Lions, Tigers, and Gummi Bears: Springing Towards Effective Engagement with Research Data Management

Nicole Vasilevsky[1], Jackie Wirz[1], Robin Champieux[1], Todd Hannon[2], Bryan Laraway[1], Kyle Banerjee[1], Chris Shaffer[1], Melissa Haendel[1]

[1] Oregon Health and Science University Library, 3181 SW Sam Jackson Park Rd, Portland, OR 97217, USA
[2] Center for Health Research, Kaiser Permanente, 3800 N Interstate, Portland, OR 97227, USA

Abstract

As part of an award from the Future of Research Communications and e-Scholarship (Force11) community organization, the Oregon Health & Science University Library implemented an outreach effort with the local research community. The three-pronged approach consisted of a data management open house, one-on-one consultations with researchers, and an interactive data wrangling session. Here we describe our efforts to understand our community's research data management needs and how the outcomes are influencing the Library's efforts to provide research data management related services.

Introduction

Research data management is the process of organizing, describing, storing, and sharing data. From planning the details of data collection to addressing long-term data plans, research data management (RDM) can affect reuse and reproducibility. As the expense of scientific research increases and the practice of team-science grows, it is essential that researchers share data across disciplines in an efficient manner. Towards these ends, many libraries have developed RDM services to assist their research communities. This report describes the approaches taken by the Oregon Health & Science University (OHSU) Library to identify strategies to support local RDM, which was initiated through '$1k challenge' funding by the Force11 consortium (www.force11.org).

Force11 is a grassroots organization that aims to improve scholarly communication through effective use of information technology and social change. RDM is an area of focus in Force11, as it specifically relates to changing modes of scholarly communication and research impact. At the 2013 Beyond-the-PDF2 conference, Force11 asked its community a simple question: what would you do with $1,000 to improve research communication? Melissa Haendel, Assistant Professor at the OHSU Library, proposed a project to investigate our researchers' data management needs and challenges. Her proposal, "Starting at Ground Zero", was designed to promote interaction between information specialists and research scientists to inform new strategies and goals for discipline specific libraries in this emerging area.

OHSU is not alone in recognizing the need to support a sophisticated research data management environment. For example, the National Institute of Health is addressing the need for improved data management through a recent request for proposals aimed to develop "training modules designed to enhance data reproducibility targeted to graduate students, postdoctoral fellows, and beginning investigators". This is in response to recent studies that demonstrated issues with reproducing previously published research (Begley and Ellis, 2012; Mobley et al 2013). Other efforts are underway to identify skills (Carlson et al 2013, http://wiki.lib.purdue.edu/display/ste/Home), develop new curriculums and resources (http://library.umassmed.edu/necdmc/index, coursera.org, http://software-carpentry.org/blog/2013/06/mozilla-science-lab-announcement.html), and provide assistance with data curation (Minor et al 2014). The OHSU Library is working towards increasing awareness of these issues as well as developing potential solutions.

Starting at Ground Zero at OHSU

In order to work closely and effectively with the research community to identify data needs and inform the development of library services, it was important to identify Library staff skills and how to best utilize these to help patrons. Our goals for the project included engaging Library staff and faculty through demonstrating the practical relationships between library and information science expertise and RDM. We organized two presentations that connected research data management to our colleagues' professional experiences and expertise. For example, relating researcher activities to classic information organization and description helped engage our colleagues.

For researchers, we hosted a Data Management Open House where we presented on various aspects of RDM. This included the role data management plays in the research cycle, reuse and reproducibility, research impact, scholarly attribution, knowledge discovery, and helping researchers understand how their "scholarly footprint," one's visible scholarly presence on the Web, can affect others' perspectives on their activities and productivity. Following the session, attendees were offered a $50 gift card to participate in consultations with Library staff. We believed the consultations would benefit researchers by giving them guidance related to specific data management questions, and benefit the Library by informing our understanding of local knowledge, needs, and challenges. Nearly 30 students, staff, and faculty from various departments attended the Data Management Open House and 11 people indicated interest in a consultation.

During private consultation, we aimed to: (1) Determine which aspects of the data required standardized metadata for sufficient reporting and reproducibility; (2) Determine if the data would be relevant for a public repository, and help researchers deposit their data; and (3) discuss RDM best practices across the data lifecycle, such as version control. For each consultation, we included at least one librarian or staff member with basic science or clinical research experience as we thought that significant subject expertise would be necessary to handle advanced questions. During the consultations, most researchers demonstrated a need for basic hands-on training and we discussed their challenges in finding, accessing, describing, and storing data. We followed up each consultation with notes and references as appropriate.

Engaging the OHSU research community was challenging. Despite positive responses during the open house event and extensive follow up by Library staff, few consultations were scheduled. However, all of the consultations pointed towards a lack of basic RDM knowledge. Student engagement was an early goal, but few students attended the open house. We concluded that we needed to target students differently, especially given that they are beginning RDM skill development for their areas of study.

We partnered with our Graduate Student Organization (GSO) to focus our engagement efforts on students and early career researchers, and collaborated on a second data management event. We worked with the student leaders to outline an approach and agenda that spoke specifically to students' goals and challenges. In contrast to the lecture-discussion based open house, here we used a hands-on case study to demonstrate best practices. We advertised the session as a "data wrangling workshop," and the GSO designed flyers that playfully targeted students' frustrations and goals using statements like: "Your lab notebook is bad, and you should feel bad... no, not really, but come to a data wrangling session" and "Would you like your hard-earned data to be: organized, effortlessly accessible, reproducible, a help instead of a hurdle towards graduating?"

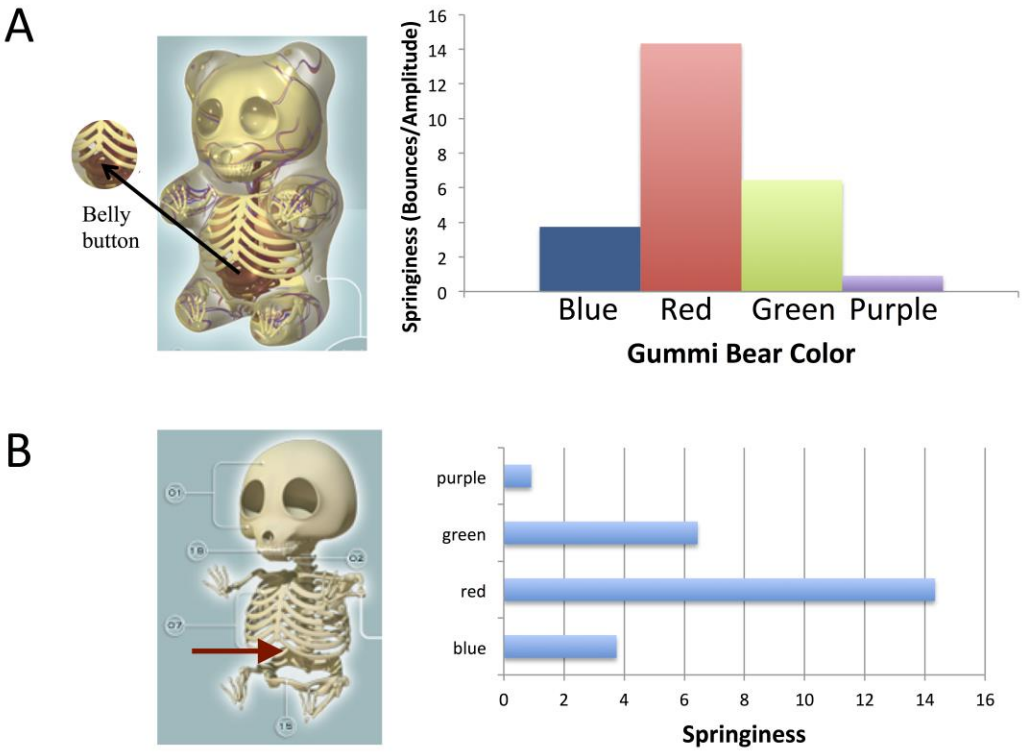
How gummi bears helped students appreciate research data management

OHSU's graduate student body includes researchers in-training across basic and clinical science

disciplines. The data they produce and manage encompasses a broad spectrum of types and techniques. While the Data Management Open House and consultations told us that most OHSU researchers share a need for essential data management education, we also learned that it was more effective to deliver this information through examples that were data and domain specific. For example, some research projects are highly computational in nature and a majority of the data is born digital. Other projects may include a variety of physical objects that must be handled and recorded in different ways. It was an interesting challenge, therefore, to design an impactful hands-on RDM activity for the diverse audience we expected at the data wrangling session.

What we needed was an exercise that was data and domain neutral as well as engaging and fun. We designed a case study that explored RDM challenges and best practices using a Gummi Bear Anatomy Model. The exercise explored RDM challenges and best practices by presenting participants with a variety of source data sets. Participants were then instructed to perform rudimentary data analysis, create a figure, and develop a methods section that captured the "springiness" of a Gummi Bear. While admittedly and intentionally whimsical, the process covered many real data activities, such as manipulation, analysis, visualization, reporting and sharing.

The data wrangling session was very well received. Attendees, which included students across our schools and institutes, worked through the Gummi Bear exercise in small groups. At the end of the session, each group shared their figure and description of their methods. Despite starting with identical 'gummi bear springiness' data, there were significant differences between the methods and results the groups produced (Figure 1). We used these differences to demonstrate and highlight how data management affects understanding, reuse, and reproducibility. At the end of the session, over half of the attendees expressed interest in the use of library services in the future, such as follow-up consultations or a lab presentation. The OHSU Library is now allocating effort to further refine educational strategies around RDM following these activities.



Hypothetical example figures from Data Wrangling Session. Participants at the Data Wrangling Info Session were asked to prepare a figure for a hypothetical publication. Starting with the image of the anatomy of a Gummi Bear, the four groups were instructed to crop the image and annotate the belly button. Additionally, they were given data quantitating the "springiness of a Gummi Bear" and asked to create a graph comparing the data between groups. Figure A and B provides examples of the types of variation obtained from the different groups, even when given the same starting data and instructions. Permission to reuse this image was granted by Jason Freeny (http://laughingsquid.com/the-anatomy-of-a-gummy-bear-by-jason-freeny/).

Discussion

The "Starting at Ground Zero" project taught us the importance of approach and delivery to support engagement of different participants within the research community. The content of the information we communicated at the Data Management Open House was directed towards senior investigators, program leaders, and administrators in understanding how RDM and their scholarly footprint relates to research impact and evaluation. The one-on-one consultations leveraged a variety of Library staff perspectives to help address the individual needs of researchers. We found that this team approach often lead to very insightful conversations and solutions. The interactive hands-on workshop was framed by direct involvement of student leaders, who had a good sense for student needs and what would resonate most with them. The activity itself allowed the students to experience first-hand how core concepts relate to their own RDM activities and develop goals for their research.

There is not a clearly defined path as to how data management training should be approached or by whom. One of our key conclusions was that data management training could be performed as a team approach, especially in a small domain specific library where there isn't a dedicated staff member in this role. Others have drawn similar conclusions, for example, in a recent talk titled "It takes a village: Strengthening data management through collaboration with diverse institutional offices" at the 2014 IASSIT meeting. Here the presenters discussed how proper data planning and management requires the participation of several institutional members, from grants coordinators, to IT administrators to librarians (http://datapub.cdlib.org/2014/06/16/it-takes-a-data-management-village/).

The OHSU Library is now defining research data management services and outreach models to our

research community based upon the outcomes of the 1K Challenge and other activities. We have a better understanding of the RDM challenges that impact OHSU researchers and we identified a few principles of engagement and interactivity that greatly enhanced our RDM outreach efforts. As a smaller, domain specific library, the OHSU Library does not have staff devoted to providing services for RDM. Use of a team-based effort that combines staff members' different backgrounds and expertise was very effective in assisting researchers, as well as supporting a learning environment within the Library. We also recognized that having deep domain expertise for any given RDM issue was not necessary, as the data management solutions were broad and could be applied across various domains. By taking a team approach, we were able to facilitate informative interactions with researchers that leveraged different areas of staff expertise. Also, the team approach necessarily leverages expertise from existing collaborations surrounding RDM topics outside the library, such as participation in the NIH BD2K program (http://bd2k.nih.gov/), the Association for Research Libraries Data Sharing group, and the Resource Identification Initiative (www.force11/org/resource_identification_initiative). Finally, the library is now actively engaged in teaching new courses and participating in graduate programs to further assist the educational mission of OHSU.

Acknowledgements
We would like to acknowledge Jason Freeny (http://laughingsquid.com/the-anatomy-of-a-gummy-bear-by-jason-freeny/) for allowing us to reuse his image, Kelly Chacón for her assistance with planning the Data Wrangling session and Matthew Brush, Maija Anderson, and Mike Grove for participation in the data management consultations.

References

Begley CG and Ellis LM (2012) Drug development: Raise standards for preclinical cancer research. Nature 483, 531–533

Carlson et al 2013 Developing an Approach for Data Management Education: A Report from the Data Information Literacy Project

Minor et al (2014) Research Data Curation Pilots: Lessons Learned. International Journal of Digital Curation

Aaron Mobley, Suzanne K. Linder, Russell Braeuer, Lee M. Ellis, and Leonard Zwelling. (2013) A Survey on Data Reproducibility in Cancer Research Provides Insights into Our Limited Ability to Translate Findings from the Laboratory to the Clinic. PLoS One. 2013; 8(5): e63221.