Chapter 8

Faculty Members as Archivists: Personal Archiving Practices in the Academic Environment

Ellysa Stern Cahoy
The Pennsylvania State University Libraries

In an era of increasingly cloud-based distributed computing and storage, the creation, management, and archiving of scholars’ personal libraries merits increased attention in the college and university environment. Faculty find, organize, share, save, and make searchable a wide range of print and online intellectual assets, including scholarly output, collections of research materials, and more informally created web-based social content. Personal scholarly archiving has become a dynamic process rather than a rote exercise at the end of a faculty member’s career.

Microsoft researcher Catherine C. Marshall with colleagues Sara Bly and Francoise Brun-Cotran identified user challenges relevant to personal archiving: accumulation (a weighty issue in the current online/offline environment), distribution (storage of items both online and offline, on a variety of media, and on different devices and computers), curation (managing and organizing files effectively and consistently)
migrating important files to "maintainable formats"), and long-term access. These challenges provide a structure for emerging personal archiving literacies for faculty and outline an imperative for instructional work for archivists and public services librarians working with faculty members.

This chapter outlines the personal archiving issues currently confronting scholars and will identify critical personal archiving literacies and instructional principles for helping scholars effectively manage, maximize, curate, and archive their scholarly materials throughout their academic career. Examples of current initiatives supporting and exploring personal archiving in the scholarly environment will be shared, as well as recommended practices for integrating personal archiving support within higher education.

**Literature Review**

Several existing studies have looked closely at personal archiving by faculty. The most significant of these studies are "Understanding Faculty to Improve Content Recruitment for Institutional Repositories," by Nancy Fried Foster and Susan Gibbons, and the Andrew W. Mellon Foundation-funded "A Multi-Dimensional Framework for Academic Support." These valuable projects provide an excellent foundation for research on faculty needs regarding personal archiving and the use of institutional repositories (IRs). While of great value, the most recent of these studies is now nearly 5 years old, and the landscape for information creation and storage has changed drastically with the widespread usage of cloud-based tools including Google Docs, Dropbox, Zotero, and Mendeley. Social media as well, including Facebook, Twitter, LinkedIn, and other personal data-driven platforms, have also become repositories for scholarly content, and scholars' social data should be considered in parallel with their more formal scholarly output.

Foster and Gibbons' article details an ethnographic, work-practice study of how faculty members at the University of Rochester conduct their research and writing. Participants in Foster and Gibbons' study reinforced Dorothea Salo's assertion that faculty's primary need with regard to IRs are sharing capabilities while versioning and building capacity for storing in-process (and eventually final) versions of materials. That study also identifies educational needs for faculty, including devising user-centered strategies for digital preservation and helping faculty build a basic understanding of the functions and benefits of utilizing IRs. A later work by Foster and David Lindahl also uses a work-practice study methodology to identify necessary features for an IR, utilizing participatory design and mapping of user needs. The authors reinforce the importance of ethnographic research in human computer interaction and interface design to uncover unanticipated solutions and limitations of current tools.

**Institutional Repository Infrastructure and Educational Services**

The literature on faculty use of IR services confirms that the existing model of an IR as a one-way, static archive must change, broaden, and become more flexible in order to attract a sustained depth of faculty users. In "Innkeeper at the Roach Motel," Salo, discussing faculty needs, asserts that "Repository software serves observed and stated faculty needs surrounding content creation and dissemination hardly at all" (author's emphasis). She continues, "'Roach motel' repositories, in which materials fixed in their final form are the only acceptable content, hold no value for many faculty." Salo emphasizes the need for IRs that assist faculty with document versioning and sharing materials (and data) in process with other faculty collaborators at the institution or elsewhere. Emphasis on actively collecting draft content (rather than passively storing finished content) is essential.

Marshall's article, "From Writing and Analysis to the Repository: Taking the Scholars' Perspective on Scholarly Archiving" brings forward recommendations that back up Salo's assertion that an IR must support collaborative, in-process authoring. Marshall takes this idea further, pushing for filtered IR synchronization with scholars' local
Critical Digital Literacies for Faculty

As mentioned earlier, Marshall, Bly, and Brun-Cottan identify critical user challenges relevant to personal archiving: accumulation, distribution, curation, and long-term access. These challenges provide the backbone for exploration of emerging digital literacies in this area and may become the instructional lifework for future academic librarians. Marshall et al. also highlight environmental factors that must be considered in tandem with personal archiving needs: malware, availability of IT support, and attitudes of users. Attitude, the affective behaviors of users, is an important factor to consider in helping faculty learn effective personal library management skills. Prior research has noted the significance of affect on learning.

“The Academic Library in a 2.0 World” stresses that the future role of the library does not reside in collection development and stewardship, but instead in supporting and enriching teaching, learning, and research. The report states:

The institution as a whole will need to embrace new critical literacies and core skills required for the teacher, learner, and researcher of the digital age. Among other things, these will encompass multimedia creation skills, conventions of behavior in new communication media, computer-aided searching and data analysis skills, new ways to develop scholarly communication, and new ways of assessing student learning. Most significantly, intensive new curricular support programs will be needed to train not only students but also faculty and staff in these skills, and the library should be a leader in this area.

The report’s focus on a user-empowered environment where the librarian’s primary role is as teacher and facilitator confirms the importance of the dual focus of our proposed study. Salo asserts that librarians “must prepare themselves to help faculty with data management throughout the research cycle.”

(desktop) collections. She notes that a good IR infrastructure “must beat email along all of these dimensions if it is to be adopted in email’s stead.” Marshall’s article was written nearly 4 years ago; we could now also substitute “Dropbox” (a synchronized cloud/desktop file storage service) for “email” in her assertion.

In “Structure, Features, and Faculty Content in ARL Member Repositories,” Holly Mercer and colleagues survey the structure and features of 72 IRs hosted by ARL libraries. The authors conclude, “Institutional repositories can, and should, continue the dual goals of preservation and access, but emphasis should be placed on the benefits of increased access to scholarship, and the repository as a destination for research information.” Two software packages are highlighted as possible facilitators for faculty collaboration and individual information management within an IR: VIVO (www.vivoweb.org) and BibApp (www.bibapp.org). These services exist outside of the repository and seek to draw together faculty research and publications for wider impact and availability.

Matthew Kirschenbaum’s report on “Approaches to Managing and Collecting Born-Digital Literary Materials for Scholarly Use” provides an exciting view into the grand future scope of scholarly archiving, where not only are texts collected, but physical computers are archived and online interfaces are preserved. At the Maryland Institute for Technology in the Humanities (MITH), the Deena Larsen Collection (mith.umd.edu/larsen) pushes the boundaries of traditional scholarly archives, with more than a dozen Macintosh Classic computers and an expansive collection of creative electronic writing. Similarly, Emory Libraries makes available Salman Rushdie’s computer desktop and files in preserved form for online browsing and exploration (findingaids.library.emory.edu/documents/rushdie1000). These creative projects cast a path for future archiving of electronic, computer-created works. Viewed together with Salo’s and Marshall’s research, the resonating concept of an IR as a dynamic, user-centered, collaborative tool preserving a wide scope of digitally created objects is obvious and apparent.
Building a user-centered repository that will be actively utilized by faculty requires a close connection to faculty needs and established scholarly behaviors. In “Content In, Content Out: The Dual Roles of the Reference Librarian in Institutional Repositories,” Barbara Jenkins, Elizabeth Breakstone, and Carol Hixson explore the importance of the subject specialist in connecting faculty with an IR, describing how reference librarians can help faculty understand not only what is in the IR, but how to easily access, search, and add to the IR.\textsuperscript{15} Library instruction, perhaps often on a more individual level tailored to faculty’s needs, is an essential element of rolling out an IR university-wide. Suzanne Bell, Foster, and Gibbons also reinforce the centrality of the reference librarian in recruiting and assisting in the process of bringing new scholarly content into the IR.\textsuperscript{16}

**Model Academic Programs/Archiving Initiatives**

Higher education and academic libraries in general are in a nascent phase of recognizing personal archiving as an imperative for proactively retaining a diversity of faculty work. A variety of universities have begun exploratory projects that explore the scope of scholars’ collections, as well as strategies for saving, annotating, and preserving a wide swath of faculty works.

SALT (Self Archiving Legacy Toolkit; sites.google.com/site/stanfordluminaryarchives) was an effort by Stanford University in 2007 to create a web-based service for faculty self-archiving.\textsuperscript{17} Described as offering “Stanford luminaries including prominent faculty, alumni, and associates a web-based personal archive for depositing, organizing, and annotating their life-work collections,” SALT’s primary purpose was to add narrative to collections. Faculty were offered the option of labeling and annotating folders and digital documents, offering users the ability to “create a narrative or thematic structure for their own work.” SALT worked collaboratively with Zotero, enabling faculty to pull directly from their own personal libraries into the archival management interface. While SALT is not currently active for a public audience, selected principles upon which it was designed are embedded in other current Stanford digital library initiatives.

A variety of higher education institutions are innovatively developing the process of collecting both print and born-digital materials from prominent faculty members. As the lifespan of technology gets shorter and shorter with each year (and with each new technology), faculty education on the migration and maintenance of core files is truly significant. In just one prominent example, the author John Updike sent 50 5½” floppy disks to Harvard’s Houghton Library just prior to his death.\textsuperscript{18} While the library has yet to place these files online, they are being kept in climate-controlled storage as the library awaits a static process for intake and digitization of born-digital materials. This challenging situation is not unique. Faculty education on file preservation and management along with development and continued revisiting of processes for managing born-digital materials are critical to the longevity of scholarly information collections.

The best-known example of online archiving of born-digital materials may be the previously mentioned Salman Rushdie Papers, a project undertaken at Emory University. For this collection, Emory archivists took the visionary step of allowing visitors to browse Rushdie’s files on his original computers, with the computer desktop, software, and file organization preserved. The computers and hardware available to visitors include one Macintosh Performa 5400/180, one Macintosh PowerBook 5300c, two Macintosh PowerBook G3 models, and one external hard drive,\textsuperscript{19} which provides a layered browsing experience that gives the user a depth of understanding of the technology in use by Rushdie as he created his works. Erika Farr, director of Born-Digital Archives at Emory, noted in an interview on the Rushdie Archive with the *New York Times*: “If you’re interested in primary materials, you’re interested in the context as well as the content, the authentic artifact. ... Fifty years from now, people may be researching how the impact of word processing affected literary output, which would require seeing the original computer images.”\textsuperscript{20}
In addition to the Rushdie archive, Emory hosts several other born-digital archives, including the collections of Lucille Clifton, Eamon Grennan, and Turner Cassity (marbl.library.emory.edu/collection-overview/digital-archives). The university’s Manuscripts, Archives, and Rare Book Library (MARBL) also has a sample “Pre-Acquisition Digital Assets Survey” for donors of born-digital collections (marbl.library.emory.edu/sites/marbl.library.emory.edu/files/preacq_digiassets_ind_donor_survey.pdf). This tool is an excellent example of how archivists can walk donors through born-digital collections and help them understand the scope and distribution of their collections. Emory archivist Laura Carroll and her colleagues discuss the intricacies and challenges of dealing with a print and born-digital collection (such as the Rushdie archive) in their article, “A Comprehensive Approach to Born-Digital Archives,” and provide additional information on how MARBL has developed its born-digital archival services and collections.21

The Stanford University Libraries have invested significant energy into acquiring and processing digital archival material and handheld media material. The Born-Digital Program is focused on these efforts and includes a Digital Forensics Lab (lib.stanford.edu/digital-forensics/home) designed to rescue and retrieve information stored on outdated technology. The Digital Forensics Lab is currently undertaking a number of born-digital projects, including the Xanadu Project collection, founded in 1960 by the father of hypertext, Ted Nelson, and the Stephen Jay Gould collection, which includes data from floppy disks and computer tapes central to the paleontologist’s seminal research. The Digital Forensics Lab is a model for peer libraries facing an increasing influx of born-digital materials in the years to come. Such a facility allows for easier retrieval and preservation of data on outdated file formats, while also providing a valuable physical presence that helps promote the importance of born-digital file migration and conservation.

Colleges and universities are also very active in bringing increased access to prominent faculty members’ print-based personal archives.

MIT has created the Edgerton Digital Collections project (edgerton-digital-collections.org), centered on the life and work of inventor and professor emeritus Harold ‘Doc’ Edgerton. While the collection is focused on digitization of primarily print material from Edgerton’s academic career, there is a social and interactive element to the online collection. Collection users are invited to contribute stories or reminiscences of Edgerton, and the collection is also crowdsourcing the transcription and identification of items in the collection in collaboration with the userbase. This is a unique way to bring a dynamic and engaging online dimension to a primarily print-based online collection.

Educational Initiatives

Colleges and universities are teaching users how to archive and preserve important information. The University of Illinois at Urbana-Champaign (UIUC) Libraries hold workshops on personal information management, as part of the Savvy Researcher series (www.library.illinois.edu/sc/services/savvy_researcher.html) offered by the libraries’ Scholarly Commons. The workshops cover the basics of information management and encourage attendees to begin thinking critically about the span, composition, and lifespan of their personal information collections. UIUC’s Scholarly Commons also offers data service walk-in hours where users can consult with an expert on how to manage their individual collections with an eye toward future data longevity.

In 2012, Columbia University held its inaugural Born Digital: Personal Digital Archiving Week.22 This series of educational events was organized by the libraries and covered several topics on archiving, including emerging technologies and support in the Columbia University Libraries for personal and research archiving. Similarly, in 2012, the University of California San Diego’s Personal Digital Archiving Day held an intensive 1-day workshop on preservation related topics, including archiving photos and creating metadata, saving email messages, and preserving web-based content.23
While the topic of personal archiving is always worthy of a special event, there are also ways to weave preservation concepts into more traditional instruction on finding and using information. Most academic libraries provide support for, and workshops on, the use of citation management tools, including (but not limited to) Zotero, Mendeley, and EndNote. These software packages help users manage their own personal libraries. Every time a student or faculty member learns how to use citation management software, there is an opportunity to embed instruction on personal information management and preservation. With Zotero, Mendeley, and EndNote, libraries save and organize only one type of information—scholarly resources—but they present a structure to help users think more broadly about their information collections, how (and where) they are accessed, as well as where they are saved (and archived). As citation management tools become even more robust and comprehensive, there will likely be opportunities for enhanced and expanded personal information management. This is the lifework of librarians now and in the relative future—helping users understand, manage, and maximize the entire scope of their individual information collections.

Future Development of Services and Resources

A next step will be for colleges and universities to begin connecting their institutional repositories to the scholarly workflow. A body of research exists documenting the low use of IRs, and reasons why faculty are (or more frequently, are not) archiving materials there. While it is important to teach faculty and other academic users how to look expansively at their information collections and cull and curate the important pieces for future access, it is also imperative that academic institutions connect their archival tools in a manner that is seamless and effortless for even the novice user. Several services in development provide promise in this area for higher education, including VIVO and BibApp. Both of these tools work in conjunction with IRs to highlight faculty research. Mercer and colleagues observed, “If repositories are marketed less for their preservation capabilities and more as platforms for experimentation in scholarly communication, additional new tools and new structures may emerge.” Similarly, the COAR (Confederation of Open Access Repositories) report, “The Current State of Open Access Repository Interoperability,” asserts that easy information exchange with other research management systems is essential. The report asks, “How can we support repository integration with information systems that support research and scholarly communication, from journal publishing platforms to Current Research Information Systems (CRIS)?”

This is, arguably, the direction of the future for IRs and the faculty workflow. Bringing personal archiving onto the scholar’s desktop, making IRs malleable and collaborative, and continuing educational initiatives that help faculty learn how to store, organize, cite, share, and archive their works throughout their academic careers is essential to this process.

At The Pennsylvania State University, a Mellon-funded project is exploring the information workflow of disciplinary faculty and faculty needs regarding the acquisition of the digital literacies essential to effective research management, robust scholarly creation, and continued navigation of the archiving process. Led by Penn State Libraries faculty, the author of this chapter, and Scott McDonald, a Penn State college of education professor, the project seeks to define the following:

- A set of design principles for archivists and librarians on how archival interfaces can be created that support not just archiving and archival access but sustained easy integration of archival practices into the online scholarly workflow
- Identification of critical digital literacies for faculty management of online scholarly workflow, and a set of recommendations for librarians on best practices for supporting faculty in developing critical digital literacies

This project focuses on information management needs at earlier stages of the research lifecycle, with the goal of developing architecture
that supports the entire online scholarly workflow, and facilitates the
development of critical literacies for faculty’s personal information
management needs. Figure 8.1 shows the researcher’s phases of infor-
mation management, creation, sharing, and archiving. With this figure
as a guiding model, the study aims to unify the all stages of the research
life cycle. This study will inform the work of user-focused librarians,
both at Penn State and beyond, as they help faculty learn how to max-
imize and mine their personal collections.

Librarians are increasingly relied upon to help faculty members
develop strategies that address their personal information management
needs. Indeed, subject librarians, with their close connections to disci-
pline-based faculty, colleges, and departments, are on the front line of
scholarly communications and are integral players in building faculty’s
understanding and acceptance of the online archiving process.29

Documenting existing practices and needs, this proposed project
will highlight gaps in existing library services and technology infra-
structures. While the initial project aims to gain a general understand-
ing of needs central to personal scholarly archiving, the proposed next
phase will begin to look at how research management software (such as
Mendeley or Zotero) can better integrate into the faculty member’s
workflow. Specifically, the primary investigators hope to design a man-
nier for institutional repositories to connect seamlessly with research
management software, allowing faculty to deposit into and draw from
IRs in a manner that is complementary, and not disruptive to the flow
of a scholar’s online work.

Conclusion

As information moves increasingly online and into the cloud, the time
is ripe for academic libraries (and colleges and universities in general)
to begin formalizing self-archiving strategies for faculty and students.
While technology-based solutions (such as institutional repositories)
are significant in the academic environment, attention must also be
paid to developing educational initiatives for users. The scholar’s work-
bench is crowded with many disparate tools, devices, and information
collections. In order for intentional, sustained self-archiving to occur,
users must learn how to draw together, assess, manage, and archive the
most important scholarly materials in their dispersed collections.
Helping faculty and students understand the impermanence of online
information and the importance of duplicate data archives is a signifi-
cant challenge. Even more challenging is the goal of creating a centrally
located, institutionally relevant place where scholars can store, share,
and self-archive.
In "The Next Generation of Academics: A Report on a Study Conducted at the University of Rochester," Ryan Randall and colleagues describe the need for an institutional repository platform that allows users to write, collaborate, self-archive, and self-publish. We are on the verge of developing such a tool that draws together the various phases of the scholarly workflow. With more study of our users’ needs, increased attention to faculty education on importance of self-archiving, and a willingness to experiment with software, we will hopefully soon achieve a more unified pathway for robust and sustained self-preservation of scholarly works.

Endnotes


3. Foster and Gibbons, "Understanding Faculty to Improve Content Recruitment for Institutional Repositories."


14. Salo, "Innkeeper at the Roach Motel."


20. Cohen, "Fending Off Digital Decay, Bit by Bit."


26. Mercer et al., “Structure, Features, and Faculty Content in ARL Member Repositories,” 342.


29. Bell, Foster, and Gibbons, “Reference Librarians and the Success of Institutional Repositories.”


---

**Chapter 9**

**Landscape of Personal Digital Archiving Activities and Research**

Sarah Kim  
University of Texas at Austin

Recordkeeping is not the type of activity that excites most people. However, since the time when records first started to be generated, recordkeeping has become one of the tasks many people do routinely. It is important to note that recordkeeping is more than just putting documents in folders and hoping they will be safe and accessible in the future. In fact, it is a human information behavior that involves complex psychological and social processes.

When people perform recordkeeping in their everyday lives, they often engage in a series of decisions, perhaps without explicit awareness, for example: which documents need be destroyed; which documents need to be kept, why, and for what purposes; for whom the documents are being kept; and where the documents will be stored, in which order or categories, and for how long. Over time, most people will likely revise their previous choices made in the recordkeeping process, corresponding with changes in their lives. Records management norms or rules at work, for instance, may influence people's recordkeeping methods in their private lives. Also, personal preferences