

Thematic Research Collections

Carole L. Palmer
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 East Daniel Street
Champaign, Illinois 61820-6211

217-244-0653
Fax: 217-244-3302
E-mail: clpalmer@uiuc.edu.

Forthcoming chapter in

Companion to Digital Humanities
Section 3. Applications

26 November 2002

Outline

Introduction

Characteristics of the Genre

Scholarly Contribution

Contextual Mass

Interdisciplinary Platform

Activity Support

The Humanities Laboratory

The Process of Collocation

Research Collections Landscape

Digital Collections Terminology

Turn in the Collection Cycle

Conclusion

Introduction

The analogy of the library as the laboratory of the humanities has always been an exaggeration. For most humanities scholars, it has been rare to find the necessary materials for a research project amassed in one place, as they are in a laboratory setting. Thematic research collections are digital resources that come closer to this ideal. Where in the past scholars produced documents from source material held in the collections of libraries, archives, and museums, they are now producing specialized scholarly resources that constitute research collections. Scholars have recognized that information technologies open up new possibilities for re-creating the basic resources of research and that computing tools can advance and transform work with those resources (Unsworth 1996). Thematic research collections are evolving as a new genre of scholarly production in response to these opportunities. They are digital aggregations of primary sources and related materials that support research on a theme.

Thematic research collections are being developed in tandem with the continuing collection development efforts of libraries, archives, and museums. These institutions have long served as storehouses and workrooms for research and study in the humanities by collecting and making accessible large bodies of diverse material in many subject areas. Thousands of extensive, specialized research collections have been established, but they are often far removed from the scholars and students who wish to work with them. In recent years, many institutions have begun to digitally reformat selected collections and make them more widely available to researchers, students, and the public on the Web.

Humanities scholars are participating in this movement, bringing their subject expertise and acumen to the collection development process. In taking a thematic approach to aggregating digital research materials, they are producing circumscribed collections, customized for intensive study and analysis in a specific research area. In many cases these digital resources serve as a place, much like a virtual laboratory, where specialized source material, tools, and expertise come together to aid in the process of scholarly work and the production of new knowledge.

This chapter is focused primarily on the thematic research collections created by scholars. The history and proliferation of the new genre cannot be examined in full within the limits of this chapter. Instead, this essay will identify characteristics of the genre and clarify its relationship to the collection development activities that have traditionally taken place in research libraries. This relationship is central to understanding how our stores of research materials will evolve in the digital realm, since thematic research collections are derived from and will ultimately contribute to such institutional research collections. The thematic collections used as examples throughout the chapter have been selected to illustrate specific features and trends but do not fully represent the variety of collections produced or under development. Although there are numerous collections available for purchase or through licensing agreements, all the examples identified here were available free on the Web as of October 2002.

Characteristics of the Genre

There are no firm parameters for defining thematic research collections (hereafter referred to as thematic collections), but there are characteristics that are generally attributable to the genre. John Unsworth (2000b) describes thematic collections as being:

electronic
 heterogeneous datatypes
 extensive but thematically coherent
 structured but open-ended
 designed to support research
 authored or multi-authored
 interdisciplinary
 collections of digital primary resources

These characteristics are common to the projects developed at the Institute for Advanced Technology in the Humanities (IATH), the research and development center at the University of Virginia directed by Unsworth. They are also broadly applicable to many of the research collections being developed elsewhere by scholars, librarians, and collaborative teams. With thematic collections, however, there is considerable synergy among these characteristics, and as the genre grows and matures additional characteristics are emerging that differentiate it from other types of digital resources.

The figure below reworks Unsworth's list of descriptors, separating content and function aspects of the collections and adding emerging features of the genre.

Content	Function
Basic Elements	
Digital	Research support
Thematic	
Variable Characteristics	
Coherent	Scholarly contribution
Heterogeneous	Contextual mass
Structured	Interdisciplinary platform
Open-ended	Activity support

Features of thematic research collections.

The first tier of features contains basic elements that are generally shared by thematic collections. In terms of content, they are all digital in format and thematic in scope. In terms of function, they are all intentionally designed to support research. The next tier of features further defines the makeup and role of thematic collections, and together they reflect the unique contribution this type of digital resource is making to research in the humanities. Unlike the basic elements, these characteristics are highly variable. They are not represented in all thematic collections, and the degree to which any one is present in a given collection is varied. Collections differ in the range and depth of content and the types of functions provided. The basic elements and all the content features, outlined below, are closely aligned with Unsworth's description. The emergent function features are explicated more fully in sections that follow.

Thematic collections are *digital* in format. While the sources may also exist as printed texts, manuscripts, photographs, paintings, film, or other artifacts, the value of a

thematic collection lies in the effectiveness of the digital medium for supporting research with the materials. For example, through advances in information technology, creators of *The William Blake Archive* have been able to produce images that are more accurate in color, detail, and scale than commercially printed reproductions and texts more faithful to the author's originals than existing printed editions (Viscomi 2002).

The contents are *thematic* or focused on a research theme. For example, a number of the IATH collections are constructed around author-based themes, including *The Complete Writings and Pictures of Dante Gabriel Rossetti: A Hypermedia Research Archive*, *The Dickinson Electronic Archive*, and *The Walt Whitman Archive*. Collections can also be developed around a literary or artistic work, such as *Uncle Tom's Cabin and American Culture*. A collection called *Hamlet on the Ramparts*, designed and maintained by the MIT Shakespeare Project, is a good example of a collection based on a narrowly defined literary theme. That project aims to bring together texts, artwork, photographs, films, sound recordings, and commentary related to a very specific literary entity, Hamlet's first encounter with the ghost. A collection theme can be an event, place, phenomenon, or any other object of study. Interesting examples outside of literary studies include the *Salem Witch Trials*, *Pompeii Forum*, and the *Waters of Rome* projects.

Like traditional library collections in the humanities, thematic collections have been built for *research support*, but the new genre is producing more specialized microcosms of materials that are tightly aligned with specific research interests and that aid in specific research processes. Some thematic collections have been designed as new virtual environments for scholarly work. For example, *Digital Dante*, a project produced at the Institute for Learning Technologies at Columbia University, has been conceived as a "place" for study and learning and a "means" of scholarly production.

The thematic framework allows for *coherent* aggregation of content. All the materials included assist in research and study on the theme. This coherence is generally anchored by a core set of primary sources. The capabilities of networked, digital technology make it possible to bring together extensive corpuses of primary materials and to combine those with any number of related works. Thus the content is *heterogeneous* in the mix of primary, secondary, and tertiary materials provided, which might include manuscripts, letters, critical essays, reviews, biographies, bibliographies, etc., but most collections are multimedia. The digital environment provides the means to integrate many different kinds of objects into a collection. In literary studies collections, multiple versions of a given text are commonly made available to aid in comparative analysis, along with additional types of media such as maps, illustrations, and recorded music. For example, *Uncle Tom's Cabin and American Culture* contains different editions of the primary text along with poems, images, films, and songs that show the context and history surrounding the primary work (Condron, Fraser, and Sutherland 2001).

The individual items in a collection are *structured* to permit search and analysis, with most projects in the humanities adopting SGML-based markup formats. Many aspects of a source may be coded, including bibliographic information, physical features, and substantive content, to produce highly flexible and searchable digital materials. The collection as a whole is further organized into interrelated groups of materials for display and to assist in retrieval. Libraries, archives, and museums have conventions for structuring and representing collections, which include systems and guidelines for applying metadata, classification schemes, and descriptors. Some of these methods are

being applied in thematic collections, but scholars are also designing new approaches and developing new standards, such as the TEI (Text Encoding Initiative) markup language for tagging scholarly material, that are more attuned with scholarly practices. As a result, there is not yet uniformity in the methods used for structuring data in digital collections, but there are ongoing efforts to align standards and a growing awareness by collection developers of the value of developing interoperable systems.

Collections of all kinds can be *open-ended*, in that they have the potential to grow and change depending on commitment of resources from collectors. Most thematic collections are not static. Scholars add to and improve the content, and work on any given collection could continue over generations. Moreover, individual items in a collection can also evolve because of the inherent flexibility (and vulnerability) of "born digital" and transcribed documents. The dynamic nature of collections raises critical questions about how they will be maintained and preserved as they evolve over time.

Scholarly Contribution

While thematic collections support both research and pedagogy, the scholarly contribution that results from the creation and use of the resources is what qualifies them as a scholarly genre. When electronic sources are brought together for scholarly purposes they become a new, second-generation electronic resource (Unsworth 2000a). Scholars are not only constructing environments where research can be done more conveniently by more people, they are also creating new research. Like other scholarship in the humanities, research takes place in the production of the resource and research is advanced as a result of it. Thus, scholarship is embedded in the product and its use. And like research generated in fields like engineering, computer science, and information science, some of the research contribution lies in the technical design, functionality, and innovation that makes new kinds of research possible.

Authorship is noted in Unsworth's description of thematic collections, but the concept does not capture the nature of the work involved in collection development. A collection is not always attributable to one author or even a few co-authors and the process of production may not generate new, original content. Some of the technical work involved in creating a collection is likely to require expertise outside that of a typical author in the humanities. Moreover, as noted above, thematic collections are inherently open-ended and therefore can be added to and altered in dramatic ways over time by new participants in the process. After a collection has been established for some time, it may not be accurate to continue to assign complete authorship to the originator, and it may prove complicated to trace authorial responsibility as it evolves over years or even decades. In many ways the work resembles that of an editor, but the activities of curators, archivists, and compilers are also applicable.

The term author does relate the significance of the purposeful organization of information, however. As Atkinson (1998) notes in reference to professional collection developers in research libraries, since "every text is to some extent a compilation of previous texts, then the collection is a kind of text—and the building of the collection is a kind of authorship" (19). Nonetheless, "creator" seems best for encapsulating the range of work involved in the development of thematic collections. The term has become common in standard schemes for describing electronic resources, such as the Dublin Core metadata element set, and it accommodates the technical, intellectual, and creative

aspects of the digital collection development process. Academic subject expertise is considered critical in the development of quality, scholarly digital resources (Crane and Rydberg-Cox 2000). Technical computing knowledge of text, image, audio and video standards, and applications is equally important. Thus, the creators of scholarly collections will need to be a new kind of scholar, or team, with a distinct mix of expertise in at least three areas—the specific subject matter and associated critical and analytical techniques, technical computing processes, and principles of content selection and organization.

Contextual mass

The creators of thematic collections are constructing research environments with contextual mass, a proposed principle for digital collection development that prioritizes the values and work practices of scholarly communities (Palmer 2000). The premise behind the principle is that rather than striving for a critical mass of content, digital research libraries should be systematically collecting sources and developing tools that work together to provide a supportive context for the research process. For libraries, this approach to collection development requires analysis of the materials and activities involved in the practices of the different research communities served (Brockman et al. 2001). Researchers are able to more readily construct contextual mass for themselves through highly purposeful selection and organization of content directly related to their specialized research area.

All collections are built through the process of privileging some materials over others (Buckland 1995), and the construction of contextual mass takes place through careful, purposeful privileging. Because of the specific scope and aims of thematic collections, creators select materials in a highly focused and deliberate manner, creating dense, interrelated collections. By contrast, in both physical and digital libraries, materials are usually separated for reasons unimportant to the scholar. For example, primary texts may be part of an isolated rare book room or special collection, while secondary works are in separate book and journal collections, with indexes, bibliographies, and handbooks kept in reference areas. Moreover, the historical, literary, and cultural treatments of a topic are likely to be further scattered across differentiated classes of subjects. When a person uses a research library collection they are interacting with a context that includes physical, institutional, and intellectual features (Lee 2000). It is a grand and scattered context compared to that of thematic collections, which tend to focus on the physical context of core primary sources and the intellectual context represented in a mix of heterogeneous, but closely associated materials.

Collections built on a contextual mass model create a system of sources, with meaningful interrelationships between different types of materials and different subjects, that work together to support deep and multifaceted inquiry in an area of research. Although many of the resources referenced in this chapter contain large, complex cores of primary materials, this is not necessary to achieve contextual mass. For instance, the *Decameron Web* project, a collection devoted to the literary, historical, and cultural context of Boccaccio's famous text, contains an established critical edition with translations and a selection of related materials, such as annotations, commentaries, critical essays, maps, and bibliographies. The pedagogical intent of the site is obvious in its content and layout, but it is simultaneously strong as a research context.

A number of existing thematic collections exemplify the notion of contextual mass in their depth and complexity, as well as in their explicit goals. The core of the Rossetti archive is intended to be all of Rossetti's texts and pictorial works, and this set of primary works is complemented by a corpus of contextual materials that includes other works from the period, family letters, biography, and contemporary secondary materials. In the Blake archive, "contextual" information is at the heart of the scholarly aims of the project. The documentation at the website explains that works of art make sense only in context. In this case this involves presenting the texts in the context of the illustrations, illuminated books in the context of illuminated books, and those in the context of other drawings and paintings. All of this work is then presented in the context of relevant historical information.

Collaboration is required in the creation of contextually rich thematic collections. Instead of being a patron of existing collections, scholars must partner with libraries, museums, and publishers to compile diverse materials that are held in different locations. For example, collections from the Boston Public Library, the New York Public Library, the Massachusetts Historical Society, the Massachusetts Archives, and the Peabody Essex Museum were melded to create the *Salem Witch Trials* collection. The *Dickinson Electronic Archives*, produced by a collective that includes four general who work collaboratively with coeditors, staff, and users, has reproduced works housed in library archives all along northeast corridor of the United States (Smith 1999).

Interdisciplinary Platform

Humanities scholars have long been engaged in interdisciplinary inquiry, but the library collections they have relied on have been structured around academic structures that tend to obscure connections between fields of research. This is partly because of the large scale of research libraries, but also because of the inherent difficulties in maintaining standard organization and access systems for materials that represent a complex base of knowledge. The continuing growth of interdisciplinary research is a recognized challenge to collection development and the provision of information services in research libraries, and recent studies have identified specific practices of interdisciplinary humanities scholars need to be better supported in the next generation of digital resources (Palmer 1996, 2002). Creators of thematic collections are beginning to address some of these processes and activities through the conscious development of interdisciplinary platforms for research.

A number of collections have been explicitly designed to be conducive to interdisciplinary research and have effectively incorporated the interests of diverse intellectual communities. The *Thomas MacGreevy Archive* aims to promote inquiry into the interconnections between literature, culture, history, and politics by blurring the boundaries that separate the different fields of study. The *Monuments and Dust*, a thematic collection focused on Victorian London, states its interdisciplinary intent in the project subtitle: "New Technologies and Sociologies of Research." An explicit objective of the project is to foster international collaboration and intellectual exchange among scholars in literature, architecture, painting, journalism, colonialism, modern urban space, and mass culture. The premise is that the aggregation of a diverse sources—images, texts, numerical data, maps, and models—will seed intellectual interaction by encouraging lines

of research to have direct bearing on each another and making it possible to discover new visual, textual, and statistical relationships within the collection.

Activity Support

The functions of thematic collections are being greatly expanded as creators add tools to support research activities. Many different activities are involved in humanities research processes, and collections are essential to many of them. Scholarly information seeking is one type of activity that has been studied empirically for some time, and our understanding of it is slowly informing the development of information retrieval systems in the humanities (Bates 1994). Other significant research activities, especially those involved in interpretation and analysis of sources, have received less attention.

Reading is of particular importance in the humanities, and the technologies being developed for collections are beginning to address the complexities of how and why people read texts (Crane, Wulfman, and Smith 2001). Scanning a text is a different activity than rereading it deeply and repeatedly over time, and it is but one stage of a larger pattern of wide reading and collecting practiced by many humanities scholars (Brockman et al. 2001). Other common scholarly activities, such as the "scholarly primitives" identified by Unsworth (2000), have yet to be adequately supported by the digital resources designed for scholars. These include the basic activities of annotating, comparing, referring, selection, linking, and discovering that are continually carried out by scholars as part of the complex processes of reading, searching, and writing. Just as materials can be structured for scholarly purposes as we transform our bodies of texts into digital format, tools can be tailored for specific scholarly tasks.

Searching collection content is a standard function of databases of all kinds, and among thematic collections there is considerable variation in the level of retrieval supported. Image searching has been a vital area of development, since many collections contain a significant number of facsimile images and other pictorial works. In the Blake archive, for instance, access to images is enhanced through extensive description of each image and the application of a controlled vocabulary. Likewise, the Rossetti archive project has been dedicated to developing methods for formally coding features of images to make them amenable to full search and analysis (McGann 1997). Imagesizer, a tool developed at IATH and provided with both the Blake and Rossetti archives, allows the user to view images in their original size or in other convenient dimensions. Inote, another IATH tool configured for the Blake project, allows viewing of illustrations, components of illustrations, and image descriptions. It also allows users to create their own personal annotations along with saved copies of an image. This, in particular, is a key advancement in activity support for collections since it goes beyond searching, viewing, and linking to assist scholars with the basic tasks of interpretation and notetaking, activities that have been largely neglected in digital resource development.

Hypertext has been a monumental advancement in the functionality of collections, and many current projects are working toward extensive interlinking among aggregated materials. The *Walt Whitman Hypertext Archive* intends to link sources to demonstrate the numerous and complex revisions Whitman made to his poems. The items associated with a poem might include the initial notes and trial lines in a notebook, a published version from a periodical, publisher's page proofs, and various printed book versions. The *Bolles Collection on the History of London*, part of the larger Perseus digital library,

exploits hypertext by presenting full historical texts on London with hyperlinks from names of towns, buildings, and people to correlate items, such as photographs and maps of places and drawings and biographies of people.

Mapping and modeling tools are valuable features of number of place-based thematic collections. The *Bolles Collection* provides electronic timelines for visualizing how time is represented within and across documents. Also, a collection of historical maps of London has been integrated with a current geographic information system (GIS) to allow users to view the same location across the maps (Crane, Wulfman, and Smith 2001). Modeling adds an important layer of data to the *Monuments and Dust* collection. A variety of VRML (Virtual Reality Modeling Language) images of the Crystal Palace have been developed from engineering plans and other sources of data to create models of the building and small architectural features such as drains, trusses, and wall panels, as well as animations of the building's lighting and other three-dimensional replicas.

The Humanities Laboratory

The thematic collections concentrating on contextual mass and activity support are coming closest to creating a laboratory environment where the day-to-day work of scholars can be performed. As with scientific laboratories, the most effective places will be those that contain the materials that need to be studied and consulted during the course of an investigation as well as the instrumentation to carry out the actual work. For humanities scholars, a well-equipped laboratory would consist of the sources that would be explored, studied, annotated, and gathered in libraries and archives for an area of research and the means to perform the reading, analyzing, interpreting, and writing that would normally take place in their offices. This conception of a collection as a site where research takes place is aligned with the notion of a collection as an information seeking and use context (Lee 2000), and as such, the focus of the collection moves beyond the theme to encompass the research questions and processes. The most successful of these sites will provide contextual mass and activity support that is not only responsive to what scholars do, but also to what they would like to be able to do.

In the sciences the virtual laboratory concept has been around for some time. Often referred to as collaboratories, they are designed as media-rich networks that link people to information, facilities, and other people (Finholt 2002). They are places where scientists can obtain resources, do work, share data, results, and other information, interact, and collaborate. Traditional laboratories that are physically located encourage interaction and cooperation within research teams, and collaboratories extend that to a distributed groups that may be as small as a work group or as large as an international research community.

Collaborative processes have not been a significant factor in technology development in the humanities. The prevalent notion that humanities scholars work alone is true to some degree. Reading, searching databases, and browsing collections are solitary activities, and most articles and books are written by individuals. However, most humanities scholars do work together in other important ways. It is common to share citations, ideas, and drafts of papers, and these interactions are dependent on strong relationships with others in an intellectual community. In this sense, collaborative activity is necessary in the daily practice of research, and it has been shown to be especially vital for interdisciplinary work (Palmer 2002).

An increasing number of thematic projects are encouraging scholars to share their research through the submission of content and by providing forums for dialogue among creators and the user community. A few initiatives are aimed at enabling collaboration among scholars, as with the case of *Monuments and Dust*, noted above. In other projects, the corpus brought together is considered a resource to foster collaboration among scholars, students, and lay readers. While it is not a thematic collection per se, *Collate* is a unique resource that provides tools for indexing, annotation, and other types of work with digital resources. This project has been designed to support collection development of digitized historic and cultural materials but also to demonstrate the viability of the collaboratory in the humanities.

The Process of Collocation

All collections, either physical or virtual, are formed through collocation, the process of bringing together related information (Taylor 1999). Collocation is a useful term because it emphasizes the purpose of collection building and can be applied to the different means used to unite materials. The collocation of research materials can take many forms. Anthologies are collocations of selected works on a subject. In a traditional archive, collocation is based on the originating source—the person or institution—and these collections are acquired and maintained as a whole. Collocation is often associated with physical location, such as when materials by the same author are placed together on shelves in a library. It is not surprising that some thematic collections have adopted the metaphor of physical collocation. For example, the *Decameron Web* describes its collection as a "specialized bookshelf or mini-library" generated from Boccaccio's masterpiece. The *Tibetan and Himalayan Digital Library* describes their collections as the equivalent of the "stacks" in a traditional library. A library catalog also provides collocation by bringing together like materials through a system of records and references.

To see a substantial portion of the works associated with a particular author or topic, it has been necessary for scholars to consult many catalogs and indexes and travel to different libraries, copying and collecting what they can along the way. In the case of fragile items, handling is limited and photocopying or microfilming may be prohibited. With thematic collections, scholars are now exercising the power of virtual collocation. By pulling together materials that are part of various works and located in repositories at different sites, they collocate deep, sophisticated collections of sources that can be used at a convenient time and place. The directory of the Rossetti archive lists pictures, poems, prose, illustrated texts, double works, manuscripts, books, biography, bibliography, chronology, and contexts, illustrating the diversity and richness that can be achieved through the collocation of digital materials.

The physical proximity of resources becomes trivial when the material is digital and made available in a networked information system (Lagoze and Fielding 1998), but the intellectual and technical work of selecting and structuring meaningful groupings of materials remains critical. This is especially true in the humanities where research often concentrates on the documents and artifacts created by or surrounding an object of study. Compared to other fields of study, the sources used are highly heterogeneous and wide-ranging, and their value persists over time, rather than dissipating through obsolescence. Over the course of a scholar's research project, certain manuscripts or artifacts in foreign

archives may be vital, but a standard edition or a popular culture Web site can be equally important (Palmer 2002). The distributed, dynamic research collections that can be created on the Web are attuned with the nature of humanities fields, which are "concerned with the construction of knowledge from sources of different types, scattered across different subject areas" (Fraser 2000, 274).

The principles that guide the collocation of research collections in libraries are different from scholarly motivations for collocation. Library collections are amassed for preservation, dispensing, bibliographic, and symbolic purposes (Buckland 1992). The process of collecting is governed first by the mission of the institution and more specifically by the selection criteria developed for research and teaching subject areas. In contrast, collections produced by scholars are customized to the research focus of a scholarly community or to the specific interests of the creators. Thus, the "principles of inclusion" for *The William Blake Archive*, that designate the illuminated books as the foundation and a strategy of adding clusters of materials based on medium, theme, or history, are idiosyncratic to that particular project. Unfortunately, explicit statements of selection criteria and long-term plans for development are not always documented for users of thematic collections.

The potential of digital collocation has been restrained by copyright concerns. It is much less complicated to digitize and redistribute sources that do not have copyright restrictions, and therefore older materials in the public domain have been more widely selected for digital collections of all kinds. Increasingly, thematic collection creators are working through copyright requirements for published works, as well as adding new, born digital sources, to build systematic and principled collections that meet their scholarly aims. Again, the *William Blake Archive* is one of the projects that offers a sound model on this front. They have gained access to valuable and important materials for their collection by working closely with museums, libraries, and collectors to address their copyright concerns.

Research Collections Landscape

Scholarly thematic collections are a new addition to the array of existing and emerging research collocations. Interestingly, many thematic collections created by scholars refer to themselves as *archives*, but conventional archives differ in important ways, especially in terms of their mission and the methods they use to organize material. The collections held in archival repositories document the life and work of institutions or individuals. An archive's role is to "preserve records of enduring value that document organizational and personal activities accumulated in the course of daily life or work" (Taylor 1999, 8). Archival collections are collocated according to provenance—the individual or corporate originator—and organized in original working order. As accumulations of materials from regular daily life, these collections may contain print and electronic documents and artifacts of any kind, including meeting minutes, annual reports, memoranda, deeds, manuscripts, photographs, letters, diaries, printed books, and audio recordings.

Thematic collections are more analogous to the *subject collections* traditionally developed in research libraries than they are to archives. Examples of research library subject collections include the Chicano Studies collection at the University of California, Berkeley and the Historical Linguistics collection at the Newberry Library in Chicago. A standard directory lists 65,818 library and museum subject collections in the United

States and Canada (Ash 1993), many of which are thematic in scope. For example, the entry for William Blake lists 20 collections, two of which are contributors to *The William Blake Archive* project. Subject collections are sometimes developed cooperatively by multiple institutions, and these tend to cover broad academic or geographic categories. Examples include the Black Studies collection at the Center for Research Libraries, a membership organization devoted to cooperative collection programs, and the East Asian collections cooperatively developed at the University of North Carolina and Duke University.

Localized collections that contain rare or valuable items may be kept as part of a *special collection*. Special collections departments develop concentrated subject and theme-based collections that include substantial primary materials. They are also the place where manuscripts, papers, and other unique, fragile, or valuable items are maintained and segregated for restricted access. As research libraries began to make materials accessible via the Web, special collections were often the first materials to be selected for digitization. Research libraries have been eager to share their treasures with a wider audience and make them more convenient to view. At the same time, offering a digital alternative decreases handling and the wear and tear on valuable materials. Digital special collections provide an important service for researchers, but they do not serve the range of scholarly functions—scholarly contribution, contextual mass, interdisciplinary platform and activity support—provided in many thematic collections.

The first digitized special collections released by the Library of Congress through their *American Memory* project in 1994 were photographic collections. The initiative has since grown to offer over one hundred online collections, many of which are thematic and multimedia. Most are based on existing special collections within the Library of Congress, but some, such as *Band Music from the Civil War Era* and the *American Variety Stage*, are thematic collections that have been collocated for the first time specifically for online presentation. Many other institutions have selected notable special collections for digitization. For instance, the library at the University of North Carolina at Chapel Hill has produced *Documenting the American South*, a digital collection based on their existing Southern Historical Collection, one of the largest stores of Southern manuscripts in the country. As research libraries continue to undertake these projects, substantial bodies of previously hidden source material is coming into public view. These digital collections make an important contribution to digital research library development and provide potential raw materials for the construction of thematic collections and other aggregations of digital content.

Digital Collections Terminology

There is little consistency in the terms used to describe digital resources, and the number of terms and the overlap between them seems to be increasing. In addition to being a new conceptualization of research collection development, the phrase "thematic research collection" is itself a new addition to the vocabulary. As discussed above, the term "archive" is being widely applied to thematic collections, and the adoption of the word has merit for this purpose, since, like traditional archives, scholarly thematic collections tend to focus on primary sources and emphasize the importance of the physical object. For example, *The Dickinson Electronic Archives* prioritizes the physical object by representing Emily Dickinson's poems, letters, letter-poems, drafts, fragments, and

manuscripts of Emily Dickinson as facsimile images. Likewise, in *The William Blake Archive* the physical nature of the artifacts is a central thrust. The digital collection represents the integration of Blake's illustrations and texts and the variations among different copies of his books, features that have not been well represented in printed editions of his work. As with most thematic collections, the actual goals of the Blake project reach beyond those of a traditional archive, where the central aim would be to preserve the physical record of production. The notion of the archive has been extended to include the catalog, scholarly edition, database, and tools that work together to fully exploit the advantages of the digital medium (Viscomi 2002).

It has been suggested that electronic archives will increasingly take the form of hypertext editions, similar to the *Electronic Variorum Edition of Don Quixote* being developed at the Center for the Study of Digital Libraries at Texas A & M University (Urbina et al. 2002). But, at present, many kinds of resources, including journal article pre-print servers and lists of links on a Web page, are being referred to as digital or electronic archives. The traditional, professional archive still holds an important place in the array of research collections, and some of these are being digitally reformatted while retaining their original aims and organizational methods. At the same time, colloquial applications of the term are increasing and new scholarly idealizations of the concept are evolving.

The vocabulary of digital resources has been further complicated by the wide usage of the term *digital library* for all kinds of digital collections, essentially blurring the distinction between collections and libraries. Of the many definitions of digital libraries circulating in the literature, several can be readily applied to thematic collections. However, a widely accepted conception in the field of library and information science clarifies the difference.

Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities. (Waters 2000)

A thematic collection is not a library in the organizational sense; it is a collection that may be developed or selected for inclusion in a digital library, or it may exist separate from any library or similar institution. A library contains a collection of collections and has an institutional commitment to services that ensure access and persistence. Because of their size and diverse user population, libraries, including digital libraries, generally lack the coherency and the functional features characteristic of thematic collections.

In the humanities, the *Perseus* project is considered an exemplar digital library. It serves as a good, albeit complex, example of the relationship between collections and libraries. The scope of *Perseus* was originally disciplinary rather than thematic, providing access to an immense integrated body of materials in Classics, including primary Greek texts, translations, images, and lexical tools (Fraser 2000). As the project has grown as both a scholarly resource and a research initiative, it has added thematic collections outside the realm of classics. The mix of subject collections within *Perseus* represents an interesting variety of collection building approaches in terms of scope and mode of creation. Three geographically oriented collections, *California*, *Upper Midwest*, and

Chesapeake Bay, have been developed in association with the Library of Congress, *American Memory* project. The library also contains an extensive body of primary and secondary materials covering the early modern period. The best example of a thematic collection within *Perseus* is the pre-twentieth century London segment based on the *Bolles Collection on the History of London*. It is a digitized re-creation of an existing special collection that is homogeneous theme but heterogeneous content. As noted previously, it interlinks maps of London, relevant texts, and historical and contemporary illustrations of the city.

The *Tibetan and Himalayan Digital Library* (THDL) is another kind of digital library/thematic collection hybrid. It is being designed to capitalize on internal collocation to create multiple collections with different structures and perspectives from an underlying base of holdings. For example the contents of the library, consisting of texts, videos, images, maps and other types of materials, are organized by space and time in the Environment and Cultural Geography. The thematic and special collections are organized by subject attributes. The categorization scheme used in the THDL is an interesting case of variant applications of digital resource terminology. Special collections are thematic, with a focus, for example, on the life or activities of an individual, while the thematic collections integrate diverse sources within broad disciplinary units such as Art, Linguistics, Literature, or Music. Subtheme collections, which are independent projects with their own content and goals, are nested within the thematic collections.

Needless to say, there is much overlap between digital library and thematic collection efforts, and variations and hybrids will continue to evolve along with the terminology. Digital research libraries will no doubt continue to acquire the collections built by scholars, collaborative teams, and institutions, while scholars' projects grow to nest and annex digital special collections. An important outcome of this activity is that expert collocation of research materials by scholars is adding an important new layer of resources to humanities research collections.

Turn in the Collection Cycle

In the past, scholars used collections for their research and contributed to collections as authors, but their role as collection builders was limited. They developed significant personal collections for their own purposes, and they collected materials through editing and publishing collected works. On the other hand, collection development has long been a significant part of the professional responsibilities of librarians, archivists, and curators. The interaction between the scholarly community and collection professionals has been an important influence on the development of research library collections. As the primary constituency of research libraries, scholars' questions, requests, and ongoing research and teaching activities have guided the collection processes at research institutions. Of course, the essential contribution of scholars has been as creators of intellectual works that make up a large proportion of research collections. Now scholars have also become creators of research collections, and this change will have an important impact on how our vast arrays of research materials take shape in the future.

Where libraries once acquired the documents created by scholars, they now need to collect the thematic research collections created by scholars. This genre has new qualities that cannot be treated in the same way as a printed or electronic single work, and

the interdisciplinary, multimedia, open-ended characteristics of the resources further complicate matters. Libraries are not yet systematically collecting the collections produced by scholars, in part because of the newness of the genre, but also because this type of meta collecting is an unfamiliar practice. In fact, most research libraries have not yet begun to systematically add and catalog non-commercial, web-based digital materials of any kind to their collections. For example, at the time of this writing, *WorldCat*, a major bibliographic database of library holdings, indicated that 263 libraries had purchased and cataloged a recent book of criticism on William Blake by William Vaughan. In contrast, only twenty-six libraries had added *The William Blake Archive* collection to their catalog. As a point of reference, thirty-four libraries had cataloged *Voice of the Shuttle*, a humanities gateway that is widely used but less similar to the scholarly creations traditionally collected by libraries than the materials in a typical thematic collection. As research libraries begin to regularly acquire and catalog thematic collections, they will be interjecting a new layer of collecting activities and causing a shift in the scholarly information transfer cycle.

The traditional cycle of document transfer as conceptualized before the advent of digital documents (King and Bryant 1971) required publishers and libraries to take documents through most steps of the process. A large part of the production and distribution of scholarly research materials still adheres to this cycle. It begins with use of a document for scholarly work, usually in conjunction with many other documents, which leads to composition by the scholar. At this point, the scholar becomes an author, in addition to an information user. A publisher handles the reproduction, or printing and distribution phase. Libraries move published documents through the circuit by selecting, acquiring, organizing, and storing them, and then by making them accessible, usually on library shelves and through representation in catalogs and indexes. Use of the materials is enhanced by assistance from reference, instruction, and access services provided by the library. The cycle is completed when the material becomes part of the research process by being accessed and assimilated by a scholar.

To some degree, libraries can treat thematic collections like the documents produced by scholars—by selecting, acquiring, and organizing them, as has been done by the twenty-six libraries that have cataloged the Blake archive. Over time, this practice could greatly expand the distribution of scholarly materials in the humanities will be greatly expanded by scholars as they take up collection selection and organization activities. Perhaps even more importantly, scholars are adding valuable features to collections as they customize them for their scholarly purposes. While research libraries strive to meet the information needs of the communities they serve, they are not equipped or charged to fully support the scholarly process. Selection criteria for collections in research libraries emphasize how to choose the best items from the universe of publications being produced relative to the research communities being served. Measurements of satisfaction, circulation, and web activity are combined with librarians' knowledge of their scholarly constituencies, which grows based on what scholars ask for and what they reveal to librarians about their interests and projects. Less attention has been paid to assessing how to prioritize materials in terms of what scholars do, what they value, or what would be the most likely to enhance specific research areas.

Scholar-generated research collections are likely to increase in number as the work of producing them becomes more widely accepted as legitimate scholarship.

Research libraries have yet to grasp how this will impact their practices. It may be some time before there is a confluence of scholar and institution generated collections. First there will need to be a wider awareness of thematic collections as an important mode of scholarly work. Scholars and scientists are producing an abundance of digital products, many of which are important, high quality compilations, and these activities are proliferating through support from funding agencies. It will be necessary for research libraries to respond to this trend in their collection development programs.

At this point in time, many research libraries are focusing on more global approaches to digital collection building by producing expansive gateways for all their user communities. At the same time, researchers are creating their own repositories and tools, highly customized to the scholarly work of their intellectual communities. Research libraries will need to fill the gap by developing mid-range collection services that actively collocate thematic collections within meaningful aggregations. The profile of a mid-level collection would look quite different from the current digital research library. It would not prioritize the top tier of scholarly journals, the major indexes, a large general set of reference materials, or disciplinary canons. Instead, it would provide access to constellations of high-quality thematic collections. Just as importantly, as collection building grows as a form of scholarly production, libraries should be providing resources to assist in this form of scholarly research. Humanities scholars who perform collection building research are not well supported by university resources, which tend to be thinly scattered across departments, computing centers, and libraries. These resources and support services would be best centralized in the library or in auxiliary research units where scholars from all fields can turn for assistance in developing quality content and tools.

Conclusion

As scholars gain mastery in digital collocation and produce innovative research environments, they are practicing a new kind of collection development. Thematic collections are conceived not only as support for scholarship but as contributions to scholarship. They provide dense configurations of materials that strongly represent the relationships between different kinds of sources and different subject areas, and they are designed and instrumented to encourage scholarly interaction with these materials. Through contextual mass, interdisciplinary platform, and activity support they add density and flexibility to previously scattered and static stores of research materials. These collections can assist in the production of new research, but they also have the potential to substantively improve the scholarly research process.

In thematic collections research materials are closely tied to the processes of inquiry, making the contours of scholarship more visible as they are inscribed into the the collection. The questions and methods that propel scholarship become part of the representation, and as scholars build the partnerships it takes to construct quality collections, networks of researchers and institutions involved in a research area become more explicit. Thematic collections are a substantive contribution to the rebuilding of research resources in the digital age, adding richness to our stores of research collections and new opportunities for humanities scholarship.

CAROLE L. PALMER

see Digital Libraries in the Humanities.

References for Further Reading

Ash, L. (1993). Subject Collections. 7th ed. New Providence, NJ: R.R. Bowker.

Atkinson, R. (1998). Managing traditional materials in an online environment: Some definitions and distinctions for a future collection management. Library Resources and Technical Services, 42, 7-20.

Bates, M. J. (1994). The design of databases and other information resources for humanities scholars: The Getty Online Searching Project report no. 4." Online & CDROM Review, 18, 331-40.

Brockman, W.S., L. Neumann, C.L. Palmer, and T. Tidline. (2001). Scholarly Work in the Humanities and the Evolving Information Environment. Washington, D.C.: Digital Library Federation and the Council on Library and Information Resources. 26 Nov. 2002 <<http://www.clir.org/pubs/reports/pub104/contents.html>>

Buckland M. (1992). Collections reconsidered. In Redesigning Library Services: A Manifesto (pp. 54-61). Chicago: American Library Association.

———. (1995). What will collection developers do? Information Technologies and Libraries, 14, 155-159.

Condron, F., M. Fraser, and S. Sutherland. (2001). Oxford University Computing Services Guide to Digital Resources for the Humanities. Morgantown, WV: West Virginia University Press.

Crane, G., and J.A. Rydberg-Cox. (2000). New technology and new roles: The need for "corpus editors. Proceedings of the Fifth ACM Conference on Digital Libraries (pp. 252-253), 2-7 June, San Antonio, Texas.

Crane, G., C.E. Wulfman, and D.A. Smith. (2001). Building a hypertextual digital library in the humanities: A case study of London. Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (pp. 426-434), 24-28 June, Roanoke, Virginia.

Finholt, T. (2002). Collaboratories. Annual Review of Information Science and Technology, 36, 73-107.

Fraser, M. (2000). From concordances to subject portals: Supporting the text-centred humanities community. Computers and the Humanities, 34, 265-278.

- King, D.W., and E.C. Bryant. (1971). *The Evaluation of Information Services and Products*. Washington, D.C.: Information Resources Press.
- Lagoze, C., and D. Fielding. (1998). Defining collections in distributed digital libraries. *D-Lib Magazine*, November. 26 Nov. 2002 <<http://www.dlib.org/dlib/november98/lagoze/11lagoze.html>>.
- Lee, H-L. (2000). What is a collection? *Journal of the American Society for Information Science*, 51, 1106-1113.
- McGann, J. (1997). The Rossetti Archive and image-based electronic editing. *Journal of Pre-Raphaelite Studies* N.S., 6, 5-21.
- Palmer, C. L., ed. (1996). Navigating among the Disciplines: The Library and Interdisciplinary Inquiry. *Library Trends* 45.
- Palmer, C. L., and L.J. Neumann. (2002). The information work of interdisciplinary humanities scholars: Exploration and translation. *Library Quarterly*, 72, 85-117.
- Palmer, C. L. (2000). Configuring digital research collections around scholarly work. Paper presented at Digital Library Federation Forum, 19 November, at Chicago, Illinois. 26 Nov. 2002 <<http://www.diglib.org/forums/fall00/palmer.htm>>.
- Smith, M.N. (1999). Because the plunge from the front overturned us: The Dickinson Electronic Archives project. *Studies in the Literary Imagination*, 21, 133-151.
- Taylor, A. (1999). *The Organization of Information*. Englewood, CO: Libraries Unlimited.
- Unsworth, J. (1996). Electronic scholarship; or, scholarly publishing and the public. *Journal of Scholarly Publishing*, 28, 3-12.
- . (2000). Thematic Research Collections. Paper presented at Modern Language Association Annual Conference, 28 December, at Washington, D.C. 26 Nov. 2002, <<http://www.iath.virginia.edu/~jmu2m/MLA.00/>>.
- . (2000). Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this? Paper presented at symposium, Humanities Computing: Formal Methods, Experimental Practice, 13 May, King's College, London. 26 Nov. 2002, <<http://www.iath.virginia.edu/~jmu2m/Kings.5-00/primitives.html>>.
- Urbina, E., R. Furuta, A. Goenka, R. Kochumman, E. Melgoza, and C. Monroy. (2002). Critical editing in the digital age: Informatics and humanities research. In J. Frow, ed. *The New Information Order and the Future of the Archive*. Conference proceedings, Institute for Advanced Studies in the Humanities, 26 March, University of Edinburgh. 26 Nov. 2002 <<http://webdb.ucs.ed.ac.uk/malts/other/IASH/dsp-all-papers.cfm>>.

Viscomi, J. (2002). Digital facsimiles: Reading the William Blake Archive. Computers and the Humanities, 36, 27-48.