LG-70-18-0092-18

IMLS National Leadership Grant for Libraries 2018

Development of an Enhanced and Expanded Data Management Training Clearinghouse

- *Abstract* -

University of New Mexico, College of University Libraries & Learning Sciences
in collaboration with *Knowledge Motifs LLC*

The University of New Mexico in collaboration with Knowledge Motifs LLC is submitting a proposal for $249,525 in support of the *Development of an Enhanced and Expanded Data Management Training Clearinghouse* project. In partnership with the ESIP Federation, DataONE, USGS, RDA-US, the NCAR Libraries, University of Notre Dame, Indiana University, the U.K.'s Digital Curation Centre, and the University of Florida, and in potential partnerships with NASA, NOAA, the Network of National Libraries of Medicine, University of Michigan, University of Minnesota, Penn State University, this project builds on the partnerships and development efforts that produced the ESIP Data Management Training Clearinghouse[1] to:

1) expand and diversify the educational content in and services of the Clearinghouse, including expansion into additional science, social science, and humanities research areas.
2) incorporate community feedback and expand upon the existing descriptive characteristics (metadata) of the content in the clearinghouse to support enhanced discovery and classification of content in the system; and
3) develop a community vetted assessment framework capability for gathering and sharing feedback on included content recipients of training, and users of the training materials (i.e. trainers).

These efforts will result in an enhanced Clearinghouse with significantly improved discovery capabilities that enables targeted discovery and reuse of existing content, and delivers well-structured and consistent feedback on materials to content developers while also providing user feedback to potential re-users.

[1] http://dmtclearinghouse.esipfed.org/about

IMLS National Leadership Grant for Libraries 2018
Development of an Enhanced and Expanded Data Management Training Clearinghouse

Statement of National Need:

In recent years there has been an increasing recognition of the importance of effective research data management planning and execution. This recognition has developed largely in response to two factors: 1) the growth of collaborative data intensive research, and 2) increasing requirements from funding agencies and publishers for documented data management and sharing plans, and for evidence of execution of those plans through availability of research data products for discovery and reuse by other users. Meeting these needs has highlighted gaps in the knowledge, skills and abilities of researchers to effectively plan for and execute full-lifecycle research data management strategies, and in the cross-disciplinary and collaboratively acquired knowledge and skills needed by librarians to support those researchers[1–3]. In an effort to fill these gaps many types, lengths, formats and modes of educational content from diverse organizations have been created for the management of research data. This variety has created an environment of growing confusion rather than opportunity for potential consumers of educational content. In particular, the growing universe of materials targets different audiences and different parts of diverse research and data lifecycle frameworks. In addition, these materials are designed for delivery in a variety of modes, and in different disciplinary focus areas. The burgeoning diversity of materials highlights the need for more effective tools for classifying and discovering educational resources that align with specific trainer/trainee needs. Furthermore, for content creators, the lack of a community agreed upon assessment framework that can inform their content creation decisions inhibits the effectiveness of their efforts. The proposed expansion and enhancement of the Data Management Training Clearinghouse is designed to address these related needs and directly relates to the key themes of the IMLS Focus Summary Report for the National Digital Platform report[4] in the areas of user engagement and crowdsourcing, and continuous professional development for library professionals and researchers. If funded, the proposed expansion and enhancement will also help the IMLS achieve the second strategic goal from the IMLS Strategic Plan of 2018 – 2022: to build capacity in libraries and museums[5]. By making more librarians aware of and involved in the training of present and future data users, not only will new public – private partnerships be built to make the work that librarians do more evident, but librarians will also have opportunities to learn more about the data needs of patrons across disciplines, thus contributing to their professional development[a] while expanding and improving their library's service portfolio. By sharing their own and reusing others' teaching and learning materials from across disciplines, library professionals and institutions will be able to share and adopt best practices and innovations in data management, but also help their managers make informed decisions about how to best manage the data that they steward as those practices change[bc].

Project Design (Work plan):

Goals, Outcomes & Assumptions

[a] IMLS Strategic Plan 2018-2022[5], Goal 2: Building Capacity: Objective 1: Support the recruitment, training, and development of library and museum staff, boards, and volunteers, helping to grow a skilled, professional workforce.

[b] IMLS Strategic Plan 2018-2022[5], Goal 2: Building Capacity: Objective 2: Encourage library and museum professionals and institutions to share and adopt best practices and innovations.

[c] IMLS Strategic Plan 2018-2022[5], Goal 2: Building Capacity: Objective 3: Identify trends in the museum and library fields to help organizations make informed decisions.

1. Build depth and expand the diversity of the inventory of educational resources in the Clearinghouse based on user needs and feedback.

Objectives:

- Establish community-developed and -supported Selection Criteria and/or Collection Development policy statements for the Clearinghouse.
- Incorporate resources into the Clearinghouse for 5 – 6 additional Science & Social Science domains.
- Add a significant number of "learning activities" to the Clearinghouse registry that provide hands-on experience with data management tools or data for disciplines represented in the Clearinghouse.
- Implement crowdsourcing events at science & social science researcher and educator, librarian/data support conferences and other venues to raise awareness, and solicit educational resources.
- Present demonstrations & workshops of the Clearinghouse contents and services to targeted audiences at venues as described above.

Assumptions:

- Adoption of mechanisms for community input, such as committed volunteer metadata and content review panels, an Advisory Board or alternative model of engaging with data, subject domain, and metadata experts who will contribute insight to strategic decisions related to the Clearinghouse educational content.
- Community input from experts will result in the development and adoption of Selection Criteria and/or Collection Development policies to use in scoping and encouraging submission of educational resources into the Clearinghouse registry.
- As the Clearinghouse becomes more well-known due to the efforts of the Advisory Board /working groups, and community outreach activities, additional educational resources will be discovered and added to the Clearinghouse.
- More substantive and user vetted Help and instructional materials for submitting to the Clearinghouse will be available online, lowering the barriers for broadened community submissions.

2. Expand discovery of resources in the Clearinghouse.

Objectives:

- Enhance baseline knowledge of Clearinghouse usage by gathering improved metrics on the use of the Clearinghouse and its search, browse and submit functions.
- Enhance discovery through external search engines through the incorporation of microformats /RDFa / microdata into the Clearinghouse web pages using the Schema.org endorsed Learning Resource Metadata Initiative (LRMI)[d] metadata schema[6].
- Conduct user testing (with submitters and searchers) on the utility, understanding and classification specificity of the descriptive metadata schema used to document the educational resources.
- Incorporate changes to the metadata schema, the Help text, and other documentation to reflect the feedback from the user testing.
- Refine the controlled vocabularies (CVs) for key elements within the LRMI metadata schema to better describe the topics important to the Science and Social Science domains of Clearinghouse resources.
- Test the impact and effectiveness of the refined CVs by soliciting user feedback from researchers, data professionals, and librarians who use the Clearinghouse search functionality.

Assumptions:

- Well-managed and vetted controlled vocabularies can be found that can be linked into the Clearinghouse registry, such as the e-Science thesaurus[e] underlying the e-Science portal for Librarians.

---

[d] http://lrmi.dublincore.net

- Workflows can be incorporated into the content management system to facilitate improvement of the CVs in a scalable fashion.
- Volunteer user experience and search engine experts can be leveraged to assess the impact and effectiveness of the refinements to the CVs.
- A sufficient user response will be obtained to provide meaningful feedback on the effectiveness of the expanded discovery environment for the Clearinghouse.

3. Improve the web interface, and add functionality and services to the Clearinghouse based on existing and ongoing user feedback.

Objectives:

- Incorporate enhancements to the Clearinghouse user interface and functionality as prioritized by existing and ongoing, formal and informal user feedback[f].
- Add services to the functionality of the Clearinghouse based on stakeholder and user feedback, e.g., user annotation and/or rating services for the resources in the Clearinghouse.
- Incorporate more methods of receiving user comments, including a component of the assessment framework described in 4) below, on the integration of targeted functionality or services into the Clearinghouse interface to achieve more specific, structured, and continuous feedback.

Assumptions:

- Mechanisms either already exist within the existing Clearinghouse platform or could be easily developed by contract and/or volunteer developers from the Open Source communities for the types of functionality and services that users request.

4. Identify, adapt or develop appropriate assessment frameworks for the educational content in the Clearinghouse in order to 1) assist users (both learners and instructors) in choosing which resources would be most appropriate for their needs (see 3 above), and 2) capture feedback on the content itself for use by content developers to improve their materials.

Objectives:

- Organize workshop(s) of participants from science and social science research data communities, data professionals including librarians, and professional educators to discuss and recommend an appropriate purpose, goals, objectives, and format(s) of assessment framework(s) that could be applied to the descriptions of the educational resources included in the Clearinghouse.
- Establish a working group to take the recommendations from the workshops mentioned above and develop or adopt one or more assessment frameworks for the Clearinghouse, e.g., by adopting the DataONE EEVA tool[g] or adapting the evaluation criteria from the CLEAN network[h].
- Incorporate the use of the assessment framework(s) adopted or developed into the Clearinghouse as a service.
- Conduct both informal and formal user testing to evaluate the effectiveness of the assessment service.
- Adapt the assessment service as suggested by the user feedback received.

Assumptions:

- The Clearinghouse inventory of educational resources is broad enough and deep enough to provide a wide range of content types to test the assessment services incorporated into the Clearinghouse.
- An interested cohort of volunteer assessment experts can be found to join in the efforts.

---

[e] http://esciencelibrary.umassmed.edu/professional-educ/escience-thesaurus

[f] See Usability Report provided as a supplemental document for the existing enhancement wishlist.

[g] https://www.dataone.org/education-evaluation

[h] https://cleanet.org/index.html

- A wide range of volunteer user testers can be found to provide meaningful feedback on the effectiveness and utility of the assessment functions / services added to the Clearinghouse.

Risks

One of the main risks associated with this project involves the reliance upon volunteers to do much of the work in the time they have to spare from their jobs, and on volunteer reliant organizations. Using the Wikipedia crowdsourcing model as a guide, project leaders are optimistic that a quality information registry for educational resources can be enhanced and sustained by data and researcher experts that have a vested interest in its success. In order to do that, however, there is a need to expand the network of experts who are aware of the Clearinghouse and to motivate them to contribute their resources, feedback and strategic advice on how to best expand and enhance the Clearinghouse and build more functionality and services. One of the important questions for a strategic Advisory Board for the Clearinghouse expansion project is to identify important motivators for participation by volunteers, such as community recognition of their expertise, recognition of the quality of their contributions by endorsement from well-reputed data management organizations, and favorable annotations or recommendations from consumers of their educational resources.

Another aspect of risk associated with this project is the unknown capacity and scalability of the technical infrastructure upon which the Clearinghouse currently relies. The data that is currently stored is small as it is all metadata about the educational resources stored elsewhere, and so the current capacity is more than adequate. In future, when the registry grows there might be a need to move to a different storage environment. More of a risk than storage, however, is the capacity of the current content management system, Drupal, to accommodate the types of functionality and services that users have begun to request within the system itself, including integration with other systems such as Github and virtual learning environments. In addition, in time, the technical infrastructure of the current host (the ESIP Federation) may change, so it will be important to ensure that the technologies and associated underlying data models used for the Clearinghouse registry and its associated services are transferable and easily replicated.

Theoretical Foundation and Current Practice

There are two key theoretical issues that could be addressed to some extent by this project:

1. Approaches to research data management skills development
2. Effectiveness of educational resources created to teach research data management skills

With regard to the first issue, there have been a number of excellent efforts to study and recommend approaches to developing research data management skills including the Carlson, *et al* study / book on skills needed by researchers[7]. In addition, there have been a number of efforts to define "research" or "data" lifecycle models to help guide the kind of activities and skill areas needed to manage research data, e.g., by the Digital Curation Centre in the United Kingdom[i], the U.S. Geological Survey Community for Data Integration's Science Support Framework[j], DataONE's Research Lifecycle[k], Force11's FAIR Data Principles[l], and the CESSDA Expert Tour Guide[m], to name a few. Other efforts to recommend curricula or skill development areas needed for research librarians, or other data "supporters" include recommendations from the Belmont Forum in their Data Skills Curricula Framework[n], for example. All of these efforts have been developed by particular communities of experts, and when educational resources have been identified and associated with

[i] http://www.dcc.ac.uk/resources/curation-lifecycle-model
[j] https://my.usgs.gov/confluence/display/cdi/CDI+Science+Support+Framework
[k] https://www.dataone.org/data-life-cycle
[l] https://www.force11.org/group/fairgroup/fairprinciples
[m] https://www.cessda.eu/Research-Infrastructure/Training/Expert-tour-guide-on-Data-Management
[n] http://bfe-inf.org/sites/default/files/doc-repository/Outline_Data_Skills_Curricula_Framework.pdf

these models (known as "frameworks" within the LRMI metadata schema[6]), the Clearinghouse provides the capability to document that association.

In terms of the second issue, the effectiveness of educational resources created to teach research data management skills, there have been some efforts to assess educational resources from the point of view of what the consumer learned from using the resource[e.g. 8]. Based on an analysis by Soyka et al.[8], one of the Clearinghouse partners, DataONE has developed an evaluation tool that could be applied to the resources in the Clearinghouse. In addition, there has been some interest expressed in bringing professional education experts into discussions of assessment frameworks in order to apply quality criteria to the educational resources included in the Clearinghouse. These criteria could prove useful not only to resource creators, but also to consumers of the resources so that they would have a better sense of how the resource intends to meet learning objectives, for example, or which learning styles are most appropriate for the resource. The assessment working group and workshops included in this project's work plan are intended to discuss these options, contribute to the approach to take for assessing the Clearinghouse's resources, and develop a strategy to implement assessment within the Clearinghouse.

Community Engagement

Community engagement will be key to the success of the efforts to enhance and expand the Data Management Training Clearinghouse. While the Clearinghouse was initially designed and implemented by participants from the Geoscience researcher and data practitioner communities (i.e., the U.S. Geological Survey, the Data Observation Network for Earth (DataONE), and the Earth Science Information Partners (ESIP) Federation), participants from other communities have become interested in participating as the Clearinghouse has become more well-known and its value better understood and appreciated. More recently, interested partners and collaborators have included those who support researchers throughout the research lifecycle including librarians, data curation specialists, data preservation specialists, and data wranglers. The Clearinghouse has been particularly well received from the research librarian communities that have been approached as it has been seen to have a great potential for helping not only the researchers whom research librarians support, but also those in the librarian profession whose job it is to teach about research data management, beginning at the undergraduate and graduate level (if not earlier) and continuing on to early career and established researchers in diverse environments. In particular, librarians are positioned at a critical communication nexus between the researcher and the data repositories that will receive and maintain the research data; thus, our focus on finding more opportunities to engage and collaborate with librarians early on in the strategic direction and development of the Clearinghouse, e.g., with RDAP, ASIS&T, medical librarians and others.

In addition, the project has received a number of suggestions for ways to help facilitate and lower barriers to teaching and learning about research data management, such as connecting through virtual teaching environments for those underserved researchers, librarians and/or data specialists who do not have the capacity within their own organizations to offer classes related to data management. Other feedback has noted the value in connecting with another key community --the educator community, e.g., the National Association of Geoscience Teachers. These kinds of professional educator communities could potentially help the Clearinghouse working groups develop assessment frameworks, or guidelines and techniques for creating quality educational resources targeted to diverse learning styles, audiences, and points during the research lifecycle when particular data management skills are needed.

Members from these communities will be identified and invited to collaborate with the Clearinghouse at workshops and/or with working, advisory, editorial or metadata review boards as appropriate.

Projected Users and User Engagement

In preliminary brainstorming about possible users of the Clearinghouse registry, project organizers discussed five generic user types (personas): Science Researchers, Science Modelers, Librarians / Curators / Information Managers, Educators, and Data Scientists/Wranglers. Given the timeframe for the initial funding and development of the Clearinghouse, only three functions were built into the first iteration of the registry:

Search, Browse and Submit. The plan was to provide relatively simple functionality in order to get the registry off the ground so that it could serve all of the generic users to some extent, but the Science Researcher / Modeler most specifically. Since the Clearinghouse launch in October 2016, the feedback that has been received has affirmed the utility of the three main functions along with suggestions for enhancing and expanding the functionality for all users. More surprisingly, however, two types of generic users have expressed more interest than anticipated in what the Clearinghouse can offer: Librarians / Curators / Information Managers, and Data Scientists / Data Wranglers. As a result, the current project leads are placing increased emphasis on adding enhancements and services that will help those users learn about research data management for themselves, as well as to teach others. User engagement efforts for these users will be directed to conference venues such as ASIS&T, RDAP, ACRL, and others.

In terms of the Research Scientist and Research Modeler users, early career researchers have expressed the most interest; thus, it seems most productive to direct efforts to engage new users in venues where the early career scientists congregate, such as the Ecological Society of America's Annual Conference, the American Geophysical Union's Fall Meeting, and other subject domains as opportunities present themselves. The project leaders will work with Advisory Board members to bring information about the Clearinghouse to those communities through their existing community connections.

By continued close affiliation with the ESIP Federation and DataONE and other collaborators at college/university libraries and professional educator organizations, the Clearinghouse project will be able to continue gathering user feedback on various aspects of the user interface to the Clearinghouse. Both organizations have interest groups and/or experts on staff who are knowledgeable about various forms of user testing who can offer advice and assistance in setting up both formal and informal user testing. These efforts will be ongoing throughout the project and will range from testing the effectiveness of the functionality of the search and browse functions by data users, evaluating the understanding of Help text and other documentation provided to facilitate the submission, and the review and publishing workflows of the librarian/data professional submissions of content to the Clearinghouse.

Breadth of Impact (underserved communities)

Most of the educational resources that are described in the Clearinghouse come from organizations and consortia that are very interested in and knowledgeable about general research data management skill development. Along with this general knowledge, there is significant familiarity with the practices of specific subject domains to use data formats, tools and practices that might vary between disciplines. Usually, unless a researcher or subject librarian knows about these organizations, it can be difficult to find the educational resources they create. The registration of these educational resources into an online location focused on cross-disciplinary and multi-organizational resources relating to data management topics can streamline resource discovery for potential users who are not familiar with the growing number of organizations that are developing and delivering data management training materials. For example, a social science librarian from a smaller undergraduate or community college may not have an academic or experiential background in a specific science discipline, so when asked to teach about or support data creation from that other science discipline, it can often be quite challenging to find or evaluate educational resources on disciplinary data management practice that they may want to incorporate into their own training activities.

One of the main goals of an expanded and enhanced Data Management Training Clearinghouse is to greatly improve the discovery of both general and discipline-specific training data management training materials. While the general topics can be used by many subject domains, user feedback has indicated that it would be very helpful to include more educational content designed to supplement general topics with hands-on learning activities or detailed exemplars of problems and solutions with a greater disciplinary focus. For this reason, collection development efforts have been and will continue to be expanded to include more science and social science subdomains, as well as supplemental learning activities that reflect the rapidly changing practices in research data management.

As a result of this additional focus, the Clearinghouse has the potential to greatly improve the breadth of its impact on research communities that are in the early stages of experiencing the complexities that come from the exponential growth in research data. The project design provides for the inclusion of more educational resources, but also for the refinement of the descriptive metadata scheme to include more specific characteristics and to craft controlled vocabulary terms that more precisely describe the educational resources coming from specific subject domains as suggested by stakeholders from those communities.

Management Plan

Administrative responsibilities for this project will be the responsibility of PI, Dr. Karl Benedict, and the University of New Mexico, his parent institution. Those areas of responsibility will include:

- Managing the overall budget and submitting reports including managing payments to subawardees, contractors, and student research associates; to participants for travel and expenses; and to organizations and businesses for workshop expenses.
- Collaborating with Co-I on project coordination, and supervision of the web application developer and research assistant.
- Participating in the project advisory board organization and development activities.

Project Coordination responsibilities for this project will be the responsibility of PI, Dr. Karl Benedict, and Co-I, Nancy Hoebelheinrich. These responsibilities will include:

- Volunteer coordination and management.
- Setup and maintenance of advisory board, working group and web application developer meetings.
- Assistance with workshop logistics, organization and presentations.
- Management of Clearinghouse operations and workflows including: serving as a liaison with the ESIP content management host site, web application developer, and with the research associates and volunteers working on metadata input, review and quality control.
- Editorial management of the Clearinghouse metadata and the publication processes.
- Participating in the project advisory board organization and development activities.
- Assistance in the development of reports, marketing and promotional materials.

Web application development for this project will be the responsibility of the contractor to be hired by the project. Areas of responsibility include:

- Working with project leaders to establish priorities, specifications, requirements, deliverables, schedule and timelines for enhancements to and expansion of the Clearinghouse.
- Providing technical project management for the enhancements to and expansion of the Clearinghouse.
- Submitting timely and accurate reports and invoices to project leaders for the work done.

Collection Development and Metadata assistance for this project will be provided by the Research Assistant hired by the project. Areas of responsibility include:

- Searching appropriate sources for suitable educational resources to add to the Clearinghouse per Selection Criteria and/or Collection Development policies developed by project leaders, based upon input from advisory board members and other experts in research data management.
- Submitting and reviewing full metadata records for the educational resources identified as suitable to be added to the Clearinghouse.
- Assisting with the refinement of the metadata scheme and controlled vocabularies used to describe the educational resources in the Clearinghouse.

Strategic advisory and community engagement activities for this project will be provided by volunteer data management, user interface / experience, and professional education experts in a number of roles including advisory board membership; metadata editorial review; working group membership on topics related to metadata refinement and review, selection criteria and/or collection development policies, assessment framework development; user interface / experience testers; workshop presenters and participants; open source developers with knowledge of content management systems such as Drupal, code and data

repositories such as GitHub, and data management tools such as DMPOnline, the DPMtol, Jupyter Notebooks, R, and others.

Resource Requirements

Currently, the Clearinghouse data is managed in the ESIP Federation's Drupal content management system and web interface which is hosted on Pantheon (cloud-based platform). The web interface provides browse and faceted search capabilities as well as the input form by which metadata descriptions are submitted to the Drupal database. The amount of storage required for the Clearinghouse is quite small as it contains only the metadata for the educational resources, not the resources themselves. Project leaders anticipate that the storage and processing requirements for the Clearinghouse database will not grow substantially even when the inventory increases considerably, and ESIP leadership have indicated that they will continue to support the small increase in storage and computational capacity that would be required for an expanded and enhanced Clearinghouse.

The ESIP Federation is contemplating a move to a different technical base for their organizational content in the next year or so which will require some advanced planning for Clearinghouse project leaders as the technical infrastructure choices are considered and decisions are made. Depending upon the capabilities of the system chosen by ESIP, the Clearinghouse may move to a different system, or distribute aspects of expanded services to other platforms such as GitHub or other open source repositories and services. Project leaders will be involved in those discussions, and will certainly seek the advice and assistance of other technical partners in the data repository / educational services communities and on the Clearinghouse Advisory Board.

Project Milestones, Evaluations, and Timeline (See *Schedule of Completion* for more details).

- Project Kickoff: July 2018.
- Form Advisory Committee, schedule and organize quarterly virtual Advisory Board meetings for the next 3 years: First three months of project
- Hire project staff each year, as necessary.
- Establish working groups on Selection Criteria / Expanded Collection Development Policy; Enhanced Metadata Model, and Development / Adoption of Assessment Framework during Year 1: October 2018.
- Bring working group recommendations to Advisory Board for endorsement, iteration, and finalization during Year 1 & beginning of Year 2.
- Implement working group recommendations into Clearinghouse web interface and service functionalities during Year 1 and Year 2.
- Plan & present three workshops on Clearinghouse functionality and services at ACRL and RDAP meetings.
- Plan & implement at least one online crowdsourcing competition / event each year of the grant.
- Quadruple the number of educational resources, and the disciplines represented in the Clearinghouse from the pre-grant total by the end of the third year (i.e., add approximately 300 resources and expand coverage to five – six subject domains).
- Plan and implement formal and informal user testing for enhancements made to the Clearinghouse by the end of each grant year.
- Develop sustainability plan with Advisory Board starting mid-year of Year 2 and completing by the end of the grant period in June 2021.
- Complete implementation of enhancements: June 2019
- Final Report: June 2021.

Dissemination Plan

Successful implementation of the project plan for an expanded and enhanced Data Management Training Clearinghouse will depend upon many voices making the Clearinghouse known as a trusted and valued resource. Besides project staff, dissemination of information about the Clearinghouse will need to be directed to two target audiences: those who want to acquire skills for and learn about research data

management practices, methods, and techniques, and those who want to teach others about those skills and practices. The Advisory Board will be comprised of various stakeholders from both of these audiences, and as such will be asked to act as ambassadors and connectors to the domains they represent. The forms of the communiques that make potential consumers aware of the Clearinghouse functionalities and services may range from conference presentations and workshops to online webinars to periodic blog and social media posts that feature recent submissions to the Clearinghouse on specific aspects of research data management or useful tools that support the data management process. Advisory Board members will be asked to recommend which venues would be most productive to reach the communities and domains they represent.

Plans to expand and enhance the content within the Clearinghouse involve a two-pronged approach to make potential submitters aware of the registry, to facilitate and lower the barriers to submission, and to solicit feedback on the submission interface and workflows. The first approach is the implementation of a schedule of online and face to face crowdsourcing activities using recommendations for methods and approaches from such organizations as the Mozilla Open Leadership Program[o] and its Open Leadership Training Program[p]. The second approach incorporates both formal and informal methods for soliciting periodic user testing and feedback into Clearinghouse operations and workflows related to the submission process.

Finally, the project will leverage the expanded discovery capabilities enabled by the use and possible extension of the LRMI metadata schema to increase both the precision and the breadth of Clearinghouse content discovery over the open Web. By ensuring that the metadata in the Drupal content management system is expressed by means of formats that can be picked up by search engines such as Google, Yahoo & Yandex, the educational resources described within the Clearinghouse will be more readily discovered beyond their home sites via the Clearinghouse.

Sustainability Plan

Moving the Clearinghouse from its current incubator status to a sustainable registry and service will require that the data communities see the need for it (already reflected in interest expressed by project collaborators), and participate in its success. Project leaders see the way to both harness the existing need, and engage the research data community by employing a combination of strategies. Such strategies include enabling crowdsourced contributions by experts using methods similar to those used by the Wikipedia Foundation and grant funding. Funding possibilities being contemplated include the formation of a separate membership organization with different tiers of paid membership depending upon contribution; endorsements and financial contributions by partner organizations such as publishers, tool developers, and data repositories, tiered service levels where clients / patrons could pay for more specialized services such as the creation of personalized educational resource recommendations; and donations. Project leaders (Hoebelheinrich) recently attended a Sustainability BootCamp organized by the Science Gateway Community Institute, an NSF funded organization focused on helping incubator organizations such as the Clearinghouse move to sustainability. From that weeklong opportunity for training and consultation with experts, it is clear that the Clearinghouse has two audiences to reach: consumers of the content in the registry (researchers and others wanting and needing to learn or teach data management skills), and suppliers of the content (educational resource creators such as librarians who have a stake in making their information known to a wider audience than their own home institution or client base). The first steps for both of those audiences is to broaden and deepen the inventory of educational resources available from the Clearinghouse, but also to add services and capabilities for making the educational resources more relevant to targeted audiences for specific data management training needs; hence this proposal to IMLS. In the future, the project leaders see the need for the completion of gap analyses to determine where other educational resources are needed, and

[o] https://mozilla.github.io/leadership-training/

[p] https://mozilla.github.io/open-leadership-training-series/

to seek creators, peer reviewers, and funding in order to fill those needs. This will also be important to do as the practices, tools, and skills needed for research data management change over time.

National Impact:

Scale of Impact

This project has the potential to achieve at least a national level of impact, as librarians, data specialists and researchers throughout the U.S. become more aware of the existence of the Clearinghouse, and increase their participation in and use of it. The impact can extend into the multiple areas of community and educational involvement in which academic libraries provide service, especially as the creation, management, and re-use of data becomes more incorporated into educational curricula. In addition, because current and future engagement with the Clearinghouse includes international organizations the project's potential impacts can scale internationally.

Anticipated Products

If funded, the products that would be created include 1) education/library-community developed Selection Criteria for educational resources that teach data management skills to audiences ranging from community college to early and late career researchers; 2) a community developed Assessment Framework for evaluating the educational resources themselves as well as the utility of the resources from the perspectives of content users; 3) metadata (both initial training resource documentation and ongoing assessment data); and 4) computer code that provides enhanced metadata-enabled search and discovery of research data management training materials, and assessment tools for capturing assessment information from users and others.

Adaptability

All of the products from this project will be generated from community developed standards or practices, will derive from open source tools and development with open licenses, and so will be easily adaptable by other domains and metadata registries.

Sustainability (of impact)

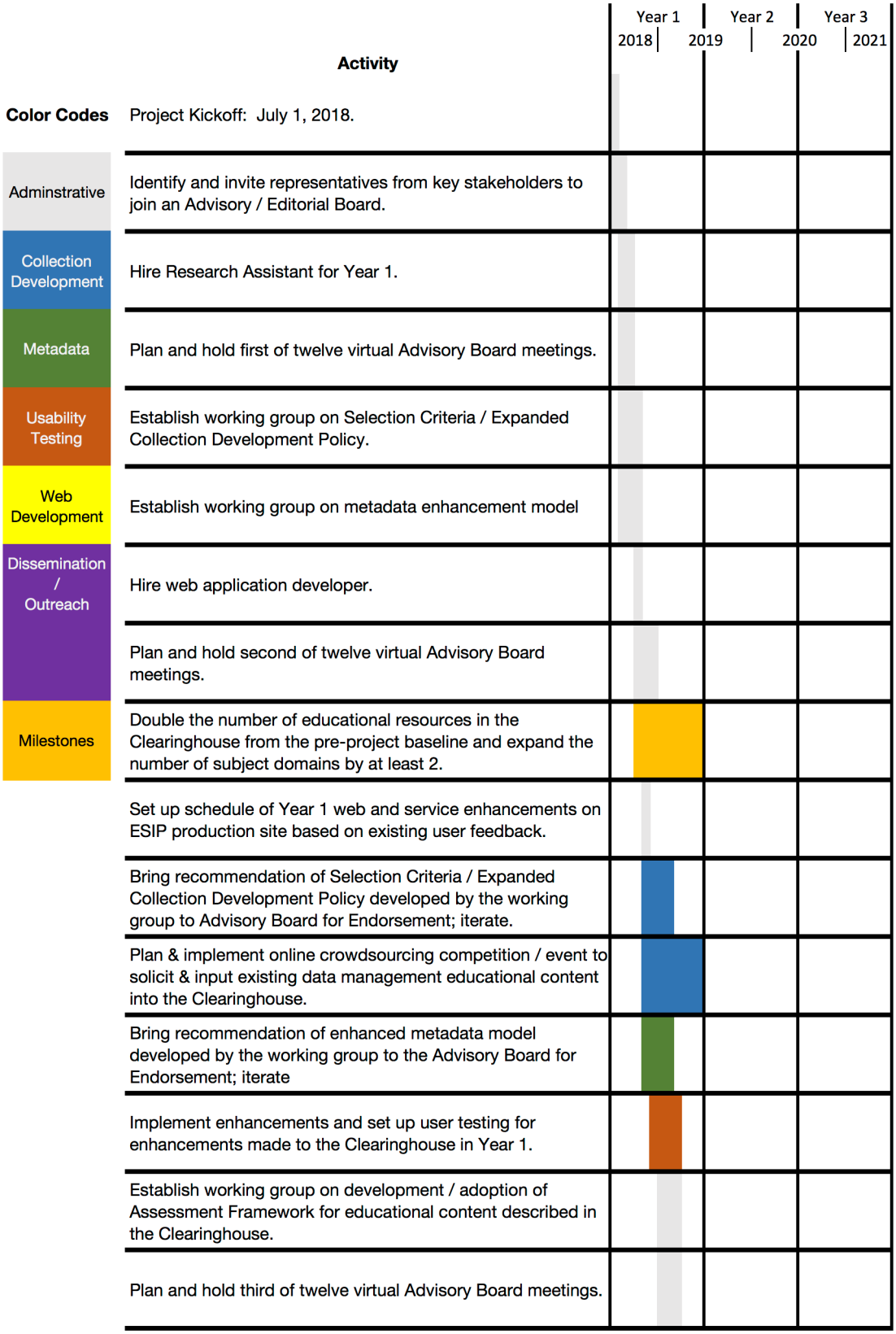
Sustaining the impacts outlined above requires ongoing demonstration of value that fosters continuing engagement from the diverse communities of users and contributors that will be involved during the project. A key element of this requirement for ongoing engagement is the development of a governance structure for the Clearinghouse that enables continuous community feedback and opportunities to contribute to the ongoing development of the Clearinghouse. This will be an explicit objective of the sustainability planning activity that will be initiated in Year 2 and completed before the end of the project. As part of this sustainability planning work, the project will leverage the network connections that the project's advisory board has to establish and strengthen linkages with organizations that have a shared interest in facilitating the effective discovery, use and assessment of data management training materials across disciplinary boundaries.

Collection and Reporting of Performance Measures with Benchmarks

The proposed project will build upon an existing successful and growing platform (http://dmtclearinghouse.esipfed.org) that provides basic discovery and browsing of research data management educational materials. This project will provide measurable improvements in the diversity, discoverability, usability, and quality of the educational materials included in the Clearinghouse by: 1) strategically expanding the holdings of the Clearinghouse to include a greater diversity of materials; 2) increasing the ability of Clearinghouse users to identify specific materials that meet their needs by improving the functionality and services of the Clearinghouse based on existing and continuously gathered user feedback, e.g., adding more effective and standardized search facets; and 3) providing a cross-Clearinghouse assessment tool that is available to all content contributors for use in capturing user (both trainer and trainee) feedback that can inform continued development and improvement of available educational content.

See* Supporting Document 1 – Bibliography *for references cited

IMLS National Leadership Grant for Libraries 2018
Development of an Enhanced and Expanded Data Management Training Clearinghouse
- ***Schedule of Completion*** -
University of New Mexico, College of University Libraries & Learning Sciences
in collaboration with *Knowledge Motifs LLC*



Activity	Year 1 2018	2019	Year 2 2020	Year 3 2021
Iterate the enhancements made to the Clearinghouse based on the user testing and feedback; complete enhancements for Year 1.				
Implement approved / endorsed Selection Criteria / Expanded Collection Development Policy within Clearinghouse workflows.				
Plan & present workshop on Clearinghouse functionality and services at ACRL (April 2019).				
Plan and hold fourth of twelve virtual Advisory Board meetings.				
Plan and hold fifth of twelve virtual Advisory Board meetings.				
Implement approved / endorsed metadata model within the Clearinghouse				
Hire Research Assistant for Year 2.				
Plan and hold sixth of twelve virtual Advisory Board meetings.				
Triple the total number of educational resources in the Clearinghouse from the pre-project baseline and expand the number of subject domains by at least 1 additional domain.				
Set up user testing for addition of enhanced metadata and search capabilities to Clearinghouse				
Set up user testing for addition of Selection Criteria / Collection Development Policies				
Bring recommendations for the Assessment Framework developed / adapted by the working Group to Advisory Board for Endorsement; iterate.				
Plan and hold seventh of twelve virtual Advisory Board meetings.				
Implement approved / endorsed Assessment Framework within Clearinghouse services and workflows.				
Plan & implement online crowdsourcing competition / event to solicit & input existing data management educational content into the Clearinghouse.				
Plan & present workshop on Clearinghouse functionality and services at RDAP 2020 (Spring 2020).				

Activity	Year 1 2018	Year 2 2019	Year 3 2020	2021
Plan and hold eiththh of twelve virtual Advisory Board meetings.				
Set up user testing for addition of Assessment Framework to the Clearinghouse in Year 2.				
Plan and hold ninth of twelve virtual Advisory Board meetings.				
Iterate the implementation of the Assessment Framework to the Clearinghouse based on the user testing and feedback.				
Set up user testing for enhancements made to the Clearinghouse in Year 2.				
Hire Research Assistant for Year 3.				
Plan and hold tenth of twelve virtual Advisory Board meetings.				
Quadruple the total number of educational resources in the Clearinghouse from the pre-project baseline and expand the number of subject domains by at least 1 additional domain.				
Plan & implement online crowdsourcing competition / event to solicit & input existing data management educational content into the Clearinghouse .				
Plan and hold eleventh of twelve virtual Advisory Board meetings.				
Iterate the enhancements made to the Clearinghouse based on the user testing and feedback.				
Plan & present workshop on Clearinghouse functionality and services at ACRL 2021 (Spring 2021)				
Plan and hold final of twelve virtual Advisory Board meetings.				
Final Report submitted by June 30, 2021.				

DIGITAL PRODUCT FORM

Introduction
The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

- ☐ Please check here if you have reviewed Parts I, II, III, and IV below and you have determined that your proposal does NOT involve the creation of digital products (i.e., digital content, resources, assets, software, or datasets). You must still submit this Digital Product Form with your proposal even if you check this box, because this Digital Product Form is a Required Document.

If you ARE creating digital products, you must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

Part I: Intellectual Property Rights and Permissions

A.1 What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

The two primary digital products that will be created by the project are metadata (both initial training resource documentation and ongoing assessment data); and computer code that provides enhanced metadata enabled search and discovery of research data management training materials, and assessment tools for capturing assessment information from users and others. While intellectual property "ownership" will be retained by creators/contributors of metadata and computer code, both sets of products will be shared with clearly defined permissions based on existing standards that maximize the potential for reuse. Specifically:

- Metadata: Creative Commons Attribution license (CC-BY 4.0 - https://creativecommons.org/licenses/by/4.0/) to maximize unrestricted reuse while maintaining acknowledgement of the contributions of the initial creator of the resource.
- Computer Code: Apache 2.0 license (https://www.apache.org/licenses/LICENSE-2.0) is a permissive OSI, FSF approved, and GPL v. 3 compatible open source license that also allows for linking to external libraries that are shared using different license models.

It is not anticipated that the project will create any new data management training materials with grant funds and rights and permissions for training materials will be retained and defined by the creators of those resources. The Clearinghouse encourages the use of permissive licenses that allow for maximal reuse of resources registered in the system, but does not require this.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

As noted above, ownership rights for contributed metadata and computer code will be retained by the creators/contributors of those digital products. The use of the CC-BY 4.0 and Apache 2.0 licenses provide a standardized method of communicating rights and obligations of users of metadata and computer code, respectively. The only requirement imposed on users of the metadata and computer code is attribution (in the case of the CC-BY 4.0 license applied to metadata) and retention of the embedded copyright notice (in the case of the Apache 2.0 license applied to computer code)

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

No such issues are anticipated in reference to the digital products that will be generated by the project. It is expected that the assessment data that will collected by the project will be anonymous and will not pose any privacy concerns. Permission to integrate anonymous feedback into the metadata collection will be obtained from contributors prior to submission of their feedback.

Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

A. Creating or Collecting New Digital Content, Resources, or Assets

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

Digital metadata created by the project will be stored in an internal format based upon the current Drupal-based internal storage model used by the Clearinghouse that uses pertinent elements from the Schema.org endorsed, and Dublin Core Metadata Initiative maintained Learning Resource Metadata Initiative (LRMI) metadata schema[1]. With funding from this project, the internal format will be extended with additional metadata elements and specifications that are identified by the project metadata enhancement working group and advisory board. While the Clearinghouse will continue to use an internal metadata model to enable search and discovery, support for applicable metadata standards such as MARC21[2]/MARCXML[3], IEEE Std. 1484.13.1-2012[4], and Dublin Core[5] will also be suggested to the metadata working group and the full advisory committee.

The project has a target of quadrupling the number of research data management training resources that are registered in the system by the end of the project. This translates into the addition of approximately 300 new resource registrations including corresponding metadata content in the Clearinghouse by the end of the project.

A.2 List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

The Data Management Training Clearinghouse is currently hosted within the Drupal-based infrastructure maintained by the Federation of Earth Science Information Partners (ESIP Federation) - http://dmtclearinghouse.esipfed.org. While the ESIP Federation is in the process of considering migration to a new hosting platform, the project will continue to be hosted within the ESIP infrastructure for the duration of the project.

While Drupal remains the target host platform (as determined by ongoing use of this platform by the ESIP Federation), PHP will be the primary programming language used to develop enhanced Clearinghouse capabilities based on the expanded metadata model, and interface elements for capturing those metadata (i.e. data registration and assessment tools). Other programming languages will be used in the development of the platform as required to accommodate any platform changes that occur during the course of the project. Code development will be managed through a public GitHub repository that will be accessible to the project team and other interested potential contributors.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

While the current metadata model is based on the internal representation managed within the Drupal platform that hosts the Clearinghouse, alternative metadata representations such as MARC21[2]/MARCXML[3], IEEE Std. 1484.13.1-2012[4], and Dublin Core[5] will be considered in consultation with the project's metadata working group and advisory committee.

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

Initial metadata entry for registered training materials will be reviewed and validated prior to publication within the Clearinghouse. The implementation of the assessment framework will provide for ongoing assessment of the registered training materials both in absolute terms, but also relative to the metadata-based classification scheme assigned to the materials. This will provide ongoing quality control for the metadata maintained in the system, providing an opportunity for revisiting and revising metadata managed in the system for resources for which the assessment process identifies gaps or inconsistencies.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

At a minimum the training resource metadata content will be exported from the Clearinghouse application as structured XML files that represent the full content of the internal metadata model implemented for the platform. Additional standard formats, such as MARC21[2]/MARCXML[3], IEEE Std. 1484.13.1-2012[4], and Dublin Core[5] will also be considered for export and preservation as determined by the project's advisory committee. The exported metadata collection will be added to UNM's Digital Repository (https://digitalrepository.unm.edu) and underlying preservation system at the end of each project year for long-term preservation, discovery and access outside of the Clearinghouse application interface.

Specifically, metadata will be archived for a minimum of 10 years at the UNM Libraries' Digital Repository after the grant ends. After this time, the data will be appraised per established collection and archival management policies for transfer to an external repository, longer-term archiving, or alternative disposition. The UNM Digital Repository is an Open Archives Initiative (OAI) compliant repository, which enables Dublin Core metadata and dataset objects to be shared and harvested by other archival and discovery systems through the OAI-PHM protocol.

The UNM Digital Repository is maintained by the UNM Libraries. Archive staff will also provide daily file integrity and format verification and will create and maintain technical and administrative metadata using the widely adopted Metadata Encoding and Transmission Standard and Preservation Metadata Implementation Strategies metadata standards. These additional metadata include digital file signatures and checksums for bitwise integrity validation and chain of custody documentation. Primary responsibility for curating and preparing the data for archiving will rest with the Libraries' Data Curation Librarian.

As the developed computer code will be managed and shared within the GitHub distributed version control system, the project will use the Zenodo repository and its integration with GitHub to enable for long-term preservation, discovery and access of target software releases that are tagged within the GitHub environment.

C. Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

The primary digital resources produced by the project will be metadata structured according to the internal data model developed within the Drupal Clearinghouse hosting platform. MARC21[2]/MARCXML[3], IEEE Std. 1484.13.1-2012[4], and Dublin Core[5] will be considered as export representations that complement these standards-based metadata models. The metadata standards used to document the archival copy of the metadata collection will be the formats supported by UNM's Digital Repository and preservation system: Dublin Core, Metadata Encoding and Transmission Standard (METS)[6], and Preservation Metadata Implementation Strategies (PREMIS)[7] metadata standards.

Code developed by the project will be maintained within a public GitHub that includes a Readme.md file for overall documentation and in-code comments that document specific areas of functionality within the code. The Zenodo repository supports the DataCite[8] minimum and recommended terms for rich metadata which are exposed through the OAI-PMH standard[9].

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

The selected source metadata model and associated standards-based representations that will be considered are intended to maximize the compatibility with the target repository (the UNM Digital Repository) for resource metadata discovery, and for more rich metadata content better aligned with the documentation of learning objects, and integration of learn object collections into library catalog systems.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

The Data Management Training Clearinghouse that will be enhanced by the project provides search and browse tools for the discovery of training materials registered in the system. The proposed enhancements will expand the metadata model upon which these tools are based, and provided additional search options beyond those currently provided. The implemented assessment framework within the Clearinghouse will further both the discovery of and quality improvements to the training material content. Metadata for the training materials placed in UNM's Digital Repository and in Zenodo will also be accessible through the OAI-PMH protocol for automated harvesting and retrieval.

D. Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

The research data management training materials registered into to the Clearinghouse through the creation of enhanced metadata will be made available through the publicly accessible search and browse tools provided by the Data Management Training Clearinghouse hosted within the Federation of Earth Science Information Partners' (ESIP Federation) Drupal web platform: http://dmtclearinghouse.esipfed.org. No special software or tools are required to use the Clearinghouse except for current standards-compliant web browsers with operational internet connections.

D.2 Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

Geographic Storage, Transformation and Retrieval Engine Version 3: A data framework for data discovery, delivery and documentation (GSToRE). http://gstore.unm.edu. While Director of the Earth Data Analysis Center, and subsequently as the Director of Research Data Services in UNM's University Libraries, the project PI (Benedict) directed the design and development of the GSToRE data management, discovery and access platform in support of four NSF EPSCoR funded research projects and the New Mexico Resource Geographic Information System.

New Mexico Resource Geographic Information System (NM RGIS). http://rgis.unm.edu/getdata/. While Director of the Earth Data Analysis Center Dr. Benedict supervised the design and development of the NM RGIS system for the discovery and access of geospatial data as the designated geospatial data Clearinghouse for the state of New Mexico.

UNM Libraries Research Data Services *Code and Coffee* workshop materials.

- https://github.com/unmrds/cc-version-control
- https://github.com/unmrds/cc-pandoc
- https://github.com/unmrds/cc-jupyter
- https://github.com/unmrds/cc-python
- https://github.com/unmrds/cc-command-line

These materials have been developed and shared through a collection of public GitHub repositories upon which an expanding series of research data management and analysis workshops have been based.

Part III. Projects Developing Software

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

This project will produce enhancements to the current Data Management Training Clearinghouse hosted within the ESIP Federation's Drupal web infrastructure. The enhancements will be designed for a combination of users seeking self-service research data management training materials and users who provide research data management training to others. These users are expected to work primarily in research organizations and institutions including public and private sectors, and academic research institutions that work in multiple disciplinary contexts. More specifically, self-service learners and recipients of data management training are expected to be

college or university students (some undergraduate, more graduate), research staff and faculty, researchers in the private sector, and training providers (including a growing population of academic librarians who are providing research data services within their institutions - a specific focus population for this project) who need to expand their knowledge and skills. Trainers seeking materials are expected to be librarians, and trainers within both larger research organizations, and educational environments of underserved populations at community colleges, and smaller public/private colleges.

The intended functionality of the enhanced Clearinghouse includes:

- An internal metadata model managed within the platform that can accommodate additional and/or more specific information for:
    - Keywords applicable to specific subject domains
    - Intended audience
    - Disciplinary focus
    - Presentation method/style
    - Types of training materials
    - Assessment information
        - Assessment of completeness and accuracy of metadata within the Clearinghouse
        - Assessment of registered training materials for quality, completeness, suitability for purpose
- Enhanced metadata capture tools to support the registration of new training materials and update existing entries for previously registered materials.
- Implementation of the defined assessment framework within the portal for continuous collection of feedback from both

A.2 List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

Educational and training material clearinghouses are typically developed as websites (or components within larger web applications) that have specialized search tools for discovering materials that meet specific user needs. Because of this model, pre-packaged clearinghouse software packages are not common, but instead lower-level web application development frameworks provide the foundation for these online resources. Based upon this common model, and upon the fact that the current Data Management Training Clearinghouse is based upon this model - using the Drupal platform that is supports the full ESIP website, some alternative web application frameworks and their underlying development languages that could be used include:

- Django - a python based web development framework that supports page templates, connections to multiple back-end database platforms, and because of its python backend has a large array of functionality available through native language support
- Ruby on Rails - a ruby based framework for accelerated website development that uses defined coding conventions (as opposed to detailed back-end configuration) to develop web assets. The broad array of standard packages available in Ruby on Rails allows for the development of robust capabilities within the environment.
- Express - an open source framework that uses javascript within the Node.js server platform to develop and manage websites. Express (and Node.js) require the installation of libraries to achieve similar levels of functionality as those provided by Django and Ruby on Rails "out of the box". Express is commonly used for light- to mid-weight web applications that don't require significant server-side computation.

A significant determinant for the current selection of Drupal for the continued development and hosting of the Clearinghouse is that it is the current platform upon which it is based and its continued used will *significantly* accelerate the development of the proposed Clearinghouse enhancements. With that having been said, there are discussions within ESIP about the possible migration of ESIP's web hosting platform to an alternative system. If this occurs, the project leadership will work with the ESIP leadership to ensure a smooth migration of the Clearinghouse and its content to the new platform while maintaining the development plan for enhancing the capabilities of the Clearinghouse.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

PHP is the primary programming language of the Drupal platform, which itself is based on an industry standard LAMP (Linux, Apache, MySQL, PHP) technology stack that is mature, broadly supported by the developer community, and broadly adopted by a wide range of large organizations for hosting their web applications. Drupal is the current hosting platform for the Clearinghouse and for accelerated development of the proposed enhancements will be retained as the platform, at least in the near term. As noted above, there are discussions of potentially changing ESIP's web hosting platform to an alternative system, but at this point those discussions are in a very early phase and no specific plans have been developed yet.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

The proposed Clearinghouse enhancements will build upon the current Clearinghouse platform and therefore extend the capabilities of this existing system.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

No additional software or system dependencies are anticipated at this time.

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

Any additional Drupal modules or other software components that are developed by the project will comply with the Drupal *API documentation and coding standards* (https://www.drupal.org/docs/develop/coding-standards/api-documentation-and-comment-standards). Additional documentation in the form of README.md files will be included in the GitHub repository that will be used for software version control, issue tracking, and public access. Clearinghouse user documentation will be provided through the online help section of the Clearinghouse (http://dmtclearinghouse.esipfed.org/help), and its associated FAQ. Both will be expanded with additional information based upon user and advisory board feedback over the course of the project.

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.

Current DMT Clearinghouse - http://dmtclearinghouse.esipfed.org. The current Clearinghouse was developed in collaboration with the U.S. Geological Survey's Community for Data Integration, the ESIP Federation, and DataONE. The project Co-I (Hoebelheinrich) was a key contributor to the planning and development of the current Clearinghouse system.

New Mexico EPSCoR Metadata Entry and Dataset Registration Tools. https://github.com/edac-epscor. As cyberinfrastructure lead for the New Mexico EPSCoR project Dr. Benedict has supervised (with John Savickas) the development of multiple software components in support of metadata capture and dataset integration into the project's data management platform - GSToRE.

Geographic Storage, Transformation and Retrieval Engine Version 3: A data framework for data discovery, delivery and documentation (GSToRE). http://gstore.unm.edu. While Director of the Earth Data Analysis Center, and subsequently as the Director of Research Data Services in UNM's University Libraries, the project PI (Benedict) directed the design and development of the GSToRE data management, discovery and access platform in support of four NSF EPSCoR funded research projects and the New Mexico Resource Geographic Information System. Two versions of the platform have been developed:

- The core GSToRE platform upon which the New Mexico Resource Geographic Information System, and NM EPSCoR Data portals are based - public release still under development prior to end of current project in June 2018.
- A modified version of the platform optimized for the management and delivery of watershed-scale model data: https://github.com/VirtualWatershed/vwp-gstore

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

While intellectual property "ownership" will be retained by creators/contributors of computer code, any developed code will be shared with clearly defined permissions based on existing standards that maximize the potential for reuse. Specifically:

- Computer Code: Apache 2.0 license (https://www.apache.org/licenses/LICENSE-2.0) is a permissive OSI, FSF approved, and GPL v. 3 compatible open source license that also allows for linking to external libraries that are shared using different license models.

License terms will be included in the project's GitHub repository using the standard model for adding license information at the repository level. Additionally, the standard Apache license model will be used in which the full license is included in the top directory

of the project repository (as a LICENSE file, the same approach used by GitHub), the corresponding NOTICE file will also be included in the same directory as the LICENSE file. A short license header will also be included in each source code file within the repository.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

The developed software will be made publicly available during development through a public GitHub repository. Major software releases (including the release that marks the end of the grant) will also be archived through a connection between the public GitHub repository and a corresponding Zenodo archival copy for which a corresponding DOI will be generated.

C.3 Identify where you will deposit the source code for the software you intend to develop: The code will developed and shared through a public code repository in GitHub within the ESIP Organization (https://github.com/ESIPFed). Tagged versions of the developed software will be archived in Zenodo through their integration with GitHub.

Name of publicly accessible source code repository: TBD - a new repository will be created for the project/clearinghouse when the project is initiated.

URL: TBD - a new repository will be created for the project/clearinghouse when the project is initiated.

Part IV: Projects Creating Datasets

A.1 Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

No data will be created.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

Not applicable

A.3 Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

Not applicable

A.4 If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

Not applicable

A.5 What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

Not applicable

A.6 What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

Not applicable

A.7 What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

Not applicable

A.8 Identify where you will deposit the dataset(s):

Name of repository: not applicable

URL: not applicable

A.9 When and how frequently will you review this data management plan? How will the implementation be monitored?

The data management plan as it relates to the data management training materials registered in the Clearinghouse and developed Clearinghouse code will be reviewed on an annual basis, with monitoring of implementation to be completed through comparison of the number of metadata records added to the Clearinghouse, completeness of those records relative to required and optional metadata elements, and the number of records packaged for long-term access and preservation through the separate UNM Digital Repository and the Zenodo platform. The PI, Karl Benedict, or designated successor if project leadership changes, will be responsible for monitoring compliance of the project with the provided data management plan.

References Cited:

1. Association of Educational Publishers & Creative Commons. LRMI 1.1. *Learning Resource Metadata Initiative* (2014). Available at: http://lrmi.dublincore.net/lrmi-1-1/. (Accessed: 15th January 2018)
2. MARC 21 Format for Bibliographic Data: Table of Contents (Network Development and MARC Standards Office, Library of Congress). Available at: https://www.loc.gov/marc/bibliographic/. (Accessed: 15th January 2018)
3. MARC 21 XML Schema. Available at: http://www.loc.gov/standards/marcxml/. (Accessed: 15th January 2018)
4. IEEE 1484.13.1-2012 - IEEE Standard for Learning Technology -- Conceptual Model for Resource Aggregation for Learning, Education, and Training. Available at: https://standards.ieee.org/findstds/standard/1484.13.1-2012.html. (Accessed: 15th January 2018)
5. DCMI: Home. Available at: http://dublincore.org/. (Accessed: 15th January 2018)
6. Metadata Encoding and Transmission Standard (METS) Official Web Site | Library of Congress. Available at: http://www.loc.gov/standards/mets/. (Accessed: 16th January 2018)
7. PREMIS: Preservation Metadata Maintenance Activity (Library of Congress). Available at: https://www.loc.gov/standards/premis/. (Accessed: 16th January 2018)
8. DataCite Schema. *DataCite Schema* Available at: https://schema.datacite.org/. (Accessed: 16th January 2018)
9. Open Archives Initiative Protocol for Metadata Harvesting. Available at: https://www.openarchives.org/pmh/. (Accessed: 16th January 2018)