Apr 6th, 12:00 AM

A Sloth of Gummi Bears: Evaluating effectiveness of research data management instruction

Rebecca Reznik-Zellen
University of Massachusetts Medical School

Lisa A. Palmer
University of Massachusetts Medical School

A Sloth of Gummi Bears: Evaluating Effectiveness of Research Data Management Instruction

Rebecca Reznik-Zellen and Lisa Palmer — Lamar Soutter Library, University of Massachusetts Medical School

FCE 3017: Research Data Management Fundamentals

Description: This course is a short (one week), for-credit elective for third-year medical students at the University of Massachusetts Medical School developed and taught by librarians. The course provides an overview of the basic principles and best practices for data management, with a focus on data lifecycle planning, security and ethics, organization and documentation, and data sharing. Course structure and content includes:

- Forty hours of instruction delivered in five three-hour, face-to-face sessions, with twenty-five hours of additional readings and assignments;
- Lecture-based curriculum featuring guest speakers, in-class activities, daily readings, and a research project;
- Active use of existing UMMS tools such as LabArchives electronic lab notebook and the DMPTool.

Students: Six third-year medical students have participated in this course over four sessions. Student performance is evaluated on participation, effort, and completion of research project deliverables. Students also submit an evaluation of the course:

- "It was such an eye-opener realizing the wealth of resources that are available to researchers through and thanks to the library. This was a great FCE with great people running it."
- "Great FCE! I would recommend this FCE to anyone planning on doing research and using metadata. Even as someone with a solid research background I learned a significant amount of new things concerning searching for data, long-term data storage, and IT security."

Lessons Learned: Based on experience and evaluation results (right column), future sessions could include:

- Pre- and post-course assessments to measure outcomes more easily;
- Additional active-learning methods, particularly for content areas where students consistently score low on the data management plans;
- A different or modified research project component to reflect more realistic research scenarios.

The Gummi Bear Population Study

To receive credit for participating in this course, students are required to conduct a fictional research project, the Gummi Bear Population Study.[1] Project deliverables include:

1. A poster showing preliminary findings for the Gummi Bear Population Study.
2. Primary research data, including a readme.txt or data dictionary documenting your dataset.
3. An NSF-Generic data management plan for a grant proposal to fund the Gummi Bear Population Study.

Using LabArchives to collect, store, and organize their data and the DMPTool to create their data management plans, the students have a hands-on opportunity to apply the concepts learned in the course.

Although brief and with a heavy tongue-in-cheek component to them, the deliverables demonstrate the extent to which the basic concepts of research data management have been understood.

[1]This approach was inspired by Vasilevsky, Nicole; Wirz, Jackie; Champieux, Robin; Hannon, Todd; Laraway, Bryan; Banerjee, Kyle; Shaffer, Chris; and Haendel, Melissa, "Lions, Tigers, and Gummi Bears: Springing Towards Effective Engagement with Research Data Management" (2014). Scholar Archive. Paper 3571.

Effectiveness of Research Data Management Instruction

Data Management Plan (DMP) Evaluation: The NSF-Generic DMP reflects the major topics covered throughout the course and is therefore a good proxy for measuring effectiveness of instruction. Two instructors rated student-created DMPs on a 5-point scale, from insufficient (1) to comprehensive (5). Scores were averaged for each student (see table).

Average DMP Score by Student
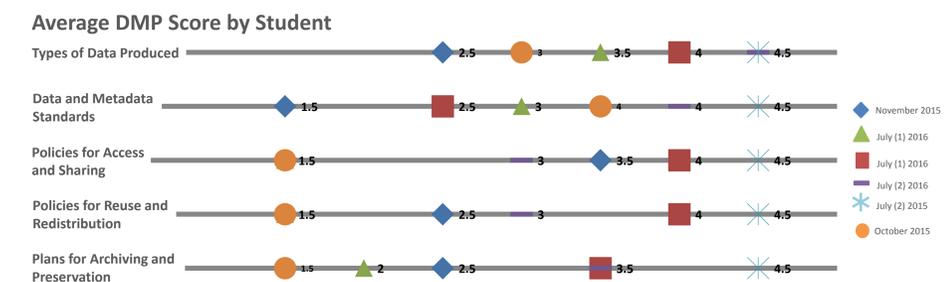
The combined average scores for all students per DMP category range between 3.7 (Types of Data Produced) and 2.8 (Policies for Reuse and Redistribution). The combined average DMP score for all categories is 3.1. Observations based on these evaluations include:

- Scores varied by session but generally improved between November 2015 and July (2), with a decrease in October. Both November and October sessions had one student only; the July session had two students each. Having more than one student per session may positively effect teaching and learning.
- Scores are consistently highest for Types of Data Produced and lowest for Policies for Reuse and Redistribution. Comfort with describing data is supported by detail in student posters, whereas inadequate description of data reuse and redistribution policies may indicate that course content or instruction methods need to be restructured.
- Brevity of details in some areas of the DMP may be due to the fictional nature of the research project.

Limitations: The course is very intense, covering a lot of content in a short period of time, which may be difficult to absorb. Other considerations:

- Course content was updated and scheduling modifications were made between each session, which may contribute to discrepancies among student scores across sessions.
- While the NSF-Generic DMP reflects the content of the course, it is not an exact match to the syllabus, which may explain some gaps.