Journal of eScience Librarianship

9-8-2017

Using Peer Review to Support Development of Community Resources for Research Data Management

Heather Soyka
Kent State University

Amber Budden
DataONE/University of New Mexico

Viv Hutchison
US Geological Survey

See next page for additional authors

Corresponding Author(s)
Heather Soyka, Kent State University, P.O. Box 5190, Kent, OH 44242, hsoyka@kent.edu

Follow this and additional works at: http://escholarship.umassmed.edu/jeslib

🌀 Part of the Information Literacy Commons, Scholarly Communication Commons, and the Scholarly Publishing Commons

Using Peer Review to Support Development of Community Resources for Research Data Management

Authors

Heather Soyka, Amber Budden, Viv Hutchison, David Bloom, Jonah Duckles, Amy Hodge, Matthew S. Mayernik, Timothée Poisot, Shannon Rauch, Gail Steinhart, Leah Wasser, Amanda L. Whitmire, and Stephanie Wright

Keywords

research data management, GitHub, peer review, academic libraries

Article Correction History

The HTML and PDF versions of this article were corrected on September 11, 2017 to change the Creative Commons license from CC BY-NC-SA to CC 1.0 Public Domain Dedication at the authors' request.

Journal of eScience Librarianship
putting the pieces together: theory and practice

Full-Length Paper

Using Peer Review to Support Development of Community Resources for Research Data Management

Heather Soyka[1], Amber Budden[2], Viv Hutchison[3], David Bloom[4], Jonah Duckles[5], Amy Hodge[6], Matthew S. Mayernik[7], Timothée Poisot[8], Shannon Rauch[9], Gail Steinhart[10], Leah Wasser[11], Amanda L. Whitmire[6], Stephanie Wright[12]

[1] Kent State University, Kent, OH, USA
[2] DataONE/University of New Mexico, Albuquerque, NM, USA
[3] US Geological Survey, Reston, VA , USA
[4] University of California, Berkeley, CA, USA
[5] Software Carpentry, Austin, TX , USA
[6] Stanford University, Stanford, CA , USA
[7] National Center for Atmospheric Research/University Corporation for Atmospheric Research, Boulder, CO, USA
[8] Université de Montréal, Montréal, Québec, Canada
[9] Woods Hole Oceanographic Institute, Woods Hole, MA, USA
[10] Cornell University, Ithaca, NY, USA
[11] University of Colorado, Denver, CO, USA
[12] Mozilla Foundation, Mountain View, CA, USA

Abstract

Objective: To ensure that resources designed to teach skills and best practices for scientific research data sharing and management are useful, the maintainers of those materials need to evaluate and update them to ensure their accuracy, currency, and quality. This paper advances the use and process of outside peer review for community resources in addressing ongoing accuracy, quality, and currency issues. It further describes the next step of moving the updated materials to an online collaborative community platform for future iterative review in order to build upon mechanisms for open science, ongoing iteration, participation, and transparent community engagement.

Abstract Continued

Setting: Research data management resources were developed in support of the DataONE (Data Observation Network for Earth) project, which has deployed a sustainable, long-term network to ensure the preservation and access to multi-scale, multi-discipline, and multi-national environmental and biological science data (Michener et al. 2012). Created by members of the Community Engagement and Education (CEE) Working Group in 2011-2012, the freely available Educational Modules included three complementary components (slides, handouts, and exercises) that were designed to be adaptable for use in classrooms as well as for research data management training.

Methods: Because the modules were initially created and launched in 2011-2012, the current members of the (renamed) Community Engagement and Outreach (CEO) Working Group were concerned that the materials could be and / or quickly become outdated and should be reviewed for accuracy, currency, and quality. In November 2015, the Working Group developed an evaluation rubric for use by outside reviewers. Review criteria were developed based on surveys and usage scenarios from previous DataONE projects. Peer reviewers were selected from the DataONE community network for their expertise in the areas covered by one of the 11 educational modules. Reviewers were contacted in March 2016, and were asked to volunteer to complete their evaluations online within one month of the request, by using a customized Google form.

Results: For the 11 modules, 22 completed reviews were received by April 2016 from outside experts. Comments on all three components of each module (slides, handouts, and exercises) were compiled and evaluated by the postdoctoral fellow attached to the CEO Working Group. These reviews contributed to the full evaluation and revision by members of the Working Group of all educational modules in September 2016. This review process, as well as the potential lack of funding for ongoing maintenance by Working Group members or paid staff, provoked the group to transform the modules to a more stable, non-proprietary format, and move them to an online open repository hosting platform, GitHub. These decisions were made to foster sustainability, community engagement, version control, and transparency.

Conclusion: Outside peer review of the modules by experts in the field was beneficial for highlighting areas of weakness or overlap in the education modules. The modules were initially created in 2011-2012 by an earlier iteration of the Working Group, and updates were needed due to the constant evolving practices in the field. Because the review process was lengthy (approximately one year) comparative to the rate of innovations in data management practices, the Working Group discussed other options that would allow community members to make updates available more quickly. The intent of migrating the modules to an online collaborative platform (GitHub) is to allow for iterative updates and ongoing outside review, and to provide further transparency about accuracy, currency, and quality in the spirit of open science and collaboration. Documentation about this project may be useful for others trying to develop and maintain educational resources for engagement and outreach, particularly in communities and spaces where information changes quickly, and open platforms are already in common use.

Introduction

Research is increasingly collaborative and data intensive, situated in a rapidly changing information environment. Gaining, maintaining, and sharing skills for research data management and data curation is critical (Tenopir, Birch and Allard 2012). Ongoing development of resources and educational modules for research data management is essential for allowing members of the research community to learn new skills.

Advancing conversations about connecting shared resource maintenance and community engagement is of interest to librarians who are often working at the intersection of those two areas. In particular, data librarians are members of a larger community of practice that is engaged with the teaching and learning of research data management (Lyon 2013; Carlson 2015). Librarians, supported by robust information infrastructure, are located at the interface between researchers and resources. They are able to interpret and provide informational instruction on research data creation, sharing, and management for local faculty and student needs, develop programming, assist with creating data management plans, and take part in instructional and research design (Briney, Goben, and Zilinski 2017; Wright et al. 2012). Without keeping these skills relevant and up-to-date in a changing information environment, there is the possibility that useful research data may not be created, shared, or reused. In this paper, we describe the use of expert peer review for community resources and the decision to move to an open git repository for sharing and version control in order to maintain ongoing accuracy, participation, and use of shared research data management educational resources.

Setting

The work that is described in this paper was completed as part of the ongoing work of the 'DataONE' (Data Observation Network for Earth) project. Established in 2009 with funding from the National Science Foundation (NSF) under the DataNet program, DataONE is now (as of 2017) in the second phase of development. DataONE has developed a sustainable, long-term network to ensure preservation of and access to multi-scale, multi-discipline, and multi-national environmental and biological science data.

One of four current Working Groups, the Community Engagement and Outreach Working Group builds effective and creative strategies for connecting with members and stakeholders of the community. One important part of this outreach is building further capacity within the community for sharing and managing research data. Providing a robust range of tools, resources, and strategies for research data management education that are current and accurate is central to the needs of the broader DataONE community.

In support of the DataONE project, a collection of research data management resources have been developed. Created in 2011-2012 by members of the CEO Working Group, one central component is the freely available Education Modules that include three complementary pieces: slides, handouts, and exercises. These were designed for research data management training and to be adaptable for use in classroom teaching.

The Education Modules were developed with the intent of covering all phases of the DataONE Research Data Life Cycle (https://www.dataone.org/data-life- cycle). Modules cover topics in areas such as data sharing, data quality, metadata, data management planning, citation, data protection, workflows, and legal issues. The modules are located on a static page in the

education section of the DataONE website (https://www.dataone.org/education-modules), where each component may be downloaded. A full list of the modules can be found in Table 1.

Table 1: List of education modules

1	Why Data Management
2	Data Sharing
3	Data Management Planning
4	Data Entry and Manipulation
5	Data Quality Control and Assurance
6	Protecting Your Data
7	Metadata
8	Data Citation
9	Analysis and Workflows
10	Legal and Policy Issues

Methods: The Peer Review Process

In October 2015, members of the Working Group developed an evaluation rubric ("Education Module Review Criteria") for outside reviewers to use. Review criteria drew on previous surveys and usage scenarios from other DataONE projects. In order to accommodate feedback from different reviewers across eleven different modules, the rubric was broadly designed to be holistic, flexible, and simple to use. The design and testing of the rubric was completed by different subsets of Working Group members with experience in designing survey instruments. The criteria sent to outside reviewers (see Figure 1) was grouped around four central questions, each with several sub-questions. The intent of the in-depth review was to explore and update the modules for accuracy, currency, and quality.

Outside peer reviewers, identified from the broader DataONE community network via members of the Working Group, were chosen for their expertise in the areas of one or more of the 11 educational modules. As Working Group members represented a variety of expertise in the DataONE community themselves, the selection process for outside reviewers was comprised of an informal conversation amongst those members, based on their knowledge of ongoing research in each topic area. This informal process allowed for the addition of reviewers as necessary when some were non-responsive. For this project, expertise was loosely defined as being recognized for having made significant contributions to the ongoing conversations around research data. Reviewers were drawn from a pool of academic, government, and industry experts in research data management and were not previously part of the project. They were all contacted by email, using a standardized request developed by the Working Group. Reviewers were asked to complete their evaluations online within one month of the request, using a custom Google form. No incentives were offered for completion of this process. Examples of the email request and rubric/Google form are found in Appendix 1 and Figure 1.

Please provide an in-depth review of the module, using the following criteria to guide your review.

1. Module successfully addresses the interests of one or more target DataONE audiences:
   - Formal educators (e.g. for use in course lectures)
   - Informal educators (e.g. for use in data management training workshops)
   - Individual learning

2. Module has conceptual integrity:
   - Concepts discussed match the overall module topic
   - Concepts discussed fit together as in a logical sequence
   - Concepts are discussed succinctly and accurately
   - Links to other topics and sources of information are logical and accurate

3. Expression of concepts is appropriate and eye-catching:
   - Concepts conveyed simply and understandably;
   - Illustrations and/or humor used effectively;
   - Imagery used appropriately;
   - Jargon terms explained succinctly;
   - Consistent slide layout and presentation

4. Please also provide any other comments, including on the following topics:
   - Are key concepts missing?
   - Do the modules reflect current up-to-date work on these topics?
   - Is the module length appropriate?

Figure 1: Education Module Review Criteria

Results

Across 11 modules, the Working Group requested 33 completed reviews from outside experts and received 22 with some reviewers submitting more than one review. Three reviews were requested for each module. Each module received at least one review; seven modules received either two or three outside reviews by the extended deadline. In order to maintain momentum on the project, the decision was made to progress with the revision process so long as each module had received at least one review. Reviews were submitted via a custom Google form or by email, and results were compiled in a Google spreadsheet. Comments on all three components of each module (slides, handouts, exercises) were compiled for analysis by the postdoctoral fellow attached to the Working Group. These reviews contributed to the revision of all modules and components by Working Group members in 2016.

Reviews

The 22 completed reviews were varied in depth, length, and attention to detail. Each section allowed an open response and many reviewers took full advantage of the space. Responses ranged in length from several sentences per prompt to hundreds of words for each sub-question. The scope of the responses also varied, with some reviewers taking broad overviews of the topical area and making general suggestions for improvement, and other reviewers taking the opposite tack, pointing out changes on a per-slide basis. However, while the free-form nature of the query was not specifically designed to elicit such a wide range of

replies, the exercise was fruitful. All of the responses contained useful, actionable information that allowed the Working Group members to reexamine, and in some cases, completely revamp the educational modules. For example, after incorporating the extensive feedback received on the two metadata modules, Working Group members decided that the updates allowed for combining the two modules into one larger module on metadata (thus reducing the number of completed modules from 11 to 10). Other reviewer notes centered on new or more useful references, suggestions for addressing learning objectives more fully, and advice about tying resources and slides together more explicitly. The rubric for reading reviewer comments is found in Appendix 2.

Deciding to move to GitHub

The peer review process took more than one year, starting in October 2015 and wrapping up in September 2016, when the educational modules were fully updated. Moving the modules to GitHub was completed in June 2017. Working Group members found the process was lengthy and somewhat cumbersome. First, the results of the outside reviews were collected and stored until Working Group members could evaluate and incorporate them with other reviews from the same module. This meant that even if a review was received immediately from a reviewer upon request in March 2016, the targeted sections or modules were not updated until six months later, in September 2016, when all revisions had been individually reviewed and incorporated. Some of this delay was tied to the nature of the Working Group, which holds two meetings per year, in the fall and spring, and its funding model. Working Group members are not funded for their work. While the group also holds monthly calls to check in, most collaborative work (including this project) is generally completed during those two working meetings each year. After further discussion, and in consideration of the potential lack of ongoing funding for these working meetings, the group decided that while the outside peer review was useful, completing this process again would not be scalable or likely possible for future maintenance or development. Therefore, finding ways to increase community engagement and build interest in shared maintenance became considerably more important for the ongoing sustainability and future of these educational materials.

Conversations related to the outside peer review process provoked the group decision in 2016 to move the modules to an open git repository (GitHub) for ongoing community engagement and transparency. Making the modules openly available to update and modify was one motivation. Other key reasons included: providing room for comments from community members, to transparently share information about recent updates, to potentially identify areas ripe for future resource development, and to give opportunity for the identification and attribution of creators and maintainers. Furthermore, GitHub is a collaborative platform that is already used by many members of the DataONE community, and it is familiar to those in the larger domain areas and research data communities.

The group used the Data / Software / Library Carpentry (the Carpentries) model of collaborative lesson development as an inspiration for this project. As with open-source software, Data Carpentry lessons allow anyone to propose changes or updates to the curriculum. Lessons are collaboratively developed and maintained by community members for Data Carpentry using GitHub. Proposals are reviewed, updated, and finally added to the core so that everyone may benefit from the shared knowledge. (Teal et al. 2015). This model was helpful for thinking about future sustainability of the DataONE Education modules, as well as

long-term maintenance. The Working Group chose this process in the hope that using GitHub for the development of lessons and having a specific community member take responsibility for future iterations will help with scalability and quick response. Additionally, having the educational modules available on GitHub allows for further collaboration with other groups and communities in the data management space that are developing educational resources and tools. Fostering these cross-project and cross-community collaborations in an open space could help build capacity, create more "buy-in" by interested stakeholders and new community members, and reduce unnecessary duplication of effort. Finally, placing the modules on GitHub can provide further insights into how and if community members are using, modifying, or suggesting updates to the resources. That capability was not available through the static webpages previously hosting the modules. Having information from users and contributors could assist in identifying problem points or future modules that could be developed.

Converting the Modules to GitHub

While the decision to migrate made sense in light of the reasons discussed above, the process of converting the PowerPoint component of the modules to slides in GitHub was time consuming and labor intensive for members of the Working Group. Several members of the group created a master slide template, put together a brief tutorial for constructing the slides, (URL removed for blind review) and led a short in-person session to teach other group members how to use a text format (Markdown) for creating the presentations. The editable code for the slide component of the modules is now available through GitHub[1] and viewable as slides[2]. Migrating the updated module content required the group to make additional choices for streamlining the slides, images, and notes in order to fit within the general restrictions of the template and space. This conversion process relied heavily on the expertise shared by members of the Working Group who already had a range of experience from intermediate to expert in working with YAML, pandoc, Markdown, and GitHub. Shifting away from proprietary formats such as PowerPoint is also congruent with best practices for sustainable research data management. As the research data community becomes increasingly more familiar and comfortable with these and other tools, similar conversations and migrations of resources may be less resource intensive.

Next Steps for the Working Group's Interaction with Modules on GitHub

As the Working Group continues to move forward with the educational modules on GitHub, a series of process decisions will need to be made. For example, how will the group allow community input? How will the group respond to community input? Will the group present the original criteria for feedback as a model moving forward, or encourage other forms of feedback in the GitHub environment? At what frequency will the group assess feedback and suggestions? These are extremely important questions the group will address in the next phase of module development, review, and maintenance.

Conclusions

In order to ensure that educational resources designed to teach skills and best practices for scientific research data sharing and management are useful, the maintainers of those

---

1   DataONE Data Management Education Module Repo: https://github.com/DataONEorg/dataone_lessons
2   DataONE Resource List: https://dataoneorg.github.io/Education

materials need to periodically evaluate and update them to ensure their accuracy, currency, and overall quality. This paper has described the process of integrating outside peer review of community resources as part of a comprehensive evaluation for addressing these concerns. It further outlines the motivations, concerns, and actions of moving the updated educational materials to an online community platform (GitHub) in order to build upon mechanisms for open science, ongoing iteration, participation, attribution, and transparent community engagement. Engaging with community members around the materials on GitHub will be an ongoing and iterative process, and the results of the move are not yet clear. However, the members of this project hope that these choices will afford more flexibility, responsiveness, and conversation around meeting the educational needs of the research data management community, broadly defined. Finally, describing and documenting this project, and particularly the impetus for changing the venue for engagement, may be a source of useful information for others, including data management librarians, who are often facing similar challenges while developing and maintaining educational resources for engagement and outreach. Generally, future funding and maintenance is a concern that is shared by many others, including librarians, who are often challenged by similar decisions about planning and where to place resources. This discussion may be most germane in communities where information changes quickly and open platforms are already in common use.

Supplemental Content

Appendix 1 and 2
An online supplement to this article can be found at http://dx.doi.org/10.7191/jeslib.2017.1114 under "Additional Files".

Acknowledgments

Funding Statement

Disclosure

The authors report no conflict of interest.

References

Baker, James, Caitlin Moore, Ernesto Priego, Raquel Alegre, Jez Cope, Ludi Price, Owen Stephens, Daniel van Strien, and Greg Wilson. 2016. "Library Carpentry: Software training for library professionals." *LIBER Quarterly* 26 (3): 141-162. http://doi.org/10.18352/lq.10176

Briney, Kristen, Abigail Goben, and Lisa Zilinski. 2017. "Institutional, Journal, and Funder Data Policies." In *Curating Research Data: Practical Strategies for Your Digital Repository*, edited by Lisa Johnston. Chicago, IL: Association of

College and Research Libraries. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838988596_crd_v1_OA.pdf

Carlson, Jake. 2015. "Future Directions for Data Information Literacy: Growing Programs and Communities of Practice." In *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, edited by Jake Carlson and Lisa Johnston. West Lafayette: Purdue University Press. http://www.thepress.purdue.edu/titles/format/9781612493527

Lyon, Liz. 2013. "What Is a Data Scientist?" Presented at the *8th International Digital Curation Conference Symposium, Amsterdam, January 15, 2013*. http://www.dcc.ac.uk/sites/default/files/documents/IDCC13presentations/LyonIDCC13.pdf

Michener, William, Dave Vieglais, Todd Vision, John Kunze, Patricia Cruse, and Greg Janée. 2011. "DataONE: Data Observation Network for Earth — Preserving Data and Enabling Innovation in the Biological and Environmental Sciences." *D-Lib Magazine* 17(1/2). http://dx.doi.org/10.1045/january2011-michener

Michener, William K., Suzie Allard, Amber Budden, Robert B. Cook, Kimberly Douglass, Mike Frame, Steve Kelling, Rebecca Koskela, Carol Tenopir, and David A. Vieglais. 2012. "Participatory Design of DataONE — Enabling Cyberinfrastructure for the Biological and Environmental Sciences." *Ecological Informatics* 11: 5-15. http://dx.doi.org/10.1016/j.ecoinf.2011.08.007

Software Carpentry. 2017. "Our Lessons." Accessed September 5. https://software-carpentry.org/lessons

Teal, Tracy., Karen A. Cranston, Hilmar Lapp, Ethan White, Greg, Wilson, Karthik Ram, and Aleksandra Pawlik. 2015. "Data Carpentry: Workshops to Increase Data Literacy for Researchers." *International Journal of Digital Curation* 10(1): 135-143. http://dx.doi.org/10.2218/ijdc.v10i1.351

Tenopir, Carol, Ben Birch, and Suzie Allard. 2012. "Academic Libraries and Research Data Services: Current Practices and Plans for the Future." *Association of College and Research Libraries (ACRL) White Paper*. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Wright, Sarah, Michael Fosmire, Jon Jeffryes, Marianne Stowell Bracke, and Brian Westra. 2012. "A Multi-Institutional Project to Develop Discipline-Specific Data Literacy Instruction for Graduate Students." Abstract IN22A-08. Presented at the *American Geophysical Union 2012 Fall Meeting, San Francisco, CA, December 3-7,2012*. http://docs.lib.purdue.edu/lib_fspres/10