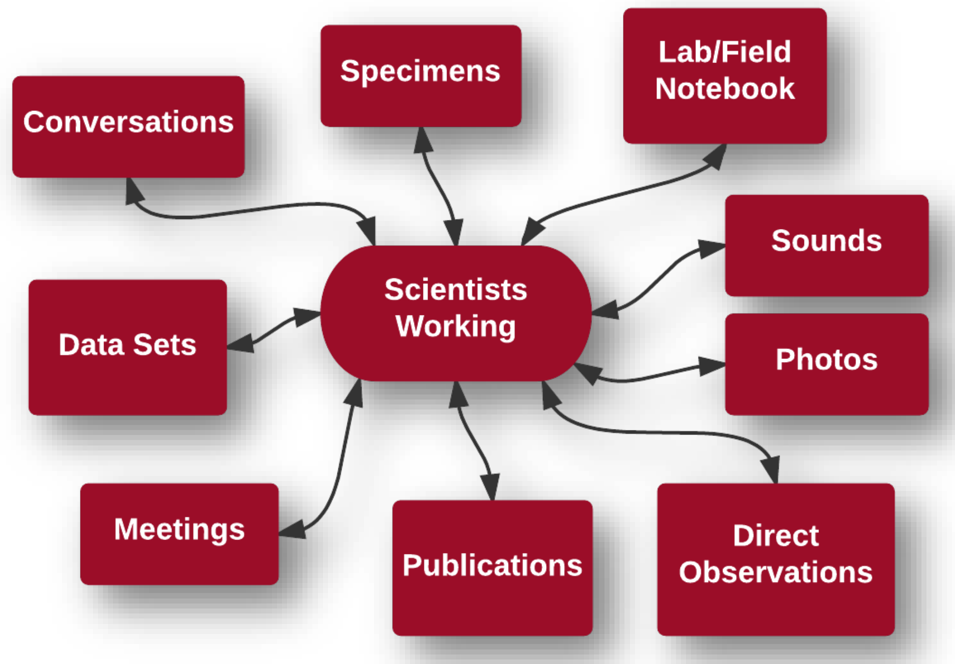


Figure 4: Expanded Formal Publication in an E-environment

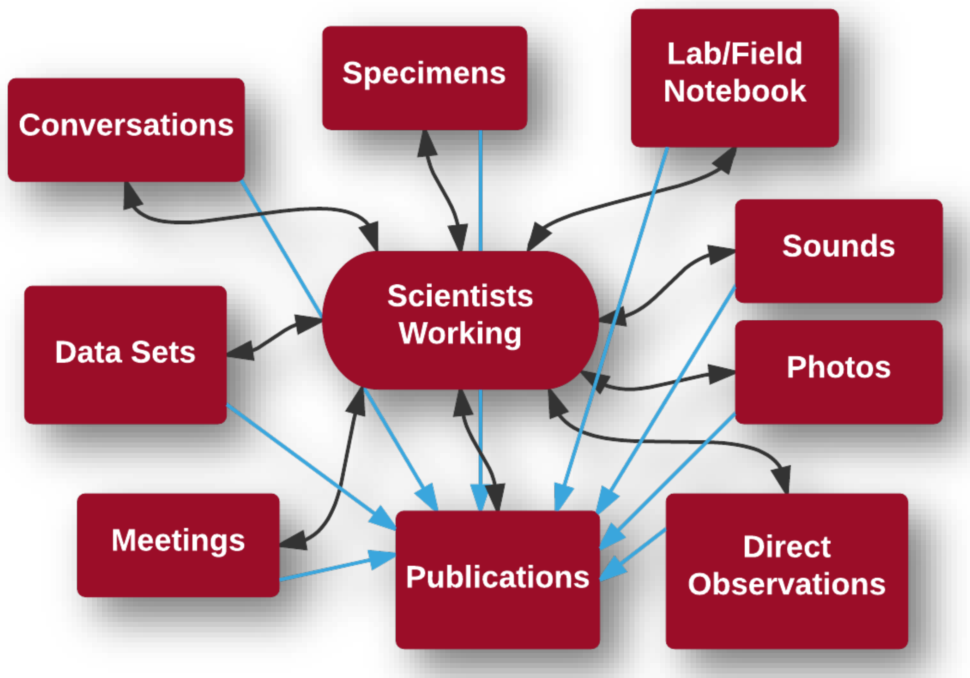
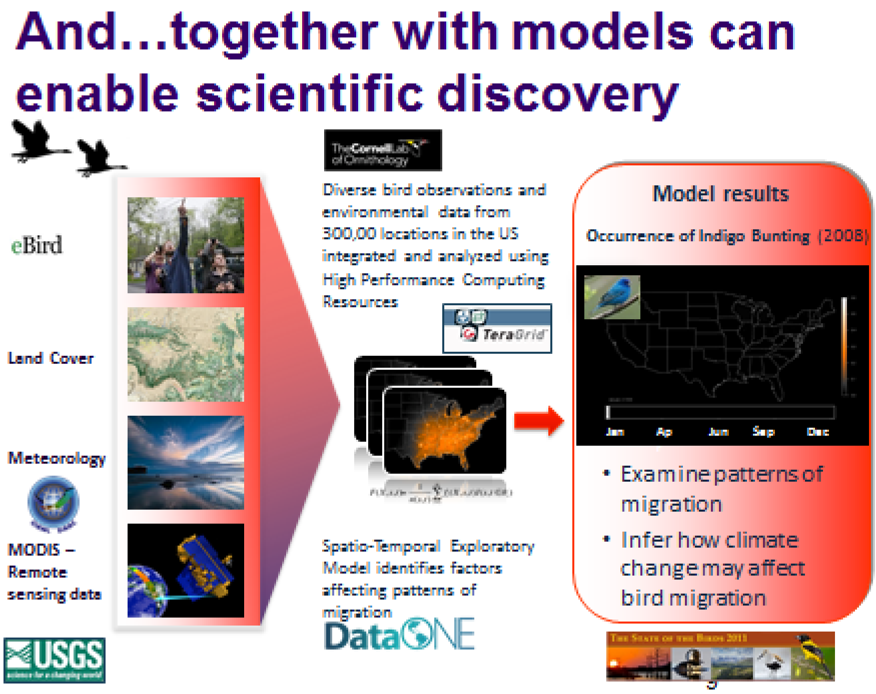


Figure 5: DataONE's Depiction of Combining Multiple Sources of Information



ticle in the center with granularity allowing use in multiple ways—bigger, smaller, or in sections.

Technology and structures are being put in place to allow this description and interlinking. For example, subject data repositories such as Dryad link articles to data, and many publishers, such as PlosONE, in turn link articles to the corresponding data sets in DRYAD. (www.plosone.org; datadryad.org). Mark-up or Digital Object Identifiers (DOIs) can be assigned at various levels of granularity to allow search and retrieval of either parts or a whole.

Sometimes researchers need an entire article or most of one, especially for catching up on a topic or when writing an article. Sometimes they need less – just a part of an article such as a paragraph that describes a method or conclusion to check a fact or a picture to use in a class or a presentation. Sometimes they need more – the data on which an article is based, to extend their own research or build on the research of others. Properly designed electronic publications can provide this level of flexibility or fluid granularity in a way not easily accomplished in traditional models.

Research highlights

This is all background that shows you my thinking on these issues; now let me give you some highlights from some research to back these assumptions. These examples come from the "Deep Indexing" study for CSA (now ProQuest), and data use, reuse, and sharing by scientists surveys for the NSF-sponsored DataONE project.

To test the desirability of direct access to tables and figures embedded in articles, we studied sixty scientists in seven universities and two institutes in US and Europe while they searched on a table and figures indexing prototype database for information relevant to their research projects. Scientists conducted over 350 searches, yielding data by direct observation, diaries, and pre and post searching surveys (Tenopir, Sandusky, and Casado, 2006.)

Research questions included:

- Do scientists need image indexing?
- What do scientists currently do with images?
- How might they use an image index?
- How effective is searching for images?
- How might image searching impact the

work of science?

Note that we did not study issues of the economic feasibility of such a service, if developed, nor whether libraries would be willing to pay for such an indexing feature. Instead, we were concerned with the possible uses (if any) that researchers could make of increased granularity in indexing of articles if it were available. Of course, any decisions CSA or ProQuest made had to take into account the financial viability of such a service.

This study found that without the existence of a special database, scientists in general search for photographs and maps more than tables, figures or graphs; use Google most often to locate images or graphical material; and consistently rate level of satisfaction with current capabilities and results of image or graphics searches as low. Some comments from the participants included: locating objects is "difficult"; and "in general, academic figures, tables, and graphs are not available to search" in current systems. However, several noted that the ability to search for figures might help them find information or data not reflected in the title or abstract of an article and help them find things previously lost in traditional abstracting and indexing tools. They generally reacted positively to the idea of such functionality and granularity to help both research and teaching.

The subjects had many suggestions of what would need to be present if the granularity of scholarly content was deconstructed into images. These suggestions or conditions for success included:

- images must be of high quality with the ability to enlarge thumbnail images;
- the context of the whole article is important and it may be dangerous to see images without the context;
- tables of contents should allow all component parts to be seen in one place and should be searchable;
- extraction of the data behind any table should be supported.

Research data deposit and access has become more common since the research described above. More recently, as part of the NSF-funded DataONE project (dataone.org), the Usability and Assessment Working Group has been surveying a variety of data stakeholders, including scientists who either need access to or create earth and environmental science data and libraries or librarians who may help provide data management services and who have constituents who need help in finding or storing data.

In a 2011 survey of scientists (Tenopir, et al, 2011), we found that although three-quarters of scientists agree with the statement "I share my data", only about a third (36%) agree that "others can access my data easily". This illustrates a gap between willingness and accessibility.

Several years later in a follow-up survey of scientists in 2014-2015 (submitted for publication), the gap had narrowed just slightly, with 78% of scientists saying they shared their data and 45% saying others can access their data. How they share their data may range from sending a data set when it is requested to uploading data into a data repository. Although most did not yet routinely upload data from preservation and sharing, in 2014, 82% of the approximately 1000 respondents said they would put at least some of their data in a central repository, and 45% would place all of their data in a central repository.

Although lack of access isn't yet seen as a major impediment to science (27% in 2011 agreed that it was a major impediment), half of them (50%) said it restricted their ability to answer scientific questions and 78% said they would use other's datasets if the data were easily accessible (Tenopir, et al, 2011).

Scientists said they do not share data in repositories for a variety of reasons, including:

- Insufficient time (45%)
- Lack of funding (34%)
- No place to put data (20%)
- Do not have rights to make data available (20%)
- Lack of standards (17%)

Figure 6: Reference support services by Academic Libraries for Research Data Management (Tenopir, Birch, and Allard, 2012).

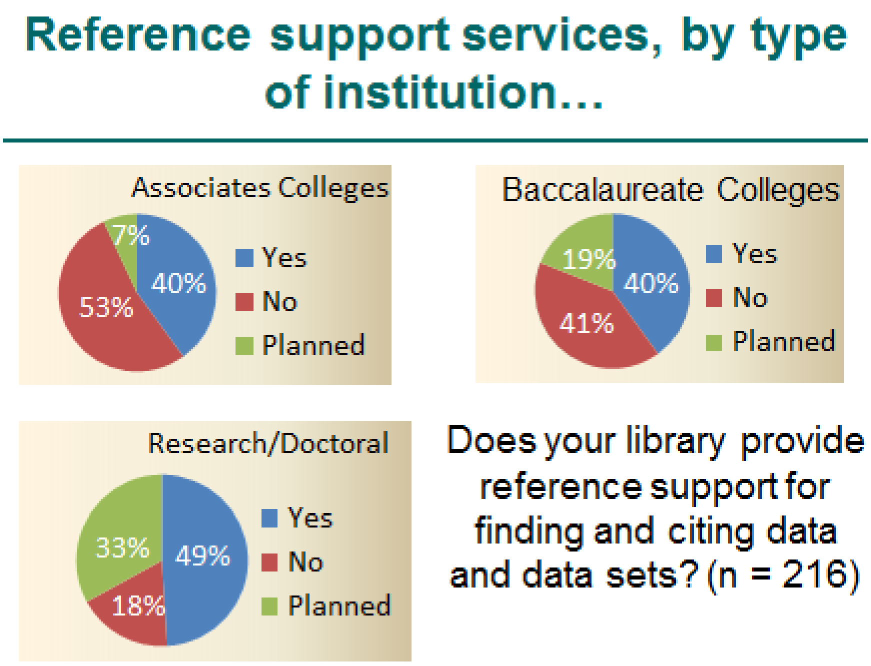


Figure 7: Technical Support Services by Academic Libraries for Research Data Management (Tenopir, Birch, and Allard, 2012).

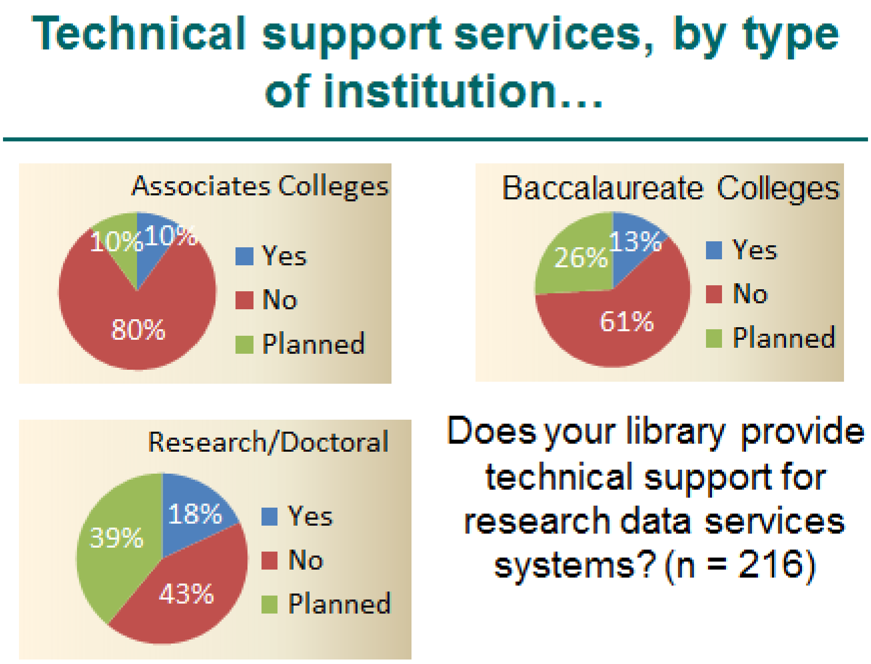
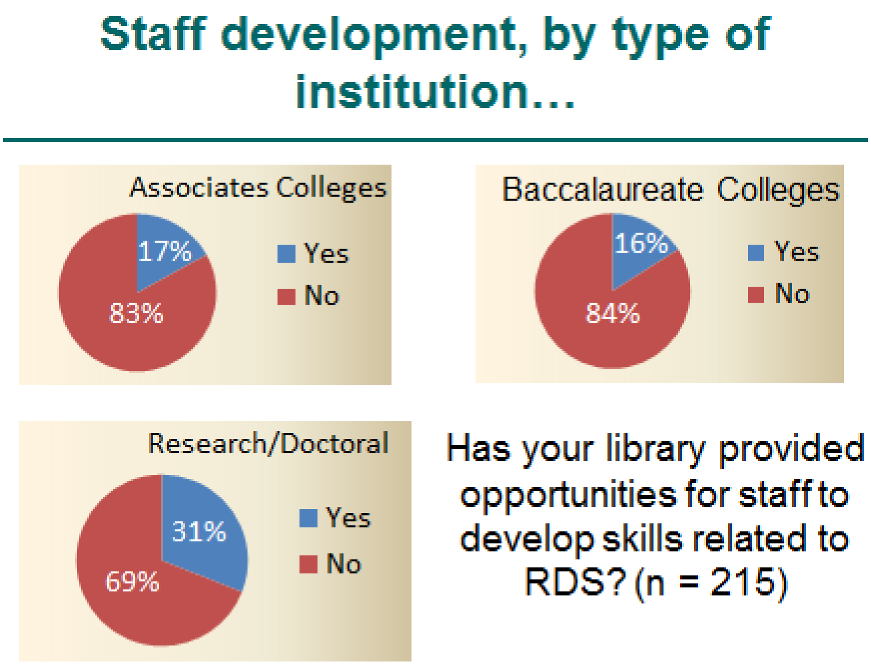


Figure 8: Staff Development in Research Data Management by Academic Libraries (Tenopir, Birch, and Allard, 2012).



- Sponsor does not require it (15%)
- Others do not need their data (13%)
- Their data should not be available (12%) (Tenopir, et al, 2011).

Some of these reasons for not sharing data cannot be easily resolved, but the major reasons of lack of time and lack of funding provide an opportunity for libraries to help.

Another survey for DataONE, assisted by the Association of College & Research Libraries, asked Directors of North American academic libraries if their libraries provided reference support for finding and citing data. A majority of research libraries said they either offered these services already, or are planning to (49% currently offer and 33% plan to offer) (Figure 6) (Tenopir, Birch, and Allard, 2012)

Far fewer research academic libraries offer technical support for data (18% of research libraries currently offer technical support, with an additional 39% saying they plan to) (Figure 7) (Tenopir, Birch, and Allard, 2012). Perhaps this is an opportunity for collaboration across campus units, across university libraries, or in conjunction with publishers, libraries, and data repositories.

An implication and opportunity for LIS educators can be found in the response to the question: "Has your library provided opportunities for staff to develop skills related to research data management?" Just 31% of the 215 research/doctoral libraries, 16% of the baccalaureate colleges, and 17% of the two-year associate degree colleges say that they provide such opportunities (Figure 8). Continuing education opportunities or inclusion of such topics in classes could provide a service from LIS programs to the changing needs of the profession (Tenopir, Birch, & Allard, 2012).

There is a clear lack of skills development opportunities regarding data management in academic libraries. Educators and libraries and professional organizations can work together to resolve this skills gap.

Conclusion

Electronic journals allow a rethinking of how scientists can communicate their research and how others will want to access the products of research to help them in their work. By building on concepts expressed in the past, we can build a granular view of formal scholarship, allowing access at many levels, in many ways, to the variety of scholarly outputs that goes into the whole picture of research. While access to component parts is an essential aspect of granularity, the data behind graphs and charts is also an important component to enlarge the typical journal article.

In conclusion, from a lifetime of research:

- access to information, data, and visuals can help scientists in many ways, but affordable/sustainable services or products need to be developed;
- when posing research questions or designing systems we need to think in many different levels of granularity, access, and utility; and
- Library and Information Science education and research can help in providing research, education, and services that help others form knowledge more effectively

References

Eliot, T.S. 1934. The Rock. *Stories of the Human Spirit*. Ed. Peter Y. Chou. Available: .http://www.wisdomportal.com/Technology/TSEliot-TheRock.html. (15 October 2014).

Garvey, W. D., & Griffith, B. C. 1967. Scientific communication as a social system. *Science*. 157(3792):1011-1016.

Garvey, W. D., & Griffith, B. C. 1972. Communication and information processing within scientific disciplines - empirical findings for psychology. *Information Storage and Retrieval*. 8 (3):123-136.

"Granularity" Def. 1. Business Intelligence Dictionary. [Online]. Accessed http://business-intelligence.com/dictionary/granularity/. (22 April 2015).

Price, Derek John De Solla. 1963. *Little Science, Big Science…And Beyond*. Columbia University Press.

Taylor, Robert. 1986. *Value Added Processes in Information Systems*. Westport. Greenwood.

Tenopir, Carol and King, Donald, W. 2004. *Communication Patterns of Engineers*. Wiley-IEEE Press.

Tenopir, Carol, Sandusky, Robert J., and Casado, Margaret M. 2006. The Value of CSA Deep Indexing for Researchers. Prepared for CSA. Available: http://works.bepress.com/carol_tenopir/1. (22 April 2015).

Tenopir, Carol, et al. 2011. Data Sharing by Scientists: Practices and Perceptions. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101. (22 April 2015). DOI: 10.1371/journal.pone.0021101.

Tenopir, Carol, Birch, Ben, and Allard, Suzie. 2012. Academic Libraries and Research Data Services. Association of College and Research Libraries. Available: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf. (22 April 2015).