Planning for Open Grants: Fostering a Transparent and Accessible National Research Proposal Infrastructure

The George A. Smathers Libraries (Libraries) at the University of Florida (UF) seeks $76,608 over 18 months to explore the scale of challenges and solutions in establishing an open grants repository. Although open methods have increased transparency in many aspects of research and practice, such as publications (e.g., open access and preprints), statistical analyses (e.g., code-sharing practices and platforms), and materials (e.g., data repositories, reporting standards), grant and fellowship applications represent a substantial investment of time and effort by multiple collaborators that remains largely opaque. This planning project will: (1) convene diverse groups of stakeholders so that a multifaceted research ecosystem is represented throughout the planning process; (2) develop a metadata schema and aggregation plan for organizing grant proposals and related documents (e.g., funder program guidelines); and (3) investigate the social and technical challenges and opportunities for creating and maintaining an open grants repository.

A. Statement of National Need

Navigating the process to secure funding and fellowships—from project ideation to submission to award—requires copious amounts of time and energy from an interconnected group of co-authors, reviewers, and grants professionals (Herbert et al. 2013). However, most of the resulting materials are not accessible or even visible to those outside the grants-making process, much less the general public. Consequently, this important piece of the research process remains opaque, hiding these materials from analysis and acknowledgement.

This inaccessibility is an obstacle for the beneficial uses of such materials, whether as examples to help guide proposal writing, or as scholarly objects documenting the questions, methods, sources, and labor that shape a research agenda or program development over time (Brennan, 2012; Collins, 2019). Although some funders and applicants post full proposals to websites or scholarly repositories, this practice is haphazard and varies depending on individual and institutional norms. Grant documents are also shared through informal networks, but this may end up reinforcing inequities through differences in awareness of and access to these networks.

An open repository of funding proposals will elevate their recognition as scholarly products, improve access for the public and other grant seekers, and bring transparency to this facet of the research process. The project frames "research" broadly; while the planning process will focus on grants and fellowships in an academic context, this includes work rooted in program or service development and evaluation alongside empirical or interpretive scholarship. The exposure of these assets has the potential to facilitate building collaborations, asset-mapping, and community development (de Farber, 2016). This concept challenges conventional practice, with researcher hesitancy interlacing both historical inertia and concerns about sharing sensitive information and being scooped. These mirror concerns about sharing data and publications. Nevertheless, shared community needs and technological infrastructure (e.g., data repositories, open access publishing platforms, and preprint servers) have enabled and shifted both practices and attitudes, and features such as controlled permissions and embargoes have helped address concerns about security, privacy, and ownership.

This complex situation merits an intentional exploration into the benefits of implementing an open repository, and the issues associated with materials' ingestion, preservation, sharing, usage, and more. This project will build upon other efforts that have established collections, including Grants@UF, which is led by project advisor Bess de Farber and shares ~400 proposals and other related grant materials voluntarily submitted by principal investigators (PIs) to the Libraries; and Open Grants, a cross-disciplinary portal co-developed by PI Hao Ye, containing ~200 proposals. IMLS funding will enable engagement with a diverse cadre of advisors, including domain experts, librarians, grants professionals, technologists, and others to ensure that future developments in this area have a robust roadmap to meet the needs of stakeholders in the pursuit of grant seeking endeavors.

Potential Use Cases

Professional development and education: Although funders often provide guidelines to accompany funding opportunities, these documents generally focus on the formatting and structure of applications, and the issues that an application should address. Providing example applications in addition to these documents will promote professional development and education for would-be applicants, just as worked examples in a textbook demonstrate how to apply the content to solve a problem. For researchers applying to funding opportunities, full grant applications can be invaluable as a planning tool—by their very nature, grant proposals are notably distinct from other forms of academic writing (Porter, 2007). This need is especially acute for early career researchers, such as graduate students, and applicants who have limited or no prior experience with grant proposals and can benefit the most from access to successful examples. Although some collections exist, full proposals are generally shared through informal networks, which may increase social inequities in awareness and access to expertise (Hu, 2019).

Landscape review and environmental scan: Grant proposals occupy a singular niche within academic writing; across many dimensions, grant proposal documents differ from conventional research publications (Porter, 2007). For example, proposals focus on future work rather than completed work, and must persuade the reader that the project is worth funding. They capture social signals related to the people involved in ideation, rationalization, justification, evidence gathering, and research design. From this perspective, a collection of grant proposals has unique value that does not exist in archives of other materials and has the potential to facilitate new modes of academic interaction and uses for research collections. For researchers interested in performing landscape reviews or environmental scans, grant proposals are a valuable and complementary resource to existing sources. The ability to understand the broader conceptual landscape in which a proposal is situated can assist users of a grant proposal database to better identify complementary work or even identify and engage with potential collaborators. As resources for grant writers, libraries would play a role in promoting awareness of a repository for this use.

History of the field(s): Like many publications, grant proposals include historical context for a specific line of inquiry. However, where most academic papers will focus on the results of a single experiment or exploration of a single question, grant proposals often describe multiple associated projects or questions related by theory, audience, or methodology. Studying and comparing the questions proposed in a grant proposal and the subsequent publications or project deliverables can produce unique insights into team decision-making and provide context to published results that are likely biased towards successful outcomes. This expanded perspective improves efforts to trace the history of a field and to map how ideas and theories evolve as hypotheses, concepts, and prototypes are generated, tested, and discarded.

Scholarly and professional network analysis: Bibliometrics uses the corpus of publications to investigate topics such as trends in research, the publication process, and authorship and citation networks. In this regard, the inaccessibility of grant proposals as scholarly works is an artificial constraint for studies. A comprehensive source of grant proposals will facilitate the study of existing topics in bibliometrics and open new avenues of research specific to the content and context of grant proposals such as total funding amounts and funding rates for different fields over time, and citation patterns within grant proposals in contrast to research articles.

Previous and ongoing work in this area

This project builds upon other efforts that have established collections, including Grants@UF, which is led by team member Bess de Farber and shares ~400 proposals voluntarily submitted by PIs to the Smathers Libraries; and Open Grants, a cross-disciplinary portal co-developed by PI Hao Ye containing ~200 proposals. Funders also engage in sharing sample applications. For example, the National Institute of Allergy and Infectious Diseases shares several examples for multiple types of common NIH research and training grants. IMLS itself shares the full narrative along with other documents for recently awarded grants in several of its programs through the Awarded Grants Search portal. In addition to those collections created through formal projects or institutional efforts, many grant proposal documents are shared via informal means, with the common purpose

of assisting researchers who are developing their own proposals. Some of the more visible efforts include GitHub repositories maintained by Jeffrey Ross-Ibarra and Yaniv Brandvain.

This listing is by no means complete. The work plan of this project includes developing an environmental scan to more systematically examine how peripheral research products and materials are collected and shared, including not only grant proposals but also datasets, data management plans, position descriptions, and gray literature. Indeed, because of their peripheral nature, such sources offer valuable lessons on the project design of workflows and standards, implementation, and community engagement. Mapping out the similarities and differences of available collections or examples will ensure that future efforts to systematically share grant proposals are better poised for success.

There is a clear utility for sharing grant proposals. Unfortunately, the existing ways in which these materials are shared is haphazard, and thus only partially fulfills the need if the searcher is knowledgeable about their availability, location, or options for using the Freedom of Information to request specific proposals submitted to federal funders. In addition to most materials remaining unshared, the lack of a searchable database is a direct obstacle for use cases which depend on systematic exploration. Finally, discrepancies in awareness of or access to these resources where they exist can also amplify existing inequities in the academic system—this is especially acute in early-career and/or first time grant applicants for whom the benefits of example grant proposals to support training and professional development are essential.

B. Project Design

Outcomes & Deliverables: Major project outcomes will not only establish a blueprint for moving forward with a potential implementation phase, but also contribute to a larger body of knowledge on the significance of grant proposals. By incorporating perspectives from funders, research administrators, librarians, and scholars, the project will document a complex funding ecosystem and identify where specific barriers to sharing might be lifted.

The central question that will guide this process is not "how can we make every grant proposal openly available?", but rather "how can we proceed strategically to facilitate sharing where it will have the most benefit?" This means assessing feasibility and cost-benefit across disciplines, funder types, and a range of different stakeholders to produce recommendations for approaches and solutions that will be most appropriate and impactful.

Deliverables (described in detail below) include:

1. **Field Report:** This document will include a detailed environmental scan and analysis of individual interviews and meeting proceedings.
2. **Ethical Engagement Plan:** This document will extend and complement functional requirements by detailing plans for long-term community outreach and ways the implemented repository will address ethical issues such as privacy.
3. **Functional Requirements:** This document will define the technologies and standards necessary to implementation, including policies, metadata and data model, interface design, and user stories.
4. **Preliminary Data:** Throughout the planning process, team members will survey and analyze available proposal metadata and digital objects as well as methods for collection. Any data collected during this phase will be shared openly.

Community Engagement: Funding will enable deep, sustained engagement with individuals and organizations over the course of the award period, including 10 in-person or virtual meetings, 15 one-on-one interviews, and attendance at three national conferences. As described in Phases 2 and 3 of the Work Plan, as well as the Diversity Plan, advisors and other planning participants will bring their unique perspectives to bear on the grants lifecycle, benefits of sharing proposals, and concerns related to open sharing. Of the virtual meetings, 4-6 will focus on the needs of grant applicants and funders in specific domains, as one way to foster communities of interest within a complex, multidisciplinary undertaking. As a guiding principle, the project team will ask advisors and interviewees how access to grant proposals could make each of these areas more equitable, and

how project design and implementation of a proposed repository can avoid replicating current biases and exclusive practices.

Evaluation and Assessment: While the ultimate goal of the project is to develop a foundation for moving forward with implementation, the project team seeks to facilitate conversations and develop resources that are valuable in and of themselves. This work should represent a snapshot of current attitudes, challenges, and possibilities for next steps. Because engagement with multiple communities is crucial to success, the project team will solicit feedback following the in-person meeting, as well as each virtual discussion. As described below, participant and broader public comments will be incorporated into each phase of the project and refinement of deliverables. The project team will rely heavily on this iterative evaluation process to ensure that by the end of the award period, final products are well-conceived and establish meaningful guidance for future implementation.

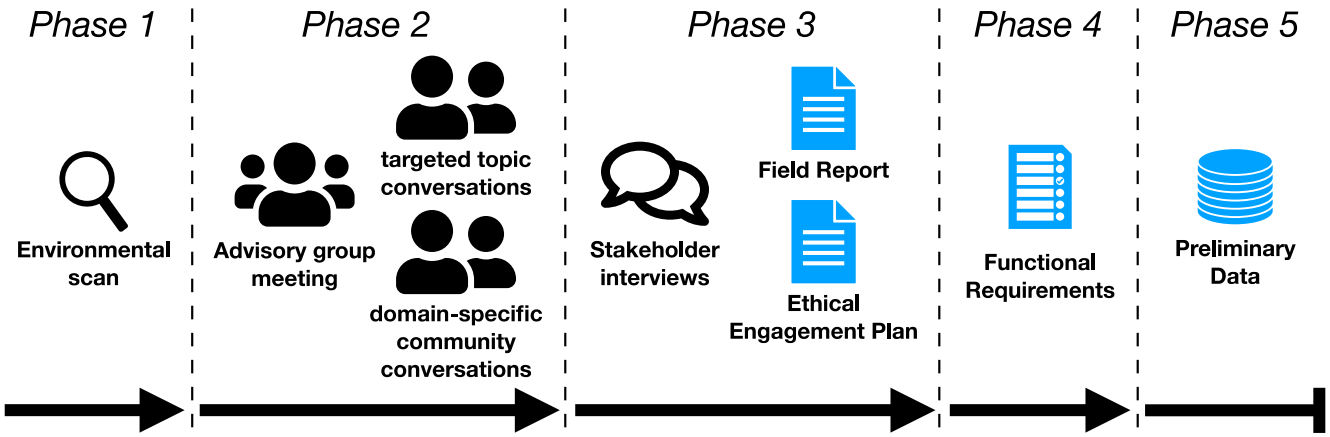

Figure 1 Project activities and deliverables.
Activities (black) and deliverables (blue) for this planning project are divided into 5 phases. Phase 1 consists of the environmental scan. Phase 2 involves the advisory group meeting and focused conversations. Phase 3 includes the stakeholder interviews and writeup for the Field Report and Ethical Engagement Plan deliverables. Phase 4 consists of developing the Functional Requirements deliverable, and Phase 5 yields the Preliminary Data deliverable. The temporal sequence of activities from left to right is approximate, with work on multiple activities and deliverables occurring simultaneously (see Schedule of Completion for details). Some icons (group meeting, sub-group meeting, and interviews) are provided by Font Awesome Free under CC BY 4.0.

Work Plan

Phase 1: Environmental scan and meeting coordination (Aug.-Dec. 2021)

Environmental Scan: As a critical part of the final field report, an environmental scan will build on the project team's existing knowledge of available literature, models, datasets, and prototypes relevant to the project. This component will jumpstart community engagement and technical planning for a future repository; it will stand alone as a contribution to the collective understanding of how proposals are organized and shared. While the PIs will be primarily responsible for this work, the project team and advisory committee will contribute by acting as an informal knowledge base to provide a common frame of reference and foundation for subsequent phases.

The scan will focus broadly on efforts to document and collect scholarly or professional outputs, especially those less likely to be formally published or receive professional recognition. These will include any initiatives dedicated to gathering grant proposals, as well as disciplinary or institutional repositories that already collect grant proposals and/or associated metadata and other related materials. Based on advisors' expertise and involvement in previous IMLS-supported work, the scan will also extend to other initiatives to aggregate peripheral scholarly outputs, such as datasets (Mannheimer), data management plan trainings (Benedict), job descriptions (Keith), and gray literature (Smith). To better document the wide range of individual practice, we will include an analysis of 25-50 example proposals shared on project or department websites.

For every example, we will annotate its scope and purpose, including guiding principles; workflows for data ingest (e.g., manual upload, API) and data sharing; descriptive and technical metadata standards, including rights and license metadata; efforts to engage and sustain user communities; roles and responsibilities of project team; institutional support and financial sustainability; and challenges or reasons for obsolescence.

The scan will feature available information about funder infrastructure, including funder guidelines, information about the application process, publicly available data about funded projects, and any processes in place to obtain sample proposals (e.g., Freedom of Information Act requests for federal agencies). We have confirmed several foundation representatives as expert advisors, and have initiated preliminary conversations on this work with funders such as the National Endowment for the Humanities.

Meeting Planning and Coordination: Based on CDC guidelines, the project team will consult with advisors to determine an in-person meeting date, with a tentative goal of February or March 2022. Available meeting rooms at the University of Florida Smathers Libraries will allow for ample social distancing with 30-40 people, with outdoor spaces to facilitate smaller breakouts. If necessary, the schedule allows sufficient flexibility to hold the meeting later in 2022.

Phase 2: Advisory meetings and community conversations (Jan.-May 2022)

External Advisory Committee Meetings: The in-person advisory committee meeting (see draft schedule, Appendix B) and follow-up remote discussions will provide opportunities to surface possibilities and methods for sharing grant proposals, as well as obstacles and challenges to implementation. Facilitated by the PIs, advisors will share their respective knowledge and experience around four key areas: (1) funding workflows and technical infrastructure; (2) community engagement in open access; (3) academic peer review and reward systems; and (4) collaborative digital library initiatives.

Each advisor (13 confirmed, 7 to be selected through an open call) agrees to:

1. Attend a 1.5 day, in-person workshop in early 2022 (with contingency plans to reschedule as necessary for mid-2022 depending on CDC guidelines).
2. Based on interest and expertise, participate in one or more additional targeted conversations via Zoom. These will explore intricacies such as community engagement to drive usage and uptake, ethical workflows for collecting, storing, and sharing materials, and defining potential use cases.
3. Where appropriate, participate in domain-specific conversations described below.
4. Review and comment on deliverables at strategic points throughout the award period.

Domain-Specific Community Conversations: To widen discussion beyond project team and advisors, remote facilitated discussions will engage with targeted, discretely-scoped communities in three topic areas and document case studies across the project deliverables. Advisory committee members will promote these opportunities throughout their networks, with a goal of bringing together about 25 people for each discussion.

Libraries and archives, especially in academic settings: Funders such as the Council on Library and Information Resources (CLIR) and IMLS have made hundreds of successful proposals openly available within the last five years. These documents offer insights into project management, technical standards, and community engagement strategies, with examples relevant to a range of audiences. Discussions with grant seekers in libraries and archives will explore how sharing proposals might encourage cross-institutional partnerships in areas of common interest such as digitization, information literacy, etc. Many researchers in these fields are also analyzing scholarly objects at scale; one conversation will be devoted to defining use cases for applying such methods to large collections of funding proposals. As a large research library, the project team's institution has invested in grants infrastructure under the leadership of Grants Manager Bess de Farber, whose team manages a public collection of proposals including fellowship applications and training materials available beyond UF. Other committed advisors include Elaina Norlin, ASERL Professional Development/DEI Program Coordinator and Sara Mannheimer, Data Librarian at Montana State University.

Graduate and postdoctoral fellowships, especially in STEM disciplines: Graduate and postdoctoral fellowship programs receive many applications (NSF awards ~ 2,000 graduate research fellowships a year),

representing a large pool of diverse proposals addressing the same call. Targeted discussions will focus on a number of facets related to sharing materials specific to these funding opportunities. For example, because these applications span a broad array of topics, an effective metadata schema and mechanisms to identify subject area and keywords will be crucial to making this resource useful for a wide array of users. Moreover, applications come from an even more diverse collection of individuals and schools (the NSF graduate research fellowship program accepts applications from both undergraduate students and graduate students in their first two years of study); thus the task of ingesting materials is complex. Finally, conversations will include challenges and opportunities stemming from motivations for sharing—applicants to these fellowship programs may be more willing to share materials because the proposed work is more exploratory relative to grant proposals that require a longer-term investment of effort and may contain confidential details.

Caribbean studies, especially in humanities and social science disciplines: As the technical host and co-founder of the Digital Library of the Caribbean (dLOC), the project team's institution has fostered an extensive and rich network in this field, guided by core principles of mutual support. DLOC itself offers one model for post-custodial, digital archiving and aggregation, with an active group of over 70 partners who share collections as well as expertise. We will encourage these partners and other Caribbean studies researchers and practitioners to consider (1) how sharing proposals might amplify and promote acknowledgement of grant-funded work across the Caribbean Diaspora; and (2) how such an effort might support less-resourced institutions—especially those based in the Caribbean—by providing examples and context for funding opportunities. Confirmed advisors include Mirerza Gonzáles Vélez and Nadjah Ríos Villarini, leaders in public humanities and digital scholarship at the University of Puerto Rico.

These communities are themselves capacious, each representing an intersection of multiple disciplines, scholarly and professional roles, and institutional types. While we will focus primarily on stakeholders within the United States during the planning phase, virtual conversations will allow us to engage international perspectives that enhance our understanding of the global research landscape. By focusing on these as case studies for a final roadmap, we will broaden the stakeholder community beyond the "usual suspects" of open access initiatives and lay a foundation for a planned, sustainable, and shared infrastructure.

Phase 3: Stakeholder interviews and meeting synthesis (June-August 2022)

Stakeholder Interviews: With input from external advisors, the project team will invite 15 individuals to participate in-depth, one-hour interviews (draft protocol in Appendix C). Documentation and analysis of these conversations will help the field better understand stakeholder perspectives and will offer more granular feedback to identify starting points for future implementation. For instance, what information in a sample proposal does a cultural heritage professional or scientific researcher seek out first? What questions are funders most interested in asking about the relationships between their funded projects and those supported by other agencies or foundations? Which pieces of information do various stakeholders consider sensitive or private? Responses to these and other questions will offer valuable insights to complement group discussion enabling the project team to develop more authentic use cases rooted in actual experience.

Deliverables: Field Report and Ethical Engagement Plan: The field report will include (1) the environmental scan, primarily completed in Phase 1; and (2) synthesized analysis of all group meetings and one-on-one interviews. The ethical engagement plan will turn these findings into actions, building on the research and discussions documented in the field report to develop concrete policies and workflows. This plan will establish parameters for any content that should not be included in the repository, such as private or culturally sensitive data; principles for ethical reuse of shared proposals (e.g., appropriate attribution); and steps to ensure an inclusive, community-driven approach to implementation. By completing these deliverables in Summer 2022, we will have ample opportunity to solicit wide community feedback, drawing on models of open review implemented for reports such as ACRL's "Open and Equitable Scholarly Communications: Creating a More Inclusive Future" (2019).

Phase 4: Functional requirements and community feedback (Mar.-Nov. 2022)

Technical planning will take an iterative approach, drawing on the environmental scan and advisory meetings to develop a detailed plan for implementation. Follow-up virtual meetings will ask subsets of the advisory committee to review specific technical components, and in Fall 2022 plans will be made openly available for feedback and collaborative online review before the final version is deposited in the UF institutional repository. Deliverable: Functional Requirements: Crucial to future implementation, refining functional requirements will require contributions from all members of the core project team, especially the co-PIs as well as Scholarly Publishing and Repository Librarian Chelsea Johnston, Metadata Librarian Xiaoli Ma, Application Developer Chris Nicolich, and Data Management Librarian Plato Smith. This document will include:

1. **Repository scope and mission statement:** What is the purpose of the repository? What content will it include, and what might it exclude?
2. **Metadata application profile:** How will proposals be described to ensure long-term discoverability and interoperability with other relevant collections?
3. **Ingest processes:** In which ways (e.g., manual upload, API) may proposals and their metadata be added to the repository? Where will it be feasible to collect both digital objects and associated metadata, and where will technical or rights-related barriers make it necessary to collect only metadata records?
4. **UI and database design:** What user stories document anticipated interactions between the platform and those sharing or accessing proposals? How will the database and user interface facilitate discovery of material that is most relevant to a particular discipline? How will users download or export data and metadata for computational analysis, following FAIR data principles (Wilkinson et al. 2016)?
5. **Sustainability plan:** Which institution(s) have capacity and expertise to steward repository development, and what are the likely costs and staffing needs? What specific roles and responsibilities will maintainers play during and beyond implementation? How will content be stored or replicated in systems best suited to long-term preservation?

Phase 5: Dissemination and data sharing (Dec. 2021-Jan. 2023)

Conference Outreach: The co-PIs plan to attend three conferences during the grant period to build engagement and obtain feedback from specific communities. Participation in the National Council of University Research Administrators (NCURA) will ensure the project team networks with institutional grants experts, with knowledge of researcher goals, technical systems, and potential obstacles. Caribbean Digital will offer an opportunity to reach scholars, librarians, and other domain experts with an interest in furthering online access to digital resources in Caribbean Studies. Digital Library Federation Forum will focus on outreach to librarians and technologists—including potential contributors to the repository as well as those experienced in developing open access infrastructure themselves.

Deliverable: Preliminary Data: This planning effort—particularly the environmental scan and technical planning components—will generate data of wide interest and potential reuse. This includes metadata records and full proposals where copyright and licensing permissions allow; grant program guidelines from a selection of foundations and federal/state agencies; and structured data documenting aspects of existing repositories and related initiatives. Data will be shared under a CC0 (equivalent to public domain) license wherever possible, with full proposals shared under Creative Commons or other licenses as indicated by the author.

University of Florida Project Team

Hao Ye, PhD, (UF) Reproducibility Librarian, former Moore Data Fellow with expertise in open source software, and data aggregation and distribution, including over 200 grant proposals in the Open Grants project. **Role:** serve as project PI, leading project team members at UF, collaborating with Collins to complete project deliverables, and supervising the graduate assistant.

Perry Collins, MLIS, (UF), Copyright and OER Librarian, former NEH program officer in Digital Humanities with expertise in intellectual property, repository development, and grants administration. **Role:** serve as project Co-PI, leading project team members, advisory groups and co-supervising the graduate assistant.

Bess de Farber, MNM, (UF) Research Administration Manager for Libraries, former program officer in Arts and Culture/Social Services/Human and Race Relations, grantsmanship books author who strategizes with teams from application inception until completion of awarded projects, and creator of the UF institutional repository collection containing over 400 grant proposals and a variety of grants training materials from the Libraries. **Role:** Advisor and Strategic Planning Expert

Chelsea Johnston, MLIS, (UF), Scholarly Publishing and Repository Librarian, manager of publishing initiatives within the UF Smathers Libraries, leader in libraries-based programs supporting open-access publishing programs, with expertise in institutional repositories, spanning multiple media formats. **Role:** Advisor and Technical Expert (repository processes)

Brian Keith, MBA/MLIS, (UF) Associate Dean Administrative Services and Faculty Affairs for Libraries, senior administrator in the UF Smathers Libraries and PI of the Association of Research Libraries Position Description Bank, a national digital repository containing over 3200 PDs and supporting a community of academic library leaders and other researchers. **Role:** Advisor and Strategic Planning Expert

Xiaoli Ma, MA/MSI, (UF) Metadata Librarian, devising and implementing strategies to enhance metadata quality and workflows for the University of Florida Digital Collections, with expertise in metadata analysis and mapping. **Role:** Advisor and Technical Expert (metadata schema)

Chris Nicolich, MS (UF) Application Developer, lead programmer and manager for the UF Smathers Libraries software development team, experience in both database and web application front-end development. **Role:** Advisor and Technical Expert (software development and implementation)

Plato Smith, PhD, (UF) Data Management Librarian, developing university-wide policy, service, and infrastructure to maximize the utility of research data, with experience in data curation, data management, restricted data, and high-performance computing. **Role:** Advisor and Technical Expert (data management)

Graduate Student, to be hired (see position description in Appendix D) **Role**: Coordinate in-person and virtual meetings, co-develop project deliverables alongside project team

External Advisors and Planning Participants

(The following are confirmed external advisors, with 7 additional advisors to be recruited via an open call.)

Miguel Asencio, MS, (FIU), Director of the Digital Library of the Caribbean **Role:** advise on feasibility aspects related to humanities and social science research areas; support aggregation of materials associated with the Digital Library of the Caribbean for the preliminary dataset.

Karl Benedict, PhD, (UNM) Director, Research Data/Services Information Technology Services for College of University Libraries & Learning Sciences **Role:** advise on environmental scan and technical planning for the scope and functional requirements report.

Mirerza Gonzáles Vélez, PhD (University of Puerto Rico, Rio Piedras), Associate Dean of Academic Affairs, College of Humanities, PI on multiple grants to promote and steward community-based knowledge in the Caribbean. **Role:** advise on outreach and engagement; ethical approaches to digital preservation and stewardship; needs of Caribbean scholars and institutions.

Beth Hodges, MSW, (FSU) Director, Office of Research Development, established this new FSU office to facilitate collaboration, service to faculty preparing fundable proposals, especially those for whom English is a second language. **Role:** advise on feasibility aspects of materials ingestion and user update, as potentially facilitated by a university office of research; provide feedback on metadata schema for the preliminary dataset.

Emily Lescak, PhD, (Code for Science and Society) Event Fund Manager, former fisheries scientist, now leading a grantmaking program in research-driven open data science, with extensive experience in program management, communicating and engaging with diverse stakeholders, and cultivating communities around new initiatives. **Role:** advise on outreach and engagement, user adoption mechanisms; support community engagement with non-profit funding organizations.

Amalia Levi, MLIS, (HeritEdge Connection, Inc) Founder and Chair, archivist and with extensive experience leading grants to promote access to Caribbean collections. **Role:** advise on feasibility and community engagement of Caribbean scholars and cultural heritage professionals from an international perspective.

Sara Mannheimer, MSIS, (Montana State University) Data Librarian, co-PI on IMLS-funded work to aggregate and unify access to research datasets. **Role:** advise on environmental scan and technical planning, especially ingest methods; support community engagement in academic libraries.

David Mellor, PhD, (Center for Open Science) Director of Policy, former project manager and biological scientist, specializing in open science practices in publication, funding, and hiring; education, and community management. **Role:** advise on shifting cultural norms in open science practices, via technological and social mechanisms; provide feedback on long-term sustainability.

Elaina Norlin, MLIS, (Association of Southeastern Research Libraries), Professional Development/Diversity, Equity and Inclusion Trainer, Coordinator and Consultant, former IMLS program officer and director of the African-American Research Library and Cultural Center in Broward County, brings expertise in funding practices, organizational development, and DEI assessment and support. **Role:** advise on needs of grant applicants in libraries and archives; provide feedback on ethical engagement plan.

Jason Rhody, PhD, (Social Science Research Council) Program Director, co-lead of HumetricsHSS, focused on values-based approaches to scholarly evaluation, expertise in collaborative funding initiatives. **Role:** advise on feasibility in humanities and social sciences; leveraging funder infrastructure.

Nadjah Ríos Villarini, PhD, (University of Puerto Rico, Rio Piedras) Associate Professor for the College of General Studies, PI on multiple grants to promote and steward community-based knowledge in the Caribbean. **Role:** advise on outreach and engagement; ethical approaches to digital preservation and stewardship; needs of Caribbean scholars and institutions.

Elizabeth Vu, PhD, (Alfred P. Sloan Foundation) Program Associate, former non-profit co-founder and biological oceanographer, with expertise in project coordination, and technological innovations for creating and disseminating scientific knowledge. **Role:** advise on technological and infrastructural needs for aggregating and sharing materials; provide funder perspective on sustainability.

Christa Williford, MLIS, (Council on Library and Information Resources) Director of Research and Assessment. **Role:** advise on needs of grant applicants in libraries and archives; provide feedback on ethical engagement plan and functional requirements.

C. Diversity Plan

Project team members will intentionally further values of diversity, equity, and inclusion through concrete action, with a focus on four major areas:

Participation and representation: Core team members and confirmed paid advisors in the planning process include a range of stakeholders whose identities and perspectives intersect across various axes of diversity and lived experiences, including racial, ethnic, and gender identities, as well as geographical location, career stage, and research domain. A call for participation (see draft in Appendix E) will be released early in the grant period to fill another seven paid advisor roles, with an explicit encouragement for applications from Black, Indigenous, and People of Color (BIPOC), first-generation graduate students, and students and faculty from Historically Black Colleges and Universities and Minority-Serving Institutions. Honoraria and travel costs have been budgeted to support full participation of all external advisors.

Meeting design and facilitation: Meetings and technical development will adhere to standards of accessibility and universal design, such as those developed by the W3C Web Accessibility Initiative. We will implement automated live-captioning services in virtual meetings. Team members will review training materials (Anti-Oppression Resource and Training Alliance, 2017, Woodley et al., 2020) for facilitating inclusive meetings and unconferences and apply them to ensure all discussion participants may contribute their voices and perspectives.

Outcomes and impact: Questions to be addressed during each phase of the project will include exploration of the ways in which the implementation plan can break through barriers commonly encountered in academic

systems, including race, ethnicity, immigration status, levels of experience interacting with grant materials, and language acquisition. What issues have arisen for those in academic fields that require infusion of grant funds to support research? How do grant guidelines and financial information forms and information impact faculty and students who are unfamiliar with US accounting systems, government sponsors, and policy language? How does a lack of access to grant proposals impact higher education students and faculty, and BIPOC students and faculty, especially? How does a lack of access to grant proposals and guidelines impact the development of professionals in academic environments? We will prioritize and directly respond to these and related questions through all deliverables, focusing on the field report and ethical engagement plan.

Graduate student development: Graduate student support will follow the Libraries' established internship model, including attention to professional development and meaningful integration into the project team.

D. National Impact

This planning endeavor will address strategic goals defined in the IMLS Strategic Plan, including Goal 2: Build Capacity, and Goal 3: Increase Public Access. More specifically, the project team will intentionally address goals of the National Digital Infrastructures and Initiatives category, as defined in the program guidelines:

Leveraging intersections between digital challenges in libraries and the work of experts in other fields to advance theory and practice: Defined by its attention to sustained collaboration, this planning project focuses on learning from one another and documenting the relevant assets, perspectives, challenges, and experiences of all participants. As grant applicants themselves with responsibility for digital stewardship and access, libraries and archives play an active and crucial role in the funding ecosystem and are well-positioned with institutional knowledge and capacity to preserve, describe, and disseminate research objects. However, the lifecycle of even the simplest proposal intersects with many different people, passing from the authors, to disciplinary experts, to institutional research administrators, to funding bodies. By gathering these stakeholders and facilitating discussion, we will better understand how experts across different fields will make use of an open grants repository, and the ways in which steps in the writing, submission, review, and award processes might enable or hamper broad sharing.

Assessing barriers to the adoption of tools and services, developing realistic and scalable solutions to those barriers, and increasing the accessibility of content and collections to a wide range of users through effective communications: A successful planning phase will produce deliverables that stand on their own as contributions to the field, using the funding landscape as a lens onto questions around open access and collaborative research. How have other repository leaders positioned their platforms not only as containers for digital objects, but also as anchors of community engagement? What incentives or disincentives do grant applicants have to share proposals now, and how might we leverage complementary initiatives such as the San Francisco Declaration on Research Assessment and HuMetricsHSS that are innovating more equitable approaches to research evaluation and credit? Although shifting academic norms is a long-term, challenging process, our aim is to complement these emerging frameworks with a concrete plan for infrastructure that embodies pathways for individuals and institutions to share, beginning with model case studies in domains that embrace open knowledge exchange.

Enhancing the sustainability, interoperability, and accessibility of digital content and collections to provide long-term value for diverse and evolving user communities: Sustainability requires an engaged set of communities whose diverse perspectives complement institutional capacity, a mindset that UF already models with its partners as the technical host for the Digital Library of the Caribbean and the Association of Research Libraries Position Description Bank. These projects, among others to be surfaced as part of the environmental scan, are inherently stronger because they draw content from a wide range of contributors. Likewise, these projects face unique interoperability challenges that we will similarly encounter as we develop and test workflows for ingesting, describing, and disseminating content from numerous sources. Our project's approach uses the planning stage as a crucial moment to define and map community needs and to prioritize potential starting points for future implementation.

Schedule of Completion

		2021					2022												2023
Phase	**Activity**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**
1	Hire graduate assistant																		
1	Conduct environmental scan																		
1	Plan advisory group meeting																		
2	Host advisory group meeting																		
2	Conduct targeted topic conversations																		
2	Conduct domain-specific community conversations																		
3	Finalize interview questions and protocol																		
3	Conduct stakeholder interviews																		
3	**-- deliverable: Field Report --**																		
3	**-- deliverable: Ethical Engagement Plan --**																		
4	**-- deliverable: Functional Requirements --**																		
5	**-- deliverable: Preliminary Data --**																		

DIGITAL PRODUCT FORM

INTRODUCTION

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to digital products that are created using federal funds. This includes (1) digitized and born-digital content, resources, or assets; (2) software; and (3) research data (see below for more specific examples). Excluded are preliminary analyses, drafts of papers, plans for future research, peer-review assessments, and communications with colleagues.

The digital products you create with IMLS funding require effective stewardship to protect and enhance their value, and they should be freely and readily available for use and reuse by libraries, archives, museums, and the public. 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

If you propose to create digital products in the course of your IMLS-funded project, you must first provide answers to the questions in **SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS.** Then consider which of the following types of digital products you will create in your project, and complete each section of the form that is applicable.

SECTION II: DIGITAL CONTENT, RESOURCES, OR ASSETS

Complete this section if your project will create digital content, resources, or assets. These include both digitized and born-digital products created by individuals, project teams, or through community gatherings during your project. Examples include, but are not limited to, still images, audio files, moving images, microfilm, object inventories, object catalogs, artworks, books, posters, curricula, field books, maps, notebooks, scientific labels, metadata schema, charts, tables, drawings, workflows, and teacher toolkits. Your project may involve making these materials available through public or access-controlled websites, kiosks, or live or recorded programs.

SECTION III: SOFTWARE

Complete this section if your project will create software, including any source code, algorithms, applications, and digital tools plus the accompanying documentation created by you during your project.

SECTION IV: RESEARCH DATA

Complete this section if your project will create research data, including recorded factual information and supporting documentation, commonly accepted as relevant to validating research findings and to supporting scholarly publications.

SECTION I: INTELLECTUAL PROPERTY RIGHTS AND PERMISSIONS

A.1 We expect applicants seeking federal funds for developing or creating digital products to release these files under open-source licenses to maximize access and promote reuse. What will be the intellectual property status of the digital products (i.e., digital content, resources, or assets; software; research data) you intend to create? What ownership rights will your organization assert over the files you intend to create, and what conditions will you impose on their access and use? Who will hold the copyright(s)? Explain and justify your licensing selections. Identify and explain the license under which you will release the files (e.g., a non-restrictive license such as BSD, GNU, MIT, Creative Commons licenses; RightsStatements.org statements). 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.

Our overarching goal is to have the results of this project be as widely used as possible. For that reason, and recognizing that various organizations may otherwise be constrained under other terms (e.g. non-commercial, share-alike), we will be using non-restrictive licenses when possible. In all cases (with the exception of data), authors will retain copyright of the works. License information will be included as a separate file accompanying each digital product.

Publications, such as reports and other creative works, will be shared under the Creative Commons Attribution 4.0 International Public License (CC-BY). Schemas and data models, as part of the Functional Requirements deliverable, will be shared under the MIT License. The Preliminary Data deliverable will be released with the CC0 license. Interview data will not be released.

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.

Most digital products will be available publicly on openly available online websites. The one exception will be the interview data, which will be housed on institutional-owned online storage, and only accessible to project team members.

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.

Not applicable for this project.

SECTION II: 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 the format(s) you will use.

Schemas for the metadata and data model of the Functional Requirements deliverable will be created and shared in XML format.

Diagrams, images, and tables may be created for the reports. These will generally be embedded in the report file (which will be in PDF format), though some may be shared independently as vector or bitmap images, in PDF, SVG, PNG, and JPG formats.

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

We will use the Oxygen XML Editor to create and test the XML-based schemas.

We will use a combination of graphic and image-editing software (e.g. Adobe Photoshop), presentation software (e.g. Apple Keynote), and programming languages and software (e.g. R and RStudio) to create diagrams, images, and tables.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG, OBJ, DOC, PDF) you plan to use. If digitizing content, describe the quality standards (e.g., resolution, sampling rate, pixel dimensions) you will use for the files you will create.

XML, PDF, SVG, PNG, JPG.

Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan. How will you monitor and evaluate your workflow and products?

Not applicable for this project.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period. Your plan should 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).

All digital products will be archived in the Zenodo repository, which will mint a DOI for each product. Data and the schemas will be archived independently, which will also enable versioning for those items. The reports will also be stored in the UF Institutional Repository.

No costs are needed for these repositories.

Metadata

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

Using Dublin Core as the starting point with additional elements from other schemes if needed to guarantee a concise but inclusive selection of Metadata Elements. Date entry will be using ISO-8601; Subjects will be using Jstor Thesaurus, the thesaurus now actively in use by Jstor to categorize scholarly publication with local addition. Proposal writers' entry will integrate with ORCID.

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

We will maintain a Git repository to collect the metadata schema and data model for this project. This will be stored on the GitHub cloud platform under a specific organizational account created for this project. The contents of this repository will be regularly archived to Zenodo to mint DOIs for attribution and for long-term preservation. At the completion of the project, the contents of the Git repository will be deposited in the UF institutional repository.

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).

We will use keywords on GitHub and Zenodo to facilitate discovery.

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, delivery enabled by IIIF specifications).

The digital content and resources for this project will be openly available online through the GitHub, Zenodo, and the UF Institutional Repository websites. GitHub and Zenodo also provide API access.

D.2. Provide the name(s) and URL(s) (Universal Resource Locator), DOI (Digital Object Identifier), or other persistent identifier for any examples of previous digital content, resources, or assets your organization has created.

Example Dataset (Portal Project - longterm ecological experiment):
GitHub URL - https://github.com/weecology/PortalData
Zenodo URL - https://zenodo.org/record/4618615
DOI (latest version) - https://doi.org/10.5281/zenodo.4618615
DOI (versioned) - https://doi.org/10.5281/zenodo.1215988

SECTION III: SOFTWARE

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.

Not applicable for this project

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

Not applicable for this project

Technical Information

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

Not applicable for this project

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

Not applicable for this project

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

Not applicable for this project

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

Not applicable for this project

B.5 Provide the name(s), URL(s), and/or code repository locations for examples of any previous software your organization has created.

Not applicable for this project

Access and Use

C.1 Describe how you will make the software and source code available to the public and/or its intended users.

Not applicable for this project

C.2 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

Not applicable for this project

URL:

Not applicable for this project

SECTION IV: RESEARCH DATA

As part of the federal government's commitment to increase access to federally funded research data, Section IV represents the Data Management Plan (DMP) for research proposals and should reflect data management, dissemination, and preservation best practices in the applicant's area of research appropriate to the data that the project will generate.

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

Data to be collected will be interview data and grants metadata. The interview data will provide feedback and grounding for the project deliverables, and may be summarized or quoted without identifying information. The grants metadata may be used by the public for exploration of content and patterns of sharing for grant proposal documents, as examples for training and professional development, and as foundational for future platforms that provide direct access to the described materials.
Interview data will be collected during Jun - Aug 2022. The grants metadata will be collected continuously throughout the project, but primarily during the environmental scan (Sep - Dec 2021) and the assembly of the Preliminary Data deliverable (Oct 2022 - Feb 2023).

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?

Yes, the interview data required approval by the UF institutional review board (IRB). If this project is funded, the protocol (see "Supportingdoc2 - interviewprotocol.docx") will be submitted for IRB approval prior to the interview phase of the project (May 2022).

A.3 Will you collect any sensitive information? This may include personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information. If so, detail the specific steps you will take to protect the information while you prepare it for public release (e.g., anonymizing individual identifiers, data aggregation). If the data will not be released publicly, explain why the data cannot be shared due to the protection of privacy, confidentiality, security, intellectual property, and other rights or requirements.

The interview data includes personally identifiable information, and interview participants will need to provide explicit consent for their identity to be shared publicly. In the absence of consent, we will anonymize individual identifiers in publicly available content (e.g. referring to occupations and institutions in broad terms - "program officer at a federal funding agency"). For confidentiality reasons, the full interview data will not be released publicly. The grants metadata will not include any sensitive information and rely on publicly available information.

A.4 What technical (hardware and/or software) requirements or dependencies would be necessary for understanding retrieving, displaying, processing, or otherwise reusing the data?

A standard web browser is sufficient for retrieving the data. Detailed usage of the data (e.g. accessing linked materials, data analysis) may require statistical or programming software.

A.5 What documentation (e.g., consent agreements, data documentation, codebooks, metadata, and analytical and procedural information) will you capture or create along with the data? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the data it describes to enable future reuse?

Consent agreements will be created alongside the interview data, and will be stored alongside it in institutionally-owned online storage.
Documentation for the grants metadata will consist of markdown-format descriptions and CSV-format data dictionaries. These files will be stored alongside the data in the GitHub-hosted repository, Zenodo, and the UF Institutional Repository.

A.6 What is your plan for managing, disseminating, and preserving data after the completion of the award-funded project?

Interview data will be maintained for three years; some portions of the interview transcripts may be kept indefinitely.
Grants metadata will be hosted on GitHub, Zenodo, and the UF Institutional Repository indefinitely.

A.7 Identify where you will deposit the data:

Name of repository:

Zenodo

URL:

https://zenodo.org/

A.8 When and how frequently will you review this data management plan? How will the implementation be monitored?

We will review this plan at the beginning of the relevant phases of this project (Sep 2021, Feb 2022, May 2022, Oct 2022). This form will be deposited in the UF Institutional Repository alongside submission of this grant proposal. The PIs Hao Ye and Perry Collins will review implementation.