

DCERC

Data Curation Workshop Evaluation

June 5-7, 2012

Location: Building FL1, room 2198 (the EOL Atrium)

National Center for Atmospheric Research, Boulder, CO

Prepared by Virgil E. Varvel Jr. for DCERC (CIRSS Report DCERC1201)

Contents

Introduction:	1
Goals	2
Conclusions and Recommendations:	2
Summary of Survey Answers:	3
Agenda:	6
Workshop Participants:.....	7

Introduction:

The Data Curation Education in Research Centers (DCERC) Summer Workshop was held at the National Center for Atmospheric Research (NCAR) in Boulder, CO prior to the start of the student internship experiences (the beginning of the Masters students' on-site research experience) and at the start of their first summer in the program. This report gives an overview of the results from a survey evaluating this workshop.

As part of this workshop, participants received an overview of NCAR including the nature of the science conducted there, and current and desired data practices. The four-day workshop was the first visit to NCAR for most and was designed to help students assimilate and relate what they had learned to the operations and goals of scientific research centers. Selected experts, including data scientists from other centers associated with DataNet, such as the National Snow and Ice Data Center (NSIDC) and Oak Ridge National Laboratory (ORNL), were invited to give presentations along with NCAR personnel. Presenters also held panel discussions on current data practices and problems and how they compared across sites. Informal forums for exchange and engagement with students and experts were a key part of the workshop, as were activities that helped students become more integrated as a cohort across the two home institutions.

Students remaining at NCAR for their internships were assigned both a science and data mentor. Students will be integrated into the existed Advanced Study Program (ASP) at NCAR, a program with a long history of mentoring young and aspiring scientists (<http://www.asp.ucar.edu>). The mentors will work 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 will be research-oriented goals for doctoral students, who will complete internships later, and practice-oriented goals for master's students.

Goals:

- Introduce students to data curation practices and issues in research center environment
 - Share expectations and outcomes for student research experience at NCAR
 - Build community among students, science and data mentors
-

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 Workshop.

Overall, the 2012 DCERC Data Curation Workshop was reviewed as a success by participants. The presentations were the most successful aspect. Participants left with an overall greater understanding of the role of research data centers in data curation and all specific topics evaluated in our survey. Even areas where they learned the most, they wanted to learn more, such as metadata, data citation, and data management in general. Unanticipated topics such as the politics of data centers and the tools of data management, which were not explicit workshop topics, were also listed as participant learning outcomes.

The workshop also provided networking opportunities, knowledge of current influential groups, and understanding of contextual aspects of data centers. Participants listed items such as 'libraries were fun' and 'how to care about data' and 'the use of data' as outcomes.

No statistical significance could be drawn however on the amount of time spent with mentors or on social activities, with some individuals wanting more and some wanting less or different interactions for each. It was clear that some individuals wanted more up front knowledge of their expectations from the workshop and more time to familiarize themselves with their potential mentors and projects prior to their experiences.

Other potential suggestions for improvements to consider included: adding a peer mentoring component (which sort of occurred in that a student who had interned at NCAR the previous summer was included in the workshop); shorter presentations or more breaks; speaker handouts (as the presentations were not available until a week after the workshop); an overview of NCAR mission and research structure (perhaps as a document prior to attending the workshop to be included with an introduction to mentors); and student providing some questions to presenter panels prior to the

workshop (although it is unknown if they would have the proper expertise, but could provide some insight into their thought processes).

Summary of Survey Answers:

Survey questions are listed below with qualitative responses listed by relevance, priority level given by respondent, or amount of repetition by respondents. I have paraphrased in some cases to merge or reduce entries.

Items Liked Most About Workshop:

- Practitioner presentations overall or mention of a specific presentation (NCAR research; Metadata; Repository Information; Data Management; Data Citation).
- Networking, opportunities to collaborate, learning from others, open discussion, and the discussion amongst the different levels (PhD, masters, scientists) of people.
- Explanations of differences between DataOne and Data Conservancy.
- Food.
- Learning about data management.
- Overall structure.
- Availability of slides.

Items Learned from Workshop:

- Contextual aspects of Data Conservancy and NCAR including the cultural aspects of a national laboratory.
- The variety of data management issues and problems (includes tensions between scientists and data manager and strategies for communication).
- Metadata.
- Databases, data collection, and best practices.
- How climate models are made and climate research.
- Data citation.
- Librarians are fun.
- Data repositories.
- Tools and sites about data management.
- 'Caring' about data and its use.
- Bringing together data curation and data management.
- Qualitative data is respected.
- Nature of scientific research.
- High need for data curators.

What They Would Like to Learn More About Now:

- Metadata
- Data capture and other data procedures.
- Repositories.
- More about current research in general.
- Atmospheric science research and data modeling, formats, and standards.

- Ways to link information and data sets.
- Data citation.
- Policy, politics, bureaucracies of a data center, current influential groups in data circles.
- Data management and data management plans.
- Economics of data.
- The curation process in general.

I now have a greater understanding of the role of research data centers in data curation: 1.2/5 -> Strongly Agree

- Especially useful to learn about differences between research centers and other institutions.
- Would like to understanding more of NCAR's effects on data curation theory and practice in other sites and settings.

I now have a greater understanding of metadata issues and initiatives: 1.8/5 -> Agree

- More on NCAR cross-group metadata similarities and differences
- Who decides what's relevant and what isn't?
- Good on general value of metadata. I'd like to see an example of metadata in use post-creation, post-search interface. How did metadata help a scientist/

I now have a greater understanding of data citation and metrics of use: 1.3/5 -> Strongly Agree

- Especially the metrics, and how that can be used to quantify the value of a dataset or research analysis.
- I hadn't considered the problem, challenges, and issues surrounding data citation and metrics.
- More understanding of examples, not much more conceptual clarity.

An appropriate amount of exposure was given to research and data activities at NCAR: 1.7/5 -> Agree

- Gary's and Toni's presentations were too technical and I don't think added much. They seemed irrelevant to me.
- Hard to fit so much information into a 3-4 day workshop, but overall, I feel like I've been exposed to a broad sense of what NCAR does.
- There was a wide range of the activities – more than I realized – in which NCAR engages.
- Still more to learn, but this was a great start.
- I'd love to hear more about NCAR. Again I'd like to see a scientists present research findings and show use of metadata.
- More!

An appropriate amount of time was provided for meeting and working with mentors: 2.0/5 -> Agree

- More informal time would be good.
- We had enough time to flesh out the project ideas and goals. More time would have been unused.
- I do not feel I've had enough time or guidance in developing my summer plans. More direct help would've been appreciated, both from mentors and faculty. More time to understand my mentor's work would've been greatly helpful too.

There were adequate opportunities for social interactions with workshop participants. -> 1.6/5 Agree

- Not enough time in the day.
- The dinner was fun, but any more would have been too much.
- Was unable to make the dinner, but it seemed there were plenty of opportunities to talk and mingle.
- I especially enjoyed the dinner at Mary's. It was so pleasant and social. Thank you.
- Lunch times and evening party were very nice.

What other topics would you like to have had covered in the workshop?

- More organizational overview of NCAR including computational infrastructure, how data practices and management have changed over time, OAIS use in archiving data, and acronym guide.
- Long term data formats.
- Semantics of data.
- More examples of what data actually look like including raw data and cleaned up data.
- I think it would have been beneficial to discuss more in depth what the researchers need in the way of data management and what obstacles they face.
- Possible jobs in the field.
- Ph.D. project overviews.

New perspectives or insights gained from this workshop.

- Insights into national center and interfaces with other national programs. Differences between research centers and other institutions.
- Perspectives of scientists and data managers including how data management/curation is highly intertwined with science and that managing data is a complex process that 'takes a village' when done properly with a lot of problems still to overcome.
- Idea of how change is being enacted for data practices
- Data management is a science of organization and many scientists are pragmatic and not organizers.
- Data managers really need a strong background in the sciences they deal with.

Suggestions for Improvement

- Keep the same but introduce MOODLE before workshop.
- Shorter presentations with more breaks and interaction time.
- Speaker notes for handouts.
- More discussion of interdisciplinary data.
- I think the only thing I would add a peer mentoring component.
- I really wish I had more time to research my mentors, and figure out what I'd be doing this summer. More advanced notification would've helped. Also more direct instruction on what we're expected to produce. [someone wanted more information despite several emails prior to the event and notification of mentors a week beforehand.]
- Start with an overview of NCAR's mission and research structure.
- Instead of Matt supplying the questions to presenters – it might be helpful to have students supply some questions based on coursework or research interests. [although their questions might not align with the workshop directives, so a unified approach might be good.]

- Discuss the career possibilities so we can know what we are working toward. [This may be more of an institutional objective and outside of the scope of this workshop.]
- Being more up front in the beginning (weeks before we arrive) about mentors and presentation expectations. [Note here that in my email to students prior to the event they were told that they would not be expected to do a PowerPoint presentation per our discussions in our DCERC meetings, and then they were expected to have slides in their final presentation, so there was some miscommunication there.]

This workshop helped you in transitioning to your internship experience. 1.5/5 -> Agree to Strongly Agree.

- Group dinner was great.
 - My mentor helped most.
-

Agenda:

Day 1: Tuesday, June 5, 2012

- 9:00 – 9:30 Introductions, overview, goals of workshop and field experiences at NCAR
- 9:30 – 10:30 Overview of data acquisition and curation data practices; role of a research data archive/centers in data curation (Steve Worley – NCAR, Bob Cook - ORNL, Lawrence Buja - NCAR)
- 10:30 – 10:45 Break
- 10:45 – 12:00 Student presentations on research interests
- 12:00 – 1:30 Lunch—Pairing of Master’s students with data and science mentors; discussions over lunch
- 1:30 – 2:30 Panel on science issues in data use and reuse (Science mentors)
- 2:30 – 2:45 Break
- 2:45 – 3:45 Panel discussion on data practices and issues and how they compare across sites (Nicole Kaplan - LTER, Karen Baker – UIUC, Joan Burkepile - NCAR)
- 5:30 – Dinner at Mary Marlino’s home – students, mentors, faculty, and staff to continue developing student projects

Day 2: Wednesday, June 6, 2012

- 9:00 – 10:00 Instituting data services; Considerations and evaluation criteria – (Ruth Duerr - NSIDC)
- 10:00 – 10:45 Metrics of data use and citation — (Bob Cook - ORNL)
- 10:45 – 11:00 Break
- 11:00 – 12:00 Metadata issues and initiatives in USGS (Viv Hutchison - USGS)
- 12:00 – 1:00 Lunch
- 1:00 – 1:45 NARCCAP and NetCDF (Toni Rosati - NCAR)
- 1:45 – 2:30 CESM/IPCC data and metadata protocols/requirements (Gary Strand - NCAR)
- 2:30 – 2:45 Break

- 2:45 – 4:00 Student/mentor meetings to flesh out summer project plans. Faculty and DCERC staff meeting to discuss DCERC annual report.

Day 3: Thursday, June 7, 2012

- 9:00 – 9:45 Data citation—developing and implementing an institution-wide approach, experiences with NCAR pilot in creating metadata (Matt Mayernik - NCAR)
- 9:45 – 10:30 NCAR Library experiences in developing an institutional repository: OpenSky (Jennifer Phillips - NCAR)
- 10:30 – 10:45 Break
- 10:45 – 12:00 Student/mentor presentations on summer project plans
- 12:00 – 1:00 Lunch
- 1:00 – 1:45 Faculty panel – Academic information science perspectives on the workshop (Suzie Allard and Carol Tenopir – UTK, Carole Palmer - UIUC)
- 1:45 – 2:15 Workshop evaluation/feedback
- 2:30 – Field trip to Research Aviation Facility for students and faculty

Day 4: Friday, June 8, 2012 for Master’s students

9:00 – Set up for IT support, bring laptops and any other devices for which access is desired. Other administrative issues such as forms and key access will be dealt with at this time also. Meet in the FL2 Library.

The workshop presentations are now posted on the NCAR library web site, at the following location:
<http://library.ucar.edu/events/dcerc>

Workshop Participants:

Last Name	First Name	Title	Institution
Allard	Suzie	Associate Professor, Associate Director School of Information Sciences	UT - CCI
Baker	Karen	PhD Student	UIUC - GSLIS
Buja	Lawrence	Director, Climate Science and Applications Program	NCAR - RAL
Cook	Robert	Scientist	Oak Ridge National Laboratory Distributed Active Archive Center
Dattore	Bob	Data Mentor	NCAR - CISL
Duerr	Ruth	Data Stewardship Program Manager	NSIDC - CIRES
Eaker	Chris	Master Student	UT - CCI
Hutchison	Vivian	Data Management Program Coordinator	USGS - Core Science Analytics and Synthesis
Johns	Erica	Master Student	UT - CCI

Hunter	Nancy	Metadata Librarian/Associate Professor	Colorado State University Libraries
Kaplan	Nicole	Information Manager	Colorado State University Natural Resource Ecology Lab
Kelly	Karon	Associate Director, IIS	UCAR - Integrated Information Services
Leohrer	Scot	Data Mentor	NCAR - EOL
Levis	Samuel	Scientist Mentor	NCAR - CGD
Marlino	Mary	Director, IIS	UCAR - Integrated Information Services
Mayernik	Matt	Data Services Specialist	UCAR - Integrated Information Services
Palmer	Carole	Professor, Director of CIRSS	UIUC – GSLIS - CIRSS
Romero-Lankao	Paty	Scientist Mentor	NCAR - RAL
Rosati	Antonia	SIParCS Program Graduate student	NCAR - CISL
Schneider	Dave	Scientist Mentor	NCAR - CGD
Sherline	Crystal	PhD Student	UT - CCI
Siddell	Kayla	Master Student	UT - CCI
Strand	Gary	Data Mentor	NCAR - CGD
Suomela	Todd	PhD Student	UT - CCI
Tenopir	Carol	Chancellor's Professor, Director of Research Director of the Center for Information and Communication Studies (CICS)	UT - CCI
Thomer	Andrea	Master Student	UIUC – GSLIS - CIRSS
Varvel	Virgil	Research Data Analyst	UIUC – GSLIS - CIRSS
Weber	Nicholas	ASP Graduate student	NCAR - ASP
Worley	Steven	Data Mentor	NCAR - CISL
Wright	Mike	Director of Strategic Initiatives	UCAR - Integrated Information Services
Young	Kate	Scientist Mentor	NCAR - EOL