10-2014

Library as Publisher of Graduate Student Data

Jessica Clemons
SUNY College of Environmental Science and Forestry, jclemons@esf.edu

NY3Rs

NY 3Rs ASSOCIATION, INC.
www.ny3rs.org

LIBRARY AS PUBLISHER INNOVATOR
PROJECT PROPOSAL

Project proposals must be submitted electronically by midnight on September 30, 2014. Submit applications to: chair@ny3rs.org.

SECTION 1:

Library/Organization: F. Franklin Moon Library, SUNY College of Environmental Science and Forestry

Contact Person: Jessica Clemons

Telephone Number: 315.470.6724

Email Address: jclemons@esf.edu

Council: ⊙ CDL   ⦿ CLRC ○ LILRC ○ METR   ○ NNYLN ○ RRL ○ SCRLC ○ SENYLRC ○ WNYLRC

Project Title: Library as publisher of graduate student data

SECTION 2:

Are you submitting more than one application?	☐ Yes ☒ No

If yes, indicate your priority for this project (1,2, etc.)   ☐ 1 ☐ 2   ☐ 3

Is this a cooperative project with other libraries or systems?   ☐ Yes ☒ No

List additional libraries here:
    Library 2:
        Contact Person:

    Library 3:
        Contact Person:

    Library 4:
        Contact Person:

SECTION 3:

Acceptance of partial grant

☐   Will not accept a partial grant

☒   Will accept a partial grant.

If you accept a partial grant, how will this affect the project?

If we are awarded a partial grant it will significantly reduce the speed with which we can respond to data management assistance and incorporation of data and related objects into our institutional repository, Digital Commons @ ESF.

SECTION 4:

<u>Funding Summary</u>

Library as Publisher Funds requested:	$5,500
+ In-kind cost share :	$1,000
Total Cost of Project:	$6,500

PROJECT NARRATIVE

SECTION 5:

<u>Project goal</u>

The purpose of this project it to publish graduate student data, particularly from individual laboratory groups, in the institutional repository Digital Commons @ ESF. Graduate students at Master and Doctoral levels generate many datasets that are all too often forgotten once the student leaves the institution. We propose having graduate students or upper-level undergraduates work with the lab groups and librarians to carefully describe and not just archive the data, but use it to communicate with future researchers.

SECTION 6:

<u>Explanation of why this project is a "Library as Publisher" project</u>

While students are only temporarily located on college campuses, their work is often part of larger, lengthier, and ongoing studies. Graduate students generate a good deal of data which can languish in various states of organization and description. If the library was to provide reliable access to this electronic data and encourage the use of data through the creation of metadata, we would show the value of these materials that may otherwise be unfindable or unusable.

It is important that the library be involved in this undertaking because when it comes to data, metadata is incredibly important and often forgotten. The library would assist these researchers to communicate their otherwise unpublished data to people who may be using it in the future. Data are often reported in journal articles but there is a sizeable amount of useful data that are not published. This project would highlight this hidden data.

SECTION 7:

<u>Project Objectives, Activities & Resources & Time Frame</u>

<u>Objective</u>
The primary objective of this project is to make graduate student data accessible with functional, helpful metadata. These data files may take form in spreadsheets, transcripts, videos, images, and a variety of other

formats. There could be several metadata schemas that are appropriate to our collections. We will be recommending the most appropriate schemas for our campus, helping students generate metadata, and evaluating the quality and completeness of the finished product to ensure usability. An exceptionally helpful document from the California Digital Library provides an overview of data management from both scientist and librarian perspectives (Strasser, Cook, Michener, Budden 2012).

Laboratory groups consist of a faculty member's advisees. For example, Dr. John Stella has several advisees seeking Master or Doctoral degrees: http://www.esf.edu/fnrm/stella/people.html
Each of these lab members conducts research in related fields. Additionally, Dr. Stella's future students may build upon or utilize current students' data. This lab group approach will encourage more datasets to be made available because more people will be participating and helping work towards a common goal. It will also showcase a usable collection with various kinds of data. This is superior to an individual data publishing approach because students will be working and learning together which is the current model in these dynamic groups. Please see appendix A for a letter of support from Dr. Stella.

Faculty data may also be included in the lab group collection, but their needs may be different. We will determine the fit of each dataset in the collection to ensure it is housed in the most appropriate place.

Activities
The primary activity for the interns would be to create a metadata/publishing checklist, review that checklist with participants to verify the information, and to assist uploading the data files into Digital Commons @ ESF (DC @ ESF). The checklist will detail what types of metadata to capture, what types of files are best for preserving and sharing, and other key pieces of information needed to create a robust collection. The checklist may be fairly simple and provide good detail that is of high quality and decreases the hurdles that some face when creating metadata (Borer, Seabloom, Jones, Schildhauer 2009).

In some instances a subject specific data repository, such as Dryad: http://datadryad.org/ or DataOne: https://www.dataone.org/ may be most appropriate. If data are deposited into those resources a link will be included in DC @ ESF to give users an additional means to locate the data files. The library will provide guidance on when to use our institutional repository and when to use a larger, open access data repository.

We will also create a research guide on publishing information and data in DC @ ESF. This will provide clarity on the project, outline the eligibility guidelines, and connect researchers with the librarians and interns who will help them. It would also be helpful to create guide to archiving data in subject-specific repositories and include some examples of what data to put in DC @ ESF versus an open access repository.

Working with the lab groups could be problematic because there may be new students and students preparing to graduate all working together. We will primarily work with the faculty leaders of these groups to establish more secure relationships and more consistent contacts. Also, we would distribute the funds to the faculty in the lab groups and they can decide how to spend the incentives within the guidelines of this grant.

If data need to be archived but not shared (due to pending publication or the sensitive nature of the data) we will consider those factors on a case by case basis. We may be able to restrict access to the files by location or establish embargo periods.

Resources
The library already has most of the software in place. We us bepress: http://www.bepress.com/ to run our institutional repository DC @ ESF: http://digitalcommons.esf.edu/.

To publish our research guides on publishing data and data repositories, we will use our subscription to Springshare's Libguides which is currently employed in the library. These guides will be made available to anyone via web. We have a workspace in the library for the interns as well as computers that they may use to work from.

Jessica Clemons will commit a significant amount of her time to this project as the institutional repository manager. She has experience creating digital collections, writing data management plans for National Science Foundation grants, and has been deeply engaged with faculty while working as a senior assistant librarian.

Time frame
If awarded this grant, we will begin attracting candidates for the internship in November. Ideally we would have several applicants to choose from in December and they would begin working on the project checklist when the academic semester begins in January. This would allow several weeks to understand the project, begin working on the checklist, and create advertisements to attract groups who may contribute data.

Near mid-March we would anticipate a few lab groups would commit to publish their data in DC @ ESF. This timeframe is a good fit for the end of classes and the graduation and departure of graduate students. We would have time to meet with the lab groups and review the checklist. The interns and Ms. Clemons may choose to work together over the summer, depending on the interns' plans. If they cannot work over the summer, we will resume the experience in the fall semester. The interns should plan to approximately 90 hours in spring semester and an additional 90 hours in the summer or fall.

References

Borer, E. T., Seabloom, E. W., Jones, M. B., & Schildhauer, M. (2009). Some Simple Guidelines for Effective Data Management. *Bulletin of the Ecological Society of America*, *90*(2), 205–214. doi:10.1890/0012-9623-90.2.205

Strasser, Carly; Cook, Robert; Michener, William; & Budden, Amber. (2012). Primer on Data Management: What you always wanted to know. UC Office of the President: California Digital Library. Retrieved from: http://escholarship.org/uc/item/7tf5q7n3

SECTION 8:

Project Evaluation Plan

Evaluating the project will be imperative for the continuation of this ambitious plan. A common complaint on college campuses is the lack of time. We are all doing too much with too little. Part of our evaluation will be how the lab groups view the time commitment required to complete the checklist. If the time commitment is deemed to be too cumbersome, we may choose to revisit the checklist. No matter how good a plan is, if there are too many barriers to complete it, the plan will flounder.

Long term evaluation will include looking at the usage statistics of datasets. Digital Commons @ ESF shows views, downloads, and search terms used to locate the items. These statistics may help us respond to users' needs in a changing online environment. We will evaluate collections periodically to determine how people are accessing this information and how we may provide the best information to help them find it.

If a data archiving class is offered on campus (more details about this in the project marketing plan) we would evaluate students' understanding and readiness to deposit data before and after the class in a brief survey. This assessment would show the impact of learning about and participating in data archiving and publishing in an open access repository. This survey may present us with enough information to publish our findings and share with colleagues beyond our institution.

SECTION 9:

Project Marketing Plan

The project will be strategically marketed to several different groups. Jessica Clemons publishes the *Research Times*, which helps faculty identify grants for their research: http://libguides.esf.edu/grants/RT. It is widely read and distributed on campus. An opportunity for assistance with managing graduate student data would be featured

prominently in the publication. She also has a relationship with the Graduate Student Association and will market this experience, both as internships and as participants, to that group at their meetings and via their facebook page and local listserv.

The Faculty Committee on Research is able to present at faculty governance meetings. It would be possible to gain a few minutes of the faculty's time during this meeting to share information about this important opportunity. This forum would not encourage questions because of the tightly scheduled agenda. We would invite interested parties to join us for a brief presentation on this topic so we can better answer the questions they may have about how to participate.

There are a few informal "lunch and learn" groups, where graduate students and faculty get together to discuss current events and ongoing research projects in departments and across lab groups. By attending these sessions, we would have an ideal audience with which to share this opportunity. The student interns may be good candidates to attend these meetings so there are more people attracting interest in publishing this data.

One of the more ambitious ideas is to create a data archiving class for graduate students. There are tentative plans to offer this course as an elective in the spring of 2015, being co-taught with library faculty and departmental faculty. If this idea comes to fruition, it would establish the perfect cohort to begin this exciting project.

SECTION 10:

Project Continuation

One of the most appealing parts of this project is the continuation of this plan after the grant money is exhausted. After the student interns work with librarians to create a reasonable, acceptable data management checklist for ESF, that list will continue to be in place. It can be readily updated as needs change. Collections that exhibit the characteristics of good data management will also be in place to serve as examples as more and more data are generated. This will aid others as they plan to archive and share their scientific data.

The data will be shared with others through these repositories and indirectly continue the project so long as the data are shared, distributed, and reused.

PROJECT BUDGET

SECTION 11:

DETAILED BUDGET

Personnel
Description:
We would like to hire 2 advanced undergraduate or graduate students to provide key areas of support for this project. Each student would be awarded a $2,000 stipend after completing approximately 180 hours of work for a rate of approximately $11.00/hour.

Grant Funds: $ 4,000
Applicant's Share of Cost: $ 0

Contractor or vendor services
Description: N/A

Grant Funds: $ 0
Applicant's Share of Cost: $ 0

Hardware, Software, Telecommunications
        Description: N/A

                                        Grant Funds: $ 0
                        Applicant's Share of Cost: $ 0

Continuing Education
        Description: N/A

                                        Grant Funds: $ 0
                        Applicant's Share of Cost: $ 0

Supplies & Materials
        Description: N/A

                                        Grant Funds: $ 0
                        Applicant's Share of Cost: $ 0

Other
        Description: Incentive to participate in data management collection and establish exemplars in the
                collection.

                                        Grant Funds: $ 2,500
                        Applicant's Share of Cost: $ 1,000

SECTION 12:

Budget Narrative: Concisely explain all budget expenditures

The student interns are critical components to this grant. Ideally we will attract an intern from the Syracuse University iSchool and one from ESF. This arrangement would be a nice complement for the creation of the metadata checklist and determining the needs of the specific lab groups. Existing library staff have the skills to advise and manage the students as they work on this project but we lack the time to do it independently. By creating this internship opportunity, students may also qualify to receive academic credit for the experience which may make it that much more attractive.

We would also like to devote $2,500 towards providing incentives to these lab groups to contribute their data to the institutional repository and/or other appropriate data repository. F. Franklin Moon Library would contribute $1,000 and we request grant funds for $1,500.

This incentive will encourage the creation of good metadata and provide high quality examples to present to other groups on campus. These incentives will be awarded in a single installment of $500 per qualifying lab group. We would stipulate that the funds shall be used for publishing papers in open access journal, presenting the work at a professional conference, or purchasing items that will assist in data gathering, such as software, lab supplies, or field gear. Incentives will be awarded on a first come, first served basis as long as the requirements are met.

The library is providing personnel and subscription costs to Libguides and Digital Commons @ ESF.

Thank you for taking the time to review this proposal.

SECTION 13:

Signature of Library, Organization or System Director

Stephen P. Weiter

Type or Print Name

9/25/14

Date

Appendix A: Letter of support

See attached file