2013

Exploring Best Practices for Research Data Management in Earth Science through Collaborating with University Libraries

Ting Wang
Lehigh University, tiw412@Lehigh.edu

Benjamin D. Branch
Purdue University, bdbranch@gmail.com

Exploring Best Practices for Research Data Management in Earth Science through Collaborating with University Libraries

Ting Wang, Lehigh University Library and Technology Services, Bethlehem, PA. Email: tiw412@lehigh.edu
Benjamin Dewayne Branch, Purdue University Libraries, West Lafayette, IN. Email: bdbranch@gmail.com

IN13A-1554
ID: 1813092

Introduction

To address the growing demand for managing earth science data, the Council on Library and Information Resources (CLIR) partners with the Library and Technology Services (LTS) of Lehigh University and Purdue University Libraries (PUL) on hosting postdoctoral fellows in data curation activity. This inter-disciplinary fellowship program funded by the SLOAN Foundation innovatively connects university libraries and earth science departments and provides earth science Ph.D.s opportunities to use their research experiences in earth science and data curation trainings received during their fellowship to explore best practices for research data management in earth science and find the answers to:

- How can a university library play a core role in STEM outreach/pipeline development and support the data curation agenda of a research university?
- Is it possible for a university library to promote a cost saving if such data curation can be effectively sustained in research data lifecycles?

Lehigh University

Background

- Lehigh Library and Technology Services (LTS): a merged organization based on the two major library facilities, the Fairchild-Martindale Library and the Linderman Library, and delivers computing, library, media, instructional technology, faculty development and distance education services to enhance teaching, learning and research.
- Earth and Environmental Sciences: a diverse department having fifteen active faculty across sub-disciplines of climate, ecology, environment, geochemistry, geochronology, geography, geomorphology, hydrology, limnology, magnetism, sedimentary, seismology, and tectonics.

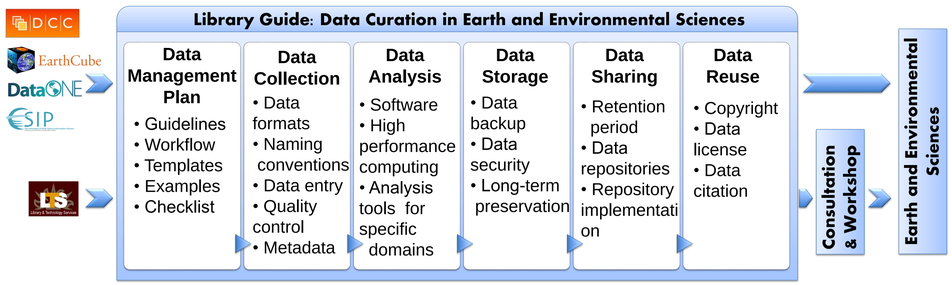
Data Management Assessment through Individual Interviews

We interviewed the faculty members and their groups in Lehigh University's Earth and Environmental Sciences department and identified a variety of data-related behaviors and needs from this data management and curation investigation.

Data Lifecycle	Data Management behaviors	Data Management Needs
Collection	From: field, lab, data centers. Type: numerical data, field and lab notes, maps, graphs.	Quality control; data documentation; and metadata standards.
Analysis	Software: ArcGIS, Excel, LabView, Matlab, R, SAS, SPSS, Stata.	Workflow and version tracking; naming conventions; data visualization; and software version compatibility.
Storage	Store at: lab computers, hard drives, university servers, cloud. Backup: automatic backup with software, manual backup.	Space limits and security; physical sample archiving; data appraisal/selection; and format migration.
Sharing	When: after publication, funding agency (2-3 years after collection). Where: domain repositories, institution repositories, on request (email or cloud).	Data submission; data maintenance; and data license, copyright, intellectual properties.
Reuse	What: raw data, intermediate and final research products, method/code, and other essential files for research reproducibility.	Research reproducibility; data search; and data citation.

Data Management Service Development

We created the library guide "Data Curation in Earth and Environmental Sciences" as the web platform to collect and share the best practices in earth science data management, and meanwhile provided individual consultation and group workshops to enhance the awareness of our program on the campus.



Purdue University

Research Questions

1) How can a university know if its geospatial community is properly supported by the university library?
2) Is Data Curation Profile (or other use cases methods) sufficient to achieve sustainable geospatial data services?
3) What challenges exist in curating geospatial data that could expand to national or global audience use?

Research Epiphanies

- Working in collaboration across campus is more powerful than isolated department research.
- Geospatial data curation is still an emerging field.
- Full data curation lifecycle support takes a team across libraries and across departments.
- Some geospatial knowledge engagement may cross into the digital humanities (**S**cience, **T**echnology, **E**ngineering, **A**rts/Humanities, Agriculture, Aerospace, **M**athematics) (**STEAM**).
- Metadata standards and compliance takes a lot of time, planning and effort.

Working to Define Data Curated Geospatial Services on a Research Campus

A Purdue University Libraries Example

- Conducted 32 geospatial data curation profile interviews
- Found geospatial data needs in earth sciences and in digital humanities areas
- Won a USDA grant from geospatial inquiry
- May generate a digital humanities grant from geospatial inquiry
- An educational data curation framework was endorsed by the GLOBE Program

PURDUE UNIVERSITY

Things to be considered

- What Does 32 interviews indicate about potential GIS and data curation services
- Should we discuss 32/60 ratios and challenges
- Should we calculate the research potential of each interview for a business model consideration
- What metrics beyond DCP would indicate a sustainable research trend

THE GLOBE PROGRAM

Data Repository 1

Ex. **Integrating Spatial Educational Experiences**
http://isee.purdue.edu/ Soils by Professor Darrell Schulze

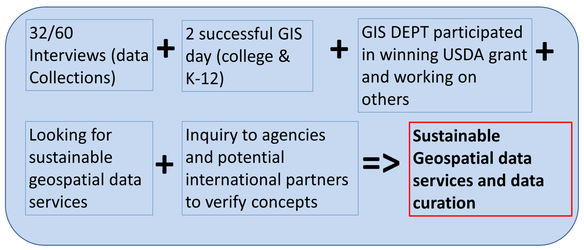
Nicole Kong –Purdue GIS repository or any Higher Ed research with GIS data

Educational Data Repository

Now learning is done for teacher where they make age appropriate.

Dewayne Branch-Pulling our reusable data sets, that prove themselves over time for Pipeline support. Higher Ed. or K-12.

Present Research Findings



Left diagram shows using multiple approaches a library can define its geospatial community for sustainable data curation consideration.

More than one means or set of inputs may be needed for sustainable geospatial data curation.

Challenges: Metadata, infrastructure, incentives, geospatial training.

Conclusions

Insights

- The processes of data management interviews and implementing data curation profiles (www.datacurationprofiles.org) provide a model for similar initiatives at other university earth science programs.
- Forming a data curation team to promote more collaborations between earth scientists and university libraries is important to develop best practices in earth science data curation.

Challenges

- Metadata issue has become increasingly critial in data curation.
- The practices and theories are still evolving in this field; not much resources are available for us to assist faculty to share their data in a cost effective manner.
- How to achieve the sustainability of data infrastructure and data services.
- How to resolve conflicts and determine priorities when library resources encounter limitation.

Future Plans

General earth science data. Lehigh and Purdue will continue the collaboration on Earth Science and Space Informatics Librarianship and together develop best practices and workflows to assist researchers in research data management. While further expanding our outreach and strengthening our connections with more earth scientists and data communities, we will also emphasize the education in data management in the future and aim to create more training opportunities for young scholars and students to raise their awareness of data management at the early stage of their research career. We may seek partnerships with the ESSI or ESIP community on this matter.

Specifications for geospatial data. As the most unique and also the most challenging component of earth science data, **sustainable** geospatial data curation will continue to be the focus of Purdue libraries. In the next step, Purdue geospatial data program will collaborate with the Globe Program and continue making efforts toward developing the metadata support for campus researchers with such need.

Acknowledgements

Work on this project was done under the auspices of the Council on Library and Information Resources (CLIR) and with active participation of the Library and Technology Services of Lehigh University and Purdue University Libraries. We would like to thank (Lehigh University) Brian Simboli, Science Librarian; Christine Roysdon, Director for Library Collections and Systems; and Department of Earth and Environmental Sciences; (Purdue University) Nicole Kong, GIS Specialist; Kristin Wegner, Project manger, The Globe Program; Steven Smith, Earth, Atmospheric, and Planetary Science Outreach Coordinator-**A GLOBE Program partner**; Darrell Schulze, Professor of Soil Science at the Agronomy Department.