Datasets, a Shift in the Currency of Scholarly Communication: Implications for Library Collections and Acquisitions

Hilary M. Davis and John N. Vickery

As the market of scholarly communication continues to evolve, a number of indicators suggest that the unit of information currency is shifting from a primary focus on journal articles to a broader emphasis on key elements of scholarly communication, namely data sets. This article examines and summarizes recent developments that have contributed to this shift in emphasis. The authors will also consider how this shift may affect some of the core functions of the collections and acquisitions processes. Serials Review 2007; 33:26–32.
© 2006 Elsevier Inc. All rights reserved.

Introduction

Herbert Van de Sompel et al. propose that "in the established scholarly communication system, the concept of a journal publication dominates our definition of a unit of communication."[1] In other words, journals and journal articles have been long recognized as the primary unit of information currency to the exclusion of most other forms of scholarly scientific information, including data sets. However, most scientific journal articles are based on empirical data or sets of data ("data sets"). Evidence suggests, however, that data sets are becoming as valuable as research articles, which in many cases could be considered as interpretations of data. In fact, "data must be regarded as a critically important part of

the publication process, with documents and data being part of a seamless spectrum."[2]

For the purposes of this discussion, data and data sets include everything from geographic information systems or geospatial (GIS) data to genomic data to any data set supporting a scholarly publication, such as census data. At this stage, this shift is clearly more immediately relevant to the sciences, but it is likely that the social sciences will also follow this trend in the foreseeable future. While this article applies mostly to research-oriented libraries, all types of libraries will need to help their communities gain access to data and data sets.

This article outlines several trends that indicate the shift in emphasis of data sets as a primary unit of information currency. The authors will briefly discuss who the stakeholders are and some of their concerns in terms of collecting and providing access to data sets. In addition, the authors will touch on some of the ways in which academic libraries may need to change collections and acquisitions functions to respond to this shift in the emerging importance of data sets. Finally, the authors will reflect on the experiences of many academic libraries transitioning from print to electronic journals and attempt to apply some of the lessons learned to potential scenarios for collecting and acquiring data sets.

Indicators of the Growing Importance of Data Sets

There are several current trends that serve as indicators that data sets are indeed becoming more valued as a unit

Davis is Collection Manager for Physical Sciences, Engineering and Data Analysis, North Carolina State University Libraries, Collection Management Department, Raleigh, North Carolina 27695-7111, USA; e-mail: hilary_davis@ncsu.edu.
Vickery is Collection Manager for Management and Social Sciences, North Carolina State University Libraries, Collection Management Department, Raleigh, North Carolina 27695-7111, USA; e-mail: john_vickery@ncsu.edu.

of information currency, including the commoditization of data sets, an increase in legislative initiatives to protect data sets, the sheer growth and manipulability of data sets, evolving publication expectations placed on authors, and public–private partnerships developing around data sets.

Commoditization

Data as commodity is one of the most compelling indicators that data and data sets are taking on a much more vital role in the information marketplace. Several companies have begun marketing products that promote access to data as a value-added service. For example, Knovel Corporation is digitizing data-intensive reference works and reselling this information in a format that allows researchers to manipulate and interact with the data. The process of making data interactive (called "Knovelizing") has recently been patented. Nature Publishing Group is also actively promoting value-added products and services based on access to data sets by linking from their online articles to the data supporting the articles. From the journal *Nature Chemical Biology*, for example, one can link directly from a chemical molecule in an article to databases like PubChem and GenBank. This connection allows readers to verify the data behind the interpretation and draw conclusions for themselves. Contextualization is becoming an integral part of publishers' services with access to and interaction with data seen as a critical component to scholarly communication. By turning scientific papers into structured data objects, readers can interact with an article in much the same way they do with databases, querying and manipulating the data contained within and supporting the article itself.[3] Data and data sets are taking on new meaning in the scholarly communication marketplace for publishers as well as consumers.

Recently Introduced Legislation

As data sets become more and more valued as commodities, it is likely that creators of data sets will want to protect their investments. Currently in the United States, factual data are not copyrightable. Only the creative display of data is protected by copyright. However, since 1996, the European Union has designated that databases as collections of facts regardless of any element of creativity can be protected against extraction and reuse.[4,5,6] Similar attempts have been made in the United States to apply blanket protections on collections of data. The Database and Collections of Information Misappropriation Act (H.R. 3261) is one example. The bill proposed to make "civilly liable any person who makes available in commerce to others a substantial part of the information contained in a database generated, gathered, or maintained by another person without authorization." One troubling question for many libraries is that there is no consideration for how fair use would apply within this proposed legislation.[7] Although this particular bill has been tabled for now, similar bills are likely to emerge in the future. The American Library Association (ALA) Web site provides a good background explanation on database

protection efforts.[8] As data sets become regarded as fiscally driven investment opportunities, there is the chance that libraries that have an obligation to collect data sets for their users will face substantial costs and legal barriers.

Growth and Organization of Data

Even while legislative efforts to control data persist, researchers continue to produce and request access to data for their work. While data are being produced at exponential rates, it is not a trivial matter for researchers to discover, access, and repurpose data sets. Lyman and Varian[9] report that in 2002 alone more than five exabytes of information were created. That equates to 37,000 times the size of the Library of Congress. With advances in computing power and technology, it is easier than ever to mine and manipulate large data sets as long as they are mounted online.[10,11] The problem for researchers and librarians, however, is that many of these data sets are scattered amongst individual researchers' desktops, in Web sites, and in open and closed digital repositories; there exists no truly organized system for discovering, accessing, and repurposing data sets. Some national and international efforts are attempting to remedy this problem by establishing cyberinfrastructure networks and programmatic policies to support archiving of digital data for the long term.[12,13,14] The proliferation of data sets drives research and as efforts to organize and share data sets take shape, users will likely expect libraries to make data sets an integral part of their collections.

Publication Requirements

In addition to efforts to organize and share data sets via cyberinfrastructure channels, some publishers are either directly or indirectly supporting efforts to organize the increasing amount of data. Consequently, more and more journals are requiring authors to submit their data to a nationally recognized repository before their manuscript will be considered for publication. The *Proceedings of the National Academy of Sciences* is one example. In their instructions to authors, it is stated that: "Before publication, authors must deposit large data sets … in an approved database and provide an accession number for inclusion in the published paper" (http://www.pnas.org/misc/iforc.shtml). In cases where no public repository exists for a particular research sector, authors are often required to provide their data as "supporting information," mounting the data online and also providing a copy of the data to the journal publisher. As such, data sets are increasingly vital for full evaluation and continuation of research agendas. This closer association with journal articles brings data sets to the fore in terms of the resources that library users need and expect to be available to them.

Public/Private Partnerships and Repackaging of Public Information

Some interesting public–private partnerships are emerging around the increasing value of data sets. These partnerships have the capacity to propel the financial

fitness of private corporations while pushing the envelope of scientific progress at the governmental level. The Human Genome Project is one of the most visible examples. The most discussed goal of the Human Genome Project was to identify all of the genes in human DNA and to sequence the entire human genome. Another major goal was to transfer related technologies to the private sector, such as drug design based on gene sequences and diagnostic testing for genetic diseases. Celera Genomics Group was one of the major private sector companies that capitalized on having a relationship with both the Department of Energy (DOE) and the National Institutes of Health (NIH) to advance their business interests as well as the goals of the Human Genome Project.[15,16] Other partnerships that are being developed on the basis of data sets are between Reed Elsevier and Pfizer Inc. to leverage the data contained in PubChem (a database of chemical data developed by the National Library of Medicine) and the partnership between Elsevier, the United States Patent and Trademark Office (USPTO), and the European Patent Office (EPO) to repackage freely available patent information in a fee-based database product.[17] Public–private partnerships such as these may result in higher quality search interfaces and value-added services of data set packages. While these features are attractive to libraries, they may also imply higher costs as opposed to data sets acquired from government entities.

While by no means a comprehensive list, these are just a few of the trends indicating that data sets are becoming more valued as a major currency of information. In each of these examples, publishers, database developers, product vendors, legislators, librarians, researchers, and the general public all hold a stake in how data can and should be shared, accessed, and further repurposed. Researchers and scientists are mostly concerned with protecting their ability to share and use data sets, whereas discussions among librarians recognize that data sets are key to knowledge discovery and are becoming just as important as the journals and books in library collections.

Stakeholders and Their Concerns

At present, the majority of the discussion about data sets is taking place among scientists and researchers, with fields such as chemistry, genomics, and mathematics at the forefront of the discussion. Researchers in these areas are clearly interested in how the massive amounts of data scattered on and off the Web will be discovered, accessed, shared, and used. The literature also points to their concern over protecting their intellectual property as it relates to data.

Regarding the potential consequences of the commoditization of data, one researcher from the National Center for Supercomputing Applications wrote: "Celera's treatment of their human genome sequence data shows the sequence's nature as an economic good as well as the possibility of it being a private good or commodity. The company can control and sell access to the data, and outside researchers have no recourse if they are denied or

cannot afford access."[18] For many researchers, having open access to share, append, and annotate data sets is critical to their work. In response to the potential database protection mandates, several researchers wrote: "For most scientists, having the right to download data cdoes not mean much if the extracted information cannot later be reutilized and republished."[19] On the other hand, one author suggested a novel way to control and limit open access to genetic data. By encoding data in MP3 music format, that data can be protected by copyright.[20] While the genetic sequences themselves are not copyrightable because they are simply factual data, the creative packaging of the gene sequences in MP3 format does offer copyright protection. Reconciliation of these two differing perspectives on access to data sets will be tricky and will likely cross national borders.

Discussions by librarians and information science advocates highlight the concerns of researchers and scientists that data sets are scattered and not well-integrated with other functions of the scholarly work environment.[21] Their vision is of a world where data sets are highly valued and highly integrated into current assemblages of information such as library collections.[22] Lynch recently stated that data software, visualization, and observation now are every bit as important and as significant as traditional monographs and journal articles to the practice of scholarship in modern research (paraphrased).[23]

Going further, some open data advocates who argue that data and data sets should be available online for free without barriers envision that "the boundaries of what would conventionally be thought of as a 'paper' or 'article' can be scaled both up and down … an article could be disassembled down to an individual marked-up component such as one atom in a molecule, or instead aggregated into a journal, collection of journals, or ultimately into the semantic Web!"[24] Along these same lines, Van de Sompel et al. wrote that "a future unit of communication should not discriminate between media types and should recognize the compound nature of what is being communicated."[25] There are various stakeholders who have differing ideas on the value and use of data sets. Libraries will be required to serve the needs of these sometimes conflicting expectations of how data sets can or should be used, shared, and preserved.

The currency of information is clearly shifting to place more value on data sets. The ability to access and reuse data sets is confounded by the amount of data and data sets being produced. Several major publishers are taking steps to link data with articles and back again to data, creating much needed and valuable services to address these issues. On the other hand, efforts are underway to pass laws that could seriously limit how data can be used. What does all of this mean for libraries and the way librarians carry out collection and acquisition functions? In the next section, the authors provide a brief overview of current business models for data sets and discuss potential implications for library collections and acquisitions functions including budgeting, selection and evaluation, negotiation, and licensing. Issues of data

preservation and training are discussed at length by others.[26,27]

Current Business Models

While there are many variations and exceptions, there are five main business models and scenarios in common use for collecting and providing access to data sets. In the following sections, these models/scenarios will be examined in terms of how they currently work and how they could evolve in the near future with regard to data sets. The authors also discuss possible models that could better represent the interest of libraries as the collection of data sets becomes more important for libraries.

Institutional Membership Model

In the institutional membership model, libraries pay an annual fee on behalf of their users to become members of an organization that collects, archives, and makes available data sets for research purposes. This fee may allow for unlimited access to the data sets collected by the organization. One possible benefit of the institutional membership model is that it frees the library from the responsibility of archiving and organizing the data. Often, organizations that operate on the institutional membership model are affiliated with academic institutions and include such examples as the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan and the Roper Center for Public Opinion Research at the University of Connecticut. Data sets are made accessible via a searchable Web interface or through a library-mediated request process. Pricing varies but Carnegie Classification or full-time equivalent measurements (FTE) appear to be the most common pricing models.

Serials Continuation Model

Another model used to provide access to data sets is similar to a serials continuation purchase. In this case, the data are often acquired on CD or other physical format and are released at regular intervals. An example of an organization that has used this model is the National Institute of Standards and Technology (NIST). For budgetary planning and acquisitions processes, this model does not present many surprises. However, the fact that data may not be accessible online or off-site has major drawbacks for library users' ability to easily access the data.

One-Time Payment Model

Another common model is the one-time payment per data set. This is perhaps the most traditional approach and corresponds to how libraries are accustomed to acquiring monographs. Organizations that offer data on a per data set basis may also provide other models such as an institutional membership. Data sets acquired through this model are often driven by users' requests. For example, a researcher may identify a data set as necessary to his or her work and petition the library to purchase it for the collection. The process is essentially the same as a monograph request; however, the cost for a data set can be much more than the cost of a monograph.

Transitional Models

The following two business models are unique in that they function in lieu of the more traditional business models discussed previously. They represent a transitional state, perhaps toward the development of a stable business model. Of particular importance is the opportunity these transitional models present for libraries to influence their development and possibly help shape them into sustainable business models which meet the long-term needs of libraries.

The first of these transitional models could be described as an *ad hoc* arrangement between data producers and libraries. This model is perhaps most common in the geographic information systems (GIS) community where a significant proportion of the data applicable to research is created by states, counties, and municipalities. Whereas some government entities may charge a fee in order to recoup costs, others may be willing to make arrangements with libraries to host the data in exchange for access. This arrangement can relieve the government or data producer of the cost of maintaining the data while the library is able to provide access for its own researchers.

The second transitional scenario is essentially the lack of a business model. In this case, the data producers simply do not have a process in place to deal with academic library needs. Organizations or data producers operating under this scenario may be new to the library market or may be transitioning from a model mentioned previously such as the one time payment model. This situation offers the greatest potential for libraries to shape the development of the data producer's business model by beginning the negotiation process on the ground floor to include not only price point and license details but the business model itself.

Hindsight and Potential Scenarios

As the relative importance of data sets to research continues to mount, the transition from print to electronic journals may be seen as analogous in terms of the effects data sets may have on the following core collections and acquisition functions: budgeting, selection, evaluation, negotiation, and licensing. Reflecting on the transition to electronic journals and applying lessons learned from that experience, libraries can begin to envision and plan how to map these core collections and acquisitions functions to data sets as an emerging unit of scholarly information. What kinds of questions should libraries consider with regard to collecting data sets and how can lessons learned from other experiences be applied to potential scenarios in acquiring and collecting data sets? The authors attempt to address some of the issues in the following sections.

One of the main questions that libraries should consider with regard to the emerging importance of data sets is "Who will control the data sets market?" It is clear that data sets will be an increasingly important asset for libraries, especially academic libraries. Hindsight with regard to the transition from print to electronic journals suggests that this trend will certainly affect core

functions in the areas of collections and acquisitions. During the transition to electronic journals, many libraries expected lower costs or even free access to content. While also expecting more flexibility in collections budgets, in reality libraries have had to pay more for online access and have had less freedom in selecting journals. Even with online only subscriptions, the annual cost is often equal to or more than the corresponding print version and many libraries find themselves locked into multi-year licensed journal packages or bundles.[28] What should libraries keep in mind as they shift towards collecting and providing access to data sets? As always, libraries should be prepared to do more with less. By now, this is a common statement but it bears repeating as libraries move into new territory. Libraries should maintain a focus on the core responsibilities of selection and evaluation. Libraries should also be aggressive negotiators and work for licenses that meet the needs of the library and its users. In the following section, the authors examine these considerations and explore how the growing importance of data sets will affect budgeting, selection, evaluation, licensing, and negotiation. At this stage in the transition, there are more questions than answers; however, looking ahead with a critical eye is important as libraries work to manage change and limit risk.

Budgeting

It is widely assumed that future budget realities will require libraries to do more with less.[29,30] The number of quality products available to researchers as well as their cost continues to increase. In order to provide access to these resources including data sets, libraries will be forced to make difficult decisions regarding allocation of funds, the costs and benefits of each new resource, and the hidden costs associated with collecting data sets.

The first question that libraries must ask is "How should funds be allocated in order to accommodate data sets as an integral part of the collections budget?" To date, data sets have typically not occupied a defined segment of libraries' collections budgets. Rather, data sets are often acquired on an *ad hoc* basis with funding based on the business model used to acquire the product. While this strategy has worked, libraries must take a comprehensive look at allocations for data sets in order to streamline their collection procedures. In determining an effective allocation model, libraries will also need to consider the costs and benefits of high-value but low-use data sets. Perhaps metrics beyond the accepted log-ins and page views should be developed to more accurately gauge the importance of data. For example, citations to data sets could be used as an indicator of the relative value to a libraries' collection.

Additionally, libraries should prepare to budget for hidden internal costs associated with collecting data sets. Looking back to the transition from print to electronic journals, there were several costs associated with electronic journals that were unnecessary for their print counterparts. For example, consider the institutional costs of link servers and resolvers, such as SFX and SerialsSolutions. Other associated costs include the

implementation of e-journal finders and ERM systems whether in-house or purchased.

At this stage, what kinds of infrastructure costs can libraries begin to plan for with regard to data sets? One example is an interface researchers use to locate and interact with data sets. It is reasonable to envision libraries developing their own interfaces, as well as external companies marketing interfaces, to serve this function. A well-known example of this for financial and economic research is *Wharton Research Data Services* which provides researchers a standardized interface supporting data from multiple producers (http://wrds. wharton.upenn.edu/). Interfaces represent only one example of the type of institutional costs that libraries will need to consider in order to accommodate researchers' demand for data set access.

Selection and Evaluation

In addition to planning an appropriate budget to accommodate data set collections, libraries must also look critically at the selection and evaluation processes. These functions represent core components of the collections process and have not yet been adequately examined in light of the growing importance of data sets to library collections.

Looking at the selection process, the most basic question libraries must ask is "Which data sets should we collect?" There are several potential drivers behind the decision to collect a particular data set or collection of data sets. The first of these, and possibly the most common to date, is a user request. Based on the request, a library can work to develop a partnership with a data producer to develop an acceptable business model if one does not already exist. Another potential driver of the selection process is the value system of data producers; a library's choice of which data sets should be collected may be limited by what the data producers consider marketable and how they decide to package data sets. A relevant parallel to describe this situation are e-journal bundles, sometimes referred to as "Big Deals," in which publishers sell large collections of journals for a single price, often under multiple-year contracts that offer little flexibility in which journals are included. The third potential driver behind any collection decision is a proactive stance on the part of libraries to begin building collections of data sets in anticipation of user needs. While libraries will likely employ a combination of these strategies, a perceived advantage of the latter driver is that libraries could begin to collect data "at the source"—that is, collect, organize, and host data sets generated by researchers at their own institutions. In doing so, libraries have the potential to exert influence over the emerging data sets market rather than waiting for commercial vendors to harvest and package the data for later re-sale.

Other issues libraries should consider with regard to the selection process range from policies to approval plans. At present, data sets do not appear to factor heavily into collection development policies. Libraries have the option of explicitly defining a policy with data set selection as a component or allowing the policy to

develop organically as data sets become more prominent as a unit of scholarly information.

As reported in a recent article in the *Chronicle of Higher Education*, the quantity of data being produced and of potential impact to scholars is growing exponentially.[31] Could libraries see a strategy similar to monograph approval plans to deal with the growing amount and importance of data sets? What parameters would a library need to consider in developing a data set approval plan? Will one of the current market leaders take on the task of data set approvals or will a newcomer enter the market?

In conjunction with selection of data sets, librarians will also need to develop mechanisms to effectively evaluate the quality and relevance of a potential data set purchase. To date, libraries have relied on the peer review system as one of the primary benchmarks of quality. It is not clear, however, that the peer review system can or should translate to data. While some have suggested that data should be peer reviewed similarly to scholarly journals, others report that peer review of data would complicate and slow down the scholarly process.[32] If, in fact, the peer review system does not apply as an effective evaluative tool for data sets, what mechanism should librarians use? Focusing data set collections on those data produced by a libraries' own institution may prove to be the most effective way to build quality collections in the absence of a peer review system.

In addition to evaluating the quality of the data itself, libraries will also be required to consider the storage and display needs of data sets. For example, the interface platform, if available, must be evaluated in terms of ease of use and the ability to manipulate the data. In addition, data format must be taken into account. Is the data in a widely recognized format? What additional software is required for researchers to make use of the data based on its format? Questions such as these add to the complexity of the evaluation process.

Licensing and Negotiation

As the importance of data sets increases, libraries should prepare to license access to data sets. In the give and take negotiations of licensing access, libraries must consider what rights are acceptable to give up in order to gain access and what rights are non-negotiable. With electronic journals, conditions such as archival access and campus wide access have proven to be important. Similar conditions should be considered equally important for data sets. Other rights and obligations necessary to consider include interlibrary loan and potential liability for misuse of data. Libraries will need to position themselves as aggressive negotiators in order to obtain licenses that are beneficial to library users and acceptable in budgetary terms.

Summary and Future Considerations

The currency of information is shifting to place more value on data sets. The authors have described several trends that point to the increasing importance of data sets; data are becoming a major commodity as publishers and vendors realign their products to focus on integration of data as a value-added service. National and international efforts to democratize access to data sets endorse the view that access to data sets should be open and collaborative. Several major publishers are taking steps to link data with articles and back again to data, creating much needed and valuable services. On the other hand, efforts are being made to create laws that could limit how data can be used. The authors have referred to several key stakeholder groups who demand and need access to data sets at all levels of research and education—from data producers to data consumers. The amount of data and data sets being developed is a significant hurdle for ensuring ease of access and reuse. The different needs and expectations placed on the value of data sets by different stakeholders are creating confounding situations with nontrivial ramifications for scholarly communication. Libraries will need to watch these developments closely as they integrate data sets in with existing library collections and services.

The authors have also discussed potential implications for library collections and acquisitions functions including budgeting, selection and evaluation, negotiation, and licensing. Currently libraries may be paying for print plus online access to society and commercial journal articles. As data sets become more integrated with journal articles, there is a likelihood that business models will soon include access to data sets along with online access. There is also the chance that publishers will not only charge for linking out to data sets from articles but may also require rights to license access to data sets altogether. The question librarians need to consider is "How much more will we be willing to pay for these access enhancements?" Many libraries are aware of the effects of the package plan or bundled deal. Is this a viable and/or desirable scenario for collecting data sets?

Many libraries are advocating for long-term preservation and access to journal content. What about archival rights to data sets? Will vendors of data sets be willing to make archival rights to data sets a possibility? If libraries purchase access to data sets, will our licenses make us liable when our users go beyond the allowed terms of use? If libraries capture data at the source and create institutional repositories of data sets, it is conceivable that libraries could gain significant control of the data sets market. Are libraries in any position to do this? At this point, there are more questions than answers. Librarians and other stakeholders will need to keep up with the emerging market of data sets as a scholarly commodity. Collecting and acquiring data sets will likely become an integral part of many academic libraries' experiences, just as with collecting and acquiring electronic journals. Collecting data sets offers many opportunities for libraries to help guide burgeoning business models of publishers who are leveraging the insights that data sets provide to research and learning. Just as users expect to find robust collections of books and journals, they will look to libraries to provide them with access to data sets as a more visible and marketable unit of currency for scholarly communication.

Notes

1. Herbert Van de Sompel, Sandy Payette, John Erickson, Carl Lagoze and Simeon Warner, "Rethinking Scholarly Communication: Building the System That Scholars Deserve", *D-Lib Magazine* 10 (2004), http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html (accessed September 12, 2006).

2. Henry S. Rzepa and Peter Murray-Rust, "A New Publishing Paradigm: STM Articles as Part of the Semantic Web," *Learned Publishing* 14, no. 3 (2001): 177–182.

3. Timo Hannay, "Database Publishing at Nature," http://db.cs.berkeley.edu/dblunch-fa2005/timo.ppt (accessed September 12, 2006).

4. Carol Ebbinghouse, "If at First You Don't Succeed, Stop! Proposed Legislation to Set Up New Intellectual Property Right in Facts Themselves," *Searcher* 12 (2004): 8–17.

5. J. Carols Fernandez-Molina, "The Legal Protection of Databases: Current Situation of the International Harmonisation Process," *Aslib Proceedings* 56, no. 6 (2004): 325–334.

6. Matthias U. Studer, "The Quest for Legal Protection of Databases in the Digital Age: Recent Developments with Focus on the United States," *Journal of World Intellectual Property* 4, no. 5 (2001): 653–702.

7. Ebbinghouse, "If at First You Don't Succeed," 13.

8. American Library Association, "Database Protection Legislation," http://www.ala.org/ala/washoff/WOissues/copyrightb/dbprotection/databaseprotection.htm (accessed September 12, 2006).

9. Peter Lyman and Harold R. Varian, "How Much Information?" http://www2.sims.berkeley.edu/research/projects/how-much-info-2003 (accessed September 12, 2006).

10. International Council for Science, Committee on Scientific Planning and Review, "Scientific Data and Information: A Report of the CSPR Assessment Panel," (2004), http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Information.pdf (accessed September 12, 2006).

11. Organization for Economic Cooperation and Development, Directorate for Science, Technology and Industry, Committee for Information, Computer and Information Policy and Working Party on the Information Economy, "Digital Broadband Content: Scientific Publishing," (2005), http://www.oecd.org/dataoecd/42/12/35393145.pdf (accessed September 12, 2006).

12. Blue Ribbon Advisory Panel on Cyberinfrastructure (Dan E. Atkins, Chair), "Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure," (2003), http://www.communitytechnology.org/nsf_ci_report/ (accessed September 12, 2006).

13. National Science Board, "Long-lived Data Collections: Enabling Research and Education in the 21st Century (Report of the National Science Board)," (2005), http://www.nsf.gov/pubs/2005/nsb0540/start.jsp (accessed September 12, 2006).

14. American Council of Learned Societies, "Draft Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences," (2005), http://www.acls.org/cyberinfrastructure/acls-ci-public.pdf (accessed September 12, 2006).

15. Human Genome Management Information System, "Human Genome Project Information," http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml (accessed September 12, 2006).

16. J. William Bell, "Our Genome in Common: Genomic Data Release Policies and the Academic Librarian," *portal: Libraries and the Academy* 3, no. 2 (2003): 293–306.

17. Elsevier Inc., "Ei Patents," http://www.ei.org/databases/ei_patents.html (accessed September 12, 2006).

18. Bell, "Our Genome in Common," 298.

19. Stephen M. Maurer, P. Bernt Hugenholtz and Harlan J. Onsrud, "Intellectual Property: Europe's Database Experiment," *Science* 294, 5543 (2001): 789–790.

20. Willem P.C. Stemmer, "How to Publish DNA Sequences with Copyright Protection," *Nature Biotechnology* 20 (2002 (March)): 217.

21. Nicola Cannata, Emanuela Merelli and Russ B. Altman, "Time to Organize the Bioinformatics Resourceome," *PLoS Computational Biology* 1, no. 7 (2005): 531–533.

22. Carole L. Palmer, "Scholarly Work and the Shaping of Digital Access," *Journal of the American Society for Information Science and Technology* 56, no. 11 (2005): 1140–1153.

23. Clifford Lynch, "ECURE 2005 Pre-note Address," http://www.asu.edu/ecure/2005/prenote/ (accessed September 12, 2006).

24. Rzepa and Murray-Rust, "A New Publishing Paradigm," 180.

25. Van de Sompel, Payette, Erickson, Lagoze and Warner, "Rethinking Scholarly Communication," http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html.

26. William L. Anderson, "Some Challenges and Issues in Managing, and Preserving Access to, Long-Lived Collections of Digital Scientific and Technical Data," *Data Science Journal* 3 (2004): 191–201.

27. Gail Hodge and Evelyn Frangakis, "Digital Preservation and Permanent Access to Scientific Information: The State of the Practice," CENDI US Federal Information Managers Group, CENDI 2004-3 (2004), http://www.icsti.org/icsti/icsti_reports.html (accessed September 12, 2006).

28. Kenneth Frazier, "The Librarians' Dilemma: Contemplating the Costs of the 'Big Deal'," *D-Lib Magazine* 7, 3 (2001), http://www.dlib.org/dlib/march01/frazier/03frazier.html (accessed September 12, 2006).

29. Martha Kyrillidou, "Research Library Trends: ARL Statistics," *Journal of Academic Librarianship* 26 (2000): 236–427.

30. Charles B. Lowry, "Re-positioning Libraries: A Consideration of the Obstacles," *portal: Libraries and the Academy* 3 (2003): vii–xi.

31. Scott Carlson, "Lost in a Sea of Data," *Chronicle of Higher Education* 52, no. 42 (2006): A35.

32. Cecelia Brown, "The Changing Face of Scientific Discourse: Analysis of Genomic and Proteomic Database Usage and Acceptance," *Journal of the American Society for Information Science and Technology* 54, no. 10 (2003): 926–938.