

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

**Matthew S. Mayernik &
Valerie Williams**

National Center for Atmospheric
Research, UCAR

Carole L. Palmer

Information School, University of
Washington

Cheryl A. Thompson

Graduate School of Library and
Information Science, University of Illinois
at Urbana Champaign

Suzie Allard & Carol Tenopir

School of Information Sciences, University
of Tennessee at Knoxville



Data Curation Education in Research Centers (DCERC)

PI: Carole L. Palmer

Co-PIs: Mary Marlino, NCAR & Suzie Allard, UT

Collaborators: Carol Tenopir, UT & Matt Mayernik, NCAR

Dates: 2010-2015

Goals

- Model for Library and Information Science education
 - Enrich curriculum and build capacity to strengthen expertise in LIS
 - Internships in research and data centers -- current state-of-the-art practices and front-line challenges
- Evaluate model
 - Effectiveness
 - Viability of scaling to network of partner data centers and iSchools

DCERC

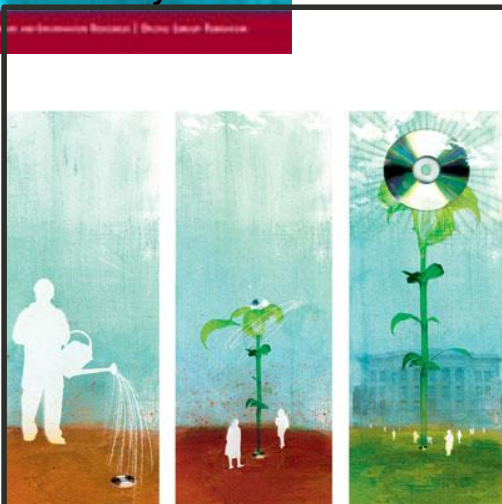


gin
n Science & Scholarship
formation Science
ampaign
and Library Science
er, Tiffany Chao, & Simone Sacchi
n Science & Scholarship

The Problem of Data

Lori Jahnke and Andrew Asher
Suzanne D. C. Kiehl
with an Introduction by Charles Henry
August 2008

Big data: How do your data grow?
Clifford Lynch



DCERC program

- Graduate level curriculum focused on data curation
 - Master's level enrichment of foundations course
 - New Doctoral seminar on research problems
- Academic research experiences with faculty mentors
- NCAR field experiences
 - **8 week internship projects guided by mentors**
 - **Begins with kick-off workshop**
 - **NCAR-wide Intern Final Presentations**



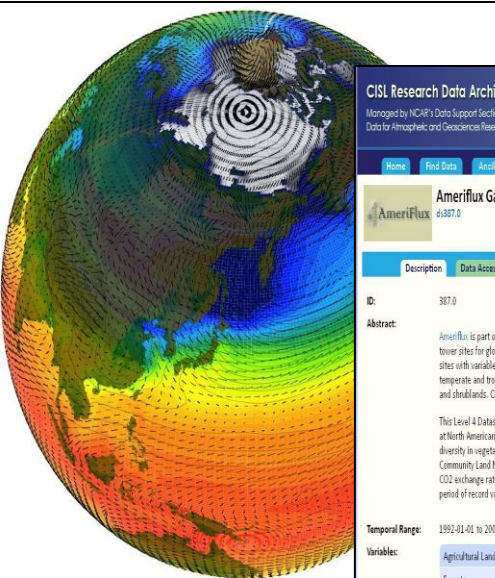
NCAR internships

- Three cohorts offered summer 2012 -2014
- 9 Masters students participated, 10 total internships
- Students paired with multiple mentors
 - Data and Science mentors
 - 16 NCAR staff members served as mentors
 - 4 data mentors served multiple years



Example student projects

Climate model metadata



Archiving data

CISL Research Data Archive
Managed by NCAR's Data Support Section
Data for Atmospheric and Geosciences Research

Home Find Data Ancillary Services About/Contact

Ameriflux Gap-Filled Data from Selected Sites
doi:10.7927/0346.0000

Description Data Access Documentation

ID: 387.0

Abstract:
Ameriflux is part of the larger network Fluxnet which endeavors to connect observations tower sites for global analysis. Ameriflux towers measure the exchange of energy, water sites with variable vegetation, disturbance records, and climatic conditions. The ecosystem temperate and tropical evergreen forests, temperate deciduous forests, woodlands, grass and shrublands. Climates in these biomes vary between tundra, temperate, tropical, and

This Level 4 Dataset accumulated from the Ameriflux Network focuses on hourly, daily, and monthly data for North American tower sites. The selection of Ameriflux sites were chosen based on their diversity in vegetation types, the past success of better land surface models, and the final Community Land Models. Ameriflux Level 4 data has been gap-filled, quality-flagged and CO2 exchange rates, heat fluxes, biogeochemical pools, calculated gross productivity, and period of record varies by individual tower site, but some sites have as much as 15 years

Temporal Range: 1992-01-01 to 2008-01-01

Variables:

Agricultural Lands	Air Temperature	Alpine/Tundra	Biogeochemical Cycle
Forests	Grasslands	Heat Flux	Humidity
Incoming Solar Radiation	Precipitation Rate	Respiration	Soil Moisture/Water Content
Soil Temperature	Solar Radiation		

Metadata harvesting, data quality, standard compliance

Sensor data, data processing, file migration

Data and metadata organization

Translator and Facilitator

Understands and articulates the needs and goals of scientist

Understands and articulates data manager needs for curation

Creates guidelines to enhance communication and efficiency between scientists and data managers

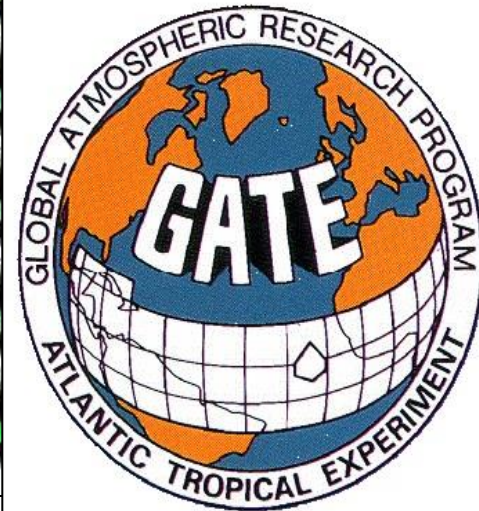
Cross-disciplinary data curation; data subsetting

Archiving temporal and spatial data



Time series data, high resolution, provenance, NetCDF

Assess analog data for digital access



International data, 50 collections, OAIS, DOI assignment

Formative evaluation questions

- Which processes are functioning well or need improvement from the student and mentor perspectives?
- To what extent are the primary activities yielding intended short term outcomes (e.g., application of theory to practice)?
- To what extent does the DCERC model improve education for students? Contribute to NCAR's operations and mission?

Internship evaluation methods

- Questionnaires
 - pre- & post-internship with students and mentors
 - post-workshop
- Weekly student reflections
- Post-internship focus groups of students and mentors
- Employer perspective interviews

Hitting the ground running

- Increase pre-summer communication
 - Conference calls orienting to NCAR organization and science
 - Project specification between students and mentors
- Streamline the kick-off workshop
 - Reduce the length but increase project discussion time
 - Reduce external speakers
- **Provide more preparation**
 - **Students: exposure to science, research projects, visit to research center**
 - **Mentors: exposure to LIS, student background, coursework, and scoping projects**

More professional development and networking

- Attending GeoData, ESIP meeting, The HDF Group, etc.
- Informal mentoring



Key considerations

in applying the model

- Multi-mentorship from scientists, data professionals
- Collaborative, peer-to-peer project design
- Recognition of LIS contributions to data management
- Formative evaluation & iterations

Acknowledgements

Carole L. Palmer, PI, Cheryl A. Thompson
University of Illinois Urbana-Champaign
Graduate School of Library and Information Science

Suzie Allard, Carol Tenopir
University of Tennessee
School of Information Sciences

Mary Marlino, Matthew S. Mayernik
National Center for Atmospheric Research

Students, Scientists and Data Mentors at NCAR

*DCERC is funded by the Institute for Museum and Library Services
Laura Bush 21st Century Librarian Program*