

Your data are an important part of your research supporting peer review, transparency and reproducibility. For publication in AGU journals, your data need to be placed in a repository that [supports discovery, preservation and citation](#) and accessibility.

By partnering with your repository early you get the benefit of incorporating your data management tasks during your research, when it is much easier, then waiting until the very end when you may be constrained by resources to prepare for publication.

Primary and processed data used for your research should be preserved. In your paper, cite these data, as well as any data you used from other sources, and include access information in the data availability statement (placed in the Open Research section). As the data for each paper may have unique challenges, we will work with you to find the best approach for your paper.

### **Considerations for data management when conducting your research:**

- [Incorporating data management into your research](#)
- [Preparing your data for preserving in a repository](#)
- [Selecting your repository](#)

For research significantly based on software (re: code, workflow, models) it might be necessary to place your software in a repository for the purpose of transparency and peer review. If you believe this is the case with your research, please contact an editor for your selected journal.

The specific process for depositing data in a repository, and getting it ready to publish is similar for most repositories, but it's best to work directly with your selected repository for specific considerations.

## Five Considerations for publication

1. [When to make your data or software available - the timing with your paper](#)
2. [Availability statement in open research section](#)
3. [Data citation](#)
4. [Software citation](#)
5. [Guidelines for research primarily based on models](#)

## Domain repositories useful to *JGR-Biogeosciences*

AGU recommends the following domain repository options by data type. This list is not meant to be comprehensive. If you have any additional recommendations, please send them to [publications@agu.org](mailto:publications@agu.org).

[Earth and Environmental data \(broad scope\)](#)

[Ecosystem CO<sub>2</sub>, water, and energy fluxes](#)

[Institutional/Project Repositories](#)

[Methane \(CH<sub>4</sub>\) and Nitrous Oxide \(N<sub>2</sub>O\) data](#)

[NASA – Funded Research \(selection of Distributed Active Archive Centers\)](#)

[NOAA - Funded Research](#)

[OMICs data](#)

[U.S. Department of Energy – Funded research](#)

[U.S. Geological Survey \(USGS\) – Funded Research](#)

[Water / Hydrologic data](#)

### Earth and Environmental data (broad scope)

**[Pangaea](#)** -Pangaea accepts any data from earth, environmental and life sciences.

When you start the data submission process, you will be redirected to the PANGAEA issue tracker that will assist you in providing metadata and uploading data files. Any communication with our editors will go through this issue tracker. For more details about the submission workflow see our tutorial.

Please note: All data and metadata are quality checked, harmonized, and processed for machine readability, which allows efficient and reliable re-usage of your data.

Depending on the extent and complexity of your data submission the editorial process and minting of DOI names might therefore take up to 8 weeks.

**[Environmental Data Initiative \(EDI\)](#)** - EDI is an NSF-funded repository accepting environmental research data and relevant processing code. Trained curators assist researchers from field stations, individual laboratories, and research projects of all sizes and actively promote and enable curation and re-use of environmental data

through outreach and training. EDI is committed to enable data that is Findable, Accessible, Interoperable, and Reusable through rich science metadata and assigning of DOIs. All data and metadata are quality checked and machine readable ensuring reliable reuse of data. [Data submission](#) is provided.

## **Ecosystem CO<sub>2</sub>, water, and energy fluxes**

**Ameriflux** - The AmeriFlux Network ensures the availability of the continuous, long-term ecosystem measurements necessary to build effective models and multisite syntheses, while maximizing insight through robust, site-specific, independent research programs. With these consistent, high-quality environmental measurements, AmeriFlux helps ensure that critical decisions are supported by the most complete understanding and data.

Independent scientists measure the flows of carbon between land and atmosphere, using a technique called eddy covariance, and then contribute their data to the AmeriFlux Network. The AmeriFlux Network works with scientists to standardize, quality check, and process data into common forms that the scientific community can use to examine crucial linkages between ecosystem processes and climate responses.

[Description of AmeriFlux data](#)

[How to upload data](#)

**ICOS** - The Integrated Carbon Observation System (ICOS) is a distributed pan-European research infrastructure producing high-quality data on greenhouse gas concentrations in the atmosphere, as well as on carbon fluxes between the atmosphere, the land surface and the oceans.

## **Institutional/Project Repositories**

**Bolin Centre Database** - The Bolin Centre Database is a storage and management facility for data collected and collated at the Bolin Centre for Climate Research. Most of the data are available with open access and can be used under the terms given in the data description. Our goal is to host all datasets produced within the Bolin Centre, to visualize the data and make the data publicly available. For inquiries contact: [bolindata@su.se](mailto:bolindata@su.se)

## **Methane (CH<sub>4</sub>) and nitrous oxide (N<sub>2</sub>O) data**

**MEMENTO** database includes methane (CH<sub>4</sub>) and nitrous oxide (N<sub>2</sub>O) data from the global ocean (both open and coastal). The MEMENTO database is administered by the Kiel Data Management Team at GEOMAR Helmholtz Centre for Ocean Research and supported by the German BMBF project SOPRAN ([Surface Ocean Processes in](#)

[the Anthropocene](#)). The database is accessible through the [MEMENTO webpage](#). A login is required to access the data. Learn more [here](#).

## **NASA – Funded Research (selection of Distributed Active Archive Centers)**

**[ORNL DAAC](#)**- The Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) for Biogeochemical Dynamics is a NASA Earth Observing System Data and Information System (EOSDIS) data center managed by the Earth Science Data and Information System (ESDIS) Project. The ORNL DAAC is operated by Oak Ridge National Laboratory (ORNL) in Oak Ridge, Tennessee, and is a member of the Remote Sensing and Environmental Informatics Group of the Environmental Sciences Division (ESD).

The ORNL DAAC is a CoreTrustSeal Certified Repository.

## **NOAA - Funded Research**

**[National Centers for Environmental Information \(NCEI\)](#)**, formerly National Climatic Data Center, the National Geophysical Data Center, and the National Oceanographic Data Center.

NOAA's NCEI hosts and provides access to one of the most significant archives on earth, with comprehensive oceanic, atmospheric, and geophysical data. From the depths of the ocean to the surface of the sun and from million-year-old ice core records to near-real-time satellite images, NCEI is the Nation's leading authority for environmental information. [Contact information](#).

## **OMICs data**

**National Center for Biotechnology Information (NCBI) [Sequence Read Archive \(SRA\)](#)** - The SRA is NIH's primary archive of high-throughput sequencing data and is part of the International Nucleotide Sequence Database Collaboration (INSDC) that includes at the NCBI Sequence Read Archive (SRA), the European Bioinformatics Institute (EBI), and the DNA Database of Japan (DDBJ). Data submitted to any of the three organizations are shared among them.

SRA accepts data from all kinds of sequencing projects including clinically important studies that involve human subjects or their metagenomes, which may contain human sequences. These data often have controlled access via [dbGaP \(the database of Genotypes and Phenotypes\)](#) .

Details for [submitting to SRA](#)

**Metagenomics Rapid Annotation using Subsystem Technology (MG-RAST)** - The metagenomics RAST server is a public resource for the automatic phylogenetic and functional analysis of metagenomes. Information [about the services](#).

### **U.S. Department of Energy – Funded research**

**ESS-DIVE** - The [U.S. Department of Energy's \(DOE\)](#) Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) is a data repository for Earth and environmental science data. ESS-DIVE stores and publicly distributes data from observational, experimental, and modeling research funded by the DOE's Office of Science under its [Subsurface Biogeochemical Research \(SBR\)](#) and [Terrestrial Ecosystem Science \(TES\)](#) programs within the Environmental Systems Science (ESS) activity.

Information for [Data Upload](#).

### **U.S. Geological Survey (USGS) – Funded Research**

**ScienceBase** - The primary data archive for recent and ongoing USGS investigations ScienceBase provides access to aggregated information derived from many data and information domains, including feeds from existing data systems, metadata catalogs, and scientists contributing new and original content. ScienceBase is a USGS Trusted Digital Repository.

[Reference here](#) for information on preparing your data for public release on ScienceBase.

**National Water Information System** - Water information is fundamental to national and local economic well-being, protection of life and property, and effective management of the Nation's water resources. The USGS works with partners to monitor, assess, conduct targeted research, and deliver information on a wide range of water resources and conditions including streamflow, groundwater, water quality, and water use and availability.

### **Water / Hydrologic data**

**CUAHSI** - Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) is a 501(c)(3) research organization representing more than 130 U.S. universities and international water science-related organizations. CUAHSI receives support from the National Science Foundation (NSF) to develop infrastructure and services for the advancement of water science in the United States.

Information about [submitting data](#).

The **Biological and Chemical Oceanography Data Management Office (BCO-DMO)**, <https://www.bco-dmo.org/> is a publicly accessible earth science data repository

created to curate, publicly serve (publish), and archive digital data and information from biological, chemical and biogeochemical research conducted globally in coastal, marine, great lakes and laboratory environments.

1. The BCO-DMO repository provides data management services at no additional cost to investigators funded through the following programs: NSF OCE Division Biological and Chemical Sections; NSF Division of Polar Programs Antarctic Organisms & Ecosystems; Gordon and Betty Moore Foundation Marine Microbiology Initiative (GBMF MMI). In certain circumstances, the office can provide data management services for a fee to investigators funded outside of these programs. Please contact the office at [info@bco-dmo.org](mailto:info@bco-dmo.org) to learn more.
2. BCO-DMO works closely with individual investigators and data originators throughout the data life cycle, from data management planning support, quality control and metadata assembly, to DOI creation and archive with appropriate national facilities. The office ensures all contributed project data and metadata are in compliance with current funder policies (i.e., NSF OCE Sample and Data Policy, NSF 17-037) and offers investigators the option to embargo data (in accordance with funder policies) until publication. Dataset DOIs are obtained and available once investigators review final curated data packages for publishing. Dataset DOIs, once generated, may be used for scholarly publication and/or funder reporting. BCO-DMO accepts scholarly publication DOIs and can link these to their respective datasets.
3. Data accepted by BCO-DMO include all project output (observational data, derived and statistical products, analysis code, software and models, and supporting documentation such as reports and calibration information). The office accepts a wide variety of data types and formats, and works to publish a non-proprietary, research-ready, data package available to new research. To contribute data to BCO-DMO, please see the "[How to Get Started Contributing Data](#)" page, located under the Resources tab of the BCO-DMO website.

## **Incorporating Data Management into your Research**

As you design and conduct your research, the data, software, model code and other outputs you and your team need and use will likely evolve. This information is commonly documented in a Data Management Plan required by your funder and should be kept up to date as you conduct your research.

To make publication straightforward, here are a few tips to follow:

1. Track the original location of data used. Make sure this is the source where the data are managed and not a copy.
2. Determine the usage license of any data you plan to use that are already published. If you can't determine the usage license, attempt to seek permission to use the data. Without clear permissions, the editor has the discretion to publish the research. Commercial data are the most problematic and are handled on a case-by-case basis. Keep in mind that usage permissions are different from "how open" and accessible the data are. Data requiring access protection and that are not fully open are still usable in your research. You, as an author, must have permission to use the data you select for your research.
3. Document all processing or changes you make to the data. This is sometimes called "the provenance" and will provide integrity to your research. For instance, if you decide to remove a data point that is an outlier, you should document that decision and explain why it was removed. If there were any steps you took to integrate several data sets, those steps and any decisions made concerning the approach should be documented. For reproducibility we suggest that you capture your steps in a script that allows you, or other readers, to rerun them, starting from the beginning, with the original data sets. Should you decide to create a script, consider publishing it in a repository designed for that purpose such as [protocols.io](https://protocols.io). Your workflow/script will be registered with a persistent identifier and can be considered an important valuable product of research promoting reproducibility and transparency.
4. Maintain a location for your data that has the "original" version, for ease of reproducibility. This will also be valuable should you decide to adjust your research design.
5. Determine which repositories you will be using and **contact them as soon as possible**. Your Data Management Plan likely has a location where your selected repository(ies) should be listed. During the writing of your proposal is the optimal time to contact them. They can clarify any specific guidance you will need for tracking and documenting your data (the metadata) using community accepted standards. They can also assist with determining any costs for depositing or curating your data, estimate the time you should plan in your schedule and assist with recommendations if you have questions about your Data Management Plan. By partnering with your repository early you get the benefit of incorporating your data management tasks during your research, when it is much easier, then waiting until the very end when you may be constrained by resources to prepare for publication.

## Preparing Your Data for Preserving in a Repository

Before you submit your work to a publisher, here are the steps to take to prepare your created (re: raw, prime) or processed (re: aggregated, synthesized, ancillary) data.

For **Created Data** going to a domain repository:

1. Get in touch with your repository as soon as possible and let them know you are preparing to deposit your data in preparation for scholarly publication. This is especially important for large or complex files, or if this is your first time preparing data for deposition to a domain repository. With their help, determine how long the data publication process usually takes; this will help with planning your publication time frame. Your data should be published as close to the time your paper is published. Some repositories prefer that the paper be published first, then your data. You are responsible for coordinating with your repository to ensure this happens smoothly.
2. Ensure you understand the data preparation guidelines recommended by your selected repository. These guidelines concern the file format, vocabulary for each column in your data, and the metadata categories that describe your data.

For **Processed Data** supporting your research findings and visualizations - These are usually aggregated/synthesized data that need to point back to the original data used in this analysis (via citation, provenance information) and may or may not be acceptable to your selected domain repository:

1. Contact the domain repository where your created data (also called raw or primary data) will be submitted. It is best if all of your data can be deposited in one repository if possible. If your domain repository won't take your process data, continue to work with them on your created data, and consider a general or institutional repository for your processed data. If your research did not include any created data, then select a general or institutional repository for our processed data.
2. Link your data products with your publication and other relevant research products. It is important that there is a link between these data and the other research objects produced. Domain repositories will usually prompt you for this information. When using general repositories, you will need to ensure that those links are captured in your metadata. Common links include your



ORCID and the DOI for your publication. Most repositories will allow you to add the DOI as a follow-on step to data deposition. The journal can provide you with your DOI at the time of acceptance. Other important links include the data from which the processed data originated. For instance, if you create a derived data product that might be useful to others, you should link the derived data with the original data files and the workflow/code that you used to get to the derived data product. This practice supports transparency and reproducibility of your research.

## Selecting Your Repository

For publishing, you need to locate a repository that provides preservation services. What this means is that:

1. The repository registers your data with a persistent identifier that is globally unique such as a Digital Object Identifier (DOI)<sup>1</sup>.
2. The data are accessible from a landing page that provides information (e.g., metadata) about your data, and preferably version controlled.

Once published in an appropriate repository, your data cannot be modified, although many repositories support version management that includes documentation on the exact changes and directional information on the landing page to alert the viewer to the most recent version. There are some datasets that are updated over time [re: dynamic data] associated with longitudinal studies or have some other special handling [e.g., genomics]. We recommend you be in contact with the repository to understand their preservation practices and how they support the community and journal requirements.

---

<sup>1</sup> A Digital Object Identifier (DOI) is an alphanumeric string assigned to uniquely identify an object. It is tied to a metadata description of the object as well as to a digital location, such as a URL, where all the details about the object are accessible. It provides an *actionable, interoperable, persistent* link

- *Actionable* – through the use of identifier syntax and network resolution mechanism (Handle System®)
- *Persistent* – through combination of supporting improved handle infrastructure (registry database, proxy support, etc) and social infrastructure (obligations by Registration Agencies)
- *Interoperable* – through the use of a data model providing semantic interoperability and grouping mechanisms

## When to Make Your Data or Software Available - The Timing with your Paper

At the time your paper is accepted, your data and software should be publicly available. If a repository will not publish your data or software until the paper is officially published, AGU may accept your commitment to publish that data or software after the paper is published. As the creator of the data, it is your responsibility to ensure that your data and software are available in this case. Failure to follow through with this commitment may be considered misconduct and could result in a retraction.

Ideally, the following is expected:

### 1. Paper submission:

- a. Data Availability Statement (required): You need a data availability statement in the Open Resources Section of your paper describing where your data are preserved. The availability statement should include the persistent identifier registered by the repository for your data. You should already be in the process of data preservation so you can provide access for peer review. Most repositories provide data access for this purpose.
- b. Data Citation (required): Include data citations in the References Section of your paper for your primary data, processed data, and any data used from another source.
  - i. **Primary Data and Processed Data**: Your selected repository should provide the ability to “reserve a DOI” before your data being published. Use this DOI in your citation. Once your data are published, it will resolve properly.
  - ii. **Data used from another Source**: These data might be located in a paper, or a repository. Cite the appropriate source. If the data are associated with a data paper, we recommend citing both the paper and the repository.
- c. Software Availability Statement (optional): If software is central to your research, you likely need a software availability statement in the Open Resources Section of your paper that describes where your software is preserved.
- d. Software Citation (optional): If software is central to your research, include a software citation in the References Section of your paper.

### 2. Paper Peer Review

- a. Data: Your data must be available for peer review. Here are options to ensure confidential access to your data.
  - i. **Preserve your data in a repository and make it available for peer review**. Depending on the repository, this can be done in a couple ways:

1. Provide a **temporary private link (“share link”)** in the last sentence of the Open Resources section of your paper. This link will not be present in your published paper as it is not a persistent link. This option allows your data to remain private until your paper is accepted. Here is an example of the format for a share link used by the MagIC repository:  
<https://earthref.org/MagIC/16724/f361947b-e8bd-4db0-8792-faafc89c6187>
2. Provide the **persistent identifier** (e.g., DOI) for your data. This option is used when your data have completed the repository submission process and is now publicly available. Using our example from the MagIC repository above, this is the DOI registered for the data: 10.7288/V4/MAGIC/16724
  - ii. **Include your data in the supplementary information of your paper, only for the purpose of peer review.** The supplement is not a repository and can only be used to support the peer review process. You must still submit your data to a repository before paper acceptance.
- b. Software: For papers where software is central to your research, your software must be available for peer review. The options for providing access to your software are the same as for data.

### 3. Paper Acceptance:

- a. Data: To the best of your ability, all data used for your paper should be accessible at the time your paper is accepted. Note the possibility that the repository policy won't allow your data to be published until your paper is published. If that is the case, AGU will accept that your data will be made available just after the moment your paper is published. It is your responsibility to coordinate with the repository to ensure availability of your data.
- b. Software: For papers where software is central to your research, your software should be accessible at the time your paper is accepted.

## Availability Statements and Examples

### Data Availability Statement:

For each dataset that supports your research, both a citation and a data availability statement must be present. The data availability statement for each data set must be included in the Open Research section of your paper indicating where readers can access the data. See the information on data citation for additional guidance. The availability statement should include an in-text citation, licensing information and access restrictions. Statements to the effect of "data available from authors" are not acceptable.

## Common templates for data availability statements

1. **For data stored in a repository:** Datasets for this research are available in these in-text data citation references: Smith et al. (2019), [with this license, and these access restrictions if any], Jones et al. (2017) [with this license, and these access restrictions if any].
2. **For data published in the literature:** Datasets for this research are included in this paper (and its supplementary information files): [citation for paper] or point to where the references are compiled.
3. **For technical reports publishing the description of a dataset and its preparation, e.g., a data paper:** Datasets for this research are described in this paper: [citation for paper, with this license, and these access restrictions if any].
4. **For theoretical papers, or most review papers:** Data were not used, nor created for this research.
5. **For data not publicly available, but available to researchers with appropriate credentials:** Data for this research are not publicly available due to [Fill in reasons]. Data are stored in this in-text data citation reference: Smith et al. (2019), [with this license, and these access restrictions if any].
6. **For data that are restricted by commercial, industry, patent, government policies, regulations or laws:** Data supporting this research are available in [cite in-text data citation reference from third party source], with [these restrictions that include information concerning required NDA, licensing, agreements], and are not accessible to the public or research community. [Provide process for how other researchers can gain access.] NOTE: If your data are in this category, the editors will determine if this statement meets the AGU data guidelines sufficiently.

### Software Availability Statement:

If your software is critical to your research, it should be preserved in a repository with both a citation and a software availability statement. The software availability statement must be included in the Open Research section of your paper indicating where readers can access the software. See the information on software citation for additional guidance. The availability statement should include an in-text citation, licensing information and access restrictions.

### Common templates for software availability statements:

1. **For software stored in a repository:** Software for this research is available in these in-text data citation references: Smith et al. (2019), [with this license, and these

access restrictions if any], Jones et al. (2017) [with this license, and these access restrictions if any].

2. **For software published in the literature as supplementary information:** Software for this research is included in this paper (and its supplementary information files): [citation for paper] or point to where the references are compiled.
3. **For software not publicly available, but available to researchers with appropriate credentials:** Software for this research is not publicly available due to [Fill in reasons]. Software is stored in this in-text citation reference: Smith et al. (2019), [with this license, and these access restrictions if any].
4. **For data that are restricted by commercial, industry, patent, government policies, regulations or laws:** Software supporting this research are available in [cite in-text citation reference from third party source], with [these restrictions that include information concerning required NDA, licensing, agreements], and is not accessible to the public or research community. [Provide process for how other researchers can gain access.] NOTE: If your software is in this category, the editors will determine if this statement meets the AGU guidelines sufficiently.

## Data citation

Your data citation(s) should include the data used in your paper. This may include data that others have created, new data as a result of your research, and processed data used for your analysis. It is especially important that new data are placed in a domain repository. For guidance on how best to format a compliant data citation along with examples, reference [ESIP's Data Citation Guidelines For Earth Science Data](#).

Examples:

1. Cline, D., R. Armstrong, R. Davis, K. Elder, and G. Liston. 2003. CLPX-Ground: ISA snow depth transects and related measurements ver. 2.0. Edited by M. A. Parsons and M. J. Brodzik. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/10.5060/D4MW2F23>. Accessed 2008-05-14. \*Reproduced from ESIP
2. Maslanik, J. and J. Stroeve. 1999, updated daily. Near-Real-Time DMSP SSMIS Daily Polar Gridded Sea Ice Concentrations, Version 1. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/10.5067/U8C09DWVX9LM>. Accessed 2019-02-14. \*Reproduced from ESIP
3. Lynch, L., M. Machmuller, C. Boot, T. Covino, C. Rithner, et al. 2019. Dissolved organic matter chemistry and transport along an Arctic tundra hillslope, Imnavait Creek

Watershed, Alaska, 2018. Arctic Data Center. <https://doi.org/10.18739/A2RF5KF5N>. Accessed 2019-02-28. \*Reproduced from ESIP

4. Moschetti, M. P., 2017, Database of earthquake ground motions from 3-D simulations on the Salt Lake City of the Wasatch fault zone, Utah: U.S. Geological Survey data release. <https://doi.org/10.5066/F7V98691>. Accessed 2019-02-28. \*Reproduced from ESIP

Data Citation Source Material: ESIP Data Preservation and Stewardship Committee (2019): Data Citation Guidelines for Earth Science Data, Version 2. [ESIP Online resource](#).

### **Guidelines for authors where research is primarily based on models**

When the primary data for the research comes from model simulations, follow these guidelines:

1. Citation of the model (most important).
  - BEST OPTION (model in repository): Cite the model using a repository that registers the version used for the paper with a persistent identifier (e.g., Digital Object Identifier) and metadata that describes the model using community standards. If a published paper has the complete description, there should be a link in the repository to the published paper. Your citation should accurately capture the authors/creators of the model.
  - GOOD OPTION (model described in paper): Cite the publication where the model is described with information about the version used for this paper.
2. Description of the model.
  - Include a description of the model in the text of the paper that is adequate to support reproducibility. If a publication describes the model thoroughly, cite that paper.
3. Information about the configuration/parameters used to run the model.
  - This information should be included in the paper text as well as providing any script/workflow used. The script/workflow should be preserved in a repository and cited. Any forcing datasets used should be described and cited.
4. Data that Supports the Summary Results, Tables and Figures.
  - BEST OPTION: Cite a package in an appropriate repository that includes scripts/workflows, provenance information, and summary files that support the research, figures and tables, consistent with

archives maintained for transparency and traceability by assessments such as the IPCC.

- GOOD OPTION: Cite files (e.g., scripts, descriptive detail) in an appropriate repository that support evaluating the research and provide the details behind the tables and figures.
- ACCEPTABLE OPTION: Provide the necessary information for transparency and traceability of the analysis using your community standards or guidance.

5. Model Output Data (optional).

- If certain model output data are instrumental to evaluating the research, then deposit these in a trusted repository. There are currently limited resources for preserving files of very large size. Selecting representative output from one or a few model runs as is recommended by a specific community may be necessary.

If the model is not open because of the sensitivity of the research or proprietary concerns, then provide as much information as possible to support evaluation of the research and reproducibility.

## Software citation<sup>2</sup>

If your research is heavily dependent on software (e.g., code, workflow, model, code packages) you may be asked by the journal editor to share your software and provide a citation to a preserved version.

### Checklist for citing software

1. Identify and cite the software (including your own) which makes a significant and specialized contribution to your academic work.
2. Check if the software has a recommended citation from the creators and use it if available. If the recommended citation is to a paper, then also cite the software directly.
3. Create as complete a citation as possible if no recommended citation is available. Include the software creator, when it was created, the title of the software (and version if available) and where the software can be accessed (preferably via a persistent identifier to an archival repository).
4. Reference the software appropriately, in compliance with citation formatting guidelines.

---

<sup>2</sup> This guidance is adapted for AGU's journals from the Software Citation Checklist for Authors developed by members of the Force11 Software Citation Implementation Working Group: Chue Hong, Neil P., Allen, Alice, Gonzalez-Beltran, Alejandra, de Waard, Anita, Smith, Arfon M., Robinson, Carly, ... Pollard, Tom. (2019, Oct. 15). Software Citation Checklist for Authors (Version 0.9.0). Zenodo. <http://doi.org/10.5281/zenodo.3479199>

## **What software should be cited?**

You should cite software that has a significant impact on the research outcome presented in your work, or on the way the research has been conducted. If the research you are presenting is not reproducible without a piece of software, then you should cite the software. Note that the license or copyright of the software has no bearing on whether you should cite it.

This may include:

- Software (including scripts) you have written yourself to conduct the research presented.
- A software framework / platform that is critical for your software, used to conduct the research, to function properly.
- Software packages, plugins, modules and libraries used to conduct your research and that perform a critical role in your results.
- Software you have used to simulate or model phenomena/systems.
- Specialist software (not considered commonplace in your field) used to prepare, manage, analyze or visualize data.
- Software being evaluated or compared as part of the research presented
- Software that has produced analytic results or other output, especially if used through an interface.

In general, you do not need to cite:

- Software packages or libraries that are not fundamental to your work and that are a normal part of the computational and scientific environment used. These dependencies do not need to be cited outright but should be documented as part of the computational workflow for complete reproducibility.
- Software that was used during the course of the research but had no impact on research results, e.g., word processing software, backup software.

## **How should software be cited?**

Software should be cited in the list of references, the same way as any other research object. To identify what citation should be used (i.e., whether it is a specific piece of text, a specific paper, or direct citation of an archive or repository), follow these steps:

1. Determine if the software developers provided a mandatory or recommended citation(s). If so, use it.
  - These citations are often found in a README file, a CITATION file, a CITATION.cff or codemeta.json metadata file, on the software's website, or in its documentation.



- In some languages and software platforms (e.g., R), a command can be used to generate the recommended citation. If there is a mandatory or recommended citation, use it.
2. If there is no mandatory or recommended citation provided, use the general principles that a reference should include the following: who, when, what, where. This is similar to the guidance for data.
- **Who:** Name the project as the author, unless the individual authorship of the software is clear (e.g., single developer).
  - **When:** The release date of the version you are using or the date you accessed/downloaded the software if using an unreleased version or one where the release date is unclear.
  - **What:** The name of the software, along with specific version / release information. This should be as specific as possible, for instance the name of a package, program or library rather than the platform or programming language it runs on.
  - **Where:** A DOI, URL or other identifier that points to the location of (ideally) a landing page for the software release, or else directly to the software itself.
    - i. This might be a DOI pointing to an archive in a digital repository, a URL pointing to the code repository, or a URL pointing to the website for the software.
    - ii. Persistent identifiers to archival repositories are preferred over URLs which may change.
3. A Software Availability Statement is required by AGU to describe the location where the software is preserved. This should include:
- Title of the software
  - Repository location
  - Additional information needed to access the software, such as sensitivity and security issues required by a government, or other entity.
    - i. \*This does not replace the citation, but further clarifies access to the software.

If you are the software author, you should follow this guidance to generate a suitable citation for your software and put your citation in your software's documentation.

Examples:

**JGR Space Physics Publication:**

**Paper:** Gallant, M.A., Mierkiewicz, E.J., Nossal, S.M., Qian L., Burns A.G., Zacharias A.R., Roesler F.L. (2019). Signatures of thermospheric-

exospheric coupling of hydrogen in observed seasonal trends of H  $\alpha$  intensity. *Journal of Geophysical Research: Space Physics*, 124, 4525-4538 <https://doi.org/10.1029/2018JA026426>.

**Software Availability Statement:**

Radiative transport executables used in this study are available from LENSES (2019a), and Pine Bluff Observatory Fabry-Perot spectrometer data are available from LENSES (2019b).

**Software Citation in Paper Reference Section:** LENSES (2019a).

lenses-lab/LYAO RT-2018JA026426: Original release.

<https://doi.org/10.5281/zenodo.2598836>

**(Improved) Zenodo Recommended Citation:** Lab for Exosphere and

Near Space Environment Studies. (2019, March 20). lenses-

lab/LYAO\_RT-2018JA026426: Original Release (Version 1.0.0).

Zenodo. <http://doi.org/10.5281/zenodo.2598836>

Please reference the Force11 Software Citation Implementation Group [Software Citation Checklist for Developers](#) for additional information.

**For Software Developers using development platforms:**

**When Using GitHub** (using embedded Zenodo connection)

GitHub is integrated with [Zenodo](#), a general repository. **GitHub** provides a step by step process to obtain [a DOI for your software](#) that supports citation. Once you have completed the process provided in the link with Zenodo, double-check your citation and make any needed updates to authors, titles or other information. In brief:

1. Finalize your software in your GitHub repo.
2. Follow [the steps provided by GitHub and Zenodo](#) to obtain a DOI.
3. Review the Zenodo citation to ensure it is correct. Update as needed.

**When Using BitBucket, GitLab, SourceForge**

These tools do not currently have a partnership with a preservation repository. If you are using these tools, we recommend making an archive file (re: tar file) of the version used for your research and placing these files in a general repository in order to preserve your work and have a proper citation in your paper. Examples of General Repositories: [Zenodo](#), [Dryad](#), [Figshare](#) or your institutional repository that has a persistent identifier registration service and provides a recommended citation.

