

Minds of Carolina

Helen R. Tibbo and Paul Jones

*School of Information and Library Science, Univ. of North Carolina at Chapel Hill
Chapel Hill, North Carolina, USA*

Abstract

While all university faculty have substantial and growing bodies of work in the published scholarly, research, and pedagogical literatures, this constitutes but a small portion of useful material they generate throughout their careers. Much valuable content remains unpublished in personal collections. Due to small staffs, lack of space, and chronic underfunding, personal and scholarly papers of only the most notable faculty will be preserved in University Archives and manuscript collections. *Minds of Carolina* is a search for means to provide access to the extensive and rich unpublished contributions of faculty and preserve this material for the foreseeable future. It is an exploration to produce a feasible way to capture the works that would otherwise die on individuals' digital desktops, in filing cabinets, or on rapidly aging media.

The *Minds of Carolina* project seeks, through a variety of research methodologies, to explore the nature of these intellectual assets and develop tools, methods, and guidance to assist faculty (and potentially a much wider audience) in identifying materials of enduring and widespread value within their personal, unpublished collections; describing and contextualizing these materials to optimize their retrieval and use; and depositing them in a durable digital archive based on long-standing archival principles and practices.

Archiving in the Digital Age

Digital preservation and archiving stand as grand opportunities and challenges of the first decade of the 21st century.⁹ Digital technologies allow us to create, manipulate, store, and make accessible all manner and amounts of information never before possible, yet these same technologies imperil the longevity of the very objects they produce and thus the "memory of the information society".² Seamus Ross has predicted that "digital archives combined with new technologies will liberalise scholarship..." but that "this vision of a rich information record just waiting to be harvested and processed by the technology-enabled researcher of the future depends upon the survival of digital data" and that "based on current experience, it is evident that not much of this digital material will survive."⁴ While digital data holds the promise of ubiquitous access, the inherent fragility and evanescence of media and files, the rapid obsolescence of software and hardware, and the need for well-constructed file systems and metadata offer little hope

of longevity for information that is not intentionally preserved.

The Need for Ubiquitous Personal Digital Archiving

In *Invest to Save*, the NSF-DELOS Working Group on Digital Archiving and Preservation observes, "the need for digital preservation touches all our lives" as we "use, trust, and create e-content and expect that this content will remain accessible..."⁴ This report further notes that "Many organizations, businesses, government agencies, and even private citizens will need digital archiving mechanisms in order to retain access to their own records in the face of constantly changing information and communication technologies." The research agenda presented in this report includes the need for designing "digital archiving tools that are inexpensive, reliable, widely available, durable, interoperable and easy to use" for individuals.⁴

At this time no self-archiving project such as *Minds of Carolina* exists and the concept is little explored in the archival literature,^{6,10} but there is widespread concern for loss of important but inherently ephemeral and fragile faculty contributions. For example, in 2003 there was a running thread on the Humanist listserv regarding the theoretical ramifications of self-archiving.⁵ In a related vein, MIT's DSpace initiative (www.dspace.org), launched in the fall of 2002, provides a structured and centralized environment in which faculty can deposit materials. While DSpace accommodates a wide variety of formats, MIT's initial community-based implementation focuses primarily on structured scholarly products such as pre-prints and does not bring archival perspectives and practices to contributors. Several other universities including Cambridge, Columbia, Cornell, Ohio State, and the universities of Rochester, Toronto and Washington are adopting the MIT model, while other institutions are developing Fedora repository software (www.fedora.info).

Clearly, there is a move towards contributor-populated digital archives, whether they are comprised of pre-prints or personal papers, and archivists need to more fully understand the motivations that lead people to deposit their materials into these untested systems, and their needs and requirements for satisfaction. Further, if they want to ensure that these repositories are historically rigorous, evidential in nature, and reliable, archivists also need to understand the nature of the materials contributors might self-select, and the ways in

which contributors might want to present that information. *Minds of Carolina* provides a rich framework within which to explore the possibilities and pitfalls of archiving in the digital age.

Preserving Our Intellectual Heritage. Why Faculty?

While all university faculty have substantial and growing bodies of work in the published scholarly, research, and pedagogical literatures, this is but a small portion of useful material they generate throughout their careers. Due to small staffs and chronic under-funding, such personal and scholarly papers of only the most notable faculty are preserved in most University Archives or manuscript repositories. There is simply neither time nor space to process and store the significant materials from the range of faculty. There is even less hope that this material, much of which presently exists in paper or other analog formats, will be digitized through present university facilities and be made available to the world at large. Perhaps most alarming, materials already in digital form today have little chance of being accessible in 2014 if active preservation measures are not undertaken now and throughout their useful lifespan.⁴

Because of limited resources, colleges and universities have never tapped the potential in personal collections of their faculty. Three factors make the continuation of such neglect and oversight unacceptable today and speak to the need for support of ubiquitous personal archiving. First, digital materials will not wait to be preserved until the end of a scholar's career. If steps are not taken early then there will be nothing for a university archivist to preserve for future generations. Second, through digitization and networked technologies, faculty have new opportunities to share significant unpublished intellectual assets, be they originally paper-based or digital, with an array of previously unimagined audiences. While print materials withstand benign neglect far better than digital or media files, access to them is always limited in that it requires researchers to travel to them or copies be made of them. Thus they remain hidden and esoteric for all but the most diligent researchers. Third, the development of digital repository software such as DSpace, Fedora, Osprey (www.ibiblio.org/osprey) promises feasible means to build durable digital archives in which individuals can deposit their own materials at low cost to the institution. While the challenges of digital preservation are far from solved, retention in a centralized, curated digital repository is much more promising than on individuals' workstations and a variety of obsolescing media.

In 1994, Margaret Hedstrom called for archivists to develop strategies to change individuals' record-keeping practices, to increase their awareness of the value of documentary evidence, and to produce tools to help in these processes. Little progress has been made on these fronts. To realize both the access and preservation potential of digital technologies, faculty must take a more active role in the management and maintenance of their own collections. Time is of the essence (NSF/LC, 2003) and distribution of the

labor-intensive process of archiving as well as repository cost control are also essential. Just as distributed computing has democratized information creation and flow, distributed or "post custodial" digital archiving, especially of personal papers, may be the most viable, and perhaps the only feasible, means to preserve the array of significant intellectual assets from university faculties.

This is not to say that the need for archival repositories will disappear. Indeed, the presence of trusted repositories is the cornerstone of much electronic record preservation research and theory. But whereas archiving has long been considered a post hoc activity to house and preserve documents of enduring value after their primary role is completed, creators of digital materials, in this case, faculty, must now become their own "archivists" if today's inherently fragile unpublished materials are to be saved in any repositories, be they personal or institutional.

Minds of Carolina: Bringing Archival Perspectives and Expertise to Self-archiving

The archivist's skills and standards are an essential part of the process of making digital information available and sustainable over time. If materials are to be preserved and made universally available in a coherent and context-rich environment that will elucidate their content and reveal their creation process, they must be managed and maintained as they would be in an archival repository. Lacking extensive archival resources, we must find ways that facilitate the creator, in this case, the university scholar, to be his or her own archivist.

Minds of Carolina is about developing tools, methodologies, and guidelines to help scholars at the University of North Carolina at Chapel Hill prepare their materials, be they initially in analog or digital form, for archiving within a trusted digital repository at UNC built on a DSpace-like architecture. *Minds of Carolina* is the search for a means to bring the extensive and rich contributions of Carolina faculty to the world and preserve this material for the foreseeable future. It is an exploration to produce a feasible way to capture the works that would otherwise die on individuals' desktops or on rapidly aging media in a matter of a few years.

The first phase of *Minds of Carolina* project has twelve main objectives:

1. *Survey the literature for any work in the area of self-archiving and interview project managers/digital library designers at all of the DSpace Federation institutions (www.dspace.org). We have begun this in the pilot study and have found other institutions have similar challenges with the DSpace software even though their repository design has not focused on individuals as contributors.*
2. *Survey the nature of materials in the personal collections that faculty have to contribute to the Minds of Carolina repository and explore how they would make such selections, that is, appraise their materials for deposit into the archive and who they would see as*

the primary audience for them. This step will provide evidence of the nature and extent of work necessary to preserve and present the most valuable contributions out of these collections. This will help us establish resource needs and the overall feasibility of the project. We anticipate a large proportion of these materials will exist in analog formats and require digitization and retroconversion to make them accessible. This step would involve some cross-campus surveys as well as in-depth interviews with a smaller number of individuals. We will survey randomly to ascertain interest in such a repository and interview selectively, especially those individuals who express a desire to deposit their materials in MOC.

3. *Explore how faculty would make selections for deposit in a MOC repository: that is, how would they appraise their materials for deposit into the archive and who would they see as the primary audience for them.* Contributors need materials to help guide them through the process of appraising their personal collections and selecting appropriate materials for archiving. Without the presence of archival gatekeepers, users will have more control over their personal papers – the *Minds of Carolina* team would like to ensure that these personally controlled and collected papers are archivally sound – that the documents have enduring value, that they are evidential in nature, and that they carry weight and meanings for future generations. We need to understand what types of selections faculty would make to 1) better understand what is valuable to them and why (not just accept traditional archival practice), and 2) create guidelines that will be meaningful and which faculty will follow while providing archival guidance in the form of prompts and questions to ask about one's materials. In step 2 we would survey what faculty have in their collections. In step 3 we are looking at what faculty creators see as being important within their accumulated materials and how and why they come to such judgments. We would also ask faculty whom would they see as the primary audiences for these materials. We know from our pilot study that this may well not be just other scholars or members of their discipline. A repository such as *Minds of Carolina* may well have extensive benefits beyond its university home.
4. *Develop a demonstration repository website.* Starting with a select and targeted first-adopters group of contributors we can build a demonstration repository that will provide tangible examples of self-archiving and information sharing within this environment. This will provide the environment in which we can test the facility of the repository software to accommodate the types of materials faculty wish to deposit, to support the relationships among the objects within a deposit or personal collection, and to provide a contextualizing structure and content necessary to make the contributions meaningful. We will also test templates for deposit and metadata creation in this environment.

Interface Design (for contributors and users): *Minds of Carolina's* two initial contributors are very different people. Their collections are different, their work is different, and the way they think about their materials is different. Notwithstanding the selection and appraisal issues inherent in these differences, we would also like to provide them with different kinds of interfaces, both on the back end, for different kinds of collections, and on the front end, for users accessing these collections.

Front end: through our discussions with our pilot study participants, we have found that the faculty member in the School of Medicine is interested in conveying very different ideas and sentiments than the faculty member from the College of Arts and Sciences. Although the underlying hierarchy and system will be the same for all contributors, we would like to provide templates for display over which the user has some level of control. This is a simple enough problem to overcome with the use of XSL stylesheets – to use stylesheets with any level of power we must develop a robust XML markup schema with a well-formed document type definition. This issue is intimately related to metadata generation.

Back End: The *Minds of Carolina* system will need to be flexible enough to provide structure for different organizational schemes. In the pilot study, one of the contributors, the Chinese translator, is providing traditional archival materials: workbooks, correspondence, publication information, etc. along with a narrative description of what these materials are and why they're significant. The other contributor, the motor neuron researcher, is providing an in-depth narrative with links and pointers to other contributed work. These collections are very different in nature, although both are providing valuable archival information. The system must have the flexibility to intake collections that have variable navigational elements, non-standardized finding aids, and different format objects.

5. *Develop guidelines and training tools to help students to work with contributors.* Our starting assumption, supported thus far in our pilot study, is that faculty will need assistance in creating their own digital archival collections. First, they will need simple, straightforward guidelines to help them make useful selections, digitize analog materials, create metadata, contextualize and explain individual digital objects, and deposit their materials. Second, they will need some ongoing human intervention and support.

SILS students are a key element in this process. Synergistically, implementation of the project relies on the knowledge and expertise they develop in SILS classes while the experience of working with *Minds of Carolina* contributors will greatly enhance the educational process. To work effectively, students in the future will need guidelines and tools developed during this first phase of the project. These tools will focus on materials appraisal, metadata creation, digitization best

- practices, transcription procedures, and end-user instruction.
6. *Develop guidelines to help contributors prepare their own materials for deposit.* Ultimately, the goal is to create an environment and resource base that will support fast and effective self-archiving. Along with creating guidelines for students who will remain a contributor resource in future phases of the project, clear and illustrative guidelines for digitization and deposit are needed for contributors. These will reflect emerging standards within the digital archiving community and established best practices and focus on content appraisal and selection, description and metadata creation, and content contextualization.
 7. *Explore with our early adopters, the best models for presentation of their materials within the repository.* The hallmark of archives has been arrangement of resources and contextualizing description in the form of finding aids. This is a labor-intensive process that we cannot expect collection contributors to complete in the traditional fashion. We will explore the efficacy of alternative models suitable to the materials and a digital environment. One model might be linking digital documents to the contributor's vita. This investigation will link the need for specific metadata and templates for easy creation and deposit with a variety of presentation templates created for the model repository
 8. *Create Minds of Carolina policy statements on issues including archiving obligations, liability, intellectual property rights, security, contributor eligibility, and appropriateness of contributions.* A number of repository policies must be established during the pilot phase of this project. These are necessary for ethical, legal, security, and preservation reasons. Such guidelines, will not only direct the activity of this project but may also have the salutary effect of guiding non-*Minds of Carolina* website producers and helping them sort through complicated and nationally important issues. In the spirit of the open source movement we will work with the Creative Commons to foster the use of their sharing licenses in lieu of traditional copyright statements when appropriate.
 9. *Develop templates for easy deposit of contributions to the Minds of Carolina repository.* The ultimate goal is to have a range of contributors, varying widely in computing expertise, to deposit their materials with *Mind of Carolina*. For this to happen we must develop very easy to use tools, along with the document preparation guidelines noted above, that will assist contributors in use of standards and best practices. Once the contributors have chosen appropriate materials, there must be some mechanism to transfer those materials into the system. This transfer, if we want to address archival issues such as keeping hierarchical and contextual information, is not a trivial problem. *Minds of Carolina* contributors need an archival system that can: intake hundreds of hierarchical documents in one go, provide metadata generation tools for collection level, group level, and object level description, and provide tools for capturing textual and contextual information.
 10. *Seek on campus partners such as the Academic Affairs Libraries and the Center for Public Service.* If this project is to be sustained long-term, it will need to involve numerous UNC-Chapel Hill partners such as the Academic Affairs Library, the University Archives, Institutional Research, the Center for Public Service, and the UNC Digital Library. Lacking any prototype, it is difficult for units to support projects in concept. While we are making contacts across campus, we believe it fairer to ask them for support and involvement once we have a prototype to demonstrate.
 11. *Develop workflow models and cost projections and assess the feasibility of self deposit for faculty within a durable university-based archive.* At this time we simply do not know how much assistance faculty will need in the self-archiving process. Once we have developed guidelines and have models of description and presentation prepared, we will work with a small group of faculty to ascertain how much assistance they need in this process and compile data to assess its overall feasibility.
 12. *Develop an evaluation component to test how effective the prototype repository and its policies and procedures are.* No project should carry forth from its pilot phase without serious evaluation of all its elements. With the assistance of other SILS faculty, expert in evaluation and user studies, we will analyze each element of the prototype and the procedures before seeking funding for the next stage of this project.

Minds of Carolina Pilot Study

The Minds of Carolina pilot study matched two faculty members with two graduate students specializing in archives. The faculty members were self-selected, as they approached the investigators before the beginning of the project, and expressed a desire to not only make their materials available online but also to preserve their work for future access and use. Both faculty members are near retirement and have large bodies of work in both digital and non-digital formats. Some of this work is well organized, while other work takes the form of notes interspersed with ephemera; some work is easily recognizable as useful to the general public and some of interest mainly to specialists. One contributor is a medical researcher – on the cellular level, thus reducing medical privacy concerns– and one is a translator of Chinese poetry. Both are exceptional researchers and teachers, but the neither of their collections fit within the current collection development plans of the Manuscripts Department or the University Archives program at UNC-Chapel Hill. Thus, we started with two contributors whom had valuable non-published records but lacked opportunity and location to provide widespread access and long-term preservation of their work.

While the project is only about six months along, we have gained extensive insight into the process and

requirements necessary for effective self-archiving. These requirements involve both technical, repository issues and human factors issues made evident in working with our contributors. Thus far we have accomplished several of our initial objectives:

1. Extensively interviewed our two contributors,
2. Assessed their existing collections,
3. Assisted with retention / digitization decisions,
4. Acquired, and where applicable, are digitizing these collections, and
5. Are beginning to deposit these collections into a demonstration system.

Our progress has drawn attention to other issues, namely: usability, design, workflow, context, and long-term preservation within this system – as well as some thoughts about the sustainability of the archiving system software and design itself.

Some Lessons from the Interviews

The project began with a series of interviews with the contributors. Our interviewers quickly realized that each contributor envisioned a different archival representation of his materials – one interested in preserving a narrative of his life's work; the other interested in the circumstances of the creative process. Each also envisioned a different audience, both of specialists and the general public. Each, although both had used archives in their research, harbored reservations about archives and had different ideas about the function and role of archives in a university setting. The two initial contributors have provided us with such different perspectives, materials, goals, and approaches that we now see a necessity for the development of a variety of metadata and presentation models to accommodate the diversity of material we expect from a campus-wide audience.. As our interviewers note in their interim project report¹¹: “The locus of that dissimilarity may be cultural (humanism versus the hard sciences), materials based (discrete units of review – a book and all of its related materials, for example – versus thematic reflections on an entire career), or related to the digital archive's intended primary audience.” Based on these observations and hypotheses, they conclude that *Minds of Carolina* will need to conduct further research to

1. Ascertain why these differences exist, through more extensive entrance interviews with a greater number of prospective contributors;
2. Advance theories as to how these narrative dissimilarities can or should be incorporated into standard digital archival practice; and
3. Develop archival-narrative templates to simplify the submission procedure for future contributors and enrich the archival experience for system users.

Further, one interviewer has developed a series of oral history questions and concepts to assist future contributors and interviewers in answering questions of assessment, retention and audience.

Brief Notes on Metadata

We have investigated several metadata schemes to underlie the *Minds of Carolina* repository. Dublin Core(<http://dublincore.org/>) is a fundamental component of creating discovery metadata for this project as it is a minimum requirement for Open Archives Initiative (OAI) repositories. Because we seek a design for a durable archive, we are also concerned with preservation metadata. Thus far, we see the preservation metadata scheme and data model developed by the National Library of New Zealand (NLNZ) to be the most promising for MOC (http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf.) While we are developing an OAI-compliant schema for use with the project based on these tools, some questions remain: How should we deal with document interrelationships and other structural elements? Is the Library of Congress' METS standard the answer? If so, how much of this work could be done by the contributors themselves?

Questions and Next Steps

In our initial uploads of materials into the DSpace environment, we noted workflow, systems architecture, and interface/interaction issues.

Workflow issues identified at this time are mostly concerned with the amount of work to be done by the contributors themselves. For example, to what degree will contributors participate successfully in the creation of complete and accurate metadata? What processes, tools, and guidance can we develop that will help assist in simplifying the metadata creation process? What types of materials will contributors select for deposit and how would this relate to typical archival appraisal judgments? Perhaps even more fundamentally, can an individual who possesses an extensive body of work select materials of enduring value that others will find useful?

Systems architecture issues reflect problems inherent to the early stages of digital durable storage and archives. In particular, the DSpace software imposes limits to the extent we can create analogs to the physical archives. Future explorations will need to focus on how much of the traditional archival descriptive infrastructure, that is, the finding aid, is necessary in the digital repository? What presentation models work well? Questions about sustainability of the DSpace software itself are also important. Will such a system, that is on one hand quite complex and dependent on several evolving software projects itself, be usable in the long term or even in the medium term? Would we be better off considering a simpler more durable software architecture, such as flat files over an SQL database?

Interface/interaction issues present both challenges and opportunities. Provision of context in which to understand individual objects and files is a hallmark of the archival profession. Lacking binding and publication data, archivists turn individual pieces of paper, and now electronic files, into understandable components in much larger collections. Presentation of archival holdings could become more

complex and interesting than those in the physical archives, if, through rich metadata, various relationships and hierarchies among objects could be presented in different ways to different archive contributors and users. While much work in digital archiving presently focuses on repository architecture, provision of contextualized access is central to true archiving. User and contributor studies lie at the foundation of durable archives. Paul Conway¹ has argued, access equals preservation in the digital world. Over time, institutions will only expend the resources necessary to maintain those materials that are used. Thus digital archive design requires functions beyond ingest, storage, and retrieval. As in print archives, long-term preservation depends upon preserving the meaning of the past for future generations.

Acknowledgements

The authors wish to thank Megan Winget and Marisa Ramirez for their unflagging efforts on the *Minds of Carolina* pilot study. They have worked closely with the initial MOC contributors and their interim report formed the basis of our preliminary findings presented here.

References

1. Paul Conway, "Overview: Rational for digitization and preservation," In *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Northeast Document Conservation Center, Andover, MA. (2000). <<http://www.nedcc.org/digital/dighome.htm>>
2. DLM-Forum, *Proceedings of the DLM-Forum on Electronic Records: European Citizens and Electronic Information: The Memory of the Information Society*. Brussels, October 18-19, 1999. European Communities, Luxembourg. (2000). <<http://europa.eu.int/ISPO/dlm>>
3. Margaret Hedstrom, "Electronic Archives: Integrity and Access in the Network Environment," In *Networking in the Humanities: Proceedings of the Second Conference on Scholarship and Technology in the Humanities, held at Elvetham Hall, Hampshire, UK, 13-16 April, 1994* (S. Kenna and S. Ross, Eds.). (1994). Bowker-Saur, London.
4. Margaret Hedstrom and Seamus Ross. *Invest to Save*. Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation. (2003). <delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/digitalarchiving/Digitalarchiving.pdf>
5. Humanist. "Self-archiving" is the subject line in Humanist 2003 issues: 16.417, 16.423, 16.429, 16.435, 16.439, 16.442.
6. Sue McKemmish, "Evidence of Me..." *Archives and Manuscripts* 24(1), 28-45 (1996).
7. National Science Foundation. Digital Government Program & Digital Libraries Program. & Library of Congress. *It's About Time: Research Challenges in Digital Archiving and Long-Term Preservation*. Final Report Workshop on Research Challenges in Digital Archiving and Long-Term Preservation. April 12-13, 2002. (2003).
8. Ross, Seamus. *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship*, NPO Preservation Guidance Occasional Papers. National Preservation Off., London. (2003). <<http://www.bl.uk/services/preservation/occpaper.pdf>>
9. Tibbo, Helen R. "On the Nature and Importance of Archiving in the Digital Age," *Advances in Computers*, v. 57, pgs. 1-67. (2003).
10. Frank Upward and Sue McKemmish, "Somewhere Beyond Custody," *Archives and Manuscripts* 22(1), 136-149 (1994).
11. Megan Winget and Marisa Ramirez, "Minds of Carolina Pilot Study. Interim Report. SILS Technical Reports TR2004-01. (2004). <http://www.ils.unc.edu/html/5_tech_report_series.shtml>

Biographies

Helen R. Tibbo is a Professor in the School of Information and Library Science (SILS) at the University of North Carolina at Chapel Hill. She earned her Ph.D. in Library and Information Science from the University of Maryland. She teaches in the areas of archival and records management, digital preservation and access, and electronic information retrieval. As well as directing the Minds of Carolina project with Paul Jones, Dr. Tibbo is principle investigator for the Managing the Digital University Desktop Project that is exploring how faculty, staff, and administrators at UNC and Duke universities manage their email and digital files.

Paul Jones received his B.S. degree in Computer Science from North Carolina State University and a MFA in Poetry from Warren Wilson College. He is director of ibiblio.org, a public digital library located at SILS. He is also a Clinical Associate Professor in the School of Journalism and Mass Communication and in the School of Information and Library Science.