In the cycle of scholarly communication, scholars play the role of both consumer and contributor of intellectual works within the stores of recorded knowledge. In the digital environment scholars are seeking and using information in new ways and generating new types of scholarly products, many of which are specialized resources for access to research information. These practices have important implications for the collection and organization of digital access resources. Drawing on a series of qualitative studies investigating the information work of scientists and humanities scholars, specific information seeking activities influenced by the Internet and two general modes of information access evident in research practice are identified in this article. These conceptual modes of access are examined in relation to the digital access resources currently being developed by researchers in the humanities and neuroscience. Scholars’ modes of access and their “working” and “implicit” assemblages of information represent what researchers actually do when gathering and working with research materials and therefore provide a useful framework for the collection and organization of access resources in research libraries.

Introduction

The dialogue among researchers and professionals concerned with the future of scholarly communication is often focused on shifts in how research articles are being published and disseminated. Undoubtedly, electronic publishing on the Internet is quickly changing the format and means of distribution of scholarly works, but there are other dimensions of the scholarly communication cycle that are also changing in important ways. If we conceive of scholarly communication as the entire set of activities involved in scholarly exchange, the resources that assist researchers in identifying and finding publications and other sources of information are clearly of great consequence. The access resources used by scholars—the indexes, directories, bibliographies, and databases, also known as reference or tertiary sources—are of particular interest because they can mediate paths of research by encouraging certain directions and introducing barriers to others. Thus, the shift to digital production of access resources may prove to be more profound than the advent of electronic journal and book publishing. After all, access resources provide the infrastructure that makes primary and secondary publications searchable or able to be gathered as raw materials for research in the first place.

Scholars generally gather large amounts of data and literature before they produce new findings or write a report on their research, and the Internet is influencing what information comes into play during this process of scholarly production. At the same time it is fostering the development of new resources for information access, a trend in which scholars are actively participating. As scholars increasingly rely on digital means of access and develop digital access resources for their own purposes, the scholarly communication landscape is becoming more variegated. The influx of these scholarly practices and products introduces complexity into the communication system, while also adding valuable new resources that are closely aligned with the ways research is conducted in different fields of study.

In this article, researchers’ digital information seeking, digital resource creation activities, and the implications of these practices for managing access resources in research libraries are examined. In accordance with Kling and McKim (2000), the aim is to understand how the roles and structures of electronic media vary from field to field, and, in keeping with their social shaping of technology (SST) perspective, demonstrate how scholarly technologies are configured and reconstituted by the communities that produce, implement, and use them. Understanding the social shaping of access resources is of particular importance because these tools of scholarship have a high degree of cognitive authority (Wilson, 1983) and act as intermediaries between scholars and our vast bodies of recorded knowledge. The influence of reference sources on research and professional practice has been studied (see, for example, Beaulieu, 2001; Bowker & Star, 1999; Myers, 1991), yet it is rarely considered in current analyses of the changing state of scholarly communication. Guédon’s (2001) historical discussion of
how the list of core scientific journals represented in the Science Citation Index (Thomson Corporation, 2005) “took on a life of its own” is one notable exception.

As background on the role of access resources in scholarship, I begin with an overview of some legacy reference resources, emphasizing the importance of their connective structures, and the current landscape of digital access that includes a range of new digital scholarly “communication forums” (Kling & McKim, 2000). I then draw on a series of qualitative studies I have conducted on the information work of scientists and humanities scholars to document specific information gathering activities influenced by digital access resources, identifying two general modes of information access evident in researchers’ information practices. To further explicate these conceptual modes of access, I examine them in relation to exemplary digital access resources being created by researchers in the humanities and neuroscience. The analysis demonstrates how researchers are currently working with and assembling digital research materials to advance their research. The conceptual models derived from the analysis provide a framework for research libraries to use in assessing access resources for addition to digital collections and for coordinating resources to support different fields of research.

Background

Access Resources and Their Connective Structures

In the past, many of the access resources available to researchers were bibliographic in nature, listing references to formally published or physically collected information. For the purposes of this discussion, it is helpful to distinguish between the resources generated through institutional efforts and those produced by scholars for publication. Institutions, such as commercial indexing and abstracting services, professional societies, libraries, archives, and other kinds of research repositories have typically developed products like indexes to journal literature, catalogs to library collections, finding aids to archival holdings, and databases with search engines. Scholars have developed access resources of a different sort—literature reviews, bibliographies, and other edited and derivative works such as handbooks and research guides. Resources produced by scholars take a more focused and selective approach to identifying research materials. They represent the expertise and judgment of the authors and are subject to a greater degree of social shaping by scholarly communities (Oxman & Guyatt, 1993; Palmer, 2001a). Works that are listed in catalogs or indexes are given visibility and recorded for posterity; those listed in scholarly reviews and guides are further distinguished as worthy contributions to a field of study.

Access resources are the means by which researchers identify recognized scholarly works within and, sometimes more importantly, outside their field of research. Standard systems of access have been developed for certain segments of scholarly exchange. For example, discipline-based journal indexing services in the sciences (e.g., Chemical Abstracts, American Chemical Society [ACS], 2005) served as a kind of blueprint for the development of similar resources across and beyond the sciences. As indexing products for research audiences proliferated, their scope of coverage and format remained much the same. That is, they tended to cover established journals using standard bibliographic records, usually supplemented with domain-specific descriptors and possibly abstracts. The journal index framework became ubiquitous across disciplines, but played a less important role in the humanities, where journals account for a much smaller percentage of the research literature. In the humanities, research library catalogs, bibliographies, and archival finding aids have been the main tools for managing access to extensive subject collections of primary materials and secondary book sources.

The connective structures, which either are embedded in or accompany these access resources, have been an important aid for navigating the terrain of scholarly communication. I use the term connective in a general sense in this discussion to refer to syndetic structures used in the practice of professional indexing as well as other manual and automated means of documenting relationships within the research literature. Connective structures link related information within indexes and catalogs, often through cross-references and listings of broader and narrower terms for descriptors or subject headings. They extend and mediate the path of information access beyond a given item to other potentially relevant documents, helping the flow and exchange of information among scholars.

In the world of print, it has been very difficult to represent the expanse of intellectual relationships within an access resource. Bibliographic indexes often made relationships between separate subject areas explicit through a system of “see” and “see also” notes, but in application, the techniques provided limited cross-referencing and sometimes conflated associations in ways that were counterproductive for researchers. For example, in one standard bibliographic guide produced throughout the 20th century, scholarly texts related to women were grouped in the same category as charm school and etiquette books, with no references back to the main bodies of associated scholarship in history, political science, and other academic fields (Palmer & Malone, 2001). Of course, when catalogs and databases became automated, connections could also be determined automatically through statistical measures of association, and with the Internet individual sources can now be interlinked directly.

The citation indexes produced by the Institute of Scientific Information (ISI, now Thomson Corporation, 2005) made an invaluable contribution as a tool for tracing scholarly connections and paved the way for citation analysis as a technique for mapping scholarly communication trends, such as the influence of specific papers, the growth of literature in a field, and interactions among researchers and domains. Citations provide tangible evidence of scholarly communication; they can be used as a guide to understanding structural relationships among researchers (Borgman, 1990; Borgman
& Furner, 2002; Callon, 1986). The structures created by citation linkages differ, of course, from the interpersonal networks formed through direct interaction between scholars (Crane, 1972; Lievrouw, 1990). The informal connections between researchers who actually talk to each other represent what might be called working scholarly communication networks. Scholars also have real working connections with the literature and research materials they engage with to produce new scholarship. On the other hand, the intellectual connections that can be traced through citations among bodies of literature represent more implicit networks, akin to what Pahre (1996) refers to as “metaphorical communities”—communication networks that emerge around ideas and concepts but do not form true communities of practice. Through their connective structures, standard access resources have generally documented this type of coarse metaphorical information network. As will be discussed later in this article, some new scholar-created access resources support working networks of information and others support implicit networks.

Digital Aggregations as Access Tools

Before the development of full-text digital databases, finding the right information usually depended on retrieval from access resources consisting of short descriptive records. Scholars found publications by manually sorting through records in bibliographies or reference books, or by searching in online catalogs and indexing databases. However, when repositories of texts and other source materials are made digital and searchable they are transformed into access resources. To some extent, in the digital environment any searchable grouping of information sources can be considered an access tool, and ideally search engines can make searching within and across tools automatic and fluid. Thus, the development of digital resources on the Internet opened a floodgate of access channels, but development of connective structures has been more limited.

Large-scale access systems are still prominent in the scholarly communication landscape. In fact, the digital environment has fostered the expansion of the access unit through bundling, federation, and interlinking of resources. For example, the three separate citation indexes for the sciences, social sciences, and the arts and humanities are now offered as one merged Web of Knowledge (Thomson Corporation, 2005). The aggregator service ScienceDirect (Elsevier Science, 1999), which bundles electronic journals for distribution, is a particularly interesting case because of its scale and potential impact on the scholarly communication cycle. With searching capability across over 1800 journals, its scope is remarkable. Moreover, this product has begun to incorporate standard access resources such as bibliographic databases and reference works into its base, moving toward an aggregation model that is likely to include formal access tools produced by both institutions and scholars. The effects of bundling on both the economic and intellectual marketplace are beyond the scope of this article; however, it is important to note that this business model nullifies the criteria previously used by research libraries for journal acquisitions. Where decisions were once based on assessments of demand, use, and quality of individual journal titles, choices are now largely driven by the availability of a limited number of all-or-nothing packages (Bell, 2001; Frazier, 2001; Packer, 2001).

Professional societies, that at one time dedicated most of their publishing activities to the production of journals, are creating new kinds of sources and aggregations that function as access resources. For example, the American Chemical Society’s Journal of Chemical Education now provides Internet access to a collection of chemistry software and a service called JCE Online (ACS, 1999) that includes interactive articles, discussion forums, data resources, and “Webware.” Collaborative efforts supported by funding agencies and advocacy organizations like the Scholarly Publishing & Academic Resources Coalition (SPARC), an alliance of institutions encouraging new business models for the scholarly communication market, are resulting in other combinations that blur commercial/noncommercial distinctions in access resource development. For example, BioOne (BioOne, 2001) is a journal aggregation service being jointly produced by scientific societies, libraries, academe, and publishers.

Digital access resources developed in research institutions are taking many forms, ranging from digital special collections, to research library gateways, and preprint servers—all of which are playing an important role in access to scholarly information. The self-archiving movement taking hold at universities is another important shift in management of access to research information. Researchers and their institutions are recognizing the value of maintaining control of research output, instead of or in addition to commercial dissemination. The institutional repository, along the model of Ginsparg’s (1991) physics preprint server or MIT’s DSpace (MIT Libraries, 2002), has been widely cited as a viable approach to the distribution of scientific results. Discipline-based resources are still common, but they are now likely to be multipurpose, such as CogNet (MIT, 2000), a SPARC supported “electronic community” for the cognitive and brain sciences, or customized for specialized areas of research, as with the four “collaboratories” on fear, aggression, affiliation, and reproduction being developed by the Center for Behavioral Neuroscience (2004). Many of these digital resources are now being further aggregated into centralized services such as the National Science Digital Library (National Science Foundation, 2001) and the range of repositories being developed based on the Open Archives Initiative Protocol (see Brogan, 2003).

Scholar-created resources are being included in these aggregations, and they include new genres of digital scholarship that differ from the reference works scholars contributed to the pool of access resources in the past. The digital archives, data repositories, collaboratories, and other Internet-based resources being produced by scholars are often direct corollaries to their programs of research and reflect the information needs of their creators. They tend to be tailored to the specific cultures of research communities, rather than adher-
ing to existing blueprints for access resources. Two types of scholar-created access resources will be discussed below, but first we will examine scholars’ use of existing digital resources. After all, while digital resource creation by scholars is widespread and increasing, the percentage of scholars involved is small compared to those who regularly access digital materials in their daily research work.

Digital Access by Scholars

The scholarly communication environment is becoming richer and more complex, with many people, organizations, and collaborative groups adding to the new digital mix. This is a relatively recent situation, and we still have much to learn about how scholars are actually using new scholarly communication forums. In the case of physicists and the arXiv.org preprint server (Ginsparg, 1991), the technology has been adopted relatively quickly because its function closely matches scholarly communication practices of that particular scientific community. In physics, preprint exchange was established long before it migrated to the digital medium, and there was early adoption of electronic documents as the preferred mode of production, in a large part because they were so easy to exchange. Nevertheless, Pinfield (2001) has found that the online preprint service is used primarily by the physicists who contribute to it and that it has not replaced formal journal publication, but instead serves as an additional means of distribution.

There is no doubt, however, that digital resources are being heavily used by scholars. The use of digital libraries has been widely studied and a recent review of that literature presents a comprehensive summary of both large- and small-scale studies (Tenopir, 2003). The longitudinal survey research by Tenopir, King, and colleagues (see, for example, King, Tenopir, Montgomery, & Aerni, 2003; Tenopir et al., 2003) has made a substantive contribution to our understanding of trends in electronic journal use. Their studies show, for instance, that levels of reading by university faculty members have increased with the availability of digital journals—with the number of readings rising more than time spent reading. This is an important distinction since qualitative studies have indicated that researchers are scanning or exploring more information and getting exposure to new sources online, but not necessarily “reading” more articles (Brockman, Newmann, Palmer, & Tidline, 2001; Palmer & Neumann, 2002). The high level of interaction with digital resources has been well documented, but we still have a limited understanding of how these materials impact the conduct of research.

Many studies in the field have examined research practices within specific disciplines (e.g., Case, 1991; Cobbledick, 1996; Cole, 1997; Hallmark, 2001; Hertzum & Pejtersen, 2000; Hurd, Blecic, & Vishwanath, 1999), or investigated general patterns of information behavior across the humanities and the sciences (e.g., Brown, 1999; Wiberly & Jones, 1989). However, this article is most closely aligned with the body of literature that broadly compares the information practices of researchers in the humanities and sciences (e.g., Bates, 1994; Covi, 1999; Ellis, 1993). It is a general comparative analysis derived from four previous studies investigating the information practices of a broad base of researchers in the humanities and sciences. Accordingly, the results are intended to inform the research library audience that serves the large and diverse population of scholars in academia.

Research Approach

This comparative analysis cumulates the results of four earlier studies to identify patterns of information seeking and use in the humanities and sciences and the specific ways digital access resources are contributing to scholars’ research processes. In the first study (Palmer, 2001b), face-to-face interviews were conducted with 23 interdisciplinary scientists in the biological, physical, behavioral, and computer sciences. Bibliometric analyses were also used to provide a detailed description of each scientist’s research domain and the breadth of the audience for their work. In the second study (Palmer & Neumann, 2002), we performed phone interviews with 25 interdisciplinary humanities scholars from a range of fields including literature, media studies, cultural studies, film, history, architecture, art history, and communication. Face-to-face follow-up interviews were conducted with 10 scholars to further elaborate on specific information activities discussed during the initial interview period. For these cases, a supplementary bibliometric analysis was also conducted to track the flow of subject areas into a scholar’s work and to compile a more complete inventory of the types of research materials used. For the third study (Brockman et al., 2001), two rounds of face-to-face interviews were conducted with 33 humanities scholars. In addition, longitudinal case studies of ongoing projects were developed for five of the scholars. Case data included additional “critical incident” interviews, observations, and analysis of documents. In the fourth study, an ongoing study of neuroscientists (Palmer, Cragin, & Hogan, 2003), we are developing case studies of 12 projects at three different laboratories. Data collection techniques include face-to-face and telephone interviews, diary entries of information activities, observation, and document analysis.

Each of the four studies was conducted using multiple methods of data collection. However, the interviews were the primary data source for the analysis provided in this article. Semistructured interviews were used across studies to generate rich, local data on the research process at a level that brings important aspects of information work into view, such as the things that make information useful, difficult to find or use, and the social linkages by which information and knowledge move and coalesce. Our approach in all four studies has been to investigate the broad range of information activities involved in researchers’ daily work, not just their use of digital resources. Following grounded theory processes of open and axial coding, we have interpreted the interview data through iterative, comparative analysis.
Confirmation and Discovery Searching

In the ongoing process of research, confirmation and discovery searching are often intertwined. Confirmation searches can lead directly to discovery searches and vice versa. Humanities scholars regularly search online catalogs and journal indexes to confirm their ideas about, and their memories of, work done in their research area. These searches may be for known items or for basic current awareness purposes, but they are also done as a way to refresh the researcher’s thinking and reaffirm the track of their research. There is high reliance on the speed and scope provided by networked digital access, and general Internet searches are regularly used to supplement searches of catalogs and indexes, but familiar print resources are still used, as well. Scientists also perform confirmation searching, but it tends to be more directed at a specific question or problem at hand. Moreover, they depend more fully on online access for both searching and accessing papers, using print sources only when absolutely necessary.

Many interdisciplinary scholars and scientists probe outside their core research area for information that is intellectually distant or from unknown sources. This kind of discovery searching is usually done through exploration of bibliographic databases and more general Internet searching. Digital access makes it easy for scholars to extend their field of exposure by exploring and acquiring information more freely, but as they draw from multiple domains to expand their intellectual base, the new ideas, theories, methods, and terms they encounter create gaps in understanding. They then need to perform further confirmation searching to interpret, verify, and anchor the new material.

In the humanities, there is a growing appreciation of the ability to interrogate the full text of large corpora, especially in literary, linguistic, and cultural fields of inquiry. Patterns discovered within bodies of electronic texts can serve as observations or “findings” to guide the formulation of research questions as well as the interpretation and analysis of texts. The ability to quickly calculate and locate word occurrences in collocated texts is influencing expectations about how research should be conducted. In a few research communities there appears to be an unstated obligation for scholars to process full-text collections in certain ways before proceeding with a project. Where once scholars were required to know about the relevant preceding work on their research topic, they may now be expected to identify patterns or themes in a body of texts. These expectations are not widespread however and seem to be taking hold slowly, considering that Ruhleder (1994) identified similar influences about a decade ago in her study of the impact of the Thesaurus Linguae Graecae (TLG) and computer-based textual analysis on scholarship in classics.

Collecting

For many humanities scholars, key primary and secondary materials are central sources of evidence, and research is conducted through interaction with these materials. Humanities scholars prize their own personal collections of research materials as well as the library collections at the institutions where they work. Their personal collections are a necessity since rereading is a significant part of their interpretive work. Any number of texts may require periodic or systematic reading, and some may be “read” for years or decades. Interestingly, local digital library collections tend to be perceived along with the Internet as one big digital blur of information, quite separate from personal or physical library collections (which may also include databases and texts accessible on computers). Interactions with information on the Internet are less informed in that scholars have little initial understanding of the provenance, surrounding collections, and cognitive authority of the materials found.

Digital access tools ease the process of locating texts; in cases where digital versions are available and acceptable for a scholar’s purposes, they may be acquired online. Currently, only a small proportion of humanities research collections have been digitized and made available on the Internet. On the other hand, traditional access resources for humanities research collections, such as manuscript catalogs and archival finding aids, are more widely available online; they have become important for preparing for research trips. Scholars regularly travel to work with and copy materials in distant libraries, archives, and museums, sometimes spending weeks, summers, and entire sabbaticals at a site, perhaps returning multiple times. Most trips require extensive preparation, and the ease and amount of work done in advance is increasing with the availability of these digital resources.

Most scientists have substantial collections of texts in the form of journals, conference proceedings, and photocopied
and PDF articles; these sources are not the focus of inquiry in their research. Nonetheless, scientists highly value the ability to collect digital papers and access them quickly on their desktop or laptop computers. The collections that are the focus of inquiry consist of the data they generate through experimentation, modeling, and laboratory study. As these data have become increasingly digital in format, some researchers are keenly interested in collecting and federating datasets for access and reuse by larger communities of scientists. A number of the scientists participating in our studies are involved in these kinds of informatics initiatives, but others are only interested in working with the data they produce locally.

**Consultation**

In conducting research, access to people and their expertise can be as important as access to literature and data. Personal communication networks are a type of access resource that has been widely acknowledged as a primary support structure for research communities in the sciences (Crane, 1972; Hertzum & Pejtersen, 2000; Price, 1963). In the humanities, where collaborative research teams are rare and scholars write almost all articles and books independently, interpersonal processes have not been considered a significant part of research. But, while it is true that humanities scholars usually work alone when reading, searching databases, browsing collections, and writing, they also frequently work with colleagues in a consultative manner, sharing citations, ideas, and drafts of papers. Of course, networked information systems have made it easier for researchers to communicate and gather information informally, but what is of particular interest in the digital environment is how often these transactions are with strangers rather than colleagues.

Scholars can readily identify kindred researchers and centers of research through the Internet and send direct queries via e-mail. As a result, they are making more cold inquiries, which appear to be paying off with advice and access to valuable information sources. For example, one humanities scholar recounted finding information on the Internet about a researcher who was working in a field she wanted to know more about. She immediately sent an e-mail asking if there was an area of linguistics relevant to her research topic. An affirmative reply came back the same day with an attached file containing a 500-item bibliography related to her area of interest.

The Internet is fostering short-term encounters that require relatively little effort but with potentially high payoff in access to new information. Elsewhere we have characterized these personal networks as “invisible constituencies” (Palmer, 2001b). They are more open, loosely organized, and heterogeneous than invisible colleges, and network associates tend to function more as ad hoc consultants than colleagues or critics. This is especially true in interdisciplinary areas of research that have not matured to the point where there is an established core of prominent experts. Invisible constituencies will likely become more common over time through the virtual connections growing among research groups, adding a new level of collegial support to the process of scholarship.

**Modes of Access**

The increased mobility of digital information has made access easier and faster while also increasing the universe of possible sources of evidence, for both scientists and humanities scholars. Nearly all of the researchers studied made consistent use of digital tools to extend the range and sometimes the amount of searching (for confirmation and discovery), collecting, and consulting. But, it is important to understand digital information use within the context of the larger research process. There are clear distinctions between the fields in how information is gathered and applied to fuel the process of inquiry. Figure 1 provides a

---

**FIG. 1.** Two modes of information access.
conceptual sketch of the different contours of access associated with the research practices of the humanities scholars and scientists studied, referred to as Mode A and Mode B, respectively. The modes represent distinguishable patterns in how scholars search, collect, and consult both online and offline; thus, they help us to better understand general variations within the system of scholarly communication. Of course, the differences between the two have much to do with scholars’ varying research questions, methods, and sources of data and evidence, and therefore there is crossover and dual use of the modes among scholars in the humanities and sciences.

Humanities projects often center on specific primary sources, archival files, or the writings of key authors or theorists, but once this core is established research takes an interpretive course where many other information sources are sought, consulted, and brought together as the core materials are studied, analyzed, and historicized. What has been most striking in our studies is the mutable intellectual course humanities scholars take as they search for background information and evidence connected to the core of a project. Paths of inquiry are long, fluid, and often overlap from project to project. Humanities scholars follow trails of leads, similar in pattern to the “berrypicking” model of searching identified by Bates (1989). A course of study may extend over years, possibly decades. A scholar may physically travel around the country, and perhaps the world, in search of information that could conceivably end up as a footnote in a journal article. Certain manuscripts or artifacts in foreign archives may be vital, but often no more so than a popular culture Web site that provides a new bit of evidence or perspective on a topic. A network of institutions is formed through the physical and virtual travels of humanities scholars as they follow their information leads. The access tools they use along the way provide important context for the many pieces of information collected. For example, a digital finding aid allows a researcher to “see” the group of works systematically collected around a particular item of interest—a context in which the next important lead may be found.

The long meandering path followed in seeking and collecting information in the humanities is different from what we have observed in our studies of scientists, who have more routine patterns of information gathering from targeted sources and domains. Work is problem centered and more bounded in time and scope. Data are generated locally, and the additional information that is brought to the problem is often to assess a preliminary finding, work out methodological or instrumentation problems, or document the recent related literature. Background literature searching tends to be deep and directed. Naturally, the search for information may at times be broad, exploratory, or reach far afield, but the process remains directed at solving a distinct problem or testing a specific idea. Furthermore, scientists’ use of older information sources is limited, both because of the need for recent scientific knowledge but also because of a reluctance to seek out print sources.

The two modes of access have distinct emphases. The work of humanities scholars is centrifugal; they work their way out into the information universe to collect a base of information in which deep inquiry can be performed through reading and writing. The information gathering process is relatively open-ended, moving outward from lead to lead. The work of the scientists is centripetal with information being pulled back to the locally generated data and results to solve the problems and questions encountered in each stage of an experiment or project. Searching, collecting, and consultation are more targeted and endpoints tend to be more defined.

**Digital Assemblage by Scholars**

Researchers have begun to create digital resources for themselves and their research communities, many of which qualify as access resources or tools. These works are significant as new contributions to the cycle of scholarly communication, but they are also important as indicators of how scholars wish to engage information technology in their research. There are, of course, many valuable scholar-created resources being added to the system of scholarly communication. The two types discussed here, thematic research collections and literature-based discovery tools, have been selected because they have emerged as important genres in the research communities we have studied and are examples of technologies for the average scholar, not the informatics specialist. Thematic research collections bring together research materials into a kind of virtual laboratory for a specific area of inquiry. According to Unsworth (2003), this new genre of scholarship reaches a larger audience than the typical humanities monograph; therefore, it could replace the book as a primary product of humanities scholarship. Literature-based discovery is a data mining approach for identifying new information or scientific relationships in the literature. The particular technique discussed here stands out from many of the large-scale, data-driven informatics approaches in that it supports traditional hypothesis-driven research practiced by a wide range of laboratory scientists (Smalheiser, 2002).

**Thematic Research Collections**

Thematic research collections, hereafter referred to simply as thematic collections, are digital aggregations of primary sources and related materials that support research on a theme (Palmer, 2004). The materials in a thematic collection may come from any number of research libraries, archives, and museum collections, with the core usually consisting of primary texts or images, supplemented with a range of other materials such as critical essays, reviews, biographies, and bibliographies. Many thematic collections have been created to support research on people—artists, authors, historical figures—such as the *William Blake Archive* (Erdman et al., 1997) or *The Complete Writings and Pictures of Dante Gabriel Rossetti: A Hypermedia Research*
Salem Witch Trials Archive (McGann, 1997). However, any event, place, phenomenon, or object of study can serve as a theme, as with the Salem Witch Trials (Ray, 2002) or the narrowly focused Hamlet on the Ramparts (MIT, 200?), which brings together texts, art, photographs, films, sound recordings, and commentary related to Hamlet’s first encounter with the ghost. The content of these collections is selective, with many scholars going to great lengths to develop solid working relationships with the libraries, museums, and other partners in digitizing or acquiring the materials and to ensure quality transcriptions and reproductions. For instance, the digital reproductions of the illuminated works in the Blake archive are more accurate than the texts and images in existing printed editions (Viscomi, 2002).

While often referred to as electronic archives by their creators, thematic collections differ in important ways from institution-based digital archives and traditional archives. They are dense aggregations of research materials, customized to support intensive study and analysis in a research specialization. However, scholars are not only constructing environments where research materials can be accessed more conveniently by more people, they are also performing their normal scholarly role of creating research products that advance the state of scholarship in the field. Like other scholarship in the humanities, research takes place in the creation of the work, and research is advanced because of it. In some cases, the technical design and functionality are important aspects of the research contribution, similar to the research products generated in fields like engineering, computer science, and information science. However, for many projects the primary goal is the creation of a new kind of multimedia scholarly edition (Hansson, Francke, Dahlström, & Gunnarsson, 2003). For example, the Rossetti project is aimed at combining the attributes of facsimile and critical editing to allow the original physical condition of a text to be represented and studied (McGann, 1996).

In these collections we can see what scholars perceive to be key content—primary source materials and related contemporary and secondary sources that fill out the historical and cultural context of the object of study. We can also see a composite of the distributed, dynamic nature of humanities fields, which are “concerned with the construction of knowledge from sources of different types, scattered across different subject areas” (Fraser, 2000, p. 274). For example, Monuments and Dust (Levenson, Trotter, & Wohl, 1997), a thematic collection focused on Victorian London, is an aggregation of diverse sources—images, texts, numerical data, maps, and models—with content of interest for scholars in literature, architecture, painting, journalism, colonialism, modern urban space, and mass culture. In this early stage of development, thematic collections are primarily supporting the searching and collection activities discussed above, by bringing together the distributed content that would be of interest along the research path of a scholar working on a given theme.

Thematic collections represent the centrifugal searching and collecting of Mode A information access. They function well as access tools for research materials, but there is still much development needed to create true virtual laboratories for the humanities. While our studies show the importance of interpersonal interaction in the humanities research process, features that support consultative or collaborative work are not yet part of most thematic collections. Moreover, unlike the “collaboratories” being built in the sciences (Finholt, 2002), many thematic collections are still not adequately equipped with tools for working with and analyzing the collected content. However, humanities computing efforts are turning toward development of technologies for basic scholarly activities, such as annotating, comparing, and reading texts (Crane, Smith, & Wulfman, 2001; Unsworth, 2000). Reading activities are particularly complex and require much more than display and printing capabilities. Scanning a text is a different activity than rereading it deeply and repeatedly over time, and both activities are part of a larger pattern of wide reading and collecting that needs to be supported to make thematic collections more effective sites for conducting research (Brockman et al., 2001).

**Literature-Based Discovery**

Literature-based discovery is a technique for mining new knowledge or identifying implicit relationships in literature databases. There are a number of technologies being developed for literature-based discovery (e.g., Hristovski et al., 2003; Sehgal, Qiu, & Srinivasan, 2003), but the tool to be discussed here, Arrowsmith, is a direct extension of Swanson’s (1986b) pioneering theoretical work that predated current computational efforts. Swanson’s approach to scientific discovery is based on the idea that new solutions to scientific problems can be derived from literature in the public record. More specifically, two sets of articles that are disjoint (contain no articles in common) can be complementary, suggesting new information of scientific interest not apparent from either set of articles taken alone (Swanson, 2003a). Swanson’s (1986a, 1987) first discovery using this technique with MEDLINE searching was a link between fish oil and Raynaud’s disease. The articles on each topic did not cite or mention each other, but when reviewed together the information suggested that patients suffering from Raynaud’s disease might benefit from dietary fish oils. This finding was later supported through clinical testing by scientists. Swanson’s next publication reported on a series of indirect linkages in the literature on migraine headaches and magnesium deficiency (Swanson, 1988), and this connection was also later corroborated in laboratory tests.

Over the past decade, Swanson has collaborated on additional discoveries and the development of the Arrowsmith technology with Neil Smalheiser, a neuroscientist (Smalheiser & Swanson, 1994, 1996, 1998). Within the field of neuroscience there is an acute awareness of the need to develop new technologies to mobilize and synthesize the extensive body of accumulated knowledge to promote new discoveries (Amari et al., 2002; Cragin, 2003; Shepherd et al., 1998). Smalheiser is directing an initiative to refine
material allows them to build a circumscribed, high-quality assemblage continuum. For example, the online component of the system in neuroscience laboratories.

Arrowsmith is a general-purpose access resource that is currently being applied to one large-scale database, MEDLINE, unlike a thematic collection (more specialized, a thematic collection collocates many previously distributed sources). The scope and diversity of the topics covered in MEDLINE are major strengths of the system, since Arrowsmith has been designed to locate information on targeted topics produced in separate scientific specializations, a process that fits well with the Mode B approach to access. The system retrieves, sorts, and presents titles of journal articles to allow the researcher to assess their fit or contribution to the research problem under investigation. It does not determine complementarity—that is left to the researcher (Swanson, 2003a). The expectation was that Arrowsmith would be used primarily for discovery purposes, but the field tests with neuroscientists have shown that the system is useful for a number of different kinds of searching (Palmer et al., 2003). It is used by the scientists to assess new findings in their lab relative to other results reported in the literature and for exploratory and directed searching both within and outside their specializations. The way that Arrowsmith filters and displays sets of records and identifies implicit links can add value to basic literature review searching, essentially bringing a discovery dimension to confirmation searching.

Literature-based discovery may eventually be able to play additional roles in the process of scientific discovery. For example, Swanson (2003b) is exploring ways to use MeSH hierarchies and subheadings to systematically produce plausible hypotheses, and Srinivasan (2003) is working on a different approach to hypothesis generation. Our team has also been experimenting with supercomputing approaches to identifying research groups with high collaboration potential. Collaborative relationships usually grow out of scientists’ informal personal networks, and the Internet and e-mail have made it much easier for researchers to locate, communicate, and work with distant colleagues. But, it is possible that by relying on personal networks and ad hoc searching researchers are missing important connections with others working in complementary fields. A discovery system or service for identifying such links would contribute to Mode B information access, providing a means for targeted contact with researchers who could add the expertise needed for innovative solutions to specific research problems.

Assemblage Networks

The two types of resources discussed above reflect scholars’ different conceptions of how digital information can best be assembled to advance their research. For the humanities scholars developing thematic collections, digital source material allows them to build a circumscribed, high-quality research base for deep, iterative study. For the neuroscientists, digital bibliographic data lends itself to mining pieces of knowledge about the brain and assessing possible associations. The resources assemble information in ways that support the research practices and modes of access of the constituent research communities, and the connections among the information assembled relate to the working and implicit scholarly communication networks introduced in the Background section of this article.

Thematic collections exemplify a working assemblage. They instantiate the work of finding and collecting materials of importance for study of the theme. The items are purposefully selected for inclusion because they are logically linked in the path of research activity. When brought together they form an assemblage that supports ongoing research work in the thematic area. Essentially, the collections contain the information that would have been gathered and worked with by visiting or borrowing materials from many libraries and archives. An Arrowsmith assemblage is not a working network of information, since materials are not connected through actual research activity. The system suggests connections between noninteractive literatures, bringing together articles that are, ideally, only connected in implicit intellectual space. The assemblage of sources is produced by the user, not the creator of the resource, in response to their hypothesis or query. Rather than capitalizing on the chain of interactions that unfolds in the process of research, the discovery technique leverages the uncoupled nature of scattered scientific knowledge.

If access resources representing working assemblages are on one end of a continuum and implicit assemblages on the other, as illustrated in Figure 2, few resources would be located at either of the two poles. A working resource would assemble known sources that are connected through research activity, and an implicit resource would assemble unknown sources that are linked only through intellectual association. An archive of a closed listserv for an established research group might sit on the far working end. All users would be aware of the associated researchers and topics, and searching would be done largely for known information or confirmation purposes. Large-scale data mining systems like Arrowsmith would be situated on the implicit end of the continuum, since they support assemblages of information from outside a working community of practice, and searching is done primarily for discovery of unknown connections. Invisible constituencies, discussed above, reflect an emergent structure that would be likely to have implicit and working characteristics. Scholars may have no working or professional relationship with the unknowns they make contact with, and their interactions may be for discovery purposes, but some of the people contacted will be part of an extended network of colleagues or will soon become part of a loose working consultation network.

Other digital access resources identified earlier in this article would also be placed toward the center of the assemblage continuum. For example, the online component of the Journal of Chemical Education (ACS, 1999) has
been developed for a specific disciplinary community but has a range of materials that would extend beyond the familiar working universe of many users and which could be of use to researchers and educators in other sciences. *ScienceDirect* (Elsevier, 1999) is multidisciplinary, general-purpose, and evolving into a meta-access resource as it adds reference sources to the journal aggregation. Because of its scope it has the potential to reveal implicit relationships for researchers (but the analysis would need to be done manually since discovery capability is not built into the system), therefore this resource would fall past the center toward the implicit pole.

Many resources, of course, can be manipulated by users for working or implicit purposes, and an assemblage that represents a working structure for one researcher may offer opportunities for discovery of implicit connections for another. Arrowsmith was designed to identify implicit structures, but it is also being used more flexibly as researchers perform searches for both known and unknown connections. Moreover, layers of structure can conceivably be added to any given resource to make working or implicit dimensions more functional. Tools can be implemented to exploit both types of structures in one resource. For example, providing reference linking to a narrow, specialized database opens it up for tracking connections beyond a working scholarly community. Adding a data mining tool to a thematic research collection would also produce this kind of hybrid assemblage.

Most researchers rely on a mix of these structures, at times finding it useful to access information that is strictly within their working scholarly community, and at other times exploring purely conjectural connections that extend outside their area of specialization. It is the job of research libraries to support both kinds of research work and organize the growing pool of access resources, or assemble the assemblages, in ways that allow researchers to easily engage in their modes of information access.

**Supporting Access in Research Libraries**

The access resources being created by researchers are scholarly products in their own right and need to be retained as part of the scholarly record. However, libraries are not yet systematically acquiring thematic research collections or literature based discovery tools, or the data archives, collaboratories, corpuses, or other digital resources being produced by scholars. This is, to some degree, because of the newness of the genres, but also because the place of these quasi-tertiary sources in the scheme of scholarly output is not yet well understood. As this layer of resources increases in mass, however, it will need to be managed as a service to researchers who can benefit from the content, structures, and digital functionality of the rich information therein.

Research libraries’ interests in access resources are different from the scholar–creator, who is much less likely to think of the works that they create within the larger multidisciplinary context of access to research materials. The scholar’s aim is to make research contributions that advance their field and their career, and the process of creating a digital resource is likely to be idiosyncratic to their particular domain and approach to research. The aim of the research library is to provide support for a broad range of academic communities through the collection and organization of many types of information. This is achieved, in part, through programmatic selection of access resources that fit the communities served and the existing base of source material. Each access resource that is selected and added to a research library collection becomes a component of a larger system of access, and each is assessed, stored, and organized for use within this larger context.

The growing layer of scholar-created access resources will require research libraries to take on added responsibilities in at least three areas: technical support for scholars, collection of scholar-bundled materials, and metaindexing of access resources. First, scholars need technical support for developing their own personal digital collections and access
resources for their research communities. Research libraries are already partners in many of these projects, and their expertise is needed in the development of this body of resources that stands to greatly affect how scholarly works are recorded and collocated in the future. Second, digital research collections are already being constructed, in part through the selection of commercially bundled materials, but now materials bundled in resources created by scholars also need to be integrated into these collections. Third, access resources of all kinds will need to be integrated into metaresources that organize their content and functions in ways that are useful for the research communities served by research libraries.

Currently, many research libraries are focusing on global approaches to digital collection building by producing expansive gateways for their entire user base. At the same time, researchers are creating their own tools and repositories, customized to their scholarly work. The organizational gap between these two requires systematic collocation and presentation of the new genres of access resources, which will by extension organize the source materials contained within these resources. For example, when the Rossetti thematic collection is made available by a research library it extends access to the entire corpus of assembled Rossetti materials. The term metaindex has been used to refer to resources that organize groups of digital access resources, such as the WWW Virtual Library Index of Web Site Indexes (Reitz, 2002). Bibliographies of bibliographies are a standard type of reference source that also provides a metaindexing function; the digital aggregations and bundles discussed in the Background section of this article are types of metaindexes as well, since they aggregate separate bodies of digital material to create a larger unit of access. A “metaindextical” is a different notion, posed by Henderson (1999) based on her study of design engineers and the role of visual representations in their work. She uses the concept of the metaindextical to articulate how assemblages of information are essential in design work. A metaindextical functions as a “gathering ground” for many kinds of information and has the capability to accumulate and generate multiple forms of knowledge. It is a structure that allows information to be both stored and manipulated to perform intellectual work. In the visual culture of design engineers, Henderson found that the mixed practice of working with both digital and analog media is central to the metaindextical function. Even as computers became common in engineering work, paper drawings remained necessary for the communication and teamwork involved in the design process.

As libraries work to aggregate and organize access resources, they will be designing metaindexes in the conventional sense, but the metaindextical criteria posed by Henderson can be used to guide the design of metaresources that more fully support the work of researchers. Metaindexes are gathering grounds that: (a) contain diverse information, (b) provide an accumulated base of knowledge, and (c) support the generation of new knowledge. The two types of scholar-created access resources discussed here can make substantive contributions to meeting these aims. In terms of the first criterion, they can harness diverse information in terms of either sources or subjects covered, or both. Concerning the second criterion, thematic collections are essentially working knowledge bases, accumulated and customized for intensive study and analysis. The literature-based discovery tool assembles an implicit knowledge base to identify possible scientific relationships. Both types of assemblages support activities, such as interpretive reading or hypothesis testing, that are important antecedents to producing new knowledge, the third criterion. However, no one resource is likely to meet all three criteria for any given group of researchers. That is the role of metaindexicals that strategically combine access resources to cover the range of activities practiced by a given research community. An effective metaindextical would look quite different from the current research library gateway. Instead of prioritizing top scholarly journals, major indexes, or a general set of reference materials, it would organize constellations of high-quality access resources, many of which will have been created by scholars and thus represent their specialized information needs and modes of working with that information.

Conclusions

Scholars are creating new networks of research materials and collegial support in the digital realm as they search for information, collect research materials, and consult with other scholars. They are also building valuable new digital resources to aid in their research. As these trends in access and assemblage continue, there will be a growing need for research libraries to provide services to assist researchers in these activities and integrate the digital resources they create into the new complex of digital access. The resources being created by scholars are an essential counterpart to the massive e-journal bundles, preprint servers, and institutional repositories under development. As Kling and McKeim (2000) demonstrated, scholarly communication development should not be based on vague notions of publication that conflate the varying activities and interests of intellectual communities. To effectively support researchers, digital library development will need to take into account the many types of research information and their divergent roles in different research fields (Agre, 1995; Borgman, 2000).

Tools created to support scholarly communication should represent the variant and vital ways that researchers gather and use information, not reestablish some minimal number of successful disciplinary constructions. However, an influx of highly specialized resources can also work to fracture scholarly communication. Just like the legacy print bibliographic systems described earlier in this article, new systems for organizing the intellectual work of scholars also have the potential to isolate segments of knowledge. For example, as thematic collections evolve into scholarly products with high cognitive authority, they may obscure intellectual relationships to domains not encompassed in a given resource.
And, while search engines can help users identify related sources, they cannot provide an information environment capable of accumulating and generating new knowledge. Metaindexicals can provide the necessary connective structures for context-rich access to research materials and a place to perform many important research activities.

As scholars populate the Internet with high-quality access resources, they are customizing and bringing clarity to the part of the digital blur they know best. Metaindexical resources will configure these individual refinements and improve researchers’ ability to navigate and work in the increasingly intricate scholarly communication landscape. And, if we design them in accordance with the actual practices and structures of scholarship, they can do much to sustain and even enhance the information infrastructure of research.

Acknowledgments

I’d like to respectfully acknowledge the late Rob Kling for initiating my work on this paper for his scholarly communication panel at the 2003 Society for the Social Studies of Science annual conference. I also wish to thank Dave Dubin, Allen Renear, and my other colleagues in the “research writing group” for their helpful comments on an early draft of the paper. Funding for the projects reported on here was received from the National Science Foundation’s CISE/IIS/DST program, the Council on Library and Information Resources and Digital Library Federation, and the University of Illinois Campus Research Board.

References


