

Keeping the Momentum: Moving Ahead with Research Data Support

Hilary Davis

North Carolina State University at Raleigh, hilary_davis@ncsu.edu

Steve Morris

North Carolina State University at Raleigh, steven_morris@ncsu.edu

Barrie Hayes

University of North Carolina, bhayes@email.unc.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/charleston>



Part of the [Library and Information Science Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at: <http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Hilary Davis, Steve Morris, and Barrie Hayes, "Keeping the Momentum: Moving Ahead with Research Data Support" (2012).

Proceedings of the Charleston Library Conference.

<http://dx.doi.org/10.5703/1288284315126>

Keeping the Momentum: Moving Ahead with Research Data Support

Hilary Davis, Associate Head of Collection Management and Director of Research Data Services, North Carolina State University

Steve Morris, Head of Digital Library Initiatives and Digital Projects, North Carolina State University

Barrie Hayes, Bioinformatics and Translational Science Librarian, University of North Carolina Health Sciences Library

Abstract

From June 2011 to early 2012 the NCSU Libraries and the UNC Libraries took part in the ARL/DLF E-Science Institute to frame a strategic agenda for supporting research data management and its broader e-science needs at our universities. We conducted an environmental scan, interviewed key researchers and administrators, and participated in capstone meetings with peer institutions. Our two institutions represent two strategies with varying degrees of divergence and convergence. At the NCSU Libraries, with no repository explicitly designed for research data, we are focusing on developing a portfolio of services and partnerships to create a “campus collaborative” of experts, tools, and training to support research data. With limited or unbalanced domain expertise, we are rethinking how subject specialists can be deployed to serve diverse research needs. At the UNC Libraries, we have an institutional repository, but recognize that it cannot serve all data management needs across campus. We, too, are developing a cooperative network of campus partners to guide researchers to various campus services at their point of need. The Carolina Digital Repository, UNC’s institutional repository, is one option among these services as is helping researchers identify disciplinary repositories where appropriate. Both institutions are particularly interested in exploring the long term possibilities of creating cultural shifts in research data stewardship by educating graduate students and early career researchers, and the ways in which regional library consortia can partner in data management support in the same way we’ve partnered on other issues. Reflecting on these two institutions’ goals, we will discuss the opportunities and challenges centered on supporting data-driven research. We’ll share our plans for next steps and invite discussion on how to respond to those opportunities and challenges in practical, achievable, sustainable, and repurposable ways with limited human, technological, and financial resources.

Introduction

From June 2011 to early 2012 the North Carolina State University Libraries (NCSU) and the UNC-Chapel Hill Libraries (UNC) took part in the ARL/DLF E-Science Institute to frame a strategic agenda for supporting research data management and its broader e-science needs at our respective universities. We each conducted an environmental scan, interviewed key researchers and administrators, and participated in capstone meetings with peer institutions.

Our two institutions represent two strategies with varying degrees of divergence and convergence. Both institutions are particularly interested in exploring the long term possibilities of creating cultural shifts in research data stewardship by

educating graduate students and early career researchers, and we are keenly interested in the ways in which regional library consortia can partner in data management support in the same way we’ve partnered on other initiatives. Reflecting on these two institutions’ goals, we discuss the opportunities and challenges centered on supporting data-driven research. We review our plans for next steps to respond to those opportunities and challenges in practical, achievable, sustainable, and repurposable ways with limited human, technological, and financial resources.

UNC’s participation in E-Science Institute coincided with a campus initiative focused on data stewardship framed around the work of a task force which conducted an environmental scan of

research data policies and trends and interviewed and surveyed researchers to gather information on the range of the activities, concerns and opinions surrounding data stewardship at UNC. At NCSU, at the commencement of the E-Science Institute, there was no single unit coordinating full-campus dialogue about research data management and discovery; instead, we observed a diverse, disorganized collection of approaches and solutions and an overall sentiment that the data management planning needs were “unfunded mandates” by funding organizations.

Insights from the E-Science Institute

The E-Science Institute enabled us to uncover some important insights about research data needs at our institutions. On a practical level, data storage, sharing, and discovery are the most immediate needs. Researchers also expressed a need to get access to data for computational analyses, and supported open data. Not surprisingly, sharing data among inter-institutional teams emerged as a major trend. Navigating the institutional review board (IRB) process also emerged as an area in need of much more campus support. A concern that was expressed at both institutions was a lack of cohesive communication across campus; many researchers are not aware of campus resources available to them. At NCSU, researchers expressed interest in more opportunities to have a shared dialogue around research data support, analytics, visualization; data management planning; and IT computing support.

On a cultural level, the amount of interest in data sharing and discovery was more developed than we expected. There has been an assumption that some of the campus project-produced data (as opposed to externally available data from organized sources) is one-off in nature or too unique or too specialized or too standardized to have a market for reuse. We expect more adoption of data sharing in disciplines where data is highly reusable (e.g., geospatial, social sciences, and genome sciences). Our experiences from conducting interviews during the E-Science Institute suggest that the nature of the actual market for reuse deserves some more study.

Having completed the ARL/DLF E-Science Institute, our primary take-aways for gaining momentum to develop research data services is to prioritize conducting interviews with campus stakeholders and to partner with a non-librarian, such as a member from Research Administration. Interviews with researchers, administrators, and librarians provided a tremendous amount of insight into trends on campus, as well as barriers and needs for data-driven, team-based research. UNC partnered with the Associate Vice Chancellor for Research, and by doing so, developed a strong advocate for the Library as a key partner for a campus-wide effort.

Motivations and Goals

We feel strongly that the opportunities inherent in becoming an integral part of the research process far outweigh the risks. Moreover, we think that not responding to the changing research environment or maintaining stasis is not an option. If not engaged, we face the risk of potentially enabling a view that the libraries do not directly impact research progress at our institutions. If libraries are not proactively engaging in the issue (at minimum, through building awareness, gaining expertise, engaging in pilot activities), we risk being unprepared at a later point if the university approaches the library about taking on a greater role in some aspects of e-research support. The bottom line in all of our work and intentions is to make it easier for faculty to write their data management plans, easier to know what best practices to follow (for all aspects of data management), easier to find appropriate storage environments, and easier to gain access to and discover datasets for their work. We are also in a good position to advocate for researchers' needs as we work to foster collaborative solutions across campus.

Campus Partners and Critical Players

Campus partnerships are absolutely critical in developing a sustainable, comprehensive, and relevant program of support for data-driven, team-based research. Both library systems are partnering with similar units at each institution. Research Administration, in particular the UNC Office of the Vice Chancellor for Research and the

NCSU Associate Vice Chancellor for Research Administration, have the potential to provide incentives for researchers to adopt data stewardship best practices and funding structures for data stewardship applications and expertise on campus. The campus Information Technology units, in particular Advanced Computing at NCSU and Research Computing at UNC, are developing robust, secure, backed-up, campus data storage environments. Both library systems are working to build partnerships at the college/school level where, to varying degrees, technical infrastructure and research administration support also exist.

The NCSU Libraries' internal partners include the Digital Library Initiatives, Collection Management, Research and Information Services, Copyright and Digital Scholarship Center, and Engineering Services. The UNC Libraries internal partners include comparable units and teams that span the University Library, Kenan Science Library, and the Health Sciences Library. Other key players include the Odum Institute, School of Information and Library Science (to develop training for data curation practitioners), the Carolina Digital Repository, and the Renaissance Computing Institute (RENCI).

Current Landscape for Research Data Support

Campus-Wide Initiatives

UNC's and NCSU's current approach for research data support can be characterized as distributed with many stakeholders working toward greater coordination.

At UNC, campus-wide discussion began in 2011 when the Office of the Provost charged a Task Force on the Stewardship of Digital Research Data to investigate stewardship needs at UNC and provide recommendations. A report was released in February 2012, and a follow-on task force has recently begun work to examine issues and recommend data stewardship practices for UNC genomics data.

At NCSU, the loose confederation between IT, Research Administration, and the Libraries, has had the specific aim of supporting NCSU grant

seekers in their need to compile a data management plan as part of their grant funding request. But in terms of research data storage, access, discovery, and sharing support, NCSU does not have an integrated system that is positioned for these functions/needs. Campus-wide dialogue is underway as of October 2012 to begin laying the groundwork, including a "Big Data" forum to bring together disparate research units across campus. The Office of Information Technology is taking the lead on exploring storage options; the Research Administration is redoubling efforts to help support for faculty in finding grants.

Library Initiatives

The UNC Libraries and the NCSU Libraries both have a long history of providing support and consultation for accessing and using commercially produced and other external data sets, such as ICPSR and geospatial data and software.

Both library systems have developed facilities to support visualization and collaboration. The UNC Health Sciences Library partnered with RENCI (in 2005) to implement a Collaboration Center featuring a high-resolution display wall to support data visualization. The NCSU Libraries is expanding a rich portfolio of visualization technologies (hardware and software) with the advent of the new Hunt Library, extending the possibilities for researchers to showcase their work and experiment with data visualization.

Repository infrastructure is an area of divergence between UNC and NCSU. UNC's institutional repository, the Carolina Digital Repository (CDR) employs Fedora Commons (Repository platform) in combination with iRODS, which serves as a distributed storage and preservation system. It was developed with the capacity to accept and provide long-term preservation and access for many different types of deposits, including research data sets. The CDR can provide a home for data that do not have another natural home. In addition to the preservation services offered through the CDR, UNC's Carolina Digital Library and Archives (CDLA) offers a wide range of project planning and management services to support UNC faculty's digital research. While the NCSU Libraries has had a DSpace-based digital

repository system in place since 2006, supporting scholarly publications, ETDs, and technical reports, support for datasets is very limited.

Both library systems support data management planning, with the UNC Libraries' support being more expansive than the NCSU Libraries' recently implemented and evolving program. The UNC Libraries' Data Management Planning Committee began work in 2010 and developed one of the earlier web guides on data management planning (deployed in December 2010). The NCSU Libraries' Research Data Committee convened in August 2012. Both NCSU and UNC web guides include campus resources, samples of data management plans and language, as well as templates. We have both implemented the DMPTool managed and developed by the California Digital Library and partners. Current data management plan consultation support comes from a combination of staff from the Libraries, the Office of Sponsored Research, and in UNC's case, also from the Odum Institute. In addition to data management planning consultation, the UNC Libraries has offered an ongoing and well-attended series of information sessions on data management topics in collaboration with the Odum Institute and other partners. Session topics have included the NSF data management plan requirement, campus data repositories, considerations for handling sensitive data, data security, the DMPTool, and most recently, research computing resources. All sessions had in-person and online attendance options, and all sessions were recorded and posted online for later consumption.

Staffing for research data support reflects areas of current convergence and divergence between our institutions. Position descriptions for some librarians are changing at both institutions to provide more focus on research data support. At the NCSU Libraries, subject specialists are expected to include research data support in their responsibilities, and one position has been adjusted to add leadership capacity for research data services. At the UNC Libraries, two positions have been revised to include more explicit research data management support responsibilities, and UNC has recently recruited a new information services head for the Kenan

Sciences Library who will have leadership responsibilities for our data management and eResearch support initiatives. The UNC Data Management Committee is developing data management education sessions for subject librarians to increase their capacity to offer basic front-line research data support. Also, integrated research support capacity is now provided by a new embedded librarian service model. In 2011, UNC's NIH funded Clinical and Translational Science Award unit, called NC TraCS, negotiated with the Health Sciences Library to expand on existing library services and to hire and manage a new librarian position completely dedicated to and funded by NC TraCS. Since June 2011, the TraCS Knowledge Management Librarian has been providing a number of grant-seeking and knowledge-management services to TraCS and its affiliated researchers. Example services include targeting funding opportunities to specific research teams and managing researcher profiles in REACH NC, the Scopus-based expertise profiling system developed and used by UNC. The NCSU Libraries' staff support for research data services is distributed amongst a group of subject specialists, library administration, and a dedicated Copyright and Digital Scholarship Center that provides guidance on copyright and intellectual property issues for campus researchers, as well as NIH-sponsored publication access. The NCSU Libraries also has a current metadata librarian position vacancy, creating a gap in being able to provide metadata consultation in support of data management planning.

Challenges

The challenges faced by both of our libraries are not unique. In both cases, but to varying degrees, UNC and NCSU are reliant on broader campus support for IT infrastructure to support practical data storage, security, access, sharing, and discovery needs. At both UNC and NCSU, coordination among campus entities is evolving. Institutional incentives are not firmly in place on either campus to help foster broad adoption of data management best practices and open data initiatives. Currently, there is no direct source of financial resources at the library or university level for e-research support. Development of university

data stewardship and ownership policies will potentially help incentivize funding to build out infrastructure, expertise, and better coordination of activities.

Compared to the UNC Libraries, the NCSU Libraries is a bit thinner on dedicated staff capacity, for example, in the areas of digital repository development and direct bioscience capacity. Neither institution currently has dedicated library staff in the organization with experience developing business/cost models around these kinds of services, so scalability of services is a concern. For both NCSU and UNC Libraries, there is a risk that we could easily get drawn out of our depth in developing domain-specific metadata.

While UNC Libraries has a robust set of repository services, including support for datasets, the NCSU Libraries does not yet have capacity in its repository environment to support datasets (e.g., no DOIs for articles or other objects such as datasets). NCSU runs the risk that the repository may be viewed as a solution to data storage problems while not being suited for that role.

Building a program of services and tools to support research data and data management planning dictates that we also establish benchmarks for value and success to justify expensive resources and time invested. However, how we measure success in supporting data-driven, team-based research is not something that we have a solid grasp on yet.

Immediate Action Areas

Both UNC Libraries and the NCSU Libraries share many of the same plans for building out services and enhancing existing services to support data-driven, team-based research. Collectively, our plans for the immediate future or one year out include an agenda of training, outreach, building a campus collaborative, and supporting tools and facilities necessary to e-research.

Over the next year, we will enhance existing tools and investigate emerging tools for relevance in supporting research data needs and data management planning on our campus. Some

examples of these tools include the DMPTool, DataCite, DataBib, and ORCID. We will investigate researchers' needs related to data sharing and publication. We will update and re-envision maintenance of our web resources/guides supporting data management planning with a broader goal of developing a strategy for learning what makes a data management plan successful (or not successful). We will develop a campus collaborative and build a communication network targeted to researchers, students, and administrators. We will deepen our understanding of stakeholders' needs via interviews with researchers, administrators, and library staff with a particular focus on interdisciplinary centers and institutes. The NCSU Libraries will also be developing a program of training, workshops, and consultation. Internal training will be targeted to library staff. Workshops are going to be targeted to graduate students and junior faculty since they are often the default IT support for their research groups. The UNC Libraries has very recently begun to pilot research services with its Carolina Center for Genome Sciences, an interdisciplinary research center which includes faculty from genetics, evolutionary biology, chemistry, biostatistics, and computer science among other disciplines. In conjunction with that pilot initiative, we are investigating interest in these initial services and other potential research services with the larger community of health and natural science researchers, beginning with those based in the new Genome Sciences Building that opened in mid-October.

Looking Ahead—The Five-Year Horizon

Both of our Libraries are looking ahead and find much convergence in our longer-term goals. Those goals center on extending the type of e-research support that has long been provided for geospatial and social sciences data into other arenas, for example, data visualization, health and natural sciences, and humanities computing. Within 5 years, we hope that the library will have become an active connector across a broad range of research areas. And we hope that by building out a campus collaborative, we can enable our partners to build on each other's experiences and

build shared services, policies, training, and expertise.

For long-term, robust support of data-driven, team-based research we hope to develop dedicated staff to lead research data services efforts and visualization services (software, hardware, and training) and to work hand-in-hand with each other, with campus IT and with Research Administration, to weave a sustained vision for supporting research data services. We also hope to see investment in staff with expertise in additional scientific domains as well as in the digital humanities.

Opportunities for Collaboration

Because both institutions face a similar set of goals and challenges, collaboration on any scale becomes crucial. The UNC and the NCSU Libraries have a long history of collaboration, focused especially on shared negotiation and licensing for electronic collections, catalog records, and a

collaborative print retention program. We also share decade-plus experience in providing data services in the area of geospatial and social sciences data, and we both have deep interest in and commitment to transforming library services to support modern research needs.

In the realm of e-research support, opportunities for collaboration include: sharing opportunities to learn from and contribute to seminars/webinars, sharing domain expertise (library and non-library), sharing policies, consulting on data management planning best practices, shared licensing where feasible, and networking among librarians who are responsible for supporting research data services.

Slides for this presentation are available at <http://www.slideshare.net/hilarymdavis/charleston2012-keeping-the-momentum-research-data-support>.

References

ARL/DLF E-Science Institute. (n.d.). Retrieved from <http://duraspace.org/e-science-institute>

DataBib. (n.d.). Retrieved from <http://databib.org/>

DataCite. (n.d.). Retrieved from <http://datacite.org/>

DMPTool. (n.d.). Retrieved from <https://dmp.cdlib.org/>

NCSU Advanced Computing Unit. (n.d.). Retrieved from <http://oit.ncsu.edu/unit-ac/ac>

NCSU Copyright and Digital Scholarship Center. (n.d.). Retrieved from <http://www.lib.ncsu.edu/cdsc>

NCSU Digital Repository. (n.d.). Retrieved from <http://www.lib.ncsu.edu/repository/>

NCSU Libraries' Data Management Planning Guide. (n.d.). Retrieved from <http://www.lib.ncsu.edu/guides/datamanagement/>

NCSU Libraries' Hunt Library Visualization Technology. (n.d.). Retrieved from http://www.lib.ncsu.edu/huntlibrary/faq_all.html#technology

NCSU Sponsored Programs and Regulatory Compliance. (n.d.). Retrieved from <http://research.ncsu.edu/sparcs/proposals/>

ORCID. (n.d.). Retrieved from <http://about.orcid.org/>

Renaissance Computing Institute (RENCI). (n.d.). Retrieved from <http://www.renci.org/> 402 Charleston Conference Proceedings 2012

Research Data Stewardship at UNC report. (n.d.). Retrieved from http://sils.unc.edu/sites/default/files/general/research/UNC_Research_Data_Stewardship_Report.pdf

UNC Carolina Center for Genome Sciences. (n.d.). Retrieved from <http://genomics.unc.edu/>

UNC Carolina Digital Library and Archives. (n.d.). Retrieved from <http://cdla.unc.edu/>

UNC Carolina Digital Repository (CDR) . (n.d.). Retrieved from <https://cdr.lib.unc.edu/>

UNC Health Sciences Library Collaboration Center. (n.d.). Retrieved from <http://hsl.lib.unc.edu/collaborationcenter>

UNC Libraries' Research Data Toolkit. (n.d.). Retrieved from <http://guides.lib.unc.edu/researchdatatoolkit>

UNC NC TraCS Institute. (n.d.). Retrieved from <http://tracs.unc.edu/>

UNC Odum Institute. (n.d.). Retrieved from <http://www.irss.unc.edu/odum/>

UNC Research Computing. (n.d.). Retrieved from <http://its.unc.edu/research/its-research-computing/>

UNC School of Library and Information Sciences. (n.d.). Retrieved from <http://sils.unc.edu/>

UNC Vice Chancellor for Research. (n.d.). Retrieved from <http://research.unc.edu/offices/vice-chancellor/index.htm>