12-1-2012

A Multi-Institutional Project to Develop Discipline-Specific Data Literacy Instruction for Graduate Students

Sarah Wright
Cornell University, sjw256@cornell.edu

Michael Fosmire
Purdue University, fosmire@purdue.edu

Jon Jeffryes
University of Minnesota - Twin Cities, jeffryes@umn.edu

Marianne Stowell Bracke
Purdue University, mbracke@purdue.edu

Brian Westra
University of Oregon, bwestra@uoregon.edu

Follow this and additional works at: http://docs.lib.purdue.edu/lib_fspres
 Part of the Curriculum and Instruction Commons, Educational Assessment, Evaluation, and Research Commons, and the Other Education Commons

A Multi-Institutional Project to Develop Discipline-Specific Data Literacy Instruction for Graduate Students

Sarah Wright, Michael Fosmire, Jon Jeffryes*,
Marianne Stowell Bracke*, Brian Westra
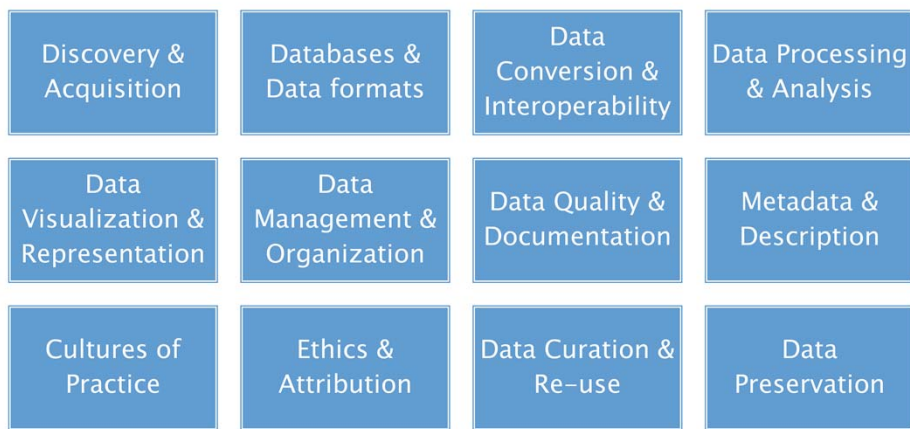
Asterisks = not here

There are five project teams: two at Purdue, which is the lead institution and one at each of the other institutions. These teams are composed of a data librarian, a subject librarian, and a faculty researcher from a science or engineering discipline. This is an IMLS-funded project to develop discipline-specific instruction for graduate students, so each team is focused on a particular STEM discipline – you can see the range of disciplines represented.

Today, I'm going to cover who we talked to, what we learned, and what kind of instruction we ended up developing to address each group's needs.

Interviewed Faculty and Graduate students in research groups
- o   highly Modified Data Curation Profile (available at our website, http://datainfolit.org)
- o   Faculty's assessment of their student needs
- o   Students' understanding of data curation issues and policies

Data Literacy – what is it?

| Discovery & Acquisition | Databases & Data formats | Data Conversion & Interoperability | Data Processing & Analysis |
| Data Visualization & Representation | Data Management & Organization | Data Quality & Documentation | Metadata & Description |
| Cultures of Practice | Ethics & Attribution | Data Curation & Re-use | Data Preservation |

Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. *Portal: Libraries & the Academy, 11*(2), 629–657.

But let's back up a bit, what is "data literacy"? I don't know how familiar you are with IL, but a key difference, and a central tenant of DIL is the recognition of researchers as producers of data, as well as data consumers.

This slide shows the major themes that make up data literacy, and are the major themes we discussed in our interviews. These are major themes - Unique technologies, techniques, and other content fall under each theme due to disciplinary differences. We also expect some themes to emerge as more important to some disciplines than others.

These themes are from research done at Purdue University – reference paper.
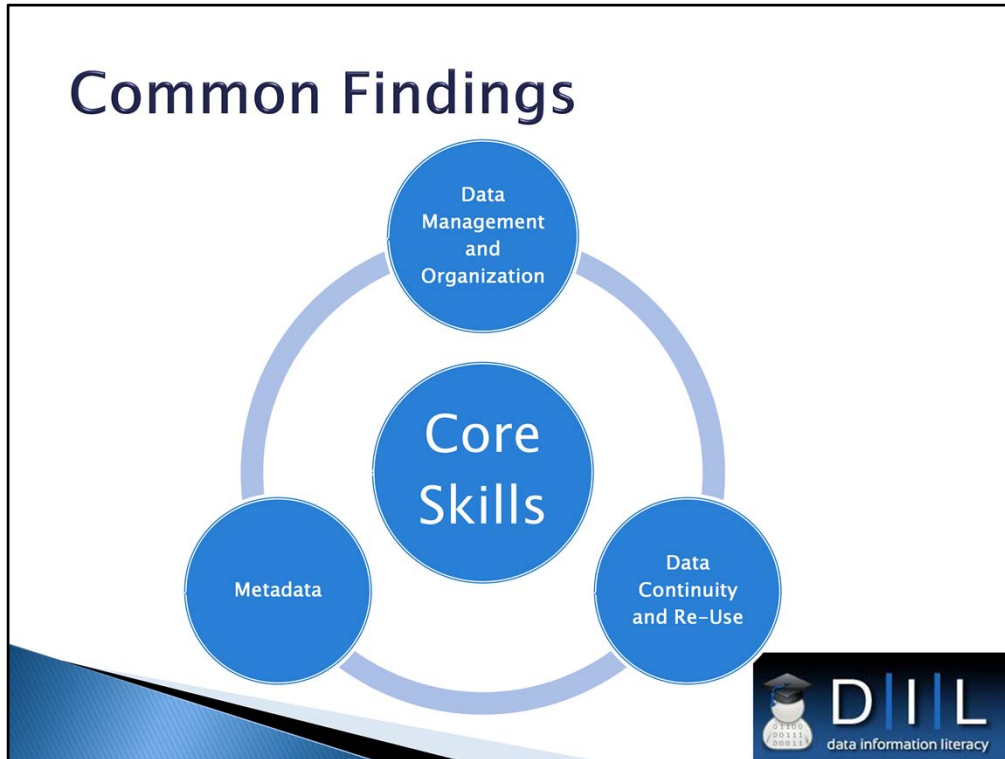
So what emerged from our interviews with faculty and graduate students? Although the academic disciplines being studied by the teams vary considerably, many similarities did emerge. These are a few core skills that we all found, and tried to address in the course of our informal interactions and formal instruction .
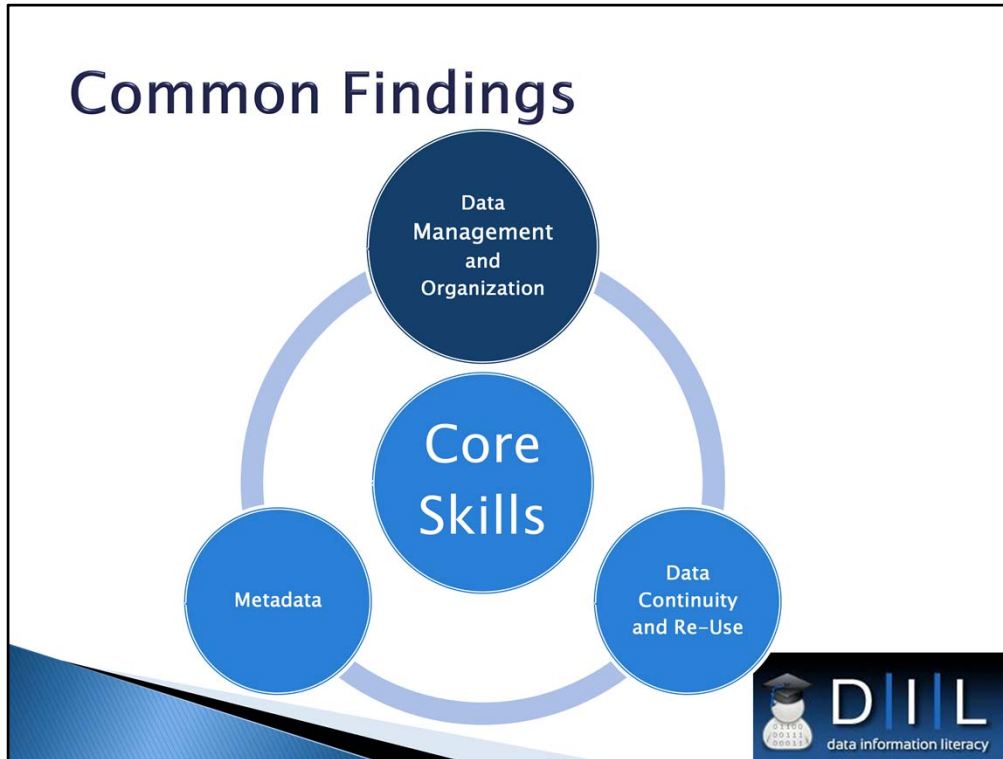
General Lack of Formal Training in Data Management/Curation: In many instances, graduate students reported a noticeable lack of training in data management and curation skills as they relate to their research practices. Although students from different institutions do report receiving basic training (usually in ethics), most of their education concerning the maintenance and upkeep of data has come either through experience in the laboratory or through word-of-mouth.# "You do it wrong, lose the data, and learn how to do it better." That's the way I learned it in my lab, too.

As data becomes increasingly digital, we have to do better. Many of you know this, and have been teaching data managemetn for a long time, but I think you'll agree that the job isn't done and we need it to spread to other disciplines.
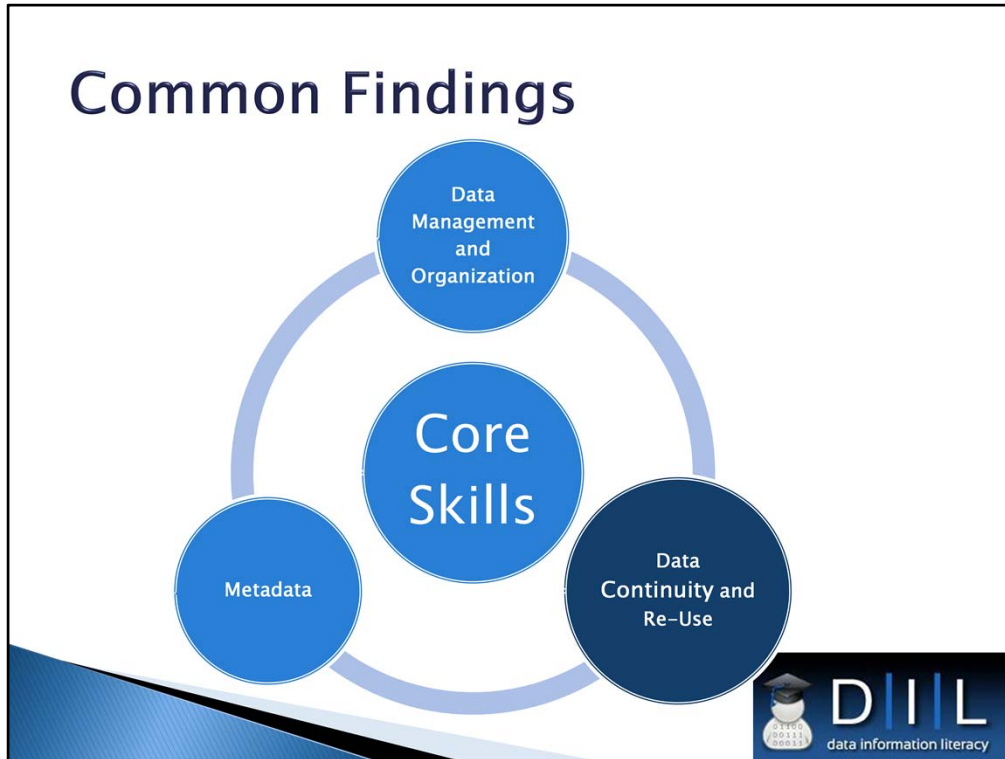
Data Continuity and Re-use: the original terminology was "data curation" but we felt that it was a little too library-specific, and doesn't really convey what is of primary concern to faculty and graduate students. What we mean by data continuity is the skills and ability to package the data so that it can continue to be used after the graduate student departs - the data remains accessible for the PI or others as grad students cycle in and out of the lab. In nearly all of our conversations, it was clear that internal re-use of data is higher priority for the research group than external re-use. So that's where we start.
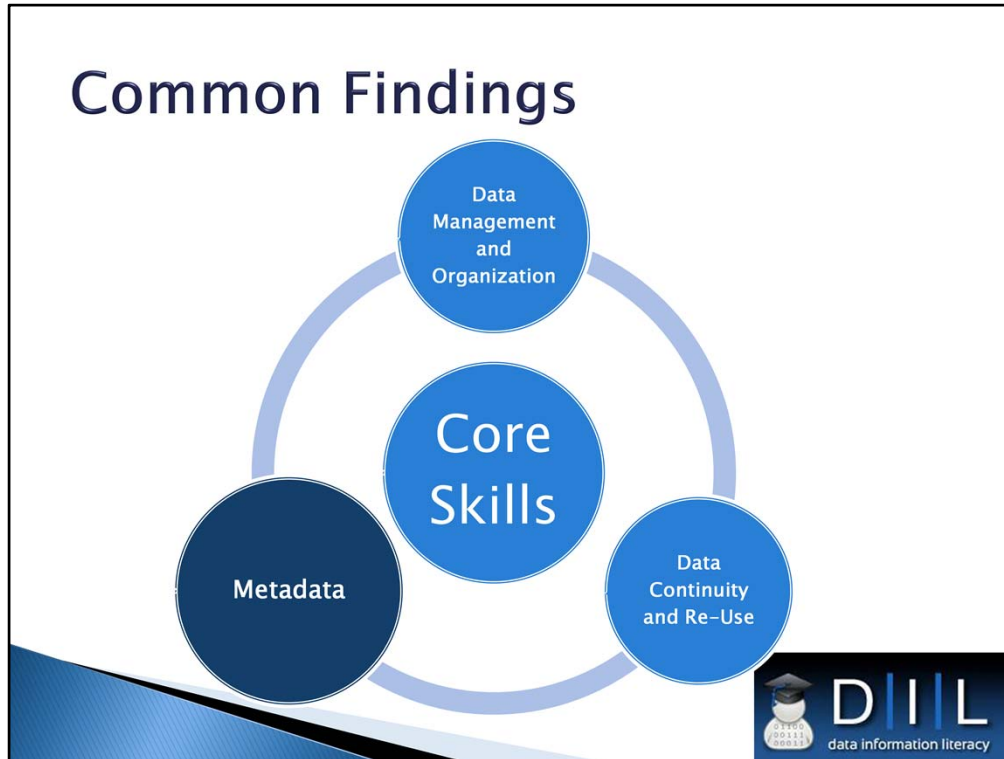
Metadata and Data Description: Despite the growing importance of metadata, two of the research groups interviewed experienced difficulty grasping the concept. I've taken to calling it "data documentation" when promoting workshops. I'm not saying that nobody understands metadata - in other interviews, metadata ranked high in importance. We did hear from several faculty that they felt they didn't have the knowledge or expertise to teach students about metadata. Many if not all of our interviewees provide some sort of metadata to the data they work with, but most don't apply it in an intentional or formal manner. So we need to make it intentional and formalize it in some way. I started simple, and gave a workshop that focused on how to write a readme file.

Unique Needs Addressed

Cornell	Minnesota	Oregon	Purdue	Purdue
Natural Resources	Civil Engineering	Ecology	Agricultural and Biological Engineering	Electrical and Computer Engineering
✓ Cultures of practice (sharing)	✓ Access and Ownership ✓ Preservation	✓ Grant close–out products	✓ SOP's ✓ Metadata for sharing	✓ Documenting code ✓ Transfer of code

D|I|L data information literacy

The groups also addressed some unique needs for their grad students.

Cornell – The faculty member I'm working with felt that one of my major contributions should be in the area of recommendations and best practices for sharing data. So I'm teaching students about options for data sharing, including the institutional repository at Cornell, subject repositories, and how to evaluate those options and to choose.

University of Minnesota – In addition to the core skills identified in the previous slides (data management, data re-use, and metadata), Minnesota is also addressing Access and ownership and preservation.

University of Oregon – the research team wrapping up a 4-year grant. This resulted in some student turnover before training, and impacted the timing and utility of the training, but it also provided opportunities for lessons learned for everyone that could translate to future projects.

Purdue (ABE) – the Purdue team's highest priority needs were the core skills; instruction was also focused on developing Standard Operating Procedures and documentation of their data, understanding external and developing internal metadata, and depositing data into a repository.

Purdue (ECE) – the research center the Purdue ECE team has partnered with has graduate students acting as T.A.s teaching undergraduate students skills in developing code for specific projects. Project span multiple semesters (and involve multiple students) so sufficient documentation of the code is needed to facilitate its transfer and use by other students, project clients and the center's administration.

Instructional Approaches

Cornell	Minnesota	Oregon	Purdue	Purdue
Natural Resources	Civil Engineering	Ecology	Agricultural and Biological Engineering	Electrical and Computer Engineering
Fall 2012: Library Workshops

Spring 2013: NTRES Special Topics (1 Credit) | Online Data Management Course

7 modules

http://z.umn.edu/datamgmt | 3 Readings

Informal Session with Research Team | 3 Workshops | Embedded into project teams

Skill Sessions |

Our instructional approaches were as varied as the teams.

Cornell – I was working with a faculty member who had no current graduate students, so we decided to create a course, but field work in the fall made scheduling a class difficult, so we shifted it to the spring. Library workshops in the fall were attended by approximately 48 graduate students. Most attended the first session, a general introduction to data management and NSF requirements, fewer attended an introduction to relational databases and a workshop on data documentation. In the spring, we're offering a 1 credit course for NTRES grad students, but we don't know enrollment yet.

Minnesota – The UofM team put together a beautiful online data management course, which is available on the web. When the students have completed the 7-module course, they will have a complete data managemetn plan.

Oregon – The Oregon librarians asked the research team to read 3 articles before they presented a session at one of the regular team meetings. The readings covered lab notebook guidelines; the case for and impact of data sharing on global change research; and an Ecology journal article that outlined simple guidelines for data management in ecology. The instruction session was discussion-based and covered basic principles and local resources for data management: file management and formats/transformations; backup and storage; data repositories & sharing; metadata; and data citation.

Purdue (ABE) – Purdue team held three workshops addressing the core skills – dm, metadata, and data continuity. The one-hour workshops included pre-work and homework to apply skills. *The students developed standard operating procedures for data handling in their research group, searched for data in a repository, and analyzed the data they found to inform how they will apply metadata to their own data when they publish it. – delete??*

Purdue (ECE) – The ECE Team embedded themselves into project teams, attending several team meetings and their design reviews. Working with the Graduate T.A.'s they developed a rubric to evaluate the work of undergraduates in documenting their code and project work. They also offered skill sessions to address expectations laid out in the rubric.
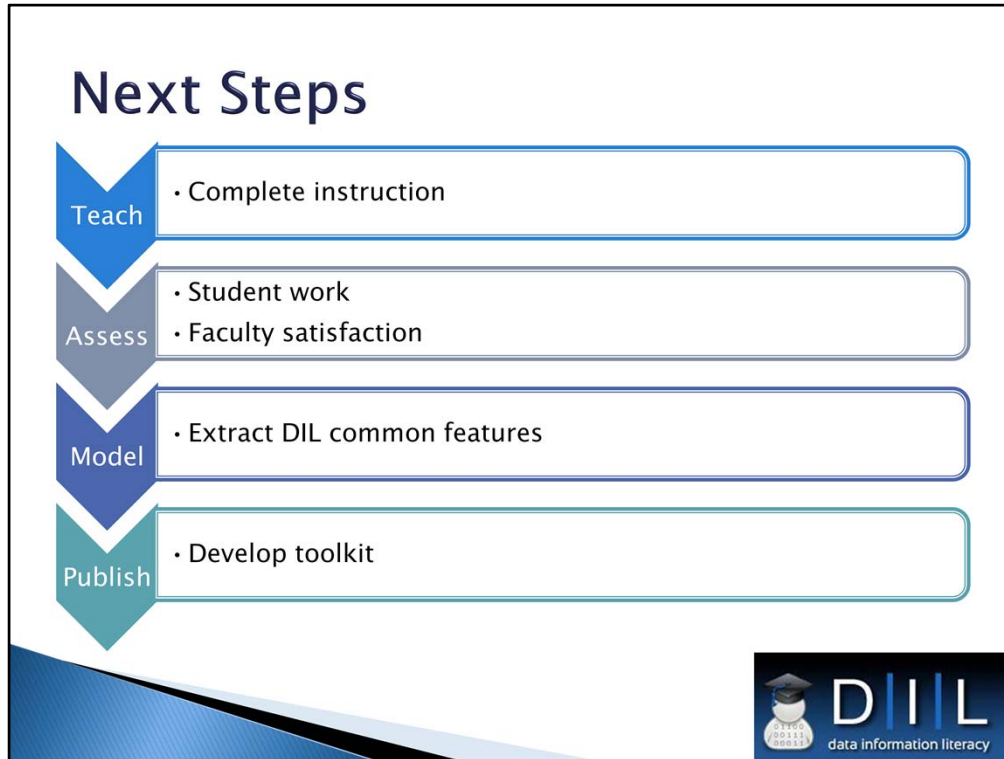
- We still have teams who are wrapping up instruction through February.
- Student achievements as well as student and faculty attitudes will be assessed to determine the relevancy and effectiveness of the instruction.
- Project teams will then conduct a collective analysis of the educational interventions to identify patterns and commonalities across their respective experiences in developing DIL programs, as well as account for the significant differences. The outcomes will be used to draft a model for other academic librarians to develop data information literacy programs of their own.
- The final model will be published after holding a symposium and gathering feedback on the draft toolkit.

DIL Project Personnel

Principal Investigator:
- Jake Carlson – Purdue University

Co-Principal Investigators:
- Camille Andrews – Cornell University
- Marianne Stowell Bracke – Purdue University
- Michael Fosmire – Purdue University
- Jon Jeffryes – University of Minnesota
- Lisa Johnston – University of Minnesota
- Megan Sapp Nelson – Purdue University
- Dean Walton – University of Oregon
- Brian Westra – University of Oregon
- Sarah Wright – Cornell University

Learn More

 http://www.datainformationliteracy.org

 @datainfolit

 https://www.facebook.com/datainfolit