Review:

ESDS-RFC-055 ESDIS Project Data Use and Citation Guidance

Status: Completed

Produced by Steve Olding, 04/11/2025

Items reviewed

1 ESDS-RFC-055 ESDIS Project Data Use and Citation Guidance

2 Introduction

1.1.1 DESCRIPTION

This document contains recommendations, with supporting narrative, in response to the ESDIS Leadership request for standardization of dataset citation practices and data use guidelines. It draws heavily from a 2023 study conducted by NSIDC, as well as the ORNL DAAC's experience and understanding of DOI and citation best practices.

3 BLUF (bottom line up front) Recommendations:

1.1.2 DESCRIPTION

1. Update the current ESDIS Project Data Guidance to be the ESDIS Project Data Use and Citation Guidance, using the changes provided at the end of this document.

2. Dataset landing pages, user guides, and CMR metadata should display citations in APA format, with the exception that all authors should be shown (APA truncates authors if there are more than 20). Where practical, dataset landing pages should also provide an option to display citations in Chicago and Harvard styles (with all authors displayed).

3. DAACs and ESDIS must ensure that bibliographic metadata in DataCite is accurate, complete, and updated.

4. A team, including representatives from the DOI Working Group, Metadata Stewardship, and CMR should determine the long-term approach to managing bibliographic data for datasets, tools, services, and other digital assets which have ESDIS-managed DOIs. This work should include what information should be in CMR (particularly for items other than datasets) and how to ensure what's in ESDIS and DataCite stores is complete, accurate, and maintained.

4 Recommendation #1

1.1.3 DESCRIPTION

The text at the end of this document is the contents of the ESDIS Data Use Guidance (formerly Project Data Use Policy). I propose to change the title to "Data Use and Citation Guidance", with the changes noted in the text below.

[Note to reviewers: Not all revisions are visible in the Jama version of this document. Please see the associated Word document for full change tracking details.]

5 Recommendation #2

1.1.4 DESCRIPTION

Regarding dataset citations, a 2023 NSIDC study (Theresa Andersen, Candida Dewes, and Leslie Goldman) and ORNL DAAC experience indicate that the most important elements are: a) supporting common citation managers (Zotero, Mendeley, EndNote, and BibTeX) and b) providing links on dataset landing pages to a citation formatter (e.g. Crosscite: https://citation.crosscite.org). While providing citations in the most common formats for the ESDIS user community (APA, Chicago, and Harvard) is useful, the specific format needed is determined by the journal where someone is publishing. Further, many users need bibliographic data in a bibliographic management tool so that they can manage the citations in documents as they are developed (such as creating a bibliography in order of appearance in the document). While the NSIDC study did not mention BibTeX specifically as a citation manager, ORNL DAAC experience is that there is substantial use within some communities, particularly modeling, though that use does appear to be less than for Zotero, Mendeley, and EndNote.

All of the citation managers noted above provide tools to import bibliographic information when provided with a DOI – either directly within the application (Zotero, Mendeley, and EndNote) or through helper web sites (BibTeX – see https://www.bibtex.com/c/doi-to-bibtex-converter/). Zotero, Mendeley, and EndNote also support importing from RIS and BibTeX formats. Particularly for EndNote and BibTeX, the process of importing a bibliographic data file (e.g. RIS or BibTeX format) is considerably simpler than updating a record using the DOI information, and some users will likely also find this true for the other citation managers. The NSIDC study indicated a nearly 3:1 user preference for a downloadable bibliographic file (38% versus 13%) rather than using the DOI to look up bibliographic information.

The NSIDC study found that the most used formats were APA (58%), Chicago (23%), and Harvard (19%). This is not surprising since APA is what AGU and EGU use. However, all these citation formats truncate long author lists. For example, APA format recommends truncating the list at 19 authors when the author list is 21 or longer. The NSIDC study and ORNL DAAC experience indicate a very strong user preference for including all of the authors on the dataset (and user guide) landing page citation and in the downloadable bibliographic file.

6 Recommendation #3

1.1.5 DESCRIPTION

Given the central role of DataCite metadata in user-facing aspects of dataset citations, ensuring the completeness, accuracy, and maintenance of this data is essential. While dataset updates that change bibliographic data are likely not common, they do happen – such as changes to an author list after publication, requests for name changes by authors, changes to DOI resolution URL (such as the change from dx.doi.org to doi.org, as well as changes in the dataset landing page URLs), changes to point of contact information, and additions to bibliographic data with changes in best practices. The current ESDIS DOI management process has options for submission using Excel, tab-delimited text, and XML.

7 Recommendation #4

1.1.6 DESCRIPTION

The CMR Collection-level metadata has a field for Collection Citations, with multiple attributes. However, the Collection Citations attributes are not the same as the attributes collected in the ESDIS DOI process. Further, a spot check of CMR Collection records indicates that a common practice is to only populate Other Citation Details with a formatted citation, leaving the other Collection Citation attributes blank. There is also a substantial heterogeneity in the style for these formatted citations, with some not being in any generally accepted style. As a result, CMR does not presently have the information to support updating DOIs, should that become a desired path forward.

ESDIS also manages DOIs for at least some documents, such as Algorithm Theoretical Basis Documents (ATBDs). There is also an emerging practice of assigning DOIs to things such as tools, services, and software (including specific software releases). There is also an emerging practice of assigning a DOI to a series (such as the MOD12A1 or ATL08 data products) along with DOIs to each specific version/instance of that series. In these cases, the Series DOI always points to the most current/recent version. See 10.5281/zenodo.10728384 for an instance of the series 10.5281/zenodo.10728200.

To reconcile these issues, I recommend that a team, centered around the DOI Working Group, but also including representation from the Metadata Stewardship and CMR teams, be formed to determine how ESDIS will manage citation metadata, at least going forward. This team should collaborate with related efforts being conducted at the NASA SMD level. It should also consider evolving best practices, such as those being recommended by the American Geophysical Union, ESIP, COPDESS[1] working groups[2]. Key questions include:

- Should citation metadata be managed in CMR? If so, should the attributes of the Collection Citation and/or the DOI registration/update templates be changed to better align with DataCite fields and current DOI metadata population best practices?
- What should the best practices be, at least going forward, for ensuring that citation metadata in DataCite is complete and current? An option would be for DAACs to ensure that the metadata in CMR is correct (depending on the outcome of the above considerations) and have CMR automatically update DOIs. If a different path is chosen, what should be done to ensure that CMR and DataCite have the same bibliographic metadata.
- In consultation with the Web Unification team, how should the results from the above considerations affect the process for and appearance of dataset landing pages (including best practices for machine interpretability of dataset landing content relating to bibliographic information, as described on the DataCite Best Practices page).
- The Metadata Stewardship team is recommending an approach for managing dataset lifecycles, including "tombstone" pages for datasets that are no longer being distributed. This team should consider that approach and determine if change should be made to the best practices (at least going forward) for ESDIS management of tombstone pages.
- What should be the practices (including any changes to CMR) for managing DOIs for documents; ESDIS-developed software, tools, and services; and other digital artifacts managed by ESDIS enterprise and/or DAACs?

8 Proposed Data Use and Citation Guidance text

See https://www.earthdata.nasa.gov/engage/open-data-services-software-policies/data-use-guidance.

9 ESDIS Project Data Use and Citation Guidance

NASA's Earth Science Data and Information System (ESDIS) Project provides open access to data and information generated under NASA sponsorship. Non-NASA data available through the ESDIS project is subject to the license arrangements of the sponsoring organization. Users are encouraged to validate the source and associated use permissions.

Data Use

NASA ESDIS content, including but not limited to images, audio, video, and computer files used in the rendition of 3-dimensional models, such as texture maps and polygon data in any format, observation data, metadata, products, information, algorithms, including scientific source code, documentation, models, images, and research results, generally are not copyrighted. You may use this material for educational or informational purposes, including photo collections, textbooks, public exhibits, computer graphical simulations and Internet Web pages. This general permission extends to personal Web pages.

NASA ESDIS content used in a factual manner that does not suggest or imply endorsement may be used without explicit permission. NASA should be acknowledged as the source of the material.

ESDIS content that is subject to usage restrictions, such as a license agreement, shall be labeled as such and the use of that data shall be in accordance with the designated license. Unless the content is marked with a use restriction or license, data provided from a NASA led mission are licensed as Creative Commons Zero (CC0). While there are no restrictions on the use of these data, data users are very strongly urged to cite the data used in their work products.

Licensing and Copyright

NASA occasionally uses copyrighted material by permission on its website. NASA's use does not convey any rights to others to use the same material. Those wishing to use copyrighted material must contact the copyright holder directly.

NASA does not license the use of NASA materials or sign licensing agreements. The agency generally has no objection to the reproduction and use of these materials, subject to the following conditions:

- NASA material may not be used to suggest or imply endorsement by NASA or by any NASA employee of a commercial product, service, or activity, or used in any manner that might mislead. Please see NASA Advertising Guidelines and Merchandising Guidelines for more information.
- It is unlawful to falsely claim copyright or other rights in NASA material.
- NASA shall in no way be liable for any costs, expenses, claims, or demands arising out of the use of NASA material by a recipient or a recipient's distributees.
- NASA does not indemnify nor hold harmless users of NASA material, nor release such users from copyright infringement, nor grant exclusive use rights with respect to NASA material.
- NASA material is not protected by copyright within the United States, unless noted. If copyrighted, permission should be obtained from the copyright owner prior to use. If not copyrighted, NASA material may be reproduced and distributed without further permission from NASA.

Citations and Acknowledgments

NASA should be acknowledged as the source of the material where applicable. Details of citation formats and practices vary by discipline and specific organizational practices (such as journal citation formats).

NASA ESDIS provides open access to data and information generated under NASA sponsorship. NASA data are freely accessible; however, when you publish these data or works based on the data, we request that you cite the datasets within the text of the publication and include a reference to them in your reference list. References should follow the specific formats of the journal or governing style guide and have enough detail to provide readers of your publication with the ability to obtain the datasets and conduct their own studies based on your work. For complete provenance and understanding of specifically which parts of data were used and how, it also may be necessary to describe in detail, within the body of the publication, exactly how the data were used.

The Earth Science Information Partners (ESIP) provides a clear, concise guidelines document that may also help you in determining the format of the data set reference. In addition, the following services provided by NASA ESDIS have additional information on data use and citation suggestions:

Global Imagery Browse Services (GIBS) Please refer to the GIBS Data Use Guidance and Acknowledgements for more information.

Open Data and the Importance of Data Citations

In addition to providing free and open access to data in accordance with NASA SMD Open Source Science Guidance, NASA's Earth Science Data Systems Program values transparency and reproducibility in scientific research, as do organizations with similar objectives such as the National Science Foundation (NSF), the Earth Science Information Partners (ESIP), the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS), and the Global Earth Observation System of Systems (GEOSS).

As such, NASA recognizes the importance of authors using NASA-provided datasets to clearly indicate which datasets were used and provide access to these datasets to readers. While there are several ways of accomplishing this objective, formally citing datasets and other research products is unambiguously among the best.

NASA's Science Mission Directorate (SMD) will continue to take steps proactively to improve its open data and data citation guidance to remain in line with the best practices of our community. Our overarching objectives are to ensure that data from NASA's Earth Science Data Systems Program can easily be accessed and that research based on NASA-supported datasets clearly cites the sources of these data.

10 References

[1] Coalition for Publishing Data in the Earth and Space Sciences – c.f. https://copdess.org

[2] See, for example:

- Stall, S., Bilder, G., Cannon, M. et al. Journal Production Guidance for Software and Data Citations. Sci Data 10, 656 (2023). https://doi.org/10.1038/s41597-023-02491-7
- https://www.agu.org/publish-with-agu/publish/author-resources/data-and-software-for-authors
- ESIP Barriers to Data and Software Sharing: The Reference Manager Gap
- Vannan, S.,Downs, R. R.,Meier, W.,Wilson, B. E., and Gerasimov, Irina V. (2020), Data sets are foundational to research. Why don't we cite them?, Eos, 101, https://doi.org/10.1029/2020EO151665.

11 RFC Author Contact Information

Bruce E. Wilson (wilsonbe@ornl.gov)

Manager, ORNL Distributed Active Archive Center for Biogeochemical Dynamics

Group Leader, Remote Sensing and Environmental Informatics

Oak Ridge National Laboratory

12 Suggested Citation

Wilson, Bruce E. (2025). Data Use and Citations Guidance. NASA Earth Science Data and Information System Standards Coordination Office. https://doi.org/10.5067/DOC/ESCO/ESDS-RFC-055