



CASE 9

Survey of the University of Pittsburgh and Association of American Universities' Websites and Physical Holdings

AUTHORS: **THERESE BARRY**
 SIRI BERDAHL
 MARY BOERGER
 KRISTIN JUSTHAM
 ROBIN C. PIKE
 MARIKA SOULSBY
 ANITA VANNUCCI

PAPER DATE: August 2008

CASE STUDY DATE: January – April 2007

ISSUE: Role of university archives in preserving web content, which may not be otherwise represented in paper holdings.

KEYWORDS: Appraisal issues, Data longevity issues, Institutional repository, Websites

Copyright by Therese Barry, Siri Berdahl, Mary Boerger, Kristin Justham, Robin C. Pike, Marika Soulsby, and Anita Vannucci.

Institutional Context

The University of Pittsburgh was founded as the Pittsburgh Academy in 1787, and has grown to include four branch campuses in Johnstown, Greensburg, Titusville, and Bradford, Pennsylvania. In 2007, the total university enrollment for the five campuses was nearly 34,000 undergraduate and graduate students. The university is divided into 21 professional schools and colleges, and it awarded over 6,200 degrees in 2007.¹

The University Library System is divided into 20 libraries, independent special collections and the Archives Service Center (ASC). The ASC serves the archival and records management function for all University of Pittsburgh records, and it contains hundreds of collections documenting the history of Western Pennsylvania.

This paper was originally completed as a group project during spring 2007 for a graduate archival studies program. The views expressed in this paper reflect our own conclusions and opinions and do not represent any institution. For the sake of brevity, conclusions rather than detailed analysis of findings have been included.²

Background

This case study compares the websites and record holdings of the University of Pittsburgh for the School of Law, the School of Information Sciences (LIS), and the Graduate School of Public Health (GSPH). The purpose of this study is to identify possible discrepancies between information generated by departments for web publishing and departmental information that is retained in the university archives. We will evaluate other Association of American Universities (AAU) public institutions' websites and their university archives holdings in order to determine what documentation professional schools are collecting.³ Finally, we will review the relevant literature in order to determine what has been written about documenting websites and about current practices of university archives.

In all, we contacted nineteen AAU institutions and inventoried the content of their websites. Thirteen of the schools we contacted responded to our inquiry. An analysis of the inventories and a comparison of their contents versus the information on the schools' websites occur in the findings section of this case study.

In his 1975 article, "The Archival Edge," F. Gerald Ham urged the profession to realize that "conceptualization must precede collection." He asked why college and university archivists do not compare "the documentation produced by institutions of higher education with the records

¹ Statistics taken from *University of Pittsburgh Fact Book 2008*, <http://www.ir.pitt.edu/factbook/fbweb08/fb08.pdf>.

² The group would like to thank Beth Kerr for her contributions.

³ Nineteen Association of American Universities were reviewed. The AAU consists of 60 universities, public and private, throughout the U.S., including two Canadian schools. These organizations have a strong focus on academic research and education. Of the schools contacted, 18 were U.S. public schools and one was a Canadian school. The U.S. schools included major East and West Coast universities, as well as schools from the South and Midwest. The postgraduate population at these schools ranges from roughly 5,300 to 23,000.

universities usually preserve, to discover biases and distortions in the selection process and provide an informed analysis on how archivists should document education and its institutions.”⁴ The following article is a discussion of our findings in an attempt to respond to this challenge. We evaluate how well each school documents itself and, in particular, how successful are the University of Pittsburgh’s self-documentation strategies. We also compare its standing with the other schools we contacted, and we assess how any gaps in the documentation process may best be addressed.

Case Methodology

In order to document an institution, one must take into consideration the entire range of its functions. In her book *Varsity Letters*, Helen Willa Samuels suggests that modern institutions’ appraisal practices must focus their analysis on *what* organizations do rather than who does a particular function. We focus on five of the seven functions that Samuels identifies as constituting modern academic institutions. We created a matrix revolving around those functions to document our findings. These functions include a university’s efforts to confer credentials, convey knowledge, foster socialization, conduct research, and sustain the institution. Each of these functions contains various sub-categories. The two categories that we did not include directly in our research were the functions dealing with providing public service and promoting culture. Aspects of these divisions were included within other sections to make the matrix more manageable.⁵ Additionally, while Samuels has suggested many subcategories, we felt it was necessary to incorporate a number of our own, such as administrative functions, public relations functions and external links, to further explain many of the records types found.

In order to make the Herculean task of reviewing each professional school’s website more manageable, we agreed that each student would review the main page and click through each link to record the types of information that was found on those pages. We decided not continue further. Though this procedure made the undertaking simpler, it must be assumed that some of the information included on the matrix does indeed exist on a school’s site but that it may only be found when further exploring its web pages. The study would have been more scientifically pure had one person looked at all of the websites, but it would not have been feasible for an individual to do a project of this size in the allotted time period.

Our appointed advisor from the University of Pittsburgh’s Archives Service Center (ASC) explained the rationale and approach for undertaking this project. By evaluating three well-documented and active schools (the School of Information Sciences, specifically the Library and Information Science program [LIS]; the School of Law; and the Graduate School of Public Health) within the University of Pittsburgh, we would create a more complete evaluation system for the types of information collected and the activities documented. We split our group of eight students into three sub-groups in order to address the schools separately. We created a web blog through Blogger to facilitate this project, which kept all members updated on individual accomplishments and allowed our two group co-leaders to post easily-accessible, weekly

⁴ F. Gerald Ham, “The Archival Edge,” *American Archivist* 38:1 (January 1975): 5.

⁵ Helen Willa Samuels, *Varsity Letters* (Metuchen, N.J.: Scarecrow Press and The Society of American Archivists, 1992), 5.

summary reports. At the end of the project, the web blog provided essential information towards documenting our process.

Our first step entailed evaluating the university archives records for the three professional schools. We were encouraged to compare our findings with the finding aids for the collections, which were frequently incomplete and contained folder names and associated dates but often excluded the record types found within the folders. Our groups were also provided with the case files detailing acquisition, if available. Of the three schools we studied, two case files existed, although one of these proved relatively incomplete.

Once box-level, document-type inventories were completed, we compared and compiled our findings within sub-groups into one record-type listing. We divided the record lists and inserted the record types into the matrix. Because we necessarily eliminated the context found within the folders by compiling the records into one list for each professional school, we found that several generic document types were difficult to place within the matrix. In order to address this problem we placed some record types into more than one category in the matrix, thereby causing some repetition.

Next, Pittsburgh's website for each professional school was evaluated. We agreed to model our approach to the website records on the archival record evaluation, but we quickly discovered that the digital records could not be compiled into one list as easily as the paper records. Each individual inserted into the matrix the records that she found on her assigned websites instead of producing a compiled list and collectively working through the record types. We did find "records" that could not be categorized within the matrix. Search queries and student, faculty, and staff logons, for example, direct users to additional information. Since we did not have the appropriate password information, we could not explore these pages for correct matrix placement.

The professional schools of LIS, Law, and GSPH produce similar administrative documentation. Because these schools serve distinctly different purposes within the university and community, we understood that comparing the other types of records would produce irrelevant results. In order to evaluate our findings accurately, the final phase of our project focused on the documentation of these professional schools within the University of Pittsburgh's peer AAU institutions. We reviewed the websites for each AAU school in a similar manner as we had done with the University of Pittsburgh's website, although not as extensively due to time constraints. Concurrently, the university archivists for each institution were contacted via e-mail and telephone, and we requested an inventory or finding aid of their holdings for each of the three types of professional schools. We added to the matrix the information we found on each AAU professional school's website and the data we received from the university archives.

Once the matrices were completed for each professional school, record types were analyzed and compared between institutions. As we had more access to the University of Pittsburgh's records, they were used as the basis for this research. A complete comparison to the other AAU institutions is somewhat incomplete because we did not have full access to their records. Finally, we analyzed the information contained in each group's matrix and we researched several conclusions about our findings.

Following completion of the matrix, we concluded that it would be helpful to know what efforts were being made to preserve university websites. Group members emailed their contacts at their assigned AAU schools. A review of the various methods occurs in the findings and conclusion sections of this paper.

Although we were aware of how Pittsburgh's archives manages website preservation (discussed below), we felt that it might be helpful to have an idea of how each department deals with the websites they create and maintain. As we did not think it wise to expend our limited time contacting each department in an attempt to acquire this information, we contacted our own school's webmaster for one example of how departments are reacting to the challenges of website preservation. Again, the findings are discussed below.

Analysis

Graduate School of Public Health

As the University of Pittsburgh was used as the basis for our research, we had more access to its physical holdings and therefore were able to develop a more complete picture of the differences and disparities between information documented by the archives and postings to the website. Missing information in the matrix can be attributed to the lack of detail in the inventories. Some schools are now attempting to convert legacy finding aids into Encoded Archival Description (EAD) in an effort to correct such problems.

Of the ten AAU institutions that we contacted, six of them sent us inventories. The inventories ranged from screen shots of listings for record series and links, to online finding aids and sample container lists. A comparison of the information that we received and the information on the websites revealed a vast discrepancy between physical holdings and the information presented in the schools' web-based information.

Findings. The findings for the schools of public health demonstrate these inconsistencies. Some informational resources are available in the archives, but little exists in the digital format. Online information seems intended to lure potential students and donors and to promote grant-funded research.

Records in Pittsburgh's ASC begin in the 1980s and include a small amount of documentation from the 1990s and the first part of the twenty-first century. The digital documentation begins in the early part of the twenty-first century. Our main concern involves the records that fall between the 1990s and the first part of the twenty-first century. It is quite possible these records remain in the offices of the Graduate School of Public Health at Pittsburgh.

According to their finding aids, public health schools tend to maintain mostly chronological records. The finding aids are incomplete and they represent merely a fragile shell of what the collections hold. There are also discrepancies between physical holdings and representations on the web. Although there is no correlation between the physical holdings and the website, neither

repository fully documents all of the functions of the university. Essential university information is missing.

Since there is no standard for documenting specific sections, each school documents functions in very different ways. Of all of the categories of records that we reviewed, however, every university appears to thoroughly document the functions of their alumni organizations.

School of Law

We were able to evaluate the finding aids and websites of ten AAU institutions. As predicted, there was little overlap between information found within university archives and information being presented on university websites. Archives often hold information pertaining to administrative functions of the school, while online records related to students directly. Finding aids contained only minimal description and vague information, making it difficult at times to compare these record types to those we had defined.

Findings. Nearly all of the records from the Pittsburgh School of Law consisted of correspondence from the dean's office, with only a few small boxes of publications and miscellany from the late 1940s through the 1970s. Also, a number of tickets and invitations documented student extracurricular life in the '50s, '60s, and '70s. Within the archives, a few items described research projects but no documentation of funding resources, staff collaborations or research data, and papers existed. This scarcity of information carried over to the other AAU institutions as well, in both the archives and the websites.

On the other hand, relatively little administrative, financial, or governance documentation was found within the archives. Information pertaining to the academic rules and regulations was more consistently found on university websites than in the archives. Policies regarding drug and alcohol use, discipline and appeals procedures and sexual harassment were not documented in either location.

The website for the University of Pittsburgh's School of Law proved heavy on information pertaining to recruitment, advising, financial aid, and alumni. The majority of information regarding degrees, programs, required courses, student activities and organizations, and extracurricular activities are found online. It seems reasonable to deduce that this information is readily available on the website because of its general nature, aimed largely at current and prospective students, while most of the records found in the archives are administrative, and saved more for evidentiary purposes. Certain aspects of governance were found online, such as affirmative action policies. Unlike most of the information in this category, records of public relations were primarily found online rather than in the archive. Additionally, some institutions used web logs to provide information to their students, another record format not present within paper archives.

School of Information Sciences (Library and Information Science Program)

Eight schools provided us with inventories or finding aids to their collections housed in their physical archives. These finding aids ranged from vague summaries (the more common occurrence) to detailed listings of their entire holdings.

One inconsistency of our findings lies in the fact that some of these programs were subsumed within larger schools of Information Science, while others were substantial enough to be considered their own school. LIS programs are also known by many other names depending on the school, including the School of Communication Information and Library Studies, the Faculty of Information Studies, and the Library and Information Studies Department, which further complicated our research.

Findings. Among the surveyed records within physical archival collections (which ranged from 1904 to 2004 with the bulk falling between 1962 and 1985), the largest gaps were found to exist in categories dealing with the daily life of those at the institution, such as advising and teaching documents, student learning and evaluation documentation, and rules and regulations. A few of these shortages persisted in the websites, specifically in the areas of student learning and evaluation documentation. The websites did show a trend of providing broader coverage than the paper archives, and their coverage of categories such as advising, admissions, teaching documentation, and rules and regulations proved especially thorough.

It makes sense that this would be the case, as capturing the current state of the institution, including day-to-day life, is essential for prospective students. It also keeps current students informed of the latest information that is pertinent to their studies. With the volatile and ever-changing nature of websites, however, this means that no constant record of daily student life exists beyond a handful of years at any point in time. An enormous amount of these record types will be lost if no effort is made to properly preserve and archive the websites.

Addressing Issues within the University Archives

To date, the ASC has taken a passive role in the collection of records. Departments are encouraged to send their records, along with a transmittal form, to the archives when they feel the records are no longer active. Since the archive has no control over which records are being sent, it is impossible to ensure that records “containing evidence and information” are maintained. There is little collaboration between the individuals sending the records and the university archivist, which raises concerns over the quality of extant documentation. The university mandate, dated October 19, 1977, does not mention how to collect or handle electronic records because these records did not exist when the policy was written. Nor do the university archives have an official policy that addresses the preservation of websites. Clearly, the university needs to revise and update its mandate to include electronic records and websites.

Additionally, no collaboration exists between website managers and the university archives. Therefore, the archive is not collecting records that are posted to the web, nor are there other individuals within the University of Pittsburgh who are working to save this information. This

again illustrates the importance of actively soliciting records rather than passively receiving them. Online records will not be preserved automatically. A system to capture these records is vitally important. Much of the information placed online is not produced in hard copy, and therefore is lost after the website is changed or updated. It cannot be stressed enough that these problems are not unique to the University of Pittsburgh, but that most of the schools that we reviewed experience a disconnect between what is posted to the web and what is retained by the university archives. Due to our findings, it may also be assumed that other university procedures for soliciting records from various departments must be examined and updated as well.

University archives serve other purposes as well. The ASC not only serves the archives and records management functions of the university but it also documents the history of Western Pennsylvania. Its collections, which have an outstanding online presence, are superlative and they attract researchers from all over the world. Furthermore, comparing web records to records found in physical archives may not be entirely fair. In order to make such a comparison, one must first think about the functions that each repository carries out. One function of a website is to serve as a public relations tool to recruit students, faculty, and donors. Another is its role as an electronic bulletin board, keeping members of a community updated on the activities of a particular group, such as a student chapter of a professional organization or an alumni association. In both of these scenarios, the scope of the posted material is limited to current and immediate concerns.

However, according to Samuels, unless a university archives fully represents its institution, it fails in its mission. Both physical and web records are important when documenting the university as a whole. For now, the records in the archives are safer in the long-term than those loaded on the web due to a number of reasons. The volatility of electronic records, the lack of cooperation of various departments, and the lack of collaboration between information professionals are all challenges that the archivist or records manager must face. Ben MacIntyre writes:

Digital information may be impossibly voluminous and convenient, but it is also vulnerable and dangerously disposable. Already a vast amount of information has been lost. CDs disintegrate in just 20 years, whereas the Domesday Book, written on sheepskin in 1086, will still be with us in another millennium.⁶

This statement effectively sums up the heart of this problem.

Separately the records maintained in the archives and those on the website may fail to document fully the functions of an institution. Together, they more accurately reflect those functions. A number of problems preclude such efforts at comprehensive documentation. Disconnect between university archivists and the creator of these records is a primary problem. Of course, this disconnect creates issues when attempting to archive all record types, but the precarious nature of web-based records exacerbates many of the concerns already relevant to more traditional

⁶ Ben MacIntyre, "History 1980–2000 Has Disappeared into the Ether. Sorry," *The Times*, March 23, 2007.

record types, particularly the nature of a website as an “evolving entity.”⁷ The function of a university website necessarily produces evolution and adaptability.

Tracking the changes made to web documents is extremely difficult, primarily because copies of previous versions are rarely, if ever, retained. The WayBack Machine at www.archive.org is one of the few attempts to preserve older versions of websites, but the immense task of successfully “archiving the academic Internet” cannot be done by an external organization like the Internet Archive.⁸ The process of documenting older web records must be a self-sufficient one, ideally with academic institutions managing their electronic records responsibly through open communication between the university archivists, webmasters, and administration. The task of retaining websites themselves is not difficult, but since archival versions of websites typically result in limited functionality, especially where hyperlinks are concerned, communication must exist between web developers and administration. Additionally, websites must be compliant with procedures for electronic archiving, which is not always the case.⁹

Communication between university archivists is essential, particularly because the individuals who typically are responsible for these websites approach their maintenance and upkeep with different concerns and priorities. We spoke with archivists from two different universities with two different ideas about preserving academic websites. The university archivist at UCLA explained that, though not carried out in a systematic fashion, UCLA has saved several of the school’s websites as digital objects.¹⁰ Typically, the content of the library’s external and internal web is backed-up daily online and to magnetic tape, and then stored at a remote location outside of Los Angeles to limit the possibility of damage or loss. Until recently, the back-ups were erased after sixty days. They now are retained permanently. A group of UCLA special collections librarians, in coordination with the head of their Digital Library Program, Stephen Davison, now create pertinent metadata. Also, the University of California Archivists Council, which is comprised of the university archivists from all ten University of California campuses, is working on an electronic records test project to automatically copy and preserve selected websites from each campus. The California Digital Library created the software for the crawler, and the project is still in the developmental stage.

Contrast this initiative with that of the University of Pittsburgh, whose webmasters have stated that keeping track of alterations to the website is not a primary concern. Our group contacted the webmaster at the university’s School of Information Sciences to determine how records are placed online and what measures are in effect to maintain them. We found that the marketing director is responsible for web content but that she does not solicit information from contributors. She reviews submissions sent to her by the staff and faculty and decides what is appropriate for the website. Certain records are never put online, including syllabi or course

⁷ Katharine A. Salzmann, “‘Contact Us’: Archivists and Remote Users in the Digital Age,” *The Reference Librarian* 85 (2004): 46.

⁸ Internet Archive: The Wayback Machine, <http://www.archive.org/web/web.php>. (Accessed April 7, 2007).

⁹ The Internet Archive: WayBack Machine FAQs, <http://web.archive.org/collections/web/faqs.html#exclusions>, (Accessed April 7, 2007).

¹⁰ The first site saved was the beta version of the initial library website, created in 1996. All of the information regarding UCLA’s projects was obtained via email from the university archivist, Charlotte B. Brown, on March 28, 2007. More information about UCLA’s Digital Library Program may be found at <http://www2.library.ucla.edu/libraries/digital.cfm>.

materials that are the intellectual property of faculty or students, patent-pending research projects or private student information.

When asked if the website was being saved in any way, the webmaster replied that backups are made daily by the systems administrator and that she also retains a local copy and backup. The webmaster insisted that there is no need to document changes as the website changes every day. This represents a striking difference in the way that records are viewed by archivists, records managers and other professionals. If archivists seriously desire to document university websites, they need to explain their goals and objectives to those currently responsible for maintaining the sites. In order to bridge this ideological gap, a mandate for interaction between these individuals that accounts for and counteracts the ephemeral nature of websites is required.

Again, the problem of how to handle born-digital records is not a predicament faced by the University of Pittsburgh alone, or even solely by universities. All types of institutions and organizations are attempting to grapple with procedures for preserving digital records. Problems inherent to these types of records should not deter universities from dealing with issues associated with their preservation. It is vital that archives maintain both physical and digital records in order to successfully document the university's function as a whole.

Conclusion

Helen Samuels observes in *Varsity Letters* that the existing collection analysis process “supports primarily a quantitative not a qualitative analysis of the holdings.”¹¹ She emphasizes that the seven functions of a university that she has identified are not intended to be prescriptive. The archivist ultimately should determine which functions that constitute an institutional documentation plan will most effectively communicate the documentary goals of the institution as well as their successful implementation.¹²

In order to devise a plan, Samuels recommends translating the functions that describe a specific institution, as well as evaluating the records that are existent in the archives and those that remain in departmental offices. We believe that archivists should next examine the individual schools' websites, to best identify those aspects that must be preserved in order to document that school. Samuels' book was written before the proliferation of the web, but we believe that this would be the next natural step in her progression. Given the volatile nature of websites and the volume of paper records produced, a well-formed institutional documentation plan would assist the university's archivist in determining which functions best represent the university and aid in the formulation of documentary goals. The articulation of documentary goals may help to direct the administration in constructing a mandate that clearly states the collection policy of the university archives regarding both physical and electronic records. This may in turn be conveyed to each school to open lines of communication regarding the preservation of websites and other records deemed appropriate for transfer to the archives.

As many of the records in the University of Pittsburgh's archives are decades old and as certain activities are not documented on the website, one can infer that Pittsburgh, and most likely many

¹¹ Samuels, *Varsity Letters*, 10.

¹² *Ibid*, 253.

other schools as well, better documented certain functions in the past. Therefore, if the archives no longer actively seeks records and does not preserve the university website, we must question if it fulfills its purpose or merely concentrates on documenting the community and the higher profile collections that bring attention and researchers to the archives.

As access to AAU institutions' records was limited, and a comprehensive inventory of entire websites was not possible due to time constraints, this is not meant to be a definitive study analyzing the documentation successes and failures of universities. It must be noted that while all of the archives and web pages document a great deal of information, the inconsistencies between them are overwhelming.

A comparison can be made between the websites and the university archives to the memory function in the human brain. Documents placed on websites are similar to experiences lodged in short-term memory. In the university, the important documents are sent to the archives for long-term storage. At the moment, however, there is a disconnect between the long-term and short-term memory banks, where vast tracts of records are uploaded to the web and then discarded, never properly sorted to determine whether they should be maintained in the archives after they are taken off-line.

When comparing the three different types of schools, more similarities than differences exist. There is a noticeable discontinuity between the archives at the examined institutions and their web pages. Each of the three sets of schools shows gaps chronologically between the physical and web holdings. The categories used to arrange the findings in the schools shows university archives are saving information that diverges from the schools' websites. Even if the two versions of documentation are combined, gaps exist. For example, two of the three schools, those of Law and Public Health, note deans' correspondence as being a substantial body of the physical holdings, in addition to faculty documentation. Administration, faculty, and students all play integral parts at a university; however, administration and faculty are represented in archives while students are under-documented. The daily activities of the schools are also missing from the archives.

Although the University of Pittsburgh was the main focus of our study simply because we had greater access to its records, the disconnect that exists between the information produced by universities on their websites and information acquired by university archives appears to be a dilemma for the majority of universities that we examined. All university archivists need to do what they can to acquire underrepresented materials for their schools. To accomplish this goal, a main objective for university archives should be to create a set of guidelines of the types of records schools should be documenting and retaining in their university repositories. A good starting point might be to locate materials that fall into these gaps between paper-based collections and web-based documentation. One of the largest gaps involves the life and activities of the students in their programs. Student life is rarely stagnant and must be documented. Curriculum, research, and funding evolve and are routinely recorded, and therefore a more proactive approach to record keeping is necessary. The steps we suggest here are not giant leaps; on the contrary, they are quite possible to accomplish. If no one takes responsibility, however, vital information will be lost forever.

Does your university archives have born-digital records?

Share how you are effectively managing these digital records by submitting a case study to *Campus Case Studies*.

Visit www.archivists.org/publications/epubs/CampusCaseStudies/.