Workshop on Scalable Institutional Data Publication Infrastructure

Statement of national need

The California Digital Library (CDL) at the University of California and the Dryad organization request $98,363 for a National Leadership Grant (NLG-L) program National Forum Grant in the National Digital Platform category for an invitational workshop to explore issues, identify impediments, and make recommendations regarding the widespread promotion and adoption of effective, scalable, and sustainable institutional data publication infrastructure.

Academic institutions deploy Institutional Repositories (IRs) to support local management of and access to PDF articles and, increasingly, research data. However, much of the institutional investment supporting research data management (RDM) is not currently being utilized effectively. Many disciplinary and non-institutional general data repositories, such as Dryad, see much higher adoption rates than IRs because of their position within the researcher's workflow (during article publishing, etc.). Unfortunately, IRs and these other repositories do not have working relationships that effectively engage institutionally-based resources such as librarians and data curator communities. Simply put, we are not working together as an integrated community to solve common RDM problems, and if we do not bridge this gap between institutions and data repositories, corporate solutions will increasingly take over the space.

The way to bridge this gap in a sustainable manner is to develop a consortial network of institutions and repositories. Working together creates global efficiencies by minimizing needless duplication of effort across institutions, freeing up time and funds and, in particular, allowing smaller institutions without significant resources (particularly staffing) to participate in and promote effective research data publishing. Our invitational workshop will bring together members of the library community to investigate and identify institutional needs and interests in a streamlined, open source data publication and curation service offering features that address research and data curator perspectives and positioned within researcher workflows. Our outputs will be a set of requirements for how data repository communities can develop a sustainable business model for and adoption of a collaborative data publishing platform that integrates data curation/stewardship more closely into researcher workflows.

Project design

The one-day workshop will be scheduled to coincide with the CNI 2018 Fall Membership Meeting, to maximize opportunities for institutional participation. We will also include institutions that do not usually attend CNI or have a voice in RDM discussions (HBCUs, non-ARL and CNI institutions). Participants will be selected to include a variety of institutional viewpoints: large (ARL and CNI) institutions, small and historically-disadvantaged institutions and liberal arts colleges, institutions that don't have an existing IR, those that have homegrown IRs, and those that are relying upon commercial alternatives. Through facilitated discussion and breakout sessions, the workshop

participants will:
- **Engage the library community** and increase the role of libraries in RDM workflows
- Identify and refine the requirements necessary to garner active **institutional support for and adoption of an open source, open data publication service** as a trusted and affordable alternative to commercial products
- Identify **institutional values** (perceived from the perspectives of Librarians, Vice Chancellors for Research, CIOs, and Data Curators) in curatorial and preservation services
- Identify **sustainable business and governance models** for these services
- Produce a set of requirements for the community for how non-institutional data repositories like Dryad can **work with and support institutions**

The recommendations of the workshop will be openly published under a permissive CC license and publicized widely through email, blogging, and other social media channels.

National impact

CDL and Dryad both have long-standing interest and experience in RDM and data publishing. CDL's digital curation team, University of California Curation Center (UC3), provides digital preservation, data curation and data publishing services, and has a history of coordinating collaborative and projects regionally, nationally, and internationally. The Dryad repository has been archiving open data supporting the scholarly literature for nearly ten years, and is integrated with the publication systems for over 120 journals. The workshop will result in a set of recommendations that, when implemented through follow-on activity, will offer the library community a full service, open source, low cost data publishing platform that can be used as a standalone solution or in coordination with existing IR strategies. The implementation of these recommendations will result in:
- An increase in the amount of open data available to the national scholarly community, through streamlined interfaces and integrations with publication workflows
- An increase in the impact of and access to that data, through the curation enhancement of and the proactive preservation management of individual datasets
- A customized, right-sized, community-defined, open source data publishing option for an institution's sole data IR offering or as a supplemental data IR offering

Budget
We are requesting $98,363 in IMLS funds to support the workshop. This will support a dedicated meeting facilitator ($8,500), provide partial travel support ($20,000) for the 20 invited participants from CNI/ARL-member institutions, full travel support ($30,000) for 12 non-CNI/ARL institutions (HBCUs, liberal arts schools, etc.), full travel support ($15,000) for 6 grant team members, $4,000 for the meeting room, $6,000 for catering, and $14,863 in UC federally-negotiated overhead (17.8% MTDC).