# LIS Programs and Data Centers: Integrating Expertise

Cheryl A. Thompson, University of Illinois at Urbana-Champaign
Matthew S. Mayernik, NCAR Library, National Center for Atmospheric Research (NCAR), University Corporation for Atmospheric Research (UCAR)
Carole L. Palmer, University of Washington
Suzie Allard, University of Tennessee at Knoxville
Carol Tenopir, University of Tennessee at Knoxville

**Abstract**
Scientific data centers have provided data services to research communities for decades and are invaluable educational partners for iSchools developing academic programs in data curation. This paper presents analyses from three years of internship placements at the National Center for Atmospheric Research and interviews with managers at prominent data centers across the country, as part of the Data Curation Education in Research Centers project. Key benefits of the internship program are identified, from the perspective of student learning and the contributions made by iSchool students to data center operations. The interviews extend the case results, providing evidence for potential data curation internship programs at data centers. The DCERC education model fosters integration of expertise across the iSchool and data center communities, enriching academic preparation with state-of-the-art practical experience in ways that are vital to the emerging data profession and its ability to meet the future demands of data-intensive research.
**Keywords:** data curation; iSchool education; workforce; data centers; research data expertise
**Citation**: Thompson, C.A., Mayernik, M.S., Palmer, C.L., Allard, S., Tenopir, C. (2015). LIS Programs and Data Centers: Integrating Expertise. In *iConference 2015 Proceedings*.

**Acknowledgements**: We wish to acknowledge funding by the Institute of Museum and Library Services (RE-02-10-0004-10) for the DCERC program, led by CIRSS at the University of Illinois, in collaboration with NCAR/UCAR Library, and the School of Information Sciences at the University of Tennessee. We also wish to express our sincere appreciation and acknowledge the contributions of Mary Marlino, Karon Kelly, Valerie Williams, and the NCAR staff members that served as data and science mentors.
**Contact**: cathmps2@illinois.edu, mayernik@ucar.edu, clpalmer@uw.edu, sallard@utk.edu, ctenopir@utk.edu

## 1 Introduction

As the iSchool community scales up research and education in data curation, extensive expertise in the curation of research data already exists in data and research centers, where data services have been evolving for decades. To date, the iSchool community has had limited interaction with these organizations. There are ad hoc cases where Library and Information Science (LIS) professionals have contributed to, and benefited from, collaborations with research data centers (Gold, 2007; Plale, et al., 2013), but the data professionals working in these facilities are largely disconnected from the information profession. Instead, their professional interactions tend to be with their counterparts at other data centers and with researchers in the fields they serve. As would be expected, they publish and attend conferences within those same spheres. The International Digital Curation Conference (IDCC) is one viable venue for interaction, but few data center professionals attend. Despite the limited overlap between these professional domains, each is experiencing new problems associated with the rapid growth in data-intensive research and open access data. .

The benefits of increasing interactions between the LIS and data center communities seem evident. Data initiatives are becoming pervasive in universities, research centers, and corporations. Many are cross-cutting efforts that bring together multiple sectors around shared problems and facilitate shared expertise and tools. Funding agencies investing in data infrastructure and data-intensive research are also increasingly expecting cross-organizational and cross-disciplinary coordination (see for example, the NSF's EarthCube initiative, http://earthcube.org). Cross-cutting data management and service initiatives require a workforce that can address end user needs, work with multiple disciplines, and advance open data practices (Atkins, 2003)—foundational principles of the information science profession (Bates, 1999). New educational programs for data curation and data science in iSchools have an opportunity to learn from, and to inform, the work in data centers, as both professional groups work to evolve best practices for the future of data services for data-intensive research.

This paper presents results from the Data Curation Education in Research Centers (DCERC) program. DCERC has developed a model for preparing Library and Information Science (LIS) master and

doctoral level students for careers in scientific data curation by advancing curriculum and providing field experiences in mature research and data centers. Based on the outcomes of student internship projects over three years and a set of interviews conducted with data center managers, we identify key factors in iSchool and data center collaboration and cross-learning, and the future transferability of the DCERC model.

## 2 Background

The growing demand for a qualified workforce to professionally manage, preserve, and ensure access to data has been recognized for many years (ACLS, 2006; Atkins, 2003; Rusbridge, 2007). New data positions and responsibilities have been established in research libraries (Maatta, 2013; Sierra, 2012; Lyon et al., 2013), as demand for data skills continues to escalate across sectors (TEKSystems, 2013, Manyika et al., 2011). In 2008, Swan & Brown identified three roles for data professionals—data librarian, data manager, and data scientist. However, the positions that have emerged are more diverse, including titles such as Data Curator, Data Engineer, and Scientific Information Specialist. As the data workforce evolves, educators face many challenges preparing professionals with the expertise required for the range of new local and global roles in the complex and ever-changing networked environment of digital data (Pryor & Donnelly, 2009; Varvel, Palmer, Chao, & Sacchi, 2011).

iSchools have been preparing information professionals for work with a variety of digital resources for many decades, with moderate advances in curriculum focused specifically on digital research data. A 2012 study of iSchools and LIS courses, identified 475 digital or data curation courses at 158 programs at 55 universities; of the courses, only 8% were data-centric, 11% were data-inclusive, and 27% were digital courses (Varvel, Bammerlin, & Palmer, 2012). Harris-Pierce & Lui (2012) focused specifically on data curation education, reporting that only a third of LIS programs offered graduate-level data curation courses. Some iSchools have also been active in providing continuing education for working professionals to develop new data skills (Renear et al., 2012; Hank, Tibbo, & Lee, 2010; Shadbolt et al., 2014).

The pedagogical value of internships is widely recognized; they provide students with vital first-hand experience that connects theory with practice. In particular, for LIS graduates with a Specialization in Data Curation, hands-on experience through internships and practicums has proven to be a key factor in their employability (Palmer, Thompson, Baker, & Senseney, 2013). However, at present programs are limited. A survey of recent graduates from 39 LIS schools in North America revealed a need for “more opportunities for practical/hands-on experience” in their programs (Marshall et al., 2010, p. 35). Moreover, arguably many of the most valuable internships for data professionals will be outside of libraries, in organization with advanced data operations. When students are placed in research libraries, it can be difficult to identify mentors with extensive expertise in data curation and eScience (Steinhart & Qin, 2012).

This paper describes a unique internship model for data curation education from the perspective of student learning and the hosting data center—the National Center for Atmospheric Research (NCAR). We highlight the workforce contributions iSchool students can contribute to established data services and the future potential for internships at other data centers.

## 3 DCERC Model

The DCERC program, funded by the Institute of Museum and Library Services (IMLS), is a partnership among the Graduate School of Library and Information Science (GSLIS), University of Illinois at Urbana-Champaign (UI); the School of Information Sciences, University of Tennessee at Knoxville (UT); and the National Center for Atmospheric Research (NCAR) Library, University Corporation for Atmospheric Research. This mutually beneficial partnership provides LIS students with unique and valuable experiences working with scientific data and mentorship by both data professionals and practicing scientists. In return, NCAR personnel are exposed to new curation perspectives, especially in relation to information organization, user services, and the broader professional landscape of data management challenges and solutions.

The DCERC model includes graduate-level coursework in data curation and research experiences at local universities for students before they begin their internship at NCAR in Boulder, Colorado. The Master’s students spend two months at NCAR during the summer, including an intensive kickoff workshop to orient the interns to the organization and its data curation activities. During the internship, students have opportunities to attend earth science and informatics seminars and other related activities at NCAR and other organizations, extending the scope of their interactions with data managers and scientists in the earth science community. (For additional information on the DCERC model, please see Kelly et al., 2013 and Mayernik et al., 2015).

Three cohorts of Master's students (2012-2014) have completed the program. The internship component was revised each year, based on preliminary results from the ongoing formative evaluation. In particular, changes were made to improve student preparation for working in a data-intensive research center environment and to increase mentor involvement in student selection and project design (Palmer et al., 2014). To date, 4 UI and 5 UT Master's students (n=9) have completed the full program. One master student completed two internships at NCAR, totaling 10 internships. In addition, two UI doctoral students have completed coursework and a 9-month (two semester) field experience at NCAR aligned with their dissertation research, with a third UI doctoral student participating in a 4-month (one semester) field experience at NCAR in Fall 2014 to complete further evaluation of the DCERC program.

## 4 Methods

This report draws on three data sources: the collection of DCERC student internship outputs, the DCERC formative evaluation, and a set of interviews with data center managers across the country who have not been involved in the DCERC program. The evidence from these sources provides complementary internal and external views on the expertise needed for the workforce developing and sustaining infrastructure and services in large data centers. In addition, the interview results address the transferability of the DCERC model to other data centers.

The student outputs include final reports and poster presentations, and related documents for the 10 master-level internships and 2 doctoral field experiences. Content analysis was conducted to identify themes related to data curation lifecycle stages, knowledge, skills, and tools. The results show an interesting balance between the knowledge and skills that students gained from the internships in relation to the unique LIS perspective and expertise that students brought to the data center.

The DCERC formative evaluation consists of data on students' satisfaction with the program, positive and negative features of the internship experience, and specific recommendations for improving the model. Students (n=8) and mentors (n=19) completed the post-internship questionnaire, soliciting feedback on the internship component. Eight focus groups were conducted with a total of 9 DCERC students and 19 mentors, exploring topics in more detail and discussing the whole DCERC model. Each focus group lasted an hour on average. Together, the evaluation results provide insights into the value of the student experience and cross-learning between data center staff and DCERC students.

The interviews were conducted with managers in research and data centers in the geosciences. A semi-structured format was used to collect information on operations and staffing for scientific data management, required knowledge and skills for the professionals working with data in their organization, and field experience opportunities for data curation students. Centers were identified using the list of federally funded research and development centers (FFRDCs), the attendee list from the EarthCube Data Facilities meeting (January 15-17, 2014), and prominent centers known to the authors. Of the 32 centers identified, 21 were selected to represent a variety of geoscience domains, organization types (e.g., government, academic, non-profit), and primary activities (e.g., data services, research and development). (See Table 1). Twenty interviews were completed between June and October 2014, averaging one hour in duration. While questions about the division of labor and staffing needs solicited the most in-depth responses, all interviewees were interested in discussing internship opportunities for data curation students. This paper focuses on the interview results related to the DCERC model transferability.

| Characteristics | Category | Number in sample (n=21) |
|---|---|---|
| **Org. Type** | Corporation | 2 |
| | Federal government | 4 |
| | Non-profit | 2 |
| | Research consortium | 1 |
| | State government | 1 |
| | Universities/University consortium | 9 |
| **Domain** | Atmospheric science | 1 |
| | Earth science | 3 |
| | Ecology | 2 |
| | Environmental science | 1 |
| | Geology | 2 |
| | Geophysical | 1 |
| | Glaciology | 1 |
| | Information/computing | 3 |
| | Interdisciplinary | 1 |
| | Oceanography | 2 |
| | Seismology | 1 |
| | Space technology | 1 |
| | Urban science | 1 |
| | Weather | 1 |
| **Activity** | Data services only | 5 |
| | Research only | 2 |
| | Research & development | 3 |
| | Research & data access | 7 |
| | Research & education | 2 |
| | Staffing agency/contractor | 1 |

Table 1. Breakdown of organization characteristics for interview sample

## 5 Internship Projects

All DCERC students completed a project during their residency at NCAR, designed by matching student interests and experience with two mentors—a data mentor and a science mentor. The dual mentorship approach offered students broad exposure to data services and science applications at NCAR. In addition, an NCAR information scientist served as a peer mentor for all students. This intensive mentorship approach resulted in successful projects that developed skills and knowledge of data management and curation, as well as the use of data in the practice of scientific research (Kelly et al., 2013). The internship projects met objectives for practical experience and concrete outcomes, greatly extending the prior from the curriculum and local research experience. Moreover, the students contributed unique knowledge and skills to the projects, infusing a new dimension to scientific data management for personnel at NCAR (Mayernik et al., 2014). Tables 2 and 3 give an overview of the projects—their focus, the tasks involved, and the knowledge or skills applied. Table 2 covers the projects focused on Curation Processes with particular datasets, and Table 3 covers projects focused on Curation Policy and Services. See the appendix for a list of project titles.

### 5.1 Student Benefits

The internship projects gave students the opportunity to apply principles, processes, and theory from their data curation coursework to practical experience in innovative research center operations. For example, they applied what they had learned about the data curation lifecycle (Higgins, 2008) to stewarding data sets through the different activities and phases of curation (Eaker, Thomer, Johns, & Siddell, 2013). Students participated in description, data transformation, preservation planning and implementation, appraisal, ingest, and access, use and reuse. Four projects involved archiving data sets, including identification of potential repositories, file transformations, digitization, and preparing and depositing submission packages. In these projects, students acted as translators between NCAR scientists and

archivists to ensure quality curation, serving in a data concierge role common in the work of practicing data professionals (Mayernik et al., 2014).

Most of the Curation Processes projects had a substantial metadata component where students interacted with scientists and other stakeholders, identifying metadata needs and auditing workflows to improve data description (Eaker et al., 2013). Students generated, standardized, and harvested metadata and documentation specific to the atmospheric research community, but also applied broader, conceptual aspects from their classroom studies, such as the relationships among different levels of metadata—project, data set, and individual files, for instance. They had responsibility for metadata harvesting, normalizing, and crosswalking between standards. Generation of data provenance was a common task. Multiple students analyzed workflows and practices by applying the Data Curation Profile toolkit (Witt, Carlson, Brandt, & Cragin, 2009). Data practices and requirements were documented using the Data Curation Profile interview protocol for gathering important information from data producers.

| Curation Processes Focus | Data Types | Tasks | Knowledge & Skills |
|---|---|---|---|
| A. Data rescue and access<br>International tropical weather field; 50 collection platforms from dozens of ships and aircraft. | observation, analog | inventorying, appraisal, metadata assessment and creation, DOI assignment, archiving, preservation planning | information organization, metadata standards, OAIS model, identifier schemes, digitization, preservation |
| B. Data access for reuse<br>Multi-decadal high temporal and spatial resolution hourly analyses of global atmospheric state. | time series, netCDF | quality assessments, interviews, metadata creation and automation, document provenance | metadata standards, data quality, programming, provenance, curation profiling, qualitative interviewing |
| C. Curation lifecycle assessment<br>Ameriflux Network data on carbon, water, and energy fluxes in North and South America. | sensor, csv | data acquisition, file transformation, quality assessment, metadata creation, archiving, preservation planning | metadata standards and compliance, data quality, programming, use requirements |
| D. Appraisal & access for reuse<br>Small data set of global soil properties. | observation, netCDF | interviews, workflow, document provenance, metadata assessment, archiving, preservation planning | needs assessment, curation profiling, provenance, OAIS model, standard compliance, preservation |
| E. Appraisal & assessment<br>International, cross-disciplinary socio-economic demographics, air pollution, and temperature readings. | observation, csv | appraisal, data subsetting, standardization, metadata creation, data movement, archiving, preservation planning | metadata standards, OAIS model, standard compliance, data quality, appraisal, preservation |

Table 2. Description of DCERC Master's projects on curation processes

Quality control was another important area of curatorial activity. Students working on Curation Processes projects got first-hand experience identifying errors or missing data and learned best practices for improving data quality, for example, by calibrating to weather conditions or sensor type. They performed metadata quality checks, assessed compliance with standards, and identified and labeled missing data points to improve usability. As is typical in scientific practice, researchers needed assistance in meeting growing expectations for more formal documentation and use of standardized metadata, instead of relying on ad hoc notes and personal communication for data sharing.

The Curation Policy and Services projects were directed toward development of curation procedures and guidelines rather than the direct curation of data. For example, levels of service were analyzed for policy development. Data citation efforts concentrated on development of guidelines and outreach materials to encourage use of data citation and digital object identifiers (DOIs) for users of NCAR data. Finally, procedural documentation was developed for archiving cross-disciplinary data and for rescue of high-risk analog data.

As a whole, the cohort of students gained experience with software, instruments, tools, and file formats well beyond what can be integrated into classroom education. DCERC interns learned about a

range of instruments and sensors used to collect atmospheric data, such as dropsondes, radiosondes, and driftsondes. They used software and systems for data processing, archiving, and analysis, including MatLab, S-PLUS, R, Drupal, and a number of in-house applications. They worked with relational databases and a variety of file formats (e.g., netCDF, csv). The projects provided opportunities to expand technical skills in areas such as querying databases, writing scripts for data processing and transfer, and more specialized curation processes like automated export of metadata into standardized interchange structures. A few students developed skills in social science methods for systematic interviewing, observation, and analysis of artifacts to understand scientific workflows and curation needs. The most critical factor in the success of the projects was students' continual interactions with scientists, data professionals, and other stakeholders in the active data center environment.

| Curation Policy & Services Focus | Tasks | Knowledge & Skills |
|---|---|---|
| F. Levels of service<br>Data service policy for an interdisciplinary and inter-organizational archive of arctic data. | literature review, case study, policy development | data services, policy, use requirements, case study method |
| G. Scientific workflow audit<br>Data work processes in science teams. | workflow audit, metadata assessments, interviewing, observations, report writing | data curation lifecycle, scientific workflow, qualitative methods and analysis |
| H. Data discovery and access<br>Preparing data archive records for access through Geoportal and Global Change Master Directory. | interviews, metadata standard crosswalks, XML template design, metadata harvesting and automation, data citation | metadata standards, XML, OAI-PMH standard, programming, data citation, policy, qualitative interviewing |
| I. Data citation and practices.<br>Outreach and education related to data citation and DOIs. | data citation and identifier schemes, DOI policy development, interviewing, instructional design | data citation and identifier standards, data analysis, learning and instructional design theories |
| J. Metadata and data citation.<br>Climate model metadata, and citation and acknowledgement guidelines assessment. | metadata harvesting and assessments, data citation review and recommendations | Metadata standards, data quality, data citation standards, use requirements, standard compliance |

Table 3. Description of DCERC Master's projects on curation policy and services

The three Ph.D. students in the DCERC program have had a different kind of experience at NCAR. Their 4 or 9-month field placements were designed to advance their information science interests and contribute to their independent doctoral research. Like the Master's students, each had multiple NCAR mentors who worked with them to achieve their goals. Most importantly, they conducted research on data curation problems only possible by being embedded in the data center environment. Their three projects leverage the NCAR experience in very different ways and have important implications for the field. One is a comparative ethnographic study of how different types of data repositories develop in the earth and environmental sciences, in response to researchers' diverse data practices with heterogeneous field data. The second is developing a framework for analyzing sustainability of common data resources in the eSciences. The case examines, and will inform the evolution of a comprehensive ocean and atmospheric dataset critical to the work of the international climate science community. The third is a study of scientific data center operations and innovations at NCAR, examining the organization of data work, positions and duties, and changes in data expertise, to inform workforce development.

### 5.2 Complementary internship activities

In addition to the individual internship projects, DCERC students had many opportunities to interact with data and library professionals in other contexts. Table 4 illustrates the range of sites and acceleration of activities that students were involved in across the three summer internship sessions. The increase in the third year was in response to the uniformly positive feedback from students in the first two summers on the value of auxiliary professional interactions. Options for professional engagement were particularly

abundant in 2014, since a number of national meetings were held in Boulder and nearby areas of Colorado. For example, DCERC students attended GeoData 2014 Workshop, funded by the NSF to facilitate coordination of geoscience data efforts (Ma, Fox, & Mayernik, 2014). They participated as notetakers during breakout discussions, in addition to attending the full meeting.

These activities allowed the students to discuss their projects with experts, participate in professional group interactions, get deeper exposure to scientific research and relevant computational systems, and learn more about the variety of settings reliant on data curation work.

| Year | Activity |
|---|---|
| **2012** | NCAR Archives – Meeting with NCAR archivist on archival collections and practices.<br>University of Colorado at Boulder (CU) and National Snow and Ice Data Center (NSIDC) – Presentations by working librarians and data professionals.<br>NCAR Visualization Lab – Demonstration of visualization techniques. |
| **2013** | NSIDC – Orientation to facility and data professionals.<br>Data Librarian meeting at the University of Colorado, Boulder – Meeting with local librarians and data professionals.<br>Science Boot Camp for Librarians – Professional meeting hosted by CU for librarians to learn about/from research faculty practices.<br>UCAR Internship Seminar Series – Weekly presentations by UCAR staff for all UCAR internship programs. |
| **2014** | UCAR Internship Ignite Session - Presentations by UCAR staff on UCAR science and engineering work.<br>National Data Service (NDS) Meeting – Two-day kickoff of new national initiative.<br>NCAR Wyoming Supercomputer Center tour – Orientation to new supercomputer facility.<br>Data librarian meeting at Colorado State University (CSU) – Meeting with local librarians and data professionals.<br>CSU Natural Resources Ecology Lab – Overview of data-related activities.<br>GeoData 2014 Workshop – National meeting.<br>Earth Science Information Partners (ESIP) meeting - Professional meeting for information and data professionals in the earth science community.<br>NCAR Director presentation – Jim Hurrell, an expert on climate change.<br>Geoscience metadata meeting – Engagement with Ted Habermann, an internationally recognized expert in metadata standards from the HDF Group. |

Table 4. Complementary internship activities for Master-level interns by year

### 5.3 Data Center Benefits

The benefits of the DCERC internships have proven to be truly reciprocal, documented in formal feedback gathered from the mentors through focus groups and questionnaires. Both data mentors and science mentors emphasized the value of the students' work, and their strong curatorial and service orientation. They welcomed the professional data curation dimension and its contribution to a range of challenges in data management and services faced by NCAR. In the Curation Processes projects outlined in Table 2, students were involved in processing, documenting and creating metadata, and archiving providing an important layer of direct hands-on work prioritized by scientists and data managers. Due to resource limitations, the data they worked with would not have been made available to secondary users, without the DCERC contributions. In the projects on Data Policies and Services, DCERC interns provided analyses and deployed technologies and resources, adding value in areas that support the functioning of the organization. The Ph.D. student contributions function at the level of research rather than practice, further highlighting the importance of information science expertise in the data enterprise. The interactions and experiences they had with NCAR mentors and personnel took many forms, ranging from interviews to examine changes and differences in repository development to direct collaboration on analyses important to the NCAR mission.

## 6 DCERC Model Sustainability and Transferability

In parallel with our ongoing formal evaluation of the DCERC program (Kelly et al., 2013), the team is examining alignment of the model with the needs of other data centers in the geosciences. The formal

evaluation of the program includes substantive feedback from students and the NCAR mentors, a particularly important aspect for assessing the fit with NCAR's initiatives and goals and the potential for sustaining the program at NCAR. The interviews, described in the methods section above, are of particular value for assessing the prospects for extending the internship model to other data centers.

The 20 interviews with data center managers indicate that there is strong interest in hosting internships for iSchool data curation students. Three respondents reported hosting iSchool interns in the past (not as part of the DCERC program) that were successful for the students and beneficial to their center. Several respondents described the potential for placing students through the existing internship programs at their institution, but as was the case with NCAR prior to DCERC, these program currently target only students in the geosciences and computer science. A few respondents were able to identify internships projects that would interest iSchool students and meet organizational needs. Areas mentioned included digitization of historical maps, collection development, data quality assessments, audit of scientific workflows, natural language processing, and data management for bioinformatics. In discussing challenges to hosting data curation interns, respondents mentioned designing mutually beneficial projects, funding concerns, and the organizational priority for supporting domain and/or computer science interns. A few respondents noted that bringing data-focused students into their organization would require new planning and arrangements beyond their existing internship program, and, more importantly, an argument would need to be made for the value of the contributions data curation students would make. One respondent recounted a case where a student from an information school was initially rejected for an internship position by the selection committee, since they did not "understand what had to be done [for the project] and what skills were needed" (ID 116).

Data centers must see tangible benefits for the resources they invest in training interns. Many data centers have very specific missions and mandates to provide operational services for particular purposes and user groups. NCAR data management teams, for example, are dedicated to providing highly reliable services and high quality data for the communities that they serve. A key challenge of setting up an internship program is aligning internship goals with the data center goals and missions. DCERC has achieved this through tight coupling of projects to current data center aims and activities. As discussed above, a goal of DCERC from the beginning has been to provide as much benefit as possible back to the data center teams and the institution.

Sustaining a program like DCERC beyond the grant-funded period is challenging. The primary obstacle with NCAR is in the cost of supporting students in residence at the site for the duration of the summer. Moreover, other data centers may have difficulty allocating staff time to support intern mentoring. A more sustainable approach would be to develop a network of LIS programs and data centers that support internships in a distributed fashion, with LIS students conducting projects at local data centers where possible. Targeted grants might support coordination among students, LIS programs, and participating data centers. We are currently investigating how such an approach might extend and sustain the DCERC model beyond our current scope.

## 7 Conclusion

The DCERC program offers a new model for data curation education that is proving to be mutually beneficial to all participating in the iterative development of both the academic and internship components of the initiative. Students gain unparalleled experience working on individually crafted projects with researchers on the front lines of data-intensive science and data professionals developing state-of-the-art data infrastructure and services. These professionals tend to have a background in domain science or engineering, and, as is typical in the field, many have "accidently" moved into data responsibilities as their careers have progressed (Pryor & Donnelly, 2009). As such, they recognize the value of having peers who are focusing their professional studies on the curation of research data. The DCERC students are providing NCAR practitioners with a new data curation approach to data management challenges and solutions (Mayernik et al., 2014; Eaker et al., 2013). The principled orientation to user-based design and services, open access, and information organization that DCERC students learn in their coursework resonates with NCAR data professionals. Moreover, their concern with supporting the entire lifecycle of data, including long-term preservation and interoperability across systems and domains, aligns with more global data and research objectives in the geosciences.

Students benefit tremendously from applying their academic coursework to practical issues in the data center environment. Their project work deepens their understanding of archiving, metadata, quality assessment and control, provenance, and data citation. Additionally, learning about the computation systems used within NCAR has been key to furthering their technical skills and their grasp of the engineering side of repository work. Interns gain direct experience with scientific systems and tools for

data collection, cleaning, analysis, and archiving; they enhance their programming skills for processing, transforming, or transferring scientific data. Mentorship and networking are at the heart of student learning during the DCERC internships. They build awareness of career opportunities and help to launch students into the professional realm. Having interacted with data professionals, data producers, and end users in the earth science community, in addition to their mentors knowledgeable about trends in the field, relevant professional associations and conferences, and potential career paths. These connections result in a network of weak ties (Granovetter, 1973) that will serve them well in advancing their careers.

As a collaboration between iSchools and a premier data center, DCERC is forging an important path toward integration of expertise from two separate communities that have much to gain from sustained and programmatic interaction. DCERC is producing the kind of cross-fertilization envisioned for the program, and the model holds promise for preparing a new generation of data professionals who are truly prepared to meet the demands of data-intensive research.

## 7.1 References

American Council of Learned Societies (ACLS). (2006). Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. Retrieved from: http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf

Atkins, D. (2003). Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. Retrieved on March 7, 2012 from: http://hdl.handle.net/10150/106224

Davenport, T. H., & Patil, D. J. (2012). Data scientist: the sexiest job of the 21st century. *Harvard business review*, *90*(10), 70-77. Retrieved from: http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/3

Eaker, C., Thomer, A. K., Johns, E., & Siddell, K. (2013). How information science professionals add value in a scientific research center. Proceedings of 2013 iConference, Fort Worth, TX.

Gold, A. (2007). Cyberinfrastructure, data, and libraries, part 2. *D-Lib Magazine*, 13(9/10). Retrieved from: http://www.dlib.org/dlib/september07/gold/09gold-pt2.html

Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 1360-1380.

Hank, C., Tibbo, H. R., & Lee, C. A. (2010). DigCCurr I Final Report, 2006-09: Results and recommendations from the Digital Curation Curriculum Development Project and the Carolina Digital Curation Fellowship Program. Chapel Hill, NC: School of Information and Library Science, University of North Carolina. Retrieved from http://www.ils.unc.edu/digccurr

Harris-Pierce, R. L., & Liu, Y. Q. (2012). Is data curation education at library and information science schools in North America adequate? *New Library World*, 113(11/12), 598-613.

Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, *3*(1), 134-140.

Kelly, K., Marlino, M., Mayernik, M. S., Allard, S., Tenopir, C., Palmer, C. L., & Varvel Jr, V. E. (2013). Model Development for Scientific Data Curation Education. *International Journal of Digital Curation*, *8*(1), 255-264.

Lyon, L., Wright, S., Corti, L., Edmunds, S., & Bennett, F. (2013). What is a data scientist? Panel presented at 2013 International Digital Curation Conference, January 14-17, 2013, Amsterdam, Netherlands.

Ma, X., Fox, P., & Mayernik, M.S. (2014). Strengthening an interagency network for geoscience data sets. Eos, Transactions, American Geophysical Union, 95(45): 411.

Maatta, S. L. (2012). A job by any other name: A few bright spots shine for the class of 2011. *Library Journal*, *136*(17), 18-25.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Retrieved from: http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation

Marshall, J. G., Morgan, J. C., Rathbun-Grubb, S., Marshall, V. W., Barreau, D., Moran, B. B. & Thompson, C. A. (2010). Toward a shared approach to program evaluation and alumni career tracking: Results from the workforce issues in library and information science 2 study. *Library Trends*, *59*(1), 30-42.

Mayernik, M. S., Davis, L., Kelly, K., Dattore, B., Strand, G., Worley, S. J., & Marlino, M. (2014). Research Center Insights into Data Curation Education and Curriculum. In *Theory and Practice of Digital Libraries--TPDL 2013 Selected Workshops* (pp. 239-248). Springer International Publishing.

Mayernik, M. S., Thompson, C. A., Williams, V., Allard, S., Palmer, C. L., & Tenopir, C. (2015). Enriching Education with Exemplars in Practice: Iterative Development of Data Curation Internships. Proceedings of the 2015 International Digital Curation Conference. http://www.dcc.ac.uk/events/idcc15/day-two-papers#sthash.nHvVMsUZ.dpuf

Palmer, C. L., Thompson, C. A., Baker, K. S., & Senseney, M. (2014). Meeting Data Workforce Needs: Indicators Based on Recent Data Curation Placements. Proceedings of 2014 iConference, Berlin, Germany.

Palmer, C. L., Thompson, C. A., Mayernik, M. S., Williams, V., Kelly, K., & Allard, S. (2014). Data Curation Education in Research Centers: Formative Evaluation Findings from Years One and Two. Proceedings of the International Digital Curation Conference, San Francisco, CA.

Plale, B., McDonald, R. H., Chandrasekar, K., Kouper, I., Konkiel, S., Hedstrom, M. L., … Kumar, P. (2013). SEAD virtual archive: Building a federation of institutional repositories for long-term data preservation in sustainability science. *International Journal of Digital Curation*, 8(2), 172–180. http://doi.org/10.2218/ijdc.v8i2.281

Pryor, G., & Donnelly, M. (2009). Skilling up to do data: whose role, whose responsibility, whose career?. *International Journal of Digital Curation*, *4*(2), 158-170.

Renear, A., Cragin, M., Heidorn, B., Palmer, C. L., Smith, L., Unsworth, J., & Senseney, M. (2012). Centuries of knowledge: Graduate School of Library and Information Science data curation education program. Final Report to IMLS. Retrieved from http://hdl.handle.net/2142/30845

Rusbridge, C. (2007). Create, curate, re-use: the expanding life course of digital research data. EDUCAUSE Australasia 2007 Advancing Knowledge Pushing Boundaries. CAUDIT. Retrieved on March 7, 2012 from: http://hdl.handle.net/1842/1731

Shadbolt, A., Konstantelos, L., Lyon, L., & Guy, M. (2014). Delivering innovative RDM training: the immersiveInformatics pilot programme. *International Journal of Digital Curation*, *9*(1), 313-323.

Sierra, T. (2012). Staffing for the future: ARL university library hiring in 2011. *Proceedings of the ARL Fall Forum,* Washington D.C.

Steinhart, G. & Qin, J. (2012). Mentoring for emerging careers in eScience librarianship: An iSchool – academic library partnership. *Journal of eScience Librarianship*, 1(3): 120-133. http://doi.org/10.7191/jeslib.2012.1017

Swan, A., & Brown, S. (2008). *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs. Report to the JISC*. Truro, Cornwall, UK, pp. 32. (July 2008).

TEKSystems. (2013). Big Data…The next frontier. IT Industry Trends. Retrieved from: http://www.teksystems.com/~/media/Files/thought-leadership/big-data-wp20130912.ashx

Varvel, V.E., Bammerlin. E.J., & Palmer, C.L. (2012). Education for data professionals: a study of current courses and programs. *Proceedings of the 2012 iConference*, 527-529. New York: Association for Computing Machinery. http://doi.org/10.1145/2132176.2132275

Varvel, V.E.Jr., Palmer, C.L., Chao, T., & Sacchi, S. (2011). Report from the Research Data Workforce Summit: Sponsored by the Data Conservancy. Champaign, IL: Center for Informatics Research in Science & Scholarship, University of Illinois. Retrieved from: http://hdl.handle.net/2142/25830

Witt, M., Carlson, J., Brandt, D. S., & Cragin, M. H. (2009). Constructing data curation profiles. *International Journal of Digital Curation*, *4*(3), 93-103.

## 7.2 Table of Tables



## 7.3 Appendix. List of DCERC Student Projects

**Baker, K.S.** (2013). *Scientific Research Group Data Management Practices and Local Data Repositories: A Comparative Study into Data Management Practices as They Develop in Relation to Data Repositories.*

**Christensen II, W. R.** (2014). *Establishing Data Management Levels of Service within NCAR: ACADIS Case Study.* https://opensky.library.ucar.edu/collections/OSGC-000-000-020-656

**Clark, J.** (2013). *Unlocking GATE: Gaining Access to Analog Data in a Digital World.* http://opensky.library.ucar.edu/collections/OSGC-000-000-018-987

**Eaker, C.** (2012). *Data Audit and Analysis: Mapping the Data Workflow from Ingest to Archive.* http://opensky.library.ucar.edu/collections/OSGC-000-000-010-444

**Gordon, S.** (2014). *Facilitating Dataset Discovery Across NCAR's Repositories: Establishing a Geoportal at NCAR's Earth Observing Laboratory.*
http://opensky.library.ucar.edu/collections/OSGC-000-000-020-752

**Hou, C.** (2014). *Discovering New Global Climate Patterns: Curating a 21-Year High Temporal (Hourly) and Spatial (40km) Resolution Reanalysis Dataset.*
https://opensky.library.ucar.edu/collections/OSGC-000-000-020-654

**Johns, E.M.** (2012). *The data lifecycle flow: For me, this time.*
http://opensky.library.ucar.edu/collections/OSGC-000-000-010-445

**Johns, E**. (2013). *Data Curation Service Development: Investigation, Outreach, and Education.* http://opensky.library.ucar.edu/collections/OSGC-000-000-018-986
**Langseth, M.** (2014). *Data Curation in the Long Tail of Science: Preparing Community Land Model Validation Data for Reuse and Preservation.* https://opensky.library.ucar.edu/collections/OSGC-000-000-020-653
**Siddell, K.D.** (2012). *Now You are Speaking My Language: Translating and Facilitating Between Researchers and Data Managers.* http://opensky.library.ucar.edu/collections/OSGC-000-000-011-152
**Thomer, A.K.** (2012). *Curating Context and Use: Pulling Scientific Workflows into the Repository*. http://opensky.library.ucar.edu/collections/OSGC-000-000-011-528
**Thompson, C.A.** (2014). *Scientific Data Workforce: Understanding Data Expertise and Talent Management at a Mature Research Center.*
**Weber, N.M.** (2013). *A Framework for Analyzing the Sustainability of Peer Produced Science Commons*.