Title	Biomedical Information and Its Users
Authors(s)	Shankar, Kalpana
Publication date	2009-12-09
Publication information	Shankar, Kalpana. “Biomedical Information and Its Users.” Taylor and Francis, 2009.
Publisher	Taylor and Francis
Item record/more information	http://hdl.handle.net/10197/4404
Publisher's statement	This is an electronic version of an article published in Encyclopedia of Library and Information Sciences, Third Edition. Taylor and Francis: New York, Published online: 09 Dec 2009; 615-623. Encyclopedia of Library and Information Sciences is available online at: http://www.tandfonline.com/doi/book/10.1081/E-ELIS3
Publisher's version (DOI)	10.1081/E-ELIS3-120043747

Biological Information and Its Users
Submitted for inclusion in the Encyclopedia of Library and Information Science

Kalpana Shankar
Assistant Professor, School of Informatics
Indiana University-Bloomington
901 E. 10th Street
Bloomington, Indiana 47408

Ph: 812-334-3686
Email: shankark@indiana.edu

KEYWORDS

biological information, cyberinfrastructure, digital libraries, scientific archives, data sharing

ABSTRACT

The purpose of this chapter is to introduce the information science reader to the wide range of data and other resources that constitute "biological information". Attention is paid to both paper and digital sources and the use of digital libraries and cyberinfrastructure for the creation, use, and re-use of information. The chapter discusses various user communities and their needs, including scientists, educators and students, policymakers, and other secondary users. The chapter concludes with challenges for data sharing, preservation, and access.

INTRODUCTION

The term "biological information" covers an enormous range of content, format, and uses. The creators and users of it are just as heterogeneous. The range of disciplines encompassed within the term "biology" is itself broad, including descriptive fields that rely on fieldwork and observation to those that are exclusively computer-based, or computationally intensive. Not surprisingly, the data generated from the results of scientific activity within these fields (which is what most biologists would call "biological information") generates a variety of information formats and storage, from

paper-based laboratory notebooks and anatomical drawings to petabytes of numeric data generated from high throughput screening, genomics, and proteomics. There is also great deal of information generated by what might be termed the context of scientific practice. These texts can provide insight into the processes that generated scientific data – grant reports, patent applications, correspondence, personnel records, and the like. This does not even include medical information, which comes with a different and attendant set of users, privacy and other concerns, and legal ramifications.

For the purposes of this entry, "biological information" will not include medical information, but will instead focus on the realm of data products from the scientific research process and contextualizing information from the conduct and organization of science within institutions. For the information scientist, biological information presents fascinating challenges for understanding and facilitating information policy, organization, sharing, and access. The goal of this chapter is to begin by discussing current practices of biological research, then lay out the varieties of information that encompass "biological information", focusing on both the data products of scientific activity (this includes results of experiments and analysis) and the documents that are created in the context of scientific research (correspondence, grant applications, and similar materials). Following that, the categories of users – scientists, educators, policy makers, and activists, will be discussed. Lastly, the chapter will outline some challenges that exist for understanding biological information – highly collaborative environments, data sharing and intellectual property regimes, proprietary formats. The conclusion will discuss potential challenges for research and practice.

INFORMATION PRACTICES ACROSS THE BIOLOGICAL SCIENCES

Until the early to mid part of the twentieth century, research progress in the biological sciences followed what has been compared to an artisanal model [1]. Primarily concentrated in government laboratories and universities around the world, junior students of the biological sciences trained with more senior scientists, then set up laboratories in other academic institutions around the world. While scientists used various instruments, the primary mechanism for recording information was the laboratory notebook which contained drawings, hand-written notes, and served as a daily aide-memoire for the scientist. To some extent, this model of research training and scientific data gathering persists and continues to be taught in the science classroom, although many digital tools exist for collecting data in a more automated fashion.

The biological sciences underwent radical changes in the latter half of the twentieth century and continue to change. The widespread use of information and communication technologies (ICTs) has made the biological sciences (indeed, all sciences) globalized and highly data-intensive in nature. As a result, one emergent trend that has influenced the nature of biological information is the advent of research teams can span countries and continents as expensive equipment is shared remotely and data sets analyzed in parallel. There have been efforts in numerous countries to develop information infrastructures that centralize data sharing and collection, standardize descriptive data (metadata), and make the use of remote tools more readily available. The emergent

sciences of genomics (the study of genes), proteomics (the study of proteins), and chemical informatics (the study of molecular and submolecular entities in living systems) belong to this data-intensive category of science.

A second important trend in the biological sciences that has been responsible for the generation of most of the data (and related and derivative information products) is the increase in the quantity and size of collaboration, often in response to emerging problems or even crisis situations (such as the SARS epidemic). Many researchers have noted that the need to gain access to instruments, unique data, and funding for increasingly complex projects requires the assemblage of individuals and teams with diverse skills [2] . As many scientists become more and more highly specialized, these teams are necessary to solve complex problems. Teams can also minimize impact on politically, environmentally, and socially sensitive regions and areas. Funding agencies such as the National Science Foundation and the European Funding Foundation often require that collaborations be formed in order to use natural and data resources in efficient ways. Collaborations also cut across geographical and institutional boundaries. Nonprofits, academic, government laboratories, and industries can work together to mobilize around problems and issues of mutual interest, hopefully with benefit to all concerned parties.

A third trend that has been instrumental in generating the explosion of biological data is the increasingly data-intensive nature of the biological sciences, aided by the introduction of information technologies with the capability to generate, manipulate, and store vast quantities of data at a high rate. Benoit [3] notes that the collaboration of computer

scientists and molecular biologists has generated an impressive collection of tools to manage data, conduct analyses, and present results that would not be possible without the integration of these disciplines. While the term bioinformatics has been reserved for the application of information technology, data mining, visualization, and other computational tools to questions in molecular biology, the other biological sciences (that is, those whose research purviews reside above the level of molecule, from the cell to whole systems) have also been aided by "informatics" and in turn have generated vast quantities of data. The resulting data (often the results of extremely intensive string manipulations [3] is stored in databases that may be local, publicly available, or commercial/proprietary.

In the last several years, the contribution of high performance computing to the scientific enterprise has been subsumed under the term "cyberinfrastructure", coined by the National Science Foundation to integrate its sponsored research efforts in data acquisition, storage, management, and other Internet-based services around scientific information [4]. The term is meant to include the human resources needed to make the technical infrastructure function, but the discourse has tended to focus on a technologically deterministic vision for Web services in science [5, 6]. In the United Kingdom and elsewhere, similar efforts have been termed e-Science [4]. Cyberinfrastructure grants have proven to be technological drivers in furthering the computationally intensive sciences. Challenges include making data formats mutually intelligible, the creation and implementation of standards, and the creation of tools that make data resources potentially useful to multiple communities of experts and

nonexperts. An important commonality across these cyberinfrastructure projects is the need for efficient storage and access to information [7]. Many in the biological communities have relied on cyberinfrastructure projects to minimize bottlenecks to data sharing, since data derived in NSF-funded projects are generally considered open access and nonproprietary. Nevertheless, because cyberinfrastructure is built upon the current Internet, it has inherited all of the technical, legal, cultural, and policy problems of the Internet, including security, language barriers, intellectual property regimes, and others. Cyberinfrastructure is not the only mechanism by which the biological sciences have been transformed. Even fields that have traditionally been "descriptive" – for example, ecology and environmental biology – have also undergone dramatic changes, aided by the use of various kinds of technologies. Mapping technologies such as Geographical Information Systems (GIS), satellite imagery, and earth-based sensor networks allow scientists to use data that has changed the nature of these "traditional" biological disciplines [8].

These three trends, when taken together, probably account for the "information explosion" in the biological sciences more than any other factors. Collaboration across disciplines and informatics together have made the biological sciences extremely data-intensive. The former has enabled scientists to use equipment and resources in distant locations and answer increasingly complex problems with the help of colleagues with very different disciplinary interests. Informatics is entwined with this collaboration, since it implies the integration of biological sciences with computational power. Of

course, policy initiatives and financial resources to back them can spur such efforts to greater success.

VARIETIES AND FORMATS OF INFORMATION

An early model by which biological information is organized is the UNISIST model, developed by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) and the International Council of Scientific Unions (ICSU), and more recently updated by Fjørdback Søndergaard et al [9]. In this model, three channels of information: tabular (or data), formal, and informal channels are defined, with three levels of information (primary, secondary, such as databases, and tertiary, such as encyclopedias and books). This makes a good departure point, but is hardly complete.

Much biological data is still recorded on paper. Many researchers in the biological sciences still use bound laboratory notebooks in the conduct of daily work. These notebooks seem antiquated and difficult to use, given the difficulty of searching across data sets and integrating them, but they are still prevalent and this is still the mainstream educational model in science classrooms at universities, which provide the foundational education for scientists [10]. Laboratory notebooks are a form of data capture that are easy to understand and use, transcend computer technology that breaks or becomes outdated, are fairly simple to maintain for short- and long-term use, and do not require any proprietary systems. They are also effective mechanisms for integrating printouts, drawings, spreadsheets, and other date formats.

Digital data, not surprisingly, is far more diverse in scope. Specialized instruments usually generate data files in proprietary formats; often, these data are further manipulated using computer-based analytical tools or integrated with other data formats. Common data formats, such as those from Microsoft and Adobe products are prevalent, but others are more specialized from manufacturers of scientific equipment. Other data, although technical, are in formats that have become standard in scientific fields: Geographical Information Systems (GIS), for example, generate data in a range of formats that are universally accepted and used by scientists. Still other data are often generated within and managed by "collaboratories" [11], closed groupware or collaborative systems that are built for or around a particular project or group that is heavily dependent on digital technologies.

These cover the general scope of primary resources in the biological sciences, but the data is analyzed and synthesized into various formats for dissemination outside of the immediate laboratory or research group. The range of publication formats is also numerous, and the reliance on different kinds of materials varies greatly by discipline and by individual preference. For example, to circumvent publishers' embargos on current research and the high costs of publications, many scientific communities created their own systems for circulating pre-prints and obtaining comments. While these have been extremely successful in the physical sciences, there has been significantly less use of pre-print or e-print servers in the biological sciences [12]. Conference papers, conference posters, monographs, reference manuals, and publications in peer-reviewed journals are

all important sources of biological information. Many of these publications may be circulated or shared in pre-print form, but scientists will also turn to secondary publishing sources to find them. Conferences and workshops will often maintain Web sites where publications from previous years can be found. In addition to general journals that may cover a wide range of scientific disciplines (such as Nature, one of its related journals, or Science), many scientists use more specialized peer-reviewed journals.

Although they may not be considered biological information per se, it is important to note another major category of documents that are important to the biological sciences: administrative materials. Scientists are responsible for obtaining grants and administering them, advising students, dealing with personnel matters, serving on peer review committees for journals and granting agencies, and depending on the status of the scientist, serving on conference and institutional, national, and other committees and boards. Many scientists may serve in multiple capacities as educators, consultants, researchers, and the like. In the context of these roles, scientists generate an enormous amount of information that may not be seen as important to the generation of biological knowledge, but are of incalculable value to other interested stakeholders with an interest in matters in the processes by which science is conducted [13].

INFORMATION AND DATA REPOSITORIES IN THE BIOLOGICAL SCIENCES

Data and other information in the biological sciences is, not surprisingly, stored in distinct and various locations, physical, virtual, and both. Laboratory notebooks and digital files tend to stay within the academic laboratories that created them, just as information in the commercial sector or government laboratories tend to stay within those institutions. In the latter case, digital data sets are not retained by the National Archives and Records Administration (NARA), but, in the best of circumstances, in the archives and records centers of these laboratories. More often, they are left in the custody of the scientists whose laboratories generated them.

Because their creators often consider these documents useless once they have been used for formal publications, they tend to be regarded without much interest for long-term maintenance and preservation [14, 15 REF National Science Board, Palmer and Heidorn]. There have been many institutional efforts to organize and contextualize data sets of long-term value as there has been some interest in repurposing them for other projects or having an easily accessible historical trail. Documents have been useful to historians and others with interest in scientific activity, so there is still interest in their long-term preservation. However, inactive documents and records often do not find themselves into archives of science for permanent retention, unless the director of the laboratory is extremely prominent. Even then, many archivists of science are reluctant to accept "raw" data, which is impossible to interpret without the aid of context in the form of the publications that resulted from them [16]. Nevertheless, there are many archives of science around the world, often focused around specific individuals, institutions, disciplines or subdisciplines, or topics. These archives may include the personal and

scientific papers of individuals and professional/academic societies, important instruments, and drawings and photographs. There have been many efforts to make the finding aids of these collections available online, but digitizing the materials themselves is difficult and expensive, so few archival scientific collections are available online in their entirety.

Scientific publications are a crucial resource for obtaining peer-reviewed research results. Publicly funded and proprietary databases index and often abstract this literature, making it accessible to individual and institutional subscribers and in the case of publicly funded databases, the broader public. Examples include the National Library of Medicine's PubMedCentral, and proprietary databases from other commercial publishers such as the Web of Science. For reasons similar to the impetus for e-print and pre-print repositories, there have been some efforts to circumvent the traditional scientific publishing model, which many critics argue favors publishers at the expense of scientists and the institutions at which they work [17, 18]. These critics charge publishers with creating a model in which universities and similar institutions subsidize research publishing, and then in turn pay for subscription rights to journals and secondary databases. However, given the imprimatur these systems lend to individual scientists in establishing their professional reputations, there has been little leeway and few incentives for them to break away from this system.

There have been some efforts to create open access models for more immediate and free access to scientific publications. One example is the Public Library of Science,

championed and co-founded by former director the National Institutes of Health Harold E. Varmus in late 2003. In addition to being a repository for publications and references, the Public Library of Science also encompasses several peer-reviewed open access journals in the biological sciences on a market-driven model in which scientists pay for publication, peer review, and online archiving.

The digital libraries movement has probably been the most influential in making biological data and other information broadly accessible and usable, particularly to nonexperts [19]. The creation of digital libraries has been instrumental in allowing educators and other nonexpert communities with interest in scientific data bypass the information overload, poor structure, and variable quality problems of scientific information available on the World Wide Web [20]. Resources in digital libraries are organized and maintained, usually around a focused topic, and can encompass a variety of information types and formats that can support searching, reuse, and other interesting re-uses of data and other scientific documents. However, appropriate learning outcomes must be in place if the technology is to be maximally useful [21].

USERS OF BIOLOGICAL INFORMATION

Most evidently, the primary users of biological information are biologists themselves. It is impossible to characterize the biological research community in any general ways. It is international, multidisciplinary, and collaborative. Biological scientists receive academic training in academic settings, but may work in universities, the private sector, in

government laboratories, for nonprofits and civil society, and even as independent scholars and researchers (although this limits the kind of work they can do). Scientists are trained in an apprentice-based model in which they ply research experiences with senior scientists, working on projects of interest to the principal investigator. These researchers can be funded through a number of mechanisms, including government and foundation grants. An emergent model is the targeted not for profit, which is often set up by an individual to target a specific disease. For example, the Bill and Melinda Gates Foundation in the United States, established by Microsoft founder Bill Gates, targets its global health funding into one disease: malaria. Since the interests of one or a few individuals essentially drive the funding from these kinds of organizations, they have the power to shape research agendas in ways that are distinct from other, traditional funding models.

Researchers can be found in other kinds of institutions as well. Although there has been significantly less biological research being conducted in the private sector, researchers in pharmaceutical companies and other biotechnology firms generate data and related information and are dependent upon multiple publication streams for disseminating results and keeping up with the current research, just like their academic counterparts.

Lastly, government laboratories and independent organizations throughout the world also employ these researchers. In the United States, the federal government administrates few biological laboratories of its own; the vast majority of its resources are disseminated in the form of grants to scientists in other institutions. Nevertheless, government

laboratories often conduct research in topics that may be of little financial interest to the private sector, such as therapies for rare illnesses, or clinical research.

Of course, biologists also work in non-Western countries; their information needs and challenges are different and often fraught with difficulties. While their needs for information may be as great as their Western counterparts, many characteristics of their situations make finding that information more difficult. Subscriptions to journals and secondary databases are often prohibitively expensive, even for institutional libraries and consortia of libraries. Most of these databases are no longer published in paper format, which would be the most useful for situations with uncertain power supplies and Internet access. Scholarly communication at conferences and workshops is of great importance to the biological community as informal networks are of important to all kinds of researchers; however, the costs of travel and often the difficulties of obtaining foreign visas make these options difficult in many situations and often marginalize scientific researchers outside of main research centers [22, 23]. To give one concrete example, the American Association for the Advancement of Science (AAAS), a non-profit organization in Washington, D.C., has argued that the United States' travel policies have made it extremely difficult for Cuban scientists to enter the U.S. to conduct legitimate research or attend important conferences [24]. These scientists, much like junior researchers and those at institutions that are considered less research-oriented, are generally deemed peripheral to the core research networks; this in turn, further marginalizes their work and thus they find it harder to get and share important information.

There are other scholarly communities with interests in studying science from social science, behavioral, and humanistic perspectives. These transdisciplinary areas of scholarship are often grouped under terms like "science, technology, and society", or "science studies"; these terms are used to denote research interests that generally focus on science as a practice and epistemology, with methodologies drawn from the social and behavioral sciences and humanities. History of science is one obvious area in which researchers rely on documents and artifacts produced by and about scientists, but philosophers, anthropologists, and sociologists of science also use similar artifacts and documents, such as journal articles, to make and situate their claims about science [25]. There are also other areas of research that may require use of biological information, such as bioethics, but there is little information on this community. In general, there has not been much research into the specific ways in which this community uses biological (or for that matter, any) information, so it is difficult to state categorically what kinds of information are of use this category of researchers. However, it is likely that their practices mirror those of other interdisciplinary fields, where informal networks often established through attendance at workshops and conferences , "footnote chasing" through publications to obtain new references, and collaboration (in some cases with biological scientists, in others with other scholars in science studies) yield useful information for the interdisciplinary scholar [26].

The introduction of digital libraries has been extremely useful for the educational community. There are a number of reasons why educators have collaborated with computer scientists, biological scientists, and information scientists to create digital

libraries. Although the biological sciences have been less affected than other scientific disciplines, concerns in the United States regarding the low interest of students in pursuing advanced degrees and subsequent careers in the sciences has led to new models of engaging students at younger ages. One way in which this challenge is being approached is through the use of real data to address real scientific problems and learn the processes of science as scientists themselves conduct it – a process often called "inquiry-based learning" [27, 28]. Much of the impetus to move towards more inquiry-based approaches to teaching science stemmed from the 1996 release of the National Science Education Standards (NSES) for grades K-12, which articulated numerous disconnects between classroom teaching and the ability of science students to connect their learning to science in the "real world". Other studies have found that these problems persist in undergraduate science education. Furthermore, the sciences have found a decline in the number of women and underrepresented minorities in the sciences (though the biological sciences have again been less affected). Some reports have suggested that one way to engage these students might be through exposing them to real world problems that can be addressed through scientific inquiry [29]. While the World Wide Web is a potential source of scientific information and today's students are familiar with the technology, in addition to the quality problems noted earlier, instructors note that the students themselves do not possess skills and aptitudes to evaluate information as critically as needed. Digital libraries have been used to support these efforts, since materials can easily be added to them to contextualize scientific data sets with age and level appropriate information. However, researchers have noted that without "pedagogical alignment", the potential of digital libraries to foster sophisticated inquiry

cannot be fully realized. Instead, they serve as yet another "information silo" in which students conduct artificially projects and teachers then move on to other lessons. Apedoe and Reeves [21] articulate seven dimensions that must be in alignment between pedagogical design and technology implementation. These include measurable outcomes and objectives, content that accurately reflects the often-poor structure and uncertainty of real scientific data, meaningful tasks, instructor support in the digital library, effective student interaction, attention to technological affordances of the library itself, and finally, appropriate assessment measures. In short, the science curriculum and specific learning activities must align appropriately with the content and organization of the digital library itself.

Another important category of users is organizations and individuals involved in policymaking and advocacy (these communities are not necessarily the same, although there may be some overlap). This arena has not been well documented by researchers in land use and planning These users may or may not have scientific training, but may be working with others who do to get help in interpreting the literature. There is also an enormous range of "policy makers", from local officials and planners [30] to members of Congress, as well as scientists and bureaucrats within government agencies. Transgovernmental organizations are also significant users of scientific information. Relying on primary evidence is considered necessary to making science-based policy in environmental affairs, health care, new biomedical technologies, as well as many other areas of science and technology policy. These users are not likely to delve into the primary literature of biology, much less the primary data sources. Therefore, task forces

and expert panels may be convened to assemble the state of the art on important topics; policy makers may also rely on information from the Congressional Research Service nongovernmental "think tanks", some of which have partisan agendas and others that do not. The National Research Council (NRC) and Institute of Medicine (IOM) are two quasi-governmental organizations that are tasked with providing expert opinion on various issues through the form of reports to the United States Congress and executive branch agencies; their reports are available to the public. Policymaking, however, is both a political and a scientific process, and that complicates the ways in which it is used. Some argue that scientific information with policy implications is not useful when specifically targeted to this population; instead, synthesized consensus on such topics should be targeted to those who have the power to influence policy-makers.

Another important group of users that rely on and even contribute to biological knowledge are interested amateurs. Some of them may be interested in data-driven approaches to social and political change. Again, these users represent a diverse set of interests, abilities, and needs. They may be geographically localized groups with an interest in environmental issues in their community [31], activist organizations that mobilize around a particular illness for personal, policy, and social change (AIDS activists are among the most well-known group, but there are many others) [32], or global movements interested in effecting change on transnational issues. Individuals may be interested in obtaining the latest scientific information for personal knowledge. They may be involved in what is often called "citizen-science" for personal reasons; these users are often involved in research efforts in ornithology, biodiversity, and other observational

fields [33, 34]. Many of the problems that plague other communities outside of the sciences persist: vast quantities of information that may or may not be valid, highly technical information, and few vetted pathways through that maze.

CHALLENGES AND CONCERNS

There are numerous open challenges to understanding the nature of biological information and making it useful. One central issue for practice and research is the issue of data sharing. The National Institutes of Health, for example, argues that data sharing is necessary to develop new analytical tools, encourage supporting studies, enable interdisciplinary research, prevent duplication of efforts. To this end, NSF and some other grantmaking agencies require a data-sharing plan in grant proposals. In spite of an ethos of information sharing, data sets in the sciences, even those with wide import and use, are often "siloed" or not preserved in useful formats. There are many reasons for this – lack of time, resources, interest, and incentives for data sharing are all cited as barriers to true information exchange [35, 36]. This varies from scientific community to community, though there have been extensive calls for data sharing in the broader scientific literature [37, 38]. It is often argued by scientists themselves that most information sharing is done in the form of peer-reviewed workshop papers, conference presentations, journal articles, and informal networks; therefore, sharing "raw data" is not feasible, since it would require setting up a dedicatee archive, responding to many individual requests, and other mechanisms that may be time-consuming. Committing

long-term resources to the maintenance of data only seems feasible in fields that are highly data intensive and "cutting-edge".

The fields of genomics and protein chemistry have publicly available databases and extensive toolkits for data analysis (examples include GenBank and RefSeq); these efforts are usually developed and implemented by various branches of the federal government, such as the National Library of Medicine's National Center for Biotechnology Information (NCBI). Data deposited to these databases is centrally maintained and curated. Many journals in molecular biology and related fields require submission of novel genetic or protein sequences before publication so that the accession number can appear in it. Europe and Japan have similar databases (EMBL and DNA Database of Japan, respectively), but work in collaboration with NCBI. Increasingly, many data are proprietary, as they are produced in the commercial sector, or in collaboration with commercial firms. The formal results of analyses of these data sets may be made public in publications, but the raw data is not made available for re-use.

Another concern is preservation and curation, not just of data, but also of the contextual information that is often of greater use to those outside of the biological communities. These are a complex set of concerns, both social and technical. Issues of digital preservation that haunt other arenas include problems with media, differing approaches, and metadata. These are issues that most laboratories do not routinely consider. Making information and data available to the scientific community requires appropriate

architectures, integration of data formats, and agreements on the funding and labor required to maintain such systems.

But to make scientific information accessible to the non-scientist, even more challenges ensue. There is certainly interest in access to such data, since at least in the United States, data sets are publicly funded resources. Digital libraries have been one approach to mitigating the access problem, but these are designed for educational users, not activists or policy making. For the researcher who needs access to the material that is produced as a by product of research, such as grant applications and letters, the problems may be one of too much information in too many places, all of it ephemeral. Much of the work that has been done on mapping formal and informal communication networks, history of science, and similar fields was only possible with the preservation of letters and documents in archives, or at least in other accessible settings and conditions. However, much of the current practice of science is done on-line: in email, word processors, and proprietary software or groupware. These systems are far less amenable to long-term preservation, creating what might be termed a "memory crisis" [39] for contemporary science. This problem will only get worse, as the kinds of technologies that scientists use become less and less amenable to digital preservation. It is still an open question, though, as to what needs to be saved at all for future use.

CONCLUSIONS

It should be clear that the term biological information is somewhat loose at best. In the research community, current focus is, not surprisingly, on the vast array of digital data being created and used, and a suite of pressing concerns with access, storage, analysis, and policy around it. But it is worth remembering that a significant amount of information is still generated on paper, in laboratories, by researchers and their students, and that this information is just as difficult to access. The ways in which these different kinds of information are generated are, not surprisingly, not always mutually exclusively. Furthermore, the kinds of information that are often of greatest interest to the nonexpert community is information that is increasingly difficult to access, or ephemeral. The users themselves are an extremely diverse group with differing needs; making biological information useful and accessible for all potential user groups is a daunting task and one that is probably impossible. Lastly, it is important to note that access to biological information is not just a set of technical problems; there are deep reasons of practice and policy that also shape storage, access, and use patterns. Most of these areas pose emerging questions of research and practice.

ACKNOWLEDGEMENTS

The author wishes to thank Jacob Warner for his research assistant and Barbara Andrews and Emily Maguire for reading preliminary drafts of this work.

REFERENCES

[1] Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. *The New Production of Knowledge*; SAGE Publications: London and Thousand Oaks, CA, 1994

[2] Hara, N. Solomon, P., Kim, S. L., & Sonnenwald, D. H. An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. Journal of the American Society for Information Science and Technology **2003**, *54* (10), 952-965.

[3] Benoit, G. Bioinformatics. In *Annual Review of Information Science and Technology*. Cronin, B. Ed.; Information Today: Medford, NJ, 2002; Vol. 39. 179-218.

[4] Atkins, D.E. (2003). Revolutionizing science and engineering through cyberinfrastructure: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. National Science Foundation. http://www.nsf.gov/od/oci/reports/atkins.pdf (accessed September 28, 2007)

[5] Gold, A. Cyberinfrastructure, data, and libraries, Part 1. D-Lib Magazine, **2007**, *13* (9/10).

[6] Almes, G., Birnholtz, J. P., Hey, T., Cummings, J., Foster, I., & Spencer, B. (2004). CSCW and Cyberinfrastructure: Opportunities and challenges, Chicago, Illinois.

[7] Heidorn, P. B., Palmer, C. L., & Wright, D. Biological information specialists for biological informatics. Journal of Biomedical Discovery and Collaboration **2003** *2*, 1.

[8] Frame, M. T. The National Biological Information Infrastructure: Building

information and geospatial technologies for biological community. Proceedings of the 2004 Annual National Conference on Digital Government Research, Seattle, WA, May 24-26, 2004; Washington, DC: Association for Computing Machinery, 1-4.

[9] Fjordback Søndergaard, T.; Andersen, J. & Hjørland, B. Documents and the communication of scientific and scholarly information: revising and updating the UNISIST model. Journal of Documentation **2003**, *59* (39), 278-320.

[10] Kanare, H. M. *Writing the Laboratory Notebook;* American Chemical Society Press: Washington, DC, 1998.

[11] Finholt, T. A., & Olson, G. M. From laboratories to collaboratories: A new organizational form for scientific collaboration. Psychological Science **1997**, *8* (1), 28-36.

[12] Lawal, I. Scholarly communication: the use and non-use of e-print archives for the dissemination of scientific information [electronic version]. Issues in Science and Technology Librarianship **2002**, *Fall*. http://istl.org/02-fall/article3.html (accessed September 28, 2007)

[13] Haas, J. K., Samuels, H. W., & Simmons, B. T. *Appraising the Records of Modern Science and Technology: A Guide;* Massachusetts Institute of Technology: Cambridge, MA, 1985.

[14] National Science Board (2005). Long-lived data collections: Enabling research and education in the 21st century. http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf (accessed September 28, 2007)

[15] Palmer, C.L., Cragin, M.H. and Heidorn, B.P Supporting biological information work: research and education for digital resources and long-lived data. Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, Chapel Hill, NC, June 11-15, 2006; Washington, DC: Association for Computing Machinery, 2006, 353.

[16] Elliott, C. A., Editor *Understanding Progress as Process: Documentation of the History of Post-War Science and Technology in the United States*; Chicago, IL: Society of American Archivists, 1983.

[17] Dewatripoint, M., Ginsburgh, V., Legros, P., Walckiers, A., Devroey, J.-P., Dujardin, M., Vandooren, F., Dubois, P., Foncel, J., Ivaldi, M., and Heusse, M.-D. Study on the economic and technical evolution of the scientific publication markets in Europe. 2006, http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf (accessed September 28, 2007)

[18] Correia, A. M. R., & Teixeira, J. C. Reforming scholarly publishing and knowledge communication - from the advent of the scholarly journal to the challenges of open access. Online Information Review, **2005**, *29* (4), 349-364.

[19] Bishop, A. P., Neumann, L. J., Star, S. L., Merkel, C., Ignacio, E., & Sandusky, R. J. Digital libraries: situating use in changing information infrastructure. Journal of the American Society for Information Science **2000**, *51* (4), 394-413.

[20] Recker, M. Perspectives on teachers as digital libraries users. D-Lib Magazine **2006**, *12* (9).

[21] Fraser, S. P., & Deane, E. M. Educating tomorrow's scientists: IT as a tool, not an educator. Teaching in Higher Education **1999**, *4* (1), 91-106.

[22] Kavulya, J. M. Digital libraries and development in Sub-Saharan Africa. Electronic Library **2007**, *25* (3), 299-315.

[23] Patil, D. J., Simon, J., & Cumberledge, S. Digital science libraries: practical approaches to supporting science in developing countries. UN Chronicle **2006**, *43* (3), 60-61.

[24] Munoz, E. (1998). The right to travel: The effect of travel restrictions on scientific collaboration between American and Cuban scientists. Washington, DC: American Association for the Advancement of Science. http://shr.aaas.org/rtt/report/report.htm (accessed September 28, 2007)

[25] Frohmann, B. The role of the scientific paper in science information systems. In Proceedings of *History and Heritage of Science Information Systems*, Philadelphia, PA, October 23-25, 1998; Bowden, M.E., Bellardo, T.H., and Williams, R.V. Eds.; Information Today: Medford, NJ, 63-73.

[26] Palmer, C. L. *Work at the Boundaries of Science: Information and the Interdisciplinary Research Process*; Kluwer Academic Publishers: Dordrecht, 1996.

[27] Apedoe, X. S., & Reeves, T. C. Inquiry-based learning and digital libraries in undergraduate science education. Journal of Science Education & Technology **2006**, *15*(5/6), 321-330.

[28] Dumouchel, B., & Demaine, J. Knowledge discovery in the digital library: access tools for mining science. Information Services & Use **2006**, 26(1), 39-44.

[29] Goldey, E. Disciplinary integration: The sciences and humanities in learning. In Invention and impact building excellence in undergraduate science, technology,

mathematics, and engineering education. Washington, DC: American Association for the Advancement of Science. http://shr.aaas.org/rtt/report/report.htm (accessed September 28, 2007)

[30] Theobald, D. M., Hobbs, N. T., Bearly, T., Zack, J. A., Shenk, T., & W., R. Incorporating biological information in local land-use decision making: designing a system for conservation planning. Landscape Ecology **2000**, *15* (1), 35-45.

[31] Castillo, A., García-Ruvalcaba, S., & R, L. M. M. Environmental education as facilitator of the use of ecological information: a case study in Mexico. Environmental Education Research **2002,** *8* (4), 395-411.

[32] Brashers, D. E., M., H. S., Neidig, J. L., & Rintamaki, L. S. Social qctivism, self-advocacy, and coping with HIV illness. Journal of Social and Personal Relationships **2002,** *19* (1), 113-133.

[33] Heidorn, P. B., Mehra, B., & Lokhaiser, M. F. Complementary user-centered methodologies for information seeking and use: systems design in the Biological nformation Browsing Environment (BIBE). Journal of the American Society for Information Science & Technology **2002**, *53* (14), 1251-1258.

[34] Card, J. J., Shapiro, L., Amarillas, A., McKean, E., & Kuhn, T. Broadening public access to data through the development of tools for data novices. Social Science Computer Review **2003**, *21* (3), 352-359.

[35] Dawyndt, P., Dedeurwaerdere, T., & Swings, J. Contributions of bioinformatics and intellectual property rights in sharing biological information. International Social Science Journal **2006**, *58* (188), 249-258.

[36] Fienberg, S. E. Sharing statistical data in the biomedical sciences. Annual Review of

Public Health **1994**, *15*, 1-18.

[37] Birnholtz, J. P. B. M. Data at work: Supporting sharing in science and engineering. International ACM SIGGROUP Conference on Supporting Group Work, Sanibel Island, FL, November 9-12, 2003; Association of Computing Machinery: Washington, DC, 2003, 339-348.

[38] Shabo, A., Rabinovici-Cohen, S., & Vortman, P. Revolutionary impact of XML on biomedical information interoperability. IBM Systems Journal **2006**, *45* (2), 12.

[39] Sherratt, T., & McCarthy, G. Mapping scientific memory: understanding the role of recordkeeping in scientific practice. Archives and Manuscripts **1996**, *24* (1), 78-85.

FURTHER READING

Abbas, J. Finding science resources online with the ARTEMIS digital library. Knowledge Quest **2003,** *31*(3), 12.

Bishop, A. P. Document structure and digital libraries: how researchers mobilize information in journal articles. Information Processing & Management **1999**, *35* (3), 255-279.

Dedeurwaerdere, T. The institutional economics of sharing biological information. International Social Science Journal **2006**, *58* (188), 351-368.

Heidorn, P. B. Publishing digital floras and faunas. Bulletin of the American Society for Information Science & Technology **2003**, *30* (2), 8-11.

Lundmark, C. BEN: The biology branch of the National Science Digital Library.

Bioscience **2003**, 53(7), 631.

Mardis, M. A., & Payo, R. P. Making the school library sticky: digital libraries build teacher-librarians' strategic implementation content knowledge in science. Teacher Librarian **2007**, *34*(5), 8-14.

Mischo, W., Norman, M., Shelburne, W. A., & Schlembach, M. The growth of electronic journals in libraries: access and management issues and solutions. Science & Technology Libraries **2006**, *26* (3/4), 31.

Murphy, J. Information-seeking habits of environmental scientists: a study of interdisciplinary scientists at the Environmental Protection Agency in Research Triangle Park, North Carolina. Issues in Science and Technology Librarianship **2003**, 38. Available: http://www.istl.org/03-summer/refereed.html

Palmer, C. L. Information work at the boundaries of science: linking library services to research practices. Library Trends **1996**, *45*(2), 27.

Saylor, J. M., & Minton-Morris, C. The National Science Digital Library: an update on systems, services and collection development. Science & Technology Libraries **2006,** *26* (3/4), 61-78.