Toward the Humanities Digital Library: Building the Local Organization

Steven Ellis

Before we can explore large-scale implementations of the digital research library, we need to know more about how new technology initiatives fit into local organizational contexts to meet the needs and expectations of local users. We must look at the conditions that have preexisted local initiatives, how organizations have responded, and be aware of the rationale behind those responses. How we explain what we do in these contexts will play a major role in determining the shape of the humanities research library of the future. The electronic text initiative at the Pennsylvania State University Libraries will be our case in point.

There has been much discussion of digital libraries, often involving major proposals for consortia and experimental "testbed" software configurations. As a parallel development, local library-based multimedia and electronic text centers have proliferated. New lines have been opened for librarians, with titles such as electronic text librarian or data services librarian.

A significant knowledge base has been acquired in the course of these local multimedia and electronic text initiatives. What we have learned about the integration of new technology into library environments represents a significant contribution toward understanding how the digital research library will affect local organizations. This paper explores the implications of this, with a principal focus on the humanities research library of the future. The digital research library cannot be effective unless its local effects are thoroughly understood.

There are many questions to be answered. For example, we need to know why certain models for technology in the humanities have been deemed appropriate and others not. Also, the ways the various initiatives serve different purposes within their institutional context is unclear. We need to know what the models have in common and how they are different. Most important, we need to know more about the way technology in the humanities has been conceptualized, and the effect this has had on the way libraries have chosen to address it and the way they might address it in the future.

It is unclear why the term electronic text might be chosen to describe one project and multimedia another. The terms ap-

Steven Ellis is the Electronic Text Librarian at Pennsylvania State University Libraries, University Park; e-mail: sre@psulias.psu.edu.
pear to refer to different things: electronic text to materials composed primarily of text, that is, of alphabetical characters, multimedia to multiple data items in different formats—text in addition to images and sound. Yet, there are many examples where the two formats have characteristics that make them indistinguishable. To coin a phrase, what’s in a name?

If the proliferation of library-based multimedia and electronic text centers is any indication, one thing is at least clear:

The emergence of library technology initiatives in the humanities furnishes many opportunities to expand and redefine the role of the library...

a tacit consensus is building that new technology in the humanities requires specialized organizational responses. However, we know very little about why technology initiatives take the shapes they do. The need for organizational change might be the result of problems and complications within the technology itself, such as the lack (or proliferation) of data-encoding standards. System incompatibility also could be a factor. In addition, it might be that the technology requires a special set of skills. Yet the extent to which factors such as these have determined organizational outcomes is still unclear.

Few humanities technology initiatives have been written about from an organizational point of view. We know very little about the conditions that give rise to new initiatives and about the rationale for the way organizations have responded to them. As a reviewer of proceedings from a recent conference on electronic text noted, many of the participants’ remarks leave an impression that their projects “emerged from combinations of happenstance, individual personal expertise, sheer force of will, or the convergence of many diverse factors.” The reviewer is right on all accounts. New technology initiatives do emerge in this way. Our task should be to put a name to those “diverse factors.” The reviewer goes on to call for a discussion of how the roles of librarians and libraries are shifting in the wake of this technology. The emergence of library technology initiatives in the humanities furnishes many opportunities to expand and redefine the role of the library and the roles of librarians. Some of the ways this redefinition may occur are addressed here, and will be a natural development of the discussion.

Issues in Humanities Technology

It is possible to offer some preliminary remarks on why electronic primary source materials are finding their way into the library. Expectations for the digital library include electronic access to content itself in addition to information that describes content. In the past, many libraries have been involved in the support of electronic text collections. It is very likely that this role will continue to grow. Industry has played a role in new technology initiatives as well, by providing electronic text and multimedia products for libraries to acquire. New texts are being distributed all the time, not only as commercial products but also in the public domain.

As a parallel development, scholars outside the library are beginning to expect the texts they study to be available in an electronic format. More and more users are coming to understand that the strengths of the computing machine may be taken advantage of in research and teaching. The impact electronic text has had on fields such as classical languages and literature has begun to be felt in other disciplines. Having access to an electronic corpus of the texts one studies has come to be considered a necessity not only for research, but also for meeting the productivity requirements of the discipline.
Despite the ubiquity of digitization, computing technology—and any technology for that matter—is rarely introduced into a local environment with a neutral effect. Human practice changes, and is changed. The introduction of new media into the library is a disruptive event. That sound recordings are now distributed on compact disc means that money has to be allocated for the purchase of compact disc players, staff have to be trained in their use, and patrons have to get used to handling the discs.

Electronic text and multimedia technology are no different. In fact, the disruption they represent is much greater. Imagine a case where all sound recordings require different playback equipment, where compact disc dealers in effect own the rights not only to the music but also to the means by which one may listen to it, and you begin to imagine (albeit with some exaggeration) the situation with multimedia and electronic text.

One need only have attempted to maintain a single multimedia workstation with access to multiple CD-ROMs to appreciate the problem. The creator of each product has made unique decisions about how to play the sound, which drivers to use to play the movies, and how to display the text. In addition to this state of affairs, products such as these draw on few standards for the way content is represented behind their operation, making cross-platform portability a problem.

For this reason, text integrity has been a central issue in electronic text and multimedia technology. It should be argued that in its most basic formulation, electronic text should conform to storage and encoding standards for preservation, allowing it to be moved easily from platform to platform. To further complicate matters, it can be argued that electronic text should accurately conform to a printed original. Many of these requirements are the direct result of what happens to a printed text when it is translated into the conventions of a computing machine. Parts of the content on the printed page, things such as margins, typeface, and spacing (the bread and butter of textual scholars), are effectively erased by the limitations of ASCII.

Much work has been done toward the accurate electronic description of printed texts using the specification for a metalanguage—the Standard Generalized Markup Language (SGML). SGML provides something like a grammar for metadata, the text that describes features of the text not readily apparent in the original, or, as in the example given above, text that is lost in the transition to the machine. SGML, as its name suggests, is far too “general” to specify how it should be used to describe a given textual event. An ordinary manuscript might have any number of features requiring special interpretation in order to implement SGML. The Text Encoding Initiative (TEI) has produced guidelines to aid this implementation. An entire culture has grown up around the problem of electronic text integrity. The values represented by this culture provide the most developed criteria for the evaluation of electronic text products, and ought to be reckoned with in benchmarking collections.

It is important to emphasize that each of the issues surrounding humanities technology has an associated set of organizational criteria. Complexities of format translate into complexities in access and service. For those libraries considering digitization projects (for example, of unique special collections material), the complexities of document analysis and TEI implementation must be addressed with high levels of staff training and evaluation. The knowledge and skill sets currently associated with the practice of librarianship will be acceptable for certain operations, but new rewards will have to be integrated into the profession to allow for others.
Organizational Models for Technology in the Humanities

To understand how humanities technology might be addressed in libraries, we need to look carefully at three extant models: the humanities computing center, the electronic text center, and the multimedia center.

Within the academic institution as a whole, the humanities computing center has provided a context for individuals seeking to explore the integration of computers and texts. There are fewer than a dozen humanities computing centers throughout the world. These centers normally reside under the administration of an academic computing center, and are staffed by individuals who have both humanities and computing backgrounds—a combination that is becoming less unique. In the same way, it was once unusual to consider the computer to be a natural component of scholarly practice.

Humanities computing centers have been associated with projects in stylometrics (the quantification of style) and concordance processing, and originated at a time when large computers and significant investments in processing time were required for such activities. Some centers have a small staff and a public service component. And although some centers have archived and made available electronic texts on a case-by-case basis, historically, many have not been involved in providing remote access to electronic texts.

Electronic text centers have found support under the administrations of both libraries and academic computing centers, and sometimes within academic departments such as English. However, the trend is for more of the activities traditionally associated with the humanities computing center to be located inside the library. The reasons for this are clear. Access, and the way it is interwoven with the necessity of standards, is an issue librarians are familiar with. Librarians, and library-based centers such as Michigan's Humanities Text Initiative (HTI), have been highly successful at providing remote access to large collections of electronic texts. Not only are these texts retrievable themselves, they are also searchable. Many of the concordance processing-based retrieval techniques originally developed within humanities computing are now, and will continue to be commonplace in the electronic library. There are roughly 30 entities that can be described as electronic text centers worldwide.

Many electronic text centers are actively engaged in text creation projects. One example is HTI's American Verse Project. Text creation, as defined here, might include consulting on the best ways to proceed with projects originating in departments outside the library. These projects may be large scale, and involve extensive decisions about how a text is to be described utilizing the specifications of the TEL. Some centers have allocated staff time for adding SGML tags.

The multimedia center represents a possible alternative to the electronic text or humanities computing center. In this model, all subject categories are grouped under the format heading of multimedia. From a collections point of view, the format of the product, presumably including text, image, and sound, would be the sole determinant for a product's inclusion in the center. Many multimedia centers also support student and faculty creation of multimedia products and projects. The distinction between electronic text center text creation and the activities supported by the multimedia center lies in the nature of the work being done on the part of the user. In the multimedia center, projects may be short term and less directed. A patron might want to digitize an image for a personal Web page, and be allowed to do so. Although electronic text centers might not actively discourage such activities, staff time is
perceived to be better spent consulting on projects with a broader impact.

**Unanswered Questions**

Despite what may be learned from these three dominant models, there are still many issues that need to be acknowledged, and at best, better understood. For example, we need to know how traditional notions of subject categories and formats determine how multimedia and electronic text are organized in libraries, and what long-term purpose a separate multimedia or electronic text area can serve.

Also, we need to know the extent to which what is considered new and strange has a force in determining local organizational outcomes. We know that electronic texts are imbedded with an authority that is different from more conventional forms of text, such as the book. To some, data associated with a computer have an aura of authenticity, whereas to others they indicate a certain undependability. How might factors such as these be influencing the treatment of humanities technology in the library structure?

In addition, we need to know how notions of technology in fields other than our own determine how technology is conceptualized and used. Many humanists mistrust computer technology and see it as indicative of a change toward productivity and efficiency in the reward structure of their work. We need to know how libraries might respond to this mistrust.

There is much to be learned not only about what kinds of work library users in the humanities currently do with computer technology, but also about the kinds of work they might do. It should be helpful to learn how a shifting definition of the humanities relates to the way technology is perceived and used.

As more libraries consider converting parts of the collection to an electronic format, there also is much to be learned about what constitutes a "state-of-the-art" project. We need to know the kinds of work that are necessary for the successful implementation of such projects. Many electronic text initiatives have significant experience in this area. And in addition to knowing what constitutes a "good" project, we need to be critical of the ways in which what is good achieves its status.

In short, we need to know more about how technology in the humanities gets conceptualized and how notions of what is acceptable are formed. The electronic text initiative at Penn State University has been a laboratory for the exploration of many of these issues. Each not only has some theoretical import, but also can be shown to have tangible organizational effects.

**We need to know . . . what long-term purpose a separate multimedia or electronic text area can serve.**

**The Electronic Text Initiative at Penn State**

**Preexisting Conditions**

Before the electronic text initiative at Penn State could emerge, a number of conditions had to prepare its way. The Penn State Libraries is an organization intensive in computing technology. Over the years, there has been an emphasis on building electronic resources, chiefly bibliographic, and on making them available through either a CD-ROM LAN or the catalog gateway.

Prior to electronic text, a Geographic Information Systems (GIS) initiative was under way. With an emphasis on spatial information, GIS data are a kind of primary source material for the social sciences. The extension of library source material GIS represented was seen as complementary to electronic text. GIS has been ahead of electronic text in its
progress toward implementation, and was seen as a potential model for the kinds of alliances that would have to be made between library and computing personnel.

Other pressures came from outside entities such as the Committee on Institutional Cooperation (CIC). The CIC is the academic equivalent of the Big Ten, and member institutions had a number of electronic text initiatives already underway, such as the LETRS (Library Electronic Text Resource Service) project at Indiana University and the HTI (Humanities Text Initiative) at Michigan. The potential for the joint purchase of core material acted as an incentive to perform on an equivalent level with CIC-wide electronic text initiatives on the horizon. Penn State would need to have an electronic text structure in place in order to participate.

The electronic text librarian would need to rely on colleagues both inside and outside the library in order to move the agenda forward.

This enthusiasm is reflective of an institutionwide climate encouraging collaboration. The Libraries had already participated in joint purchase agreements and joint access projects, mostly for bibliographic resources. It also was perceived that such activity is not unique to the CIC. The University of Virginia, for example, had struck up a statewide partnership under the umbrella of the Virtual Library of Virginia (VIVA) and the State Council of Higher Education to facilitate the purchase of some of its resources.

Although the imperatives of interinstitutional competition were strong, it also is possible to point to a number of direct stimuli within the institution. Several vocal faculty in departments such as English and religious studies periodically queried the libraries faculty and administration with electronic text requests. Selectors would find electronic text publishers' flyers forwarded to them by outside faculty asking that a purchase be made. These faculty were interested in using the computer to study the texts they were working on. Discussions began as to how best to meet these needs.

Organizational Response

With growing user demand, and the perceived need to participate as a full partner in CIC activities, a new line was opened for an electronic text librarian, a faculty appointment at Penn State. The electronic text librarian would hold a public service position and have selection responsibilities for electronic text. No other provisions were immediately made, however, with the intention to allow the candidate the opportunity to define the position through the interview process and for the long term.

This open-ended structure would foster a gradual and collaborative approach to electronic text. The electronic text librarian would need to rely on colleagues both inside and outside the library in order to move the agenda forward. In addition, such an approach would not immediately disturb the current structure of resource allocation. It was foreseen that service expectations might be placed out of reach by incurring costs that were too high and too early in an untried area.

Little work had been done across campus to inform students and faculty about the potential value of electronic text. The term would mean very little to most people. The few colleagues who had expressed interest would be valuable allies, but would not constitute a large enough user base to justify further expenditures.

Thus, it would be necessary to: (1) put together a larger network of colleagues from many departments who would express enthusiasm for electronic text and actively use it in their research and teaching; (2) explain the value of electronic text and humanities-based technology to a
large group of potential users in a short period of time; and (3) create a physical space in order to demonstrate and explore the various uses of electronic text—an electronic text center.

On the whole, it would be difficult to justify expenditures without a significant user base, and it would be difficult to build a user base without having something for them to use. A balance would have to be struck up, and maintained.

A task force of humanities librarians was organized to help build the network. The task force also had an implicit responsibility for organizing the support of colleagues within the library. The selection of the librarians who would participate in this group was based on professional interests, as well as organizational imperatives. These librarians would compose the professional staff of the new humanities library planned as part of the construction of a new wing onto the present building. It was expected that electronic text would be part of this new humanities library.

The task force was effective in navigating internal funding channels, and eventually was able to have certain equipment and collection resources shifted to electronic text. Most important, however, the task force widened the network of contacts by arranging introductions for the electronic text librarian all over campus. These contacts were then able to exert friendly pressure on behalf of an electronic text agenda.

As a parallel development to the liaison effort, special funds were allocated for a small Electronic Text Center. Initially, this foundational center would be used for demonstration purposes, with the ability to expand over time as needs required. Office space was secured and two multimedia-capable machines (a Macintosh and a PC) were purchased, along with a high-speed CD-ROM tower and a scanner. The scanner would allow for experimentation in imaging and OCR for electronic text creation.

The Arts Library, located within the larger structure of the Pattee Library, was chosen as the logical location for the physical space of the center. Arts was in need of another librarian who would do public service. The Electronic Text Center would be in need of Arts Library staff time to handle things such as scheduling and entrance to the center. A partnership would benefit both interests.

Why was electronic text chosen as the appropriate descriptive term for this initiative? Why was the electronic text model chosen over those of multimedia and humanities computing? The term would have a descriptive utility that the others would not. It had recently been given a national focus with the establishment of the Center for Electronic Texts in the Humanities (CETH), a joint project of Rutgers and Princeton Universities. CETH would be working toward cataloging and providing access to a large collection of electronic text materials. CETH also had inaugurated a two-week summer seminar for the exploration of electronic text issues. In addition, many of the other CIC initiatives were "electronic text-like" in the kinds of service they were providing. The term electronic text was generally perceived to be more inclusive than its rivals. Multimedia might not necessarily include an electronic text body of expertise, for example, whereas electronic text could easily be extensible to multimedia.

Preliminary Outcomes
Although an initial lack of funds, staff, and even a ready-made electronic text agenda might, on the face of it, have seemed to create a problematic situation, these gaps instead opened up a series of possibilities. A careful balance has been maintained; needs and expenditures have maintained an even course.

Much was to be gained by having control over how electronic text was to be described to colleagues, both inside and outside the library. This control over defi-
nitions allowed for an open approach to the sorts of research projects and practices that might be addressed by humanities technology. It also has allowed for expansion into nontraditional areas, including access to multimedia products based on film criticism and archeology—formats and subjects not usually associated with the electronic text frame.

In talking with faculty across campus, it has been important to acknowledge that early efforts at bringing together computers and texts are perceived to have yielded mixed results. Many of these early efforts reveal unreasonable expectations about the computer’s capacity for manipulating natural languages—expectations that are still commonplace today. And it has been important to note that exercises associated with humanities computing, such as stylometrics and authorship attribution, generally have fallen out of favor.

Certain other early decisions greatly influenced the acceptance of electronic text in the library. Basing the selection of the librarians who would participate in the electronic text task force on their future role in the new humanities library helped organize the electronic text agenda around a set of goals that were easily defined, and part of the libraries’ strategic plan.

Outside the library, the meetings facilitated by the existing network of humanities librarians and other colleagues have allowed for the growth of a large electronic text network. The necessity of making these new contacts through existing contacts has added immeasurably to the legitimacy of the electronic text effort.

In addition to meetings with faculty, there have been numerous class visits. Seminars have been given on subjects ranging from literature on the Internet to electronic culture and privacy. Electronic text has been integrated into course syllabi as assignments that require electronic text resources. A graduate research course in English is being designed with the resources of the Electronic Text Center in mind. A partnership among the library, the Academic Computing Center, and the English Department has recently allowed for Electronic Text Center staff, and the appointment of staff has opened up the possibility of being engaged more vigorously in electronic text creation and access projects.

On the whole, the liaison effort has had the effect of creating new users whose interests might not have led them previously to consider using computer technology in their work. Although most had experience using word processors, and some with using electronic mail, few had considered how an electronic corpus might be beneficial. For example, using a large textual database, within one hour a faculty member from the English Department was able to find roughly half the unattributed quotations he had been collecting for years.

Without the liaison effort, those faculty who had extensive computing experience might not have considered the library to be the natural place where research and teaching with technology could be explored. All they needed was a little encouragement.

Toward a Definition of the Humanities Digital Library

Technology in the humanities library is evolving from printed texts and bibliographic databases toward a fully integrative and immersive scholarly environment that might include “texts” of any number and kind. As it becomes apparent that this new technology represents something of a disruption, the question
becomes how to make sense of it in local organizations. Our goals certainly should be to do our best to include as many people in the process as possible, and to understand and be accountable to their concerns. To do this, we must be aware of how we conceptualize and explain what we do. We need to be aware that these explanations have direct organizational effects.

First and foremost, it must be acknowledged that although the term electronic text may represent any number of artifacts and practices (markup, or text description, database programming, multimedia, hypermedia, etc.), any new initiative proceeding under the sign of electronic text is simply providing a local context for new technology in the humanities. Our goal should be the total integration of electronic techniques and products into the humanities library of the future. Otherwise, the library will always be a kind of two-class environment: the new technology area on the one side, and the way things used to be done on the other.

Second, it must be acknowledged that although statements about the potential value of the digital library might not be lost on librarians, such statements might be lost on our colleagues outside the library. Library staff should be aware not only of how this environment might work (thenuts and bolts), but also of how it is perceived by the academic community as a whole. The value of the digital library is not self-evident, and librarians should not be surprised if their colleagues outside the library do not see it as inherently good. This has become abundantly clear in building the support structure necessary to the electronic text initiative at Penn State. Our colleagues want to know why what we are doing is important, and how it will affect their work. An appropriate approach must be developed that takes this ambivalence into account.

Simply saying something is good because it enhances “access” is not enough. We need to become experts not only in how we provide access to new material, but also why certain kinds of access are important. And we need to be able to explain this within the context of local initiatives involving any number of checks and balances. (Large infusions of cash will not solve this problem.)

One way to accomplish this is to be aware of the ways in which our field is tied to trends in other disciplines. We need to be aware of how our field participates in broader societal debates. An expansion of electronic resources in the humanities might be best understood within the context of widespread digitization trends. These trends are the concern of scholars from English to engineering. It is not just growth in the number of sources in the humanities that might concern us, for example; the shift in the sort of material that constitutes a “source” ought to concern us as well.

Electronic text might not only include just the set of practices pertaining to the textual scholar, but also a whole range of artifacts and activities that reflect the changing research interests on the part of all “humanities” scholars. Practitioners of new disciplines such as science studies and cultural studies, with their concerns for the emancipatory potential of technology, as well as its potential for manipulation and social control, are not only interested in using texts to further their own research, but also in the ways notions of “text” are made in the first place. Again, the “inherent value” of the digital library will not be apparent to these scholars; we have to be careful that how we think about what we do does not privilege one group of practitioners and exclude another.

Thus, the humanities library of the future must provide an environment for exploring the way technology may be used and understood, in addition to providing basic information access. The accomplishment of this environment hinges on how we explain what we do.
locally. In addition to being interested in our "texts," scholars of the future will be interested in the machines and technical structures that surround, organize, and disseminate the "books" of the future. The humanities library of the present must acknowledge these trends, and embrace them.

Notes

1. I use the term "digital library" to refer to any library initiative for the extension of library collections into the electronic domain. For our purposes here, these resources are mostly primary sources in the humanities. For the best collection of articles representing the discussion about digital library technology and constortia, see Communications of the ACM 38 (Apr. 1995). Many of the initiatives presented here have a science and technology information bent, and are recipients of the NSF/ARPA/NASA grant. Although large and well-funded, they are still very much in the planning stage. Humanities resources will, nevertheless, be an integral part of the digital library. For example, humanities resources are included prominently (as they should be) in Peter Graham's award-winning "Requirements for the Digital Research Library," College and Research Libraries 56 (July 1995): 331–39.


17. Similar growth in use patterns has been reported by other electronic text initiatives. See for example David M. Seaman, "A Library and Apparatus of Every Kind: The Electronic Text Center at the University of Virginia," and Beth Forrest Warner and David Barber, "Building the Digital Library: The University of Michigan's UMlibText Project," both in Information Technology and Libraries 13 (Mar. 1994): 15–19, 20–24 respectively.