

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

Narrative

Summary

Software is both a copyrightable creative work¹ and foundationally important to the future of scholarly research, yet software citation is not ubiquitous. Despite increasing acceptance of general software citation principles, “technical and community challenges exist that make moving beyond general acceptance to implementation and common use difficult.”² As a result, metadata to facilitate software citation goes unrecorded, software goes uncited, and software authors continue to be divorced from their contributions to science and human cultural heritage. If this status quo persists, uncited software will become increasingly difficult to find, access, and build upon, which will prevent software from being “Findable, Accessible, Interoperable, and Reusable” (FAIR)³ in the future.

Addressing the challenges that currently impact software citation implementation will require action from throughout the scholarly communication ecosystem and digital preservation landscape, along with cooperation from software authors and users. This situation presents an opportunity to enable a future for FAIR software through software citation by leveraging intersections between digital challenges in libraries and archives, and the work of experts in other disciplines to advance theory and practice. The John G. Wolbach Library at Harvard-Smithsonian therefore requests \$99,372 from the IMLS to bring together stakeholders and experts representing the many forms of labor and expertise needed to ensure that software citations support software authors, preservationists, and software users alike. Funding from the IMLS will ensure that this diverse community of people are given the opportunity to work together to build on past progress to develop a specific plan of action that can address software citation implementation on all fronts.

¹ Albert, K., Bouquin, D., Farber, A., Hoover, R. (2019). Copyright Guide for Scientific Software.

<https://doi.org/10.5281/zenodo.3581326>

² Katz, D. S., Bouquin, D., Chue Hong, N. P., Hausman, J., Jones, C., Chivvis, D., Clark, T., Crosas, M., Druskat, S., Fenner, M., Gillespie, T., Gonzalez-Beltran, A., Gruenpeter, M., Habermann, T., Haines, R., Harrison, M., Henneken, E., Hwang, L., Jones, M. B., ... Zhang, Q. (2019). Software Citation Implementation Challenges. ArXiv E-Prints, <https://arxiv.org/abs/1905.08674>

³ Lamprecht, A. Garcia, L., Kuzac, M., Martinez, C., Acila, R., Martin, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P.A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J., Chue Hong, N., Goble, C., Capella-Gutierrez, S. (2019).

Towards FAIR principles for research software. Data Science.

<https://datasciencehub.net/paper/towards-fair-principles-research-software>

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

Statement of National Need

Software has been a crucial intellectual contribution to scientific and social progress for decades, but practices that enable machine-actionable citations have not been consistently applied to software itself. Instead, studies have shown that people mention software in their papers using a variety of inconsistent and ambiguous mechanisms. For example, a recent twenty year case study analyzing software citation behavior in astronomy found that 109 different aliases (including alternate names, phrases, and publications) were used to identify just nine software packages in papers published by a single publisher, and hundred of papers mentioned software without giving any recognizable form of attribution to software authors.⁴ Similarly, a study examining software citation in a random sample of biology publications⁵ found, “diverse and problematic practices” noting that, “informal mentions are very common, even in high impact factor journals and across different kinds of software.” The same study found that software citations, “often fail to accomplish many of the functions of citation.” These studies highlight the fact that it is not a lack of motivation that prevents software from being cited, rather that software citation practices do not currently meet the criteria laid out in the Software Citation Principles.

Software Citation Principles

Recognizing a critical need for software citation, in 2016, a FORCE11 working group developed the Software Citation Principles.⁶ FORCE11 working groups are international, interdisciplinary collaborations involving a variety of diverse stakeholders including publishers, librarians, archivists, funders, repository developers, and others who contribute meaningfully to shaping the future of scholarly communication. The FORCE11 Software Citation Principles established core values and motivations for software citations stating, “software should be considered a legitimate and citable product of research” and citations should, “enable credit and normative, legal attribution to all contributors to the software.” Further, the Principles laid out the need for persistent, unique software identification to ensure access to specific software and its associated

⁴ Bouquin, D., Chivvis, D., Henneken, E., Lockhart, K., Muench, A., Koch, J. (2020). Credit Lost: Two Decades of Software Citation in Astronomy. The Astrophysical Journal Supplement Series: ApJS. <http://doi.org/10.17605/OSF.IO/P9GJR>

⁵ Howison, J., & Bullard, J. (2016). Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature. Journal of the Association for Information Science and Technology, 67(9), 2137–2155. <https://doi.org/10.1002/asi.23538>

⁶ Smith, A. M., Katz, D. S., & Niemeyer, K. E. (2016). Software Citation Principles. PeerJ Computer Science, 2, e86. <https://doi.org/10.7717/peerj-cs.86>

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

metadata; this robust form of software identification is essential for software preservation and scientific reproducibility, in addition to software citation.⁷

Implementing the Software Citation Principles

The Software Citation Principles outlined how software citations should work and provided stakeholders in this context with a target. However, the Principles are not in their own right a solution to the issues impacting software citation implementation. Instead, the Software Citation Principles should be seen as the first step toward a goal. An important next step was taken by a group called the FORCE11 Software Citation Implementation Working Group (SCIWG). The SCIWG has worked to assess and document the predominant challenges that prevent the software citation principles from becoming reality— these include technical and cultural challenges that libraries and archives are strategically well-situated to address. Specifically, libraries and archives can help resolve technical issues impacting software identification and metadata standardization (e.g., adoption of CodeMeta⁸ and CFF⁹), particularly metadata availability, granularity, storage, and interoperability. For example, the SCIWG’s report on software citation implementation challenges¹⁰ outlines the following:

“Repositories archive both data and software from computational workflows, often in integrated packages... These integrated packages often contain software, but that software is typically not specifically designed for reuse. Rather, it implements the particular computations needed for an analysis or project... Challenges surrounding these mixed data and software packages include:

- Embedding of software (e.g. scripts, Jupyter notebooks) with data inputs, outputs, and other computation artifacts in mixed packages that are

⁷ Software preservation: A stepping stone for software citation [blog]. (2018). Software Heritage.

<https://www.softwareheritage.org/2018/06/25/software-preservation-for-software-citation/>

⁸ Chue Hong, N., Druskat, S., & Slaughter, P. (2018). codemeta: Minimal metadata schemas for science software and code, in JSON-LD. CodeMeta. <https://github.com/codemeta/codemeta/blob/master/codemeta.json>.

⁹ Druskat, S., Chue Hong, N., Haines, R., & Baker, J. (2018). Citation File Format (CFF)—Specifications. Zenodo. <https://doi.org/10.5281/zenodo.1405679>

¹⁰ Katz, D. S., Bouquin, D., Chue Hong, N. P., Hausman, J., Jones, C., Chivvis, D., Clark, T., Crosas, M., Druskat, S., Fenner, M., Gillespie, T., Gonzalez-Beltran, A., Gruenpeter, M., Habermann, T., Haines, R., Harrison, M., Henneken, E., Hwang, L., Jones, M. B., ... Zhang, Q. (2019). Software Citation Implementation Challenges. ArXiv E-Prints, <https://arxiv.org/abs/1905.08674>

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

assigned a single DOI in which the software may not be individually referenceable...

- Metadata for packages derives primarily from data repository communities but includes sections on software used...
- Need for documenting the provenance relationships between software, data, and products that describes the lineage of computational workflows
- Software in repository packages that import other software packages... but does not document these dependencies in metadata”

The SCIWG’s report goes on to further emphasize how essential archival repositories are to enabling software citation by defining what it means to “publish” software: “...to achieve wide adoption of software citation, more repositories should support software publishing... Repositories become the ‘publishers’ of each software version, in a similar way that repositories have become the ‘publishers’ of datasets...” This is to say that for software to be citable using a persistent identifier, software must be archived. Software citation using persistent identifiers will not be possible without the active involvement of libraries and archives.

In addition to helping resolve technical challenges, the potential for libraries and archives to address cultural issues impacting software citation is difficult to overstate, particularly because community-specific software citation practices need to be developed, “within the context of existing scholarly communication and software development norms.”¹¹ As facilitators, resource providers, stewards, and innovators, librarians and archivists can help guide how digital preservation platforms address their community’s specific needs, bridge divides between their communities and other key collaborators, and develop new tools that support authors.

Interdisciplinary Collaboration Required

Although libraries and archives have the opportunity to help lead efforts to address software citation implementation challenges, they cannot solve all of the problems on their own. Disciplinary communities, scholarly publishers, indexers, funders, and institutions all play critical roles that influence the technical, sociocultural, and economic factors that determine how citations work, what citations mean, and who

¹¹ Katz, D. S., & Chue Hong, N. P. (2018). Software Citation in Theory and Practice. In J. H. Davenport, M. Kauers, G. Labahn, & J. Urban (Eds.), *Mathematical Software – ICMS 2018* (pp. 289–296). Springer International Publishing. https://doi.org/10.1007/978-3-319-96418-8_34

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

citations benefit. For instance, indexers (e.g., Google Scholar, the NASA Astrophysics Data System) will need to distinguish software from other research artifacts and provide metrics and authorship information across software versions— this may require indexers to request additional metadata from the publishers, which will in turn will require publishers to determine how software is internally identified their systems, update relevant policies, and provide guidance to authors on repositories that meet their criteria for software deposition. Similarly, funders will need to define requirements for depositing grant-funded software and metadata in archives, and determine how software should be addressed in proposals, research outcomes, and research output management plans. Meanwhile, institutions will need to determine the software citation metrics required to measure researchers' output, including their software contributions, to support hiring and promotion-related decision-making (in addition to incorporating software citation metadata standards into their repository workflows and metadata schemas).

In short, librarians, archivists, and scholars investigating software citation need to be given the opportunity to spend significant time working directly with publishers, funders, and institutions enacting relevant policies, as well as indexers exposing software citation metadata, for the challenges presented by software citation implementation to be addressed. Continuing to work as disparate groups in discipline-specific silos will limit the collaboration that we now know is essential to implementing the Software Citation Principles.

Project Design

We propose to host a two-day in-person workshop to develop an action plan for addressing and piloting solutions to challenges outlined by the FORCE11 Software Citation Implementation Working Group. Emphasis will be placed on ways for libraries and archives to address software citation implementation challenges, but this will not limit the experience or ideas that participants bring to the table. We aim to gather key contributors and leaders in areas relevant to software citation and will not limit the workshop to FORCE11 members, domestic participants, or people with appointments within libraries and archives. We plan to work in collaboration with people actively addressing software citation issues through organizations like DataCite, the Research Software Alliance (ReSA), the Software Preservation Network (SPN), RDA's Software Source Code Interest Group, and the US Research Software Sustainability Institute conceptualization project, in addition to international institutions (e.g., the UK Software Sustainability Institute, the British Library) and others. Having

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

diverse and differing perspectives on the software citation problem space is imperative to developing an achievable set of tasks to respond to known obstacles and anticipate emerging needs.

The workshop will result in a published action plan in the form of a white paper with a timeline and anticipated outcomes for deliverables associated with the plan. The plan could potentially frame a future IMLS proposal and clarify where funding and other resources are needed. The workshop will be evaluated through attendee feedback on satisfaction with the workshop design, as well as the resulting action plan.

Workshop Design

While the first half our proposed one-year funding period would be dedicated to refining plans for the workshop itself and identifying stakeholders to participate (see “Schedule of Completion”), the general approach we plan to take with the workshop is outlined here:

- Day 1
 - Presentations will be given to summarize both complete and ongoing work done to address known software citation challenges. Brief presentations will also outline the primary challenges (e.g., metadata availability) in more detail.
 - Smaller discussion groups will be organized by specific challenge categories and given the opportunity to brainstorm technical, social, and educational interventions to tackle their specific areas of interest.
 - Groups will be asked to:
 - Outline how their approaches could be converted into discrete actions and projects
 - Define the obvious hurdles that would prevent the execution of their ideas
 - Explain how their approaches would help or conflict with other approaches to the same or other challenges
 - Estimate the technical (and other) expertise, time, and financial resources needed to execute their ideas
 - Small groups will be asked to summarize their work using a template provided by workshop organizers that will be shared with the rest of the group
- Day 2

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

- Three quarters of the day will be spent discussing the ideas proposed the day before
 - Attendees will be asked to:
 - Evaluate and prioritize ideas that address the same challenges
 - Lay out a series of mutually supporting approaches
 - Identify potential project leaders
 - Lay out a timeline for the execution of ideas for which resources are available
 - Define steps for acquiring resources that are still needed
- The final part of the day will be dedicated to discussing needs that will not be met by the action plan to make clear the action plan's limitations.

Writing plan

The primary deliverable of the funding being sought through this proposal would be a published action plan for addressing barriers that impact software citation. In order to ensure both the completion of this document and its usability, we would establish "Report Leaders" who would take on the primary responsibility for writing up workshop outcomes. Report Leaders would be charged with drafting the initial action plan, circulating the draft to workshop attendees for their review and iterative feedback, and submitting the finalized version for publication. Attendees would volunteer to be Report Leaders prior to the workshop and would be given stipends for their efforts. Report leaders would meet for an additional half day after the two day workshop to finalize both writing logistics and the action plan's layout.

Results dissemination

Open access publication of the workshop's resulting white paper will ensure that contributions resulting from this project are documented for broad consumption. We plan to archive a copy of the final white paper using Zenodo, as well as submit a copy of the final document to arXiv so it can be indexed appropriately. The published white paper will subsequently be circulated through institutional blogs, newsletters, and social media channels to share information with the wider public and target audiences. The target audiences for the white paper will include the communities identified above (e.g., ReSA, SPN) as well as (but not limited to) the following:

- Librarians and archivists

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

- Repository managers and curators
- Publishers
- Journal and proceedings editors
- Scholars of research (e.g., information scientists, network science, team science)
- Administrators and others interested in supporting FAIR outputs

In addition to publishing the white paper, key project staff will attend relevant conferences (e.g., FORCE2021, RDA Plenary, Super Computing 21, SIAM Computational Science and Engineering) to present the workshop's outcomes. To ensure our ability to present at relevant conferences, we have requested supplemental funding for key project staff to travel to meeting locations. Presentations at meetings will amplify information shared through the white paper and enable direct in-person feedback on the action plan itself.

Diversity Plan

By offering to fund both international and domestic attendee travel and lodging, our goal is to make participation equally achievable for attendees with varying levels of financial support from their employers. We also commit to incorporating equity and inclusion into our considerations when inviting participants and planning any presentations associated with the workshop. Furthermore, we plan to discuss issues that may be particularly challenging to women and underrepresented groups in the open source community and encourage participation from early career participants. We will enforce a code of conduct throughout the workshop's planning and execution.¹²

Assessment

To assess the impact of the workshop and resulting action plan, we will develop a set of survey assessments for workshop attendees and Report Leaders. Feedback from attendees will help us further understand the limitations of using a two-day workshop to catalyze action on software citation implementation, as well as inform decisions about how to improve the workshop's format, and the value of the action plan itself. Using a survey to get feedback from Report Leaders directly will help us improve the collaborative writing, review, and publication processes for future collaborations.

¹² The John G. Wolbach Library's existing Code of Conduct is available here and will be updated as appropriate for the proposed workshop: <https://library.cfa.harvard.edu/code-of-conduct>

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

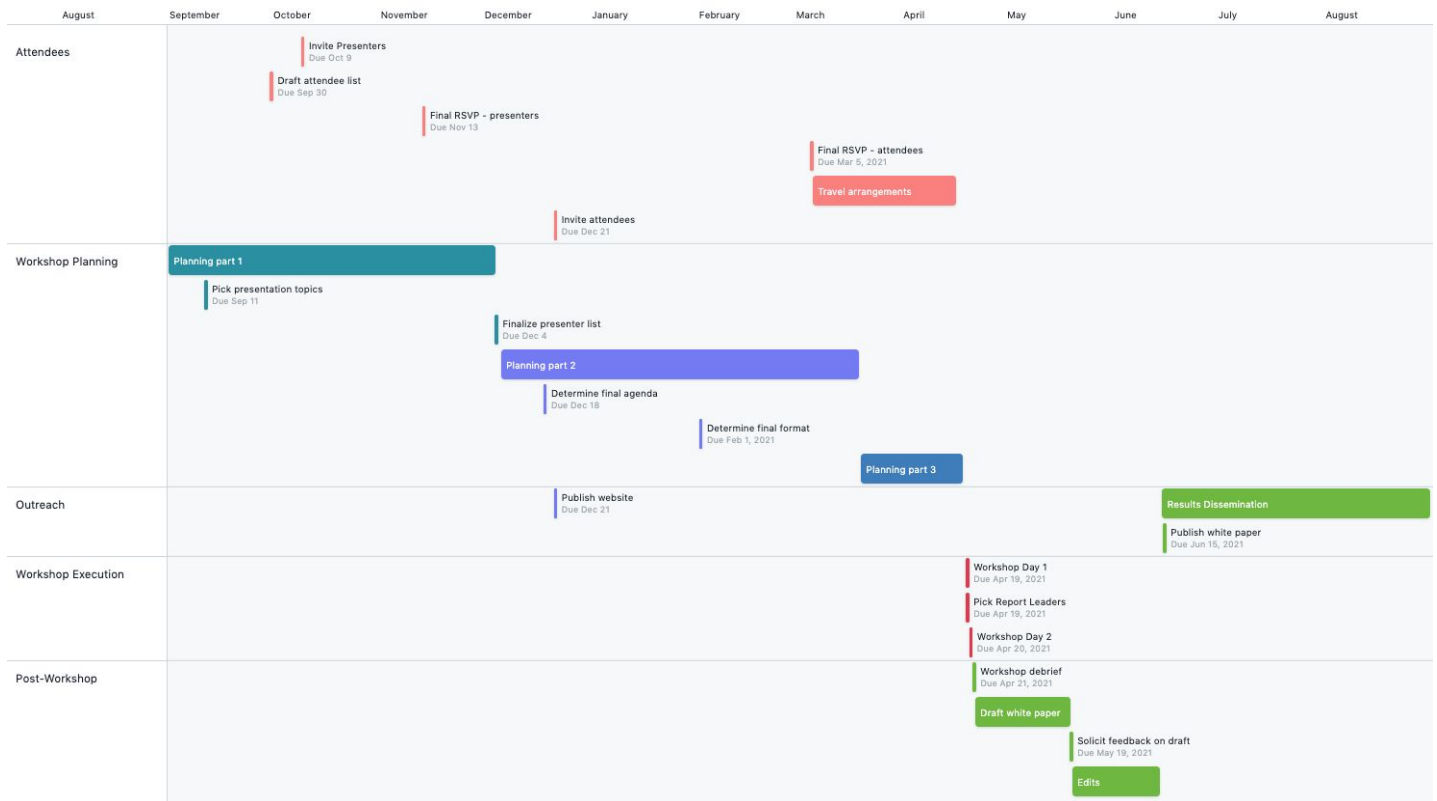
National Impact

By taking an interdisciplinary approach to software citation implementation we will be able to amplify the impact of work already done and underway by workshop participants, and establish new collaborations between libraries, archives, and previously disparate groups. By creating an executable action plan, we will be able to instigate changes to systems and norms that impact the international community— citations have no boundaries and software is developed in global collaborations. Most importantly, software citation implementation will move forward in libraries and archives. Right now, software archiving is a relatively nascent practice, and archived software is difficult to cite. As a result, only heavily used software and software developed by people with the ability to actively, effectively, and continually advocate for the importance of their work will have software that is consistently cited and connected to new research. Less privileged software may be deposited in an archival repository, but associations between that software and work that could demonstrate its importance (citations) are often missing, ambiguous, or difficult to establish. It is often hard for the people who wrote that software to demonstrate its value and give their work context. When collections are not contextualized and valued, they cannot be prioritized for curatorial care by institutions with resources to invest in collection preservation and maintenance. Software (like any other physical or digital object) can be deposited in an archive and still be lost if no one knows why it matters. Developing an action plan to directly confront challenges to software citation implementation will help shift the current landscape so software can be linked to the people who created it and provide provenance for its future use and care.

Software Citation Implementation: Action Plan Development

Daina Bouquin, Head Librarian of Harvard-Smithsonian Wolbach Library

Schedule of Completion



Activity	Start date	Due date	Notes
Planning part 1	2020-09-01	2020-12-04	Pick presentation topics, draft attendee list, presenter invites
Planning part 2	2020-12-05	2021-03-19	Finalize presenters, format, agenda, website, invite attendees
Travel arrangements	2021-03-05	2021-04-16	Work with attendees
Planning part 3	2021-03-19	2021-04-18	Final logistics
Workshop Day 1		2021-04-19	Pick Report Leaders
Workshop Day 2		2021-04-20	
Workshop debrief		2021-04-21	Determine write up plan
Draft and edit white paper	2021-04-21	2021-06-14	Solicit feedback on draft
Results dissemination	2021-06-15	2021-08-31	Publish final white paper, present at conferences



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.

The copyright of digital assets will be held by their authors and released under the Creative Commons license CC BY-NC-SA 3.0 US (<https://creativecommons.org/licenses/by-nc-sa/3.0/us/>)

We will release assets under CC BY-NC-SA 3.0 US to ensure that authors are given credit for their work, and that people are able to copy and redistribute the material in any medium or format, and remix, transform, and build upon the material for any non-commercial purpose.

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.

Harvard University will not apply access restrictions to digital products or impose additional limitations on their reuse.

We will list the CC BY-NC-SA 3.0 US license on all digital assets and share the Harvard Terms of Use (<https://hwp.harvard.edu/terms-use>)

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.

We do not plan to capture any information that would involve any privacy concerns or require obtaining additional rights or raise cultural sensitivities. Participation in our planned post-workshop survey for attendees will be done anonymously.

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.

We plan to produce the following digital assets:

Meeting notes - From workshop planning, the workshop itself, and post-workshop activities

White paper - Describes the resulting action plan (PDF/A)

Surveys - One for workshop attendees, one for Report Leaders (Google Forms); Survey results (CSV)

Conference presentations - These will take the form of slide decks and will be saved printable documents (PDF/A)

Workshop handouts may also be produced.

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.

The project team will use the Open Science Framework to aggregate working documents and files associated with the project.

An institutionally managed instance of G Suite (Google Drive, Google Forms, Google Slides) will be used to write the proposed white paper, develop the proposed survey, and create presentations for disseminating results.

The Wolbach Library's GitHub organization may be used for version control of shared documents (e.g., meeting notes) and any other digital assets.

Zenodo will be used to archive and document digital assets. Our white paper will also be deposited on arXiv.

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.

The white paper and presentation slides will be released as archival PDFs (PDF/A).

Additional documentation will be written in GitHub renderable markdown (MD).

Survey results will be saved as CSV files.

Workflow and Asset Maintenance/Preservation

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

For all meetings associated with the proposed workshop and the subsequent writing of the proposed white paper, a note taker will be responsible for recording minutes and depositing minutes in the Wolbach Library's GitHub organization. This workflow will allow meeting attendees to give feedback on GitHub and will make audits of changes simple.

The project lead will be responsible for organizing any Google Docs and Forms associated with the project into folders and integrating both the Google Drive and GitHub organization's content into the Open Science Framework. Centralizing documents on the Open Science Framework will make it clear which documents are serving which purpose.

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).

The Wolbach Library is dedicated to the long term stewardship of research and historical material in astronomy and its related fields. Wolbach will ensure that all digital assets are deposited in digital archives (i.e., Zenodo, arXiv) with appropriate metadata to enable their future reuse and/or reformatting.

Wolbach Library's intention with all assets within its collection is to maintain those assets in perpetuity; the digital assets produced by this project would be maintained using the same practices as those applied to the library's archival collections.

Prioritization of collections for reformatting is a necessary component of digital preservation, at minimum though, the project team aims to ensure the assets produced by this grant are available continuously for at least 10 years without needing significant curatorial attention.

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).

We will produce metadata for our digital assets by creating a data dictionary for the survey results. We will generate descriptive, administrative, and preservation metadata for at the point of deposit into the archival repository Zenodo. We will also generate metadata upon submitting our white paper to arXiv.

Metadata will include a CITATION.MD file in addition to other markdown files (e.g., README.MD).

We plan to also tag the resulting digital assets schema.org elements.

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

We plan to archive the metadata files associated with our digital assets along with the digital assets themselves.

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).

By depositing our digital assets on Zenodo and submitting our white paper to arXiv, we will be ensuring indexible access to metadata using the Zenodo and arXiv APIs, in addition to creating records that are indexible by Google and disciplinary indexes.

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).

Open publication on Zenodo and arXiv of the workshop's resulting white paper will ensure that contributions resulting from this project are documented for broad consumption. The published white paper will subsequently be circulated through institutional blogs, newsletters, and social media channels to share information with the wider public and target audiences.

The target audiences for the white paper will include the communities identified above (e.g., ReSA, SPN) as well as (but not limited to) the following: librarians and archivists; repository managers and curators; publishers; journal and proceedings editors; scholars of research; administrators and others interested in supporting FAIR outputs

In addition to publishing the white paper, key project staff will attend relevant conferences to present the workshop's outcomes. Presentations at meetings will amplify information shared through the white paper and enable direct in-person feedback on the action plan itself.

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.

The following recent example is a publication created in collaboration with the Software Preservation Network and the Harvard Cyberlaw Clinic:

Albert, K., Bouquin, D., Farber, A., Hoover, R. (2019). Copyright Guide for Scientific Software.
<https://doi.org/10.5281/zenodo.3581326>

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.

N/A

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.

N/A

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.

N/A

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

N/A

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

N/A

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

N/A

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

N/A

Access and Use

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

N/A

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

Name of publicly accessible source code repository:

N/A

URL:

N/A

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.

To assess the impact of the workshop and resulting action plan, we will develop a set of survey assessments for workshop attendees and Report Leaders.

Feedback from attendees will help us further understand the limitations of using a two-day workshop to catalyze action on software citation implementation, as well as inform decisions about how to improve the workshop's format, and the value of the action plan itself. Using a survey to get feedback from Report Leaders directly will help us improve the collaborative writing, review, and publication processes for future collaborations.

We plan to disseminate the workshop survey after the workshop in spring 2021, and the survey for Report Leaders will be circulated in early summer 2021.

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?

No IRB approval is required.

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.

No PII will be collected.

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

Will export survey results into CSVs and metadata will be recorded in markdown files. Opening CSV and markdown files will require only a text editor.

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?

The exported survey results will be archived with data dictionaries (i.e., codebooks) in markdown format to describe the survey results and their provenance.

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

The Wolbach Library is dedicated to the long term stewardship of research and historical material in astronomy and its related fields. Wolbach will ensure that all datasets are deposited in a digital archive (i.e., Zenodo) with appropriate metadata to enable its future reuse and/or reformatting.

Wolbach Library's intention with all assets within its collection is to maintain those assets in perpetuity; the digital assets produced by this project would be maintained using the same practices as those applied to the library's archival collections.

Prioritization of collections for reformatting is a necessary component of digital preservation, at minimum though, the project team aims to ensure the assets produced by this grant are available continuously for at least 10 years without needing significant curatorial attention.

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?

This data management plan will be reviewed and revised at the end of the funding period and incorporated into the Wolbach Library's general collection procedures— these include an annual review of collection development and maintenance strategies.