Enriching Education with Exemplars in Practice: Iterative Development of Data Curation Internships

Matthew S. Mayernik
NCAR Library
National Center for Atmospheric Research

Cheryl A. Thompson
University of Illinois at Urbana
Champaign

Valerie Williams
University Corporation for Atmospheric Research (UCAR)

Suzie Allard
School of Information Sciences
University of Tennessee, Knoxville

Carole L. Palmer
Information School
University of Washington

Carol Tenopir
School of Information Sciences
University of Tennessee, Knoxville

Abstract

Partnerships between educational programs and research centers are vital to meeting the escalating workforce demands in data curation. They offer a platform for educators to increase their knowledge of current best practices and emerging challenges in the field. Student internships can be key to the success of these partnerships, not just for the students who gain authentic experience in facilities that excel at data intensive research and data services. Such partnerships provide an effective platform for rich and mutually beneficial engagement among educators, data professionals, scientists, and students. This paper reports on results from the Data Curation Education in Research Centers (DCERC) program aimed at developing a model for data curation education featuring field experiences in exemplar research centers. A strength of the DCERC model is its emphasis on facilitating mutual exchange of information among the DCERC program mentors and students. This model has evolved as a result of iterative and gradual refinements to the program model based upon information gathered from the formative evaluation. These refinements not only resulted in improved outcomes for the program participants but also, we believe, a more sustainable model for the program that leverages the knowledge base of the research scientists and students through peer-to-peer learning, rather than a traditional expert to trainee model. This paper describes formative evaluation findings that shaped the development of the DCERC program. We conclude with a discussion of the critical features of this model for the development of similar programs, and a data curation workforce that is able to accommodate and adapt to emergent data needs in a variety of environments.

Received 16 January 2015 | Accepted 10 February 2015

Correspondence should be addressed to Matthew S. Mayernik, P.O. Box 3000, Boulder, CO, 80307-3000. Email: mayernik@ucar.edu

An earlier version of this paper was presented at the 10th International Digital Curation Conference.

The International Journal of Digital Curation is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see http://creativecommons.org/licenses/by/2.0/uk/
Introduction

The opportunities and challenges presented by the proliferation of digital data have led to data curation work becoming more professionalized in many institutional sectors (Higgins, 2011; Maatta, 2013; Weber, Palmer, and Chao, 2012). In response, data curation education is becoming more established. Universities are developing data curation curricula as part of formal academic programs, primarily in library and information science (Harris-Pierce and Liu, 2012; Varvel, Bammerlin, and Palmer, 2012). Professional development and training options are increasing for information professionals working in the field (Erdmann, 2014), as well as for researchers within specific domains (see for example, DataONE, 2012; DHDC\(^1\); ESIP, 2014).

Experience in real-life work settings is a key aspect of professional education. Internships in data curation provide students with vital first-hand experience and increase employability in academic research institutions, government agencies, and private companies (Palmer, Thompson, Baker, and Senseney, 2014). They are particularly important for new professionals, because of the formative stage of the profession and the resulting dynamic nature of the work. Interns apprentice in specific data curation skills and tasks, in particular settings, and make personal contacts that can help sustain continued professional development (Kim, Addom, and Stanton, 2011; Steinhart and Qin, 2012).

The Data Curation Education in Research Centers (DCERC) program is a partnership among the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign (UIUC), the School of Information Sciences at the University of Tennessee at Knoxville (UTK), and the NCAR Library in the National Center for Atmospheric Research (NCAR), a research center based in Boulder, Colorado. The goal of the program is to develop a sustainable and transferable model for preparing the next generation of leaders in scientific data curation through coursework in academic programs and strategic internship opportunities and mentoring for students.

Funded by the U.S. Institute for Museum and Library Services (IMLS), the DCERC program supported field experiences for Master’s and Ph.D. students from 2012-2014. This paper focuses on the nine Master’s internships, awarded to nine students, four from UIUC and five from UTK, over three summer sessions. As described in detail in Kelly et al. (2013), the central components of the program include:

- Focused curriculum, including a graduate-level core course in data curation completed by students from both universities;
- Academic research experience through assistantships with faculty mentors;
- Internship kickoff workshop with experts from multiple scientific data centers presenting and engaging with students for multiple days;
- Eight week data curation internships with students projects guided by both data and science mentors;
- Final poster presentation of internship projects at NCAR.

\(^1\) Digital Humanities Data Curation Institute (DHDC): [http://www.dhcuration.org/institute/](http://www.dhcuration.org/institute/)
NCAR data professionals manage and curate multiple data collections that are extensive and widely used in the atmospheric and related sciences. Data curation work has been central to the development of the atmospheric sciences for many decades (Edwards, 2010), and NCAR data management efforts extend back to the 1960s (see for example Jacobs and Worley, 2009). Thus, the DCERC students coming to NCAR have been able to experience data curation work in a mature research and data center environment.

LIS students, however, bring a new set of knowledge and skills, and a unique professional orientation to data curation work (Mayernik et al., 2014). Integrating them into a disciplinary data center setting, such as NCAR, has required careful collaboration with the mentors that accounts for their different professional orientation. This paper outlines the evolution of the DCERC program, emphasizing the formative evaluation process and iterative adjustments to increase benefits for both students and NCAR mentors. We also share our perspective on how to optimize the DCERC model based on progress made on design and implementation of the program.

**Evaluation Methods**

The primary aim of the formative evaluation was to assess and improve the design of the DCERC model for data curation education. It was designed to address the following questions:

1. How do different stakeholders perceive key components of the program (e.g. strengths, weaknesses, benefits, value-added, etc.)?

2. What processes are functioning well or need improvement from the student perspective? From the mentor perspective?

3. What do participants actually do in the program? What are the primary activities (in detail)? What do they experience? To what extent are those experiences yielding the intended short-term outcomes?

4. To what extent does the DCERC model appear to be improving education for students? Contributing to NCAR’s operations and mission?

Data to address these questions was collected through multiple instruments and sources over the three year period, outlined in Table 1.

**DCERC Internship Evolution**

This section outlines key aspects of the three successive years of summer internships, including recruitment of students and mentors, students and mentor preparation, identification of student projects, and additional student activities during the eight week summer session. In particular, we identify lessons and adjustments made as the program progressed from year to year.
Table 1. Data collection activities for formative evaluation.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Respondents</th>
<th>Information Gathered</th>
<th>Data Collection Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student applications for DCERC program</td>
<td>Students</td>
<td>Student applications; Number of students completing program and student professional activities</td>
<td>2012-2014 (annually)</td>
</tr>
<tr>
<td>Foundation of Data Curation course questionnaire</td>
<td>Master’s students</td>
<td>Usefulness of course, learning experience, recommendations</td>
<td>2012-2014</td>
</tr>
<tr>
<td>Pre-internship questionnaire</td>
<td>Students</td>
<td>Demographics, motivation for participation, self-reported knowledge of data curation, expectations for program</td>
<td>2012-2014</td>
</tr>
<tr>
<td>Kick-off workshop questionnaire</td>
<td>Workshop attendees</td>
<td>Satisfaction with workshop; positive, negative, and most useful aspects</td>
<td>2012, 2014</td>
</tr>
<tr>
<td>Internship reflections</td>
<td>Students</td>
<td>Reflections on learning and changes knowledge during internship</td>
<td>2012-2014</td>
</tr>
<tr>
<td>Post-internship questionnaire</td>
<td>Students</td>
<td>Satisfaction and value of program, recommendations</td>
<td>2012-2014</td>
</tr>
<tr>
<td></td>
<td>Mentors</td>
<td>Value of program, evaluation of student</td>
<td>2012-2014</td>
</tr>
<tr>
<td>Post-internship focus group</td>
<td>Students</td>
<td>Student background and motivations, mentorship experience, workforce issues, program impact</td>
<td>2012-2014</td>
</tr>
<tr>
<td></td>
<td>Mentors</td>
<td>Motivations, program structure and impact, experience with student</td>
<td>2013-2014</td>
</tr>
<tr>
<td>Employer perspective interviews</td>
<td>Data center managers in geosciences</td>
<td>Employer perspective on data workforce needs and willingness to host interns</td>
<td>2014</td>
</tr>
</tbody>
</table>
Internship Year 1, 2012

The initial design of DCERC designated recruiting Master’s students from UTK and PhD students from UIUC. The Master’s internship program was to support three UTK students for two successive summers, after their first and second academic year of coursework. The initial cohort of UTK Master’s students was specially recruited with a full scholarship to be part of the DCERC program. The recruitment was focused on finding a diverse group of students with a range of knowledge, educational backgrounds, and work experiences from UTK. Resources allowed for the addition of one data curation student from UIUC in the first internship cohort. The students accepted into the program had backgrounds including life science, engineering, and social science.

Three of the first data mentors were recruited from a NCAR ad hoc committee on data citation, which had representatives from a number of data management groups within the organization (Mayernik et al., 2012). The NCAR Library, which was coordinating the committee, was well positioned to assess interest in serving as a DCERC data mentor. The fourth data mentor worked as a staff member under another member of the data citation working group. The science mentors were recruited in a more ad hoc fashion. One was a past collaborator with the NCAR Library on a data-related project, and the others were identified based on their scientific research areas.

Local DCERC team members had informal meetings with the mentors identified to outline the goals and scope of the internships, possible projects, and mentorship expectations. One additional formal meeting was held with most of the mentors shortly before the beginning of the internships. To facilitate the matching of students with potential mentors, students were asked to fill out an internship application form and write a statement of interest, in which they indicated information/data science topics of interest, as well as science/engineering topics of interest. In the weeks before the internship began, these application materials were provided to the mentors.

Pairings of data mentors and science mentors with students aimed to align a mentor’s organizational position and topical interests with that of the students’ interests. No decisions about internship projects were made at this stage, but pairings were made to be conducive to likely projects. Students received little preparation directly focused on their NCAR internships prior to coming to NCAR. They were informed of their mentors shortly before arriving in Boulder and did not interact with their mentors in advance. Organizing the internship projects was one of two goals of the two and a half day kick-off workshop. The second goal was to introduce the students to NCAR and related activities and areas of expertise. The workshop program consisted primarily of science and data-focused presentations by NCAR staff and invited experts. The students gave a short presentation on the first day to outline their background and interests, and a good amount of time was allocated for one-on-one student-mentor meetings.

During the internships, DCERC team members in the NCAR library met on a weekly basis with the four students to provide opportunities to share experiences and discuss any challenges they were encountering. They also organized a meeting with professionals from local organizations for students to interact with a broader range of professionals working with scientific information or data. The final component of the program was poster presentations of the internship projects in a dedicated DCERC poster event at NCAR.
Year 1 feedback and adjustments

Formative evaluation activities carried out in Year 1 are noted in Table 1. The following feedback contributed to how the program was adjusted for Year 2:

- Students described their mentors as great, accessible, supportive, and enhancing the learning experience.
- Mentors appreciated their student’s LIS perspective.
- Students valued professional development opportunities.
- The kick-off workshop, while informative, was overwhelming.
- Student project organization needed improvement.
- Students needed better orientation to NCAR and mentors.
- Mentors had limited understanding of student skills.
- Internship timelines were very tight.

Based on the first year’s assessment, the following strategic adjustments were planned for the second year’s cohort of students and mentors:

- Do more project development before the students arrive, including:
  - Earlier pairing of students with mentors,
  - Discuss possible projects with students and mentors.
- Streamline kick-off workshop to focus on student projects.
- Provide more NCAR activities for students to:
  - Learn about NCAR science and engineering,
  - Interact with students from other NCAR internship programs.
- Provide additional professional development opportunities.

Internship Year 2 - 2013

As noted above, the initial model for DCERC was that the three UTK Master’s students would come to NCAR for two successive summers. This model had to be adjusted for Year 2, however, because two of the three UTK students were placed in professional positions sooner than expected. We believe this is largely due to their DCERC experience. One new student from UIUC who had completed the academic component applied and was accepted into the program, giving a cohort of two students for Year 2. Two data mentors were identified from Year 1 data mentors. As in Year 1, the data mentors were provided with the students’ application materials in the weeks before the internship began.

In Year 2, our goal was to identify the student projects before the students arrived at NCAR. The DCERC staff at NCAR met with the two students multiple times via phone in the months preceding the internship, and iterated with the data mentors and the students to develop appropriate projects. As the projects shaped up, two science mentors were identified. One science mentor was suggested by a data mentor. Since the other
With the projects identified, the students were provided with relevant resources, such as documents and web sites related to their future projects. The returning student spoke with both of her mentors about her likely projects prior to her arrival to NCAR. The new student was introduced to her data mentor over email, but did not have additional discussion with her mentor(s) prior to her arrival.

Since the cohort included one returning student and one new student, the DCERC internship did not start with a full-fledged kick-off workshop like Year 1. However, a one-day introductory meeting was held in order to facilitate student and mentor interactions and internship project planning. Since the projects were largely specified prior to the students arriving at NCAR, the goal of this meeting was to accelerate the pace of students into the beginning stages of their projects.

Similar to Year 1, NCAR Library staff met with the students on a weekly basis about their projects, and organized meetings with local professionals working with scientific information or data. Building on positive feedback from students in Year 1 about professional development opportunities, in Year 2 the students attended the 2013 Western Science Boot Camp for Librarians meeting, which took place in Boulder.

**Year 2 feedback and adjustments**

A number of lessons were drawn from the findings that emerged out of the formative evaluation activities for Year 2.

- Projects had better alignment between student and mentor goals and student skills.
- Students valued the additional professional development opportunities.
- Mentors appreciated the students’ analytical approaches.
- Mentors noted the significant time commitment.
- Data mentors stressed importance of computing skills, which LIS students may not have.

Although the Year 2 internships proceeded more smoothly than Year 1, the Year 2 lessons learned again helped us identify a number of adjustments that could be made for Year 3:

- Give mentors opportunity to provide feedback on student applications.
- Develop projects early, before the students arrive at NCAR.
- Provide additional professional development opportunities.
- Work with mentors to identify the variety of skills involved in data curation.

**Internship Year 3, 2014**

For the third DCERC summer, budgets at UTK and UIUC allowed for four new students to come to NCAR. To recruit new students, DCERC staff developed and circulated an announcement at UTK and UIUC as advertisement, and also encouraged promising students to apply. Students were recruited for their interest in data curation,
relevant coursework, and interest in an NCAR internship. Four students submitted applications.

DCERC staff identified three data mentors from previous years, and asked them to review the student applications. The data mentors were asked to rank the applicants on three categories: “Preparation for internship,” “Likelihood for success at NCAR,” and “Your interest in working with them”. When all applicants received interest from at least one data mentor in this ranking process, all four were accepted into the program. DCERC staff then identified a fourth data mentor, and made mentor-pairings based on the mentors’ rankings of the applicants.

During this application process, DCERC staff at NCAR met with the data mentors to discuss possible internship projects. Three of the four data mentors had experience with previous DCERC internships. Science mentors were recruited by data mentors, and by NCAR Library staff. One student was not paired with a science mentor because the student’s project was focused on metadata and systems, and no distinct “science” component was appropriate.

One new effort in the months leading up to the summer internships was to organize a series of phone calls for an NCAR Library staff member to meet with the four students. These calls focused on 1) introducing the students to the NCAR and UCAR organization, 2) introducing the students to the data archiving activities within NCAR, and 3) answering any questions.

As in Year 1, NCAR hosted a kick-off workshop during the first week of the internship. The workshop was 1.5 days long, and largely followed the format of the Year 1 kick-off workshop, but adjustments were made to the structure and content based on the feedback from Year 1. For instance, the workshop was shorter in duration, and had a smaller set of speakers. The advanced preparation of the students and projects also changed the goals of the workshop. Discussions between students and mentors focused on how to get projects running quickly.

Similar to previous years, NCAR Library staff organized weekly meetings for the students to share experiences, and meetings with local scientific information or data professionals. The students also were given opportunities to participate in a wide range of professional development activities, including two geoscience informatics meetings that coincidentally took place in, or within driving distance from, Boulder, Colorado, while the students were at NCAR. Finally, multiple events took place within UCAR for interns, giving the students opportunity to learn more about their internship site, and interact with students in other internship programs.

**Year 3 lessons learned**

Formal evaluation for the Year 3 internships is still ongoing. Preliminary feedback from the students and mentors indicates that the adjustments made following the Year 1 and 2 internships have helped to smooth the process of identifying good student and mentor pairs, as well as to get projects organized and launched quickly. The numerous additional professional development opportunities were well received, and enabled the students to make helpful professional contacts. This positive feedback on the Year 3 internship experiences will be folded into the evaluation of the program as a whole.
Optimizing the DCERC Internship Model

With three years of DCERC internships now finished, and iterative adjustments made at each stage, we are able to outline how to optimize the DCERC internship model to provide positive and productive experiences for students and mentors. We focus on three key issues that have been central to operating data curation internships within a research and data center: 1) hitting the ground running, 2) aligning expectations, and 3) situating the internships in a larger professional context. The discussion of these issues derives from the assessment of the feedback and data gathered through the various evaluation activities discussed above.

Hitting the Ground Running

Central to the evolution of the DCERC internship model has been an increased emphasis on tailoring student preparation to increase their ability to initiate their projects once the internships begin. Preparing students for the DCERC internship involved a number of facets. Students benefited from having more knowledge of the internship site, such as the background on the mission and structure of hosting organization. Many research and data centers, including the data management teams at NCAR, are driven by particular missions, and serve particular communities. In Year 1 of the DCERC internship program, the kick-off internship workshop was intended to introduce the students to NCAR and its activities. In subsequent year, more information about NCAR and the data facilities was provided to the students prior to the internships. This additional background information enabled the students to understand how their projects would be situated within the NCAR organizational structure and broader mission.

Another important set of adjustments related to the pre-internship student-mentor communication. Facilitating communication between the students and mentors about projects prior to the internship was very beneficial to both groups. Mentors were able to use this pre-internship communication to organize more streamlined projects, to prepare their physical space (e.g. prepare offices and computers), and to bring other colleagues into a project where appropriate. Students in Year 2 and 3 were more able to jump right into project tasks.

Finally, the kick-off workshop was refined to be more directly relevant to student projects. Whereas the kick-off workshop for the Year 1 covered a wide array of topics in data curation, the kick-off workshop for Year 3 was organized to specifically address topics that would be relevant to the student projects. The length of the workshop and number of speakers for the Year 3 workshop were about half of that from Year 1. One aspect of the workshops that did not change was considerable time for students and mentors to meet. The goals for the student-mentor discussions did shift from project organization in Year 1 to project launch in Year 3.

Managing Expectations

A big hurdle that must be addressed with mentors who are based in disciplinary research and data centers is their lack of familiarity with LIS programs and students. Data center professionals typically come to their positions with scientific or technical backgrounds, and have little awareness of the LIS student backgrounds, course work, skills, and career goals. DCERC staff had numerous informal meetings with prospective mentors.
to discuss appropriate projects and work tasks. As our program evolved, we increasingly included the data mentors in the evaluation of the student applications, and facilitated student-mentor discussions students before the internships begin.

Serving as a DCERC mentor did require an investment in time by both data and science mentors. While we have not attempted to determine actual hours spent mentoring DCERC students, a number of mentors commented that the time commitment was larger than originally anticipated. To ensure that this time investment was worthwhile, DCERC emphasized finding student-mentor pairings that were “win-win”, namely where both sides benefit from the students’ projects. Student projects were organized to both a) match students’ interests and career goals, and b) provide benefit the mentors’ work. Benefits to the mentor might be new knowledge, new technology, new data collections, or new processes.

Situating the Internships in a Larger Professional Context

A key practical goal of DCERC has been to train data curation professionals. An important emphasis throughout the DCERC program has been to provide the students with opportunities to understand the larger context of the data profession. This has included organizing professional development and networking opportunities, providing considerable informal mentoring, and adjusting the formal mentor structure.

Professional development opportunities complement the students’ formal education and internship projects by enabling the students to learn about the broader professional circles in which data curators work. As discussed in the Year 1-3 narratives above, we organized opportunities for the students to attend professional meetings of geosciences information and data experts. In addition, we tried to take advantage of the strong network of geosciences organizations in Boulder, CO, area by setting up meetings for the students to interact with local professionals from other organizations. These professional interactions allow students to present their work in informal fashion to new people, and to see how their work within one organization fits in a wider network of professional practice.

DCERC also emphasized giving the students opportunities to reflect on their experiences via informal weekly meetings with NCAR DCERC staff and their weekly reflections in Moodle, a content management software. Being able to synthesize and reflect on experiences and events is a critical part of becoming a new professional. The meetings with NCAR DCERC staff also provided an opportunity to discuss the outside professional development meetings and conferences.

Conclusions

The primary lessons learned in operating the DCERC program over three summers relate to the themes of mentorship, partnership, exchange, and iteration. The multi-mentorship model, however mentor roles are defined, pushes students and mentors towards projects that are collaborative in nature, and require alignment of multiple interests. Such collaborative projects are essential for the students to experience, as they are a regular feature within data curation workplaces. The most effective mentor combinations in DCERC have existed where one of the mentors recruits the second.

DCERC mentors have been partners in the development of the internship program, and have been critical sources of feedback throughout the program. In emphasizing
mutual benefit for students and mentors, we have encouraged and facilitated partnership between the students and mentors. The different professional perspectives that the mentors and students brought to the internship projects were initially a hurdle to surmount, but in the end provided an opportunity for exchange of ideas, concepts, and techniques for addressing data management challenges. Because of this exchange, the mentors and students worked on a collaborative peer-to-peer basis as much as they worked on a traditional master-to-apprentice basis.

Finally, the iterative nature of the DCERC program has been critical in adjusting different components of the internship program, from the student and mentor preparation, to the internship project implementation. The evaluation data have enabled us to be responsive to problems, inefficiencies, or opportunities as they arose.

These themes – mentorship, partnership, exchange, and iteration – should help with the development of future data curation internship programs in the future, especially as data curation educational programs continue to build on the expertise already in place within research and data centers.

Acknowledgements

The DCERC program is funded by IMLS grant # RE-02-10-0004-10. We thank the many NCAR staff who have served as mentors, as well as Mary Marlino and Karon Kelly of the NCAR Library for their essential contributions to the program.

References


