

# DCERC

## Data Curation Internship Evaluation

Part 1 (Survey and Anecdotal Responses)

Summer, 2012

Location: National Center for Atmospheric Research, Boulder, CO

Prepared by Virgil E. Varvel Jr. for DCERC (CIRSS Report DCERC1202)

### Contents

Introduction: .....	1
Goals .....	3
Summary of Findings.....	3
Survey Responses: .....	5
Student Responses , n=3.....	5
Mentor Responses , n=4 .....	7
Non-Response Analysis .....	9
Additional Metrics Specifically Mentioned in Proposal .....	10
Conclusions and Recommendations:.....	10

### Introduction:

One of the primary aspects of the Data Curation Education in Research Centers (DCERC) program is the incorporation of a data internships or field experiences into the educational program. The internship experiences will be evaluated and analyzed in three ways. First, an end-of-event survey was voluntarily completed by the students and mentors. Second, anecdotal input by students and faculty will be integrated into the on-going case study. Third, the internship is being formally evaluated by an external evaluator through a focus group. The first aspect of this evaluation informed in parts by the other aspects is reported on here.

For the Masters students, these internships occurred at NCAR during the summer of their first year and a future internship will occur during their second years (thus at the end of the educational years). This year represents the first year for the students. The MS internships will each last 8 weeks. The first internship began immediately following a data curation workshop. Through June and July of 2012, four students took part in internships at NCAR.

NCAR, the students were also integrated into the existed Advanced Study Program (ASP), a program with a long history of mentoring young and aspiring scientists (<http://www.asp.ucar.edu>). While in residence, students shared their experiences with classmates and faculty on a regular basis through an

online forum as well. They also held meetings with the local DCERC coordinators, Karon Kelley and Matthew Mayernik, to report on their progress.

Each student was paired with a data and science mentor. These mentors were designed to provide multiple educational inputs for the students. The pairings were made through a collaborative process based on student interest statements, mentor interests, and meetings help with the DCERC team. The pairings are documented in a separate DCERC Student Research Interests and Mentor Pairing document.

The mentors worked collaboratively with the student's academic advisor to guide and monitor the student's activities and experiences and to ensure alignment with the educational goals of the student and the requirements of the home institutions. These are research-oriented goals for doctoral students and practice-oriented goals for master's students. Their projects covered a wide range of topics based upon the students' interests and included:

- Creation of a cross discipline data management plan
- Assessment and potential improvement of the climate data guide
- Conducting a data audit on the data lifecycles of two groups in the Earth Observing Laboratory
- Following data through its lifecycle at NCAR

Two final deliverables were required for this first summer. Each student prepared a poster presentation for a closing reception for the program on July 26th. The reception was attended by all involved in DCERC at NCAR, the project PIs, and interested guests, with a visiting scholar in data curation providing a keynote talk. The posters were all well done and well received by the audience. The group of students is also required to submit a collaborative poster proposal to the 2012 International Digital Curation Conference, due in October 2012. A first draft has been circulated to the students' faculty advisors. [UPDATE: All students submitted posters or papers that were accepted to IDCC.]



picture obscured in public version

## Goals:

The internship aspect of the DCERC program encompassed several of the programs goals. These included:

- Introduce students to data curation practices and issues in a research center environment.
- Build community among students, science mentors, data mentors, and institutional faculty.
- Foster collaboration and synergy for data management practice and future internship experiences with mentors.
- Specific impacts addressed from proposal with how they can be accomplished in brackets:
  - Train a core base of experienced graduates who have a first-hand understanding of contemporary data-intensive research environments, and who can assume leadership positions in cyberinfrastructure R&D and education. [through educational activities including internships and weekly meetings.]
  - Build effective collaborations among scientists and LIS students that focus on using and preserving data resources for solving scientific problems. [through interaction among students and mentors and continued collaborations .]
  - Contribute new perspectives and findings to the literature in this emerging field, with research results, documentation of curation best practices, and advances in practitioner training. [through conference reports based on findings from this experience by the researchers and the student poster proposals.]

Through the findings of this evaluation, we will address how well the internships have met these goals.

## Summary of Findings

The findings of the survey with additional evidence when available are presented here. The sections on outcomes specifically speak to the goals above; however, all sections speak in some ways to the quality of the DCERC program.

**Overall Value of Internship Experience:** From everyone's perspective, the internship was a valued experience. Both students and mentors rated the overall experience as extremely valuable and strongly agreed that they would recommend the internship experience to other students or colleagues.

**Overall Value or Quality of the Intern:** The mentor's rated the overall quality of their student interns as excellent. They also strongly agreed that the internship fully lived up to their expectations. The mentors were also pleased at the synergism that existing in which they gained valuable data management work through the intern as well as data curation knowledge and insight into their own projects from the interns. One problem occurred that is only hinted at in the comments as 'citation' where a student did not understand the scholarly rules of acknowledgements and authorship on works resulting from the internship, but these issues were resolved. Additional Unix knowledge would also have been helpful in one group, but was not known to have caused any issues producing required work results. At the same time, another student was highly praised for her attitude and work in comments on the survey.

**Overall Value or Quality of the Mentors:** Although the mentors were not directly rated by Likert scaled questions (We do not directly survey this aspect so as to maintain mentor buy-in. Mentor quality is

internally monitored by NCAR staff through their qualitative assessment of performance.), we can assess their positive value/quality through unprompted comments received from the students. Every student entered an added comment stated that the staff and/or mentors were good, accessible, helpful, committed, and/or recommended. The unprompted nature of these comments adds weight to their value.

**Intern Preparedness:** In both the student and mentors' opinions, the interns were adequately prepared for their placements. We will word future surveys to ask if they were well prepared in order to evaluate program improvement. One intern did feel that additional preparations would have been helpful based on open-ended responses. See the improvements section below for more including suggestions.

**Student Outcomes:** A primary outcome from the student perspective will be whether the internship experience makes them more employable. Future tracking of the students will demonstrate this employability factor, but in terms of perceptions, the students unanimously strongly agreed that the internship experience made them more employable. They also agreed that the work that they were performing as part of their internships was closely related to their career objectives, thus demonstrating an alignment with project designs and student goals. The students also all had a self-perceived strong agreement that they learned a great deal during their internship.

As part of this experience, the students also had to create tangible research products in the form of a poster presentation at the end of the internship experience. These posters are currently being refined with the support of the students' institutional faculty mentors into conference quality contributions to the field that will be submitted for the iConference 2013.

The skills learned, enhanced, and performed by the students during their internships were wide and varied, encompassing many data curation topics. These skills and knowledge elements included: data access and acquisition; data ingest; data appraisal; research and assembling of datasets; data transformation and format conversion; data visualization; data management of scientific data sets; usage of scientific tools related to atmospheric science; metadata collection, audit, and coding; archiving and preservation; data recovery; research skills such as interviewing, problem solving, and report writing; content management through Drupal; web analytics; patience and time management; organization and documentation; and collaboration skills in a scientific setting. Mapping these skills as well as the deliverables produced by the students (final project posters) to the DigCCurr Matrix of Digital Curation Knowledge and Competencies (<http://www.ils.unc.edu/digccurr/digccurr-functions.html>), one sees that of twenty-four competencies, students gained experiences in eleven directly. They gained: access – discovery and retrieval of data; advocacy and outreach – synergistic activities with NCAR personnel; analysis and characterization of digital objects – metadata activities; archival storage – various archival activities; data management – many respects such as reports, queries, and updates; identifying, locating, and harvesting – data access and retrieval of aggregated datasets; ingest – exposure in some projects; selection, appraisal, and disposal – appraisal in one project; transfer – specifically in one project; transformation – specifically in one project; use, re-use, and adding value – not done in particular, but supported by student work in some projects; and although not in Matrix terms, working with mentors and scientific workforce developed collaboration experience.

**Mentor Outcomes:** In addition to getting an educated data curation student for eight weeks who could perform data assembly, management, quality control, etc., the data and research scientists appeared pleased at additional outcomes from this internship experience. A synergistic relationship existed in which not only the mentee benefited, but also the mentor in intellectual ways. The mentors learned how data curators work and about data curation in general. They developed data management knowledge and learned to think about their data in new ways. They also saw the value in a continued internship program.

**Improvements and Recommendations:** No matter how positive a program is, there is always room for improvement, and the willingness of participants to share ideas demonstrates the open atmosphere of this program. Several ideas were shared that could improve future iterations.

One student cited several times how s/he would have liked more advanced notice about his/her project and mentor placement. These were made known one week prior to the internship start. The students then had this time to work on a project plan for the summer. More time would be helpful, but it is also difficult to create the mentor pool and create matches. Often, internship assignments are not made until interns arrive at their location; however, the DCERC program is planning to try to make these matches earlier in the future when possible. In addition to these matches, bringing the students a higher level of familiarity with the vocabulary and science of NCAR prior to even the pre-internship workshop could be accomplished through reading materials prior to arriving on site.

Another student desired a longer internship period. We do not believe this is possible with funding limitations and remembering that the students will actually return next summer for a second aspect to their internship. At the same time, a longer internship could also be considered employment.

A recurring opinion from a few appears to be the desire to make this experience a regular part of the NCAR process, and we hope that this also occurs.

## Survey Responses:

Survey questions are listed below sorted by category with qualitative responses listed by relevance, priority level given by respondent, or amount of repetition by respondents. Some responses are paraphrased in some cases to merge or reduce entries. Quotation marks represent preserved entries that contain particularly telling remarks. Slight grammatical or spelling corrections may have been made. Following this list of responses, a summary of findings is presented that addresses the goals above.

## Student Responses , n=3

### *Overall*

- Overall value of internship experience (extremely valuable...not valuable at all)  
Average = 1.00 = Extremely Valuable; StDev = 0
- Overall quality of supervision (excellent...poor)  
Average = 1.67 {1,1,3} = Above Average; StDev = 1.15

- I would recommend this internship to other students. (strongly agree...strongly disagree)  
Average = 1.33 {1,1,2} = Strongly Agree; StDev = 0.58

### Preparation

- Did you feel adequately prepared for this placement? (yes, absolutely...no, not at all)  
Average = 2.33 {1,2,4} = yes, mostly to yes and no; StDev = 1.53

### Outcomes

- This internship has made me more employable (strongly agree...strongly disagree)  
Average = 1 {1,1,1} = strongly agree; StDev = 0
- The work you performed as a part of your internship was closely related to the career objectives you had prior to participating in the program. (strongly agree...strongly disagree)  
Average = 2 {1,2,3} = Agree; StDev = 1
- I learned a great deal during this internship. (strongly agree...strongly disagree)  
Average = 1 {1,1,1} = Strongly Agree; StDev = 0
- What skill(s) did you enhance and/or develop as a result of your internship?
  - Scientific data management
  - Scientific tool usage
  - Metadata
  - Archiving
  - Data Transformation
  - Working with scientists
  - Patience and Time Management
  - Organization and Documentation
- What were the 3 most common activities you performed as part of your internship?
  - Data management
  - Data transformation
  - Data appraisal
  - Metadata creation
  - Interviewing
  - Report writing
  - Problem solving
  - Documentation
- Please describe the most positive aspect(s) of your internship (1-3 points).
  - "It was nice to see first-hand how scientists manage data."
  - "It reinforced my interest in offering data management training one day."
  - "It was neat to see how scientists use visualization tools for data QC, especially open source software."
  - "I had a chance to follow the complete data curation process rather than doing one piece of several project. I had a chance to see the big picture and appreciate the time and energy it takes to do the job right."
  - "I felt that everyone got something out of the experience."
  - Mentorship
  - Work
  - Community/collaboration

- Please describe any negative aspect(s) of your internship (1-3 points).
  - “It was difficult and sometimes overwhelming, but my mentor always stepped in to help me fix problems that I could not solve.”
  - “The time here was too short. I would like to have several months to a year at NCAR because I think I could learn more and do more if I had more time.”
  - “While i understand this is a pilot program, having known who the mentors were and the area of science we would be dealing with about a month before we arrived would have been excellent.”
  - “There was a level of familiarity with the vocabulary and science that could have been accomplished during this time to make the first bit of the internship a more fluid experience.”
  - None

### *Recommendations*

- Do you have any added comments or recommendations for improving the internship aspects of this program?
  - “My mentors were good. They were accessible and helpful.”
  - “Both of my mentors were great. They were supportive and made it clear that they were available to help me. They also gave me enough space to work independently and learn on my own. They each had a unique approach and different tools with which to assist me. “
  - “All the staff were really great and supportive as well. I always felt that my learning came first and they were here for me, and I was there to contribute as much as I could as well. The activities outside of work were really fun, too. It was all in all a fantastic experience and I cannot wait to come back next summer.”
  - “I would recommend my data mentor for years to come. He was very committed to making sure that I got out of the experience what I was hoping to, and in turn, I helped him to see where some of his archive could be more user friendly. I just think that the level of preparation could have been better, but now after having been there for a summer, I think we all know how we would like to continue at NCAR.”

## **Mentor Responses , n=4**

### *Overall*

- Please rate the overall value of this internship experience. (extremely valuable...not valuable at all)  
Average = 1.4 {1,1,1,2,2} = Excellent; StDev = 0.55
- Please rate the overall quality of your student intern during this internship experience (excellent...poor)  
Average = 1.4 {1,1,1,2,2} = Excellent; StDev = 0.55
- I would recommend this internship experience to other colleagues? (strongly agree...strongly disagree)  
Average = 1.2 {1,1,1,1,2} = Strongly Agree; StDev = 0.45

### Preparation

- The student was adequately prepared for this placement (strongly agree...strongly disagree)  
Average = 1 {1,1,1,1,1} = Strongly Agree; StDev = 0
- Note for next survey: Need to add a question asking if the mentor felt that s/he felt adequately prepared [no one specifically mentioned the need for this, but as a research I can see the need.]

### Outcomes

- The internship fully lived up to my expectations. (strongly agree...strongly disagree)  
Average = 1.2 {1,1,1,1,2} = Strongly Agree; StDev = 0.45
- Please describe the most positive aspect(s) of this internship experience (1-3 points).
  - "I learned how the new generation of data curators thinks, design solutions, and how they perceive the expected audience of their work."
  - "I learned something about 'data curation' an area of which I was not aware."
  - "- From my selfish point of view, the assembly of data that are useful for me and my colleagues."
  - "- Collaborating with the intern and with the data mentor was a pleasant experience."
  - "- I was thankful that as science mentor I did not have to dedicate too much of my time. The greatest time commitment occurred in the first week. Then approximately 1 hour per week."
  - "[Student Name] came in wanting to learn everything that she could about the data lifecycle. She asked lots of good questions and sometimes challenged the "why" behind the way we curate data. This led to some good discussions where we both learned from each other. Also, [Student Name] comes from a discipline that is outside of our target community and she was able to give some good feedback on how we can make our archive more useful to users outside of the target discipline."
  - "Explaining my work revealed various shortcomings and inadequacies in the processes involved, which will be improved and therefore make my work more efficient and productive."
- Please describe any negative aspect(s) of this internship experience (1-3 points).
  - "The student had insufficient computer experience, in the unix environment, to fully accomplish all the steps necessary in end to end data curation. "
  - None x2
  - "My email address was wrong on the contacts sheet, and I reported it on day one. However, my intern was emailing me at the wrong address and got stressed because she thought that I was ignoring her."
- What were the 3 most common activities that the intern performed as part of the internship?
  - Evaluation of data collections
  - Downloading data
  - Reorganizing data
  - Posting the reorganized data
  - Researching and assembling a dataset
  - Format conversion
  - Assisted archiving
  - Preparation of documentation
  - Learned Drupal



- Google analytics
- Had to learn citation 'stuff'.
- “Learning about the types of metadata that we capture and how it allows us to curate data so that users can find datasets, we know what we have in our archive, we can preserve the data for the long-term, and we can recover from a catastrophic data loss.”
- “Actually adding a dataset to our archive and going through the full process of metadata collection.”
- “Performed a metadata audit efficiently and correctly.”
- Regular meetings
- Learned the various technical aspects of my work very quickly

### *Recommendations*

- Do you have any added comments or recommendations for improving the internship aspects of this program?
  - “I think this was a very worthwhile experience and I hope NCAR continues to participate. Data curation is a complex process and it seems that it is in a data archive's best interest to foster the development of future curators if we want to be able to preserve data that are often irreplaceable. The idea of both a science mentor and a data mentor is very smart. It reinforces for the student the need to be in contact with the users and to incorporate user feedback as part of the data curation process. From a mentoring standpoint, it was refreshing to step away from the day---to---day "auto---pilot" work that I do and take the time to explain it to someone new. It reinforced the "why" behind much of our operations and also helped to highlight areas where we can continue to improve. One recommendation I would have for the interns would be to get exposed to some Unix and some simple scripting. Parts of the data archival process which are repetitive and tedious can be sped up if one knows how to write a script to do many operations in a single process.”
  - “I greatly enjoyed working with my intern and look forward to this program becoming a regular part of the interactions NCAR has with the library sciences community.”
  - No further comments.

### **Non-Response Analysis**

Not all participants responded to the survey. The student who did not respond to the survey did respond to an email request and stated that s/he “highly valued the program and the opportunity and would complete the survey soon”; however, s/he did not eventually complete the survey. Mentor time is highly valuable, and thus they were not pressed to complete the survey after one reminder by email and one verbal reminder by an NCAR representative. Two non-responders did provide some constructive verbal feedback that was incorporated into the case study. As above, they valued the experience, but also provided suggestions for improvement such as that we clarify citation of results expectations and provide more training on academic collaborative processes. Despite a few moments of conflict that were resolved, the overall experience was positive and both mentees desired to continue the mentorship experience. Their mentee also returned positive survey responses.

## **Additional Metrics Specifically Mentioned in Proposal**

Number of NCAR personnel involved in education: 1 data mentor and 1 science mentor per student plus 2 NCAR personnel involved in meetings and overall management = 10 NCAR personnel involved in student education related to internships.

Hours and types of mentoring conducted with students: Each student took part in meetings with NCAR faculty 1 hour each specifically concerning internship activities. The students were also required to post weekly updates to the program MOODLE communication forum. Students also met with their home institution faculty advisors on a weekly basis for 1 hour on average to discuss academic goals, personal discussion to resolve any conflict or answer programmatic questions, and to discuss their final poster project.

Student Contributions to NCAR operations: For this metric, see specific notes that point to such contributions in the analysis section below. Students made particular contributions in data management, data organization, data transformation, data practice, and curation education.

## **Conclusions and Recommendations:**

Based on the summary of survey answers presented below, the following conclusions and recommendations can be made regarding the 2012 DCERC Data Curation Internships.

Clearly the first DCERC internship was a success from both the intern and mentor perspectives judging from quantitative and qualitative responses to the evaluation survey. Overall they both viewed it as a quality experience in which both found significance in the other participant and in which a synergistic relationship existed permitting outcomes all participants saw as valuable. These outcomes included extensive data curation knowledge and experience, real-world scientific data practice, and data-intensive training for the students that addressed the goals of the DCERC program. The program also will contribute new perspectives and findings in this field through the poster outputs of the students. Furthermore, the mentors also benefited from this program through the collaborative relationships built with the students, building their data curation knowledge and practice, potentially improving NCAR data practice overall. These relationships could help to institutionalize the program over time and build further internship partnerships elsewhere.

The only recommendations are to potentially improve early information delivery to students with regards to the internship experience and to provide additional scientific citation guidance to avoid misunderstanding on shared authorship of findings. Early student-mentor matching might also provide for the possibility of pre-internship student skill acquisition that could facilitate the intern experience such as Unix knowledge that would have facilitated greater student involvement in one lab. These are minor recommendations on an otherwise successful first summer to this program for which all parties agreed should not only be included, but several cited should potentially be institutionalized for future partnerships.