Journal Production Guidance for Software and Data Citations

27 December 2022

Journal Production Guidance for Software and Data Citations

Abstract

Software and data citation are emerging best practices in scholarly communication.

This article provides structured guidance to the academic publishing community on how to implement software and data citation in publishing workflows. These best practices support the verifiability and reproducibility of scientific results; sharing and reuse of valuable data and software tools, and attribution to the creators of the software and data.

While data citation is increasingly well-established, software citation is rapidly maturing. With the current intensive use of software, including specialized tools and models for scientific research problems, the research community has begun to recognize that software, as a key research result and resource, requires the same level of transparency, accessibility, and disclosure as data.

Software and data that support scientific results should be preserved and shared in scientific repositories for discovery, transparency, and use by other researchers. These goals can be supported by citing these products in the Reference Section of papers and effectively associating them to the software and data preserved in scientific repositories. Publishers need to mark up these references in a specific way to enable downstream processes, specifically those that enable automated attribution.

Academic publishers wishing to stay current with best practices in the field are encouraged to follow the guidance provided here.

Introduction

Software and data citation are emerging best practices in scientific communication that provide excellent support for scientific results validation, reproducibility, credit, sharing and reuse of valuable tools and resources. Data citation has been possible with some rigor since the establishment of DataCite in 2009 (Brase, 2009), and was recommended by a comprehensive report of a CODATA-ICSTI task group in 2012 (Uhlir P., 2012). It has become increasingly adopted since the introduction of the “Joint Declaration of Data Citation Principles” (Data Citation Synthesis Group, 2014) in 2014 and its endorsement by 125 publishers, academic institutions, and funders. The enthusiasm for these principles prompted an effort to provide a clear set of practical guidelines for publishers to begin to implement data citations. In 2018 “A data citation roadmap for scientific publishers” (Cousijn et al., 2018) was published, incorporating work from several groups via workshops and including major participation by representatives of Elsevier, Springer-Nature, PLOS, eLIFE, JISC, EMBO Press, CrossRef, and Wiley. However, even when following this guidance, several publishers were finding that machine-readable software and data citations from their own published papers were not preserved intact when the paper was published, and that downstream services and linking were not enabled. Furthermore, it should be noted that software

usage and citation, while similar to data citation, has certain key differences that should be reflected in publication workflows. These are complex and involve authors, reviewers, editors, and other infrastructure services. Here we provide updates on this guidance to enable automated attribution and credit for software and data used in published papers.

Funders and advisory groups are now beginning to require data and software citation in order to connect publications with their supporting software and data (Wellcome, 2017). Notably, the Intergovernmental Panel on Climate Change (IPCC) enhanced the traceability of its recent Assessment by the implementation of the FAIR Guidelines (Pirani et al., 2022), which extends IPCC's Assessment process by the documentation, long-term preservation, and citation of the assessed digital information, including both data and software. Recently, the White House Office of Science and Technology (OSTP) Memorandum "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research" (Nelson, 2022) has directed federal agencies to develop implementation plans around open access for research results attributed to grants, including to the software and data.

Recommended software citation practices have more recently been clarified by the FORCE11 Software Citation Implementation Working Group in "Recognizing the value of software: a software citation guide" (Katz et al., 2021). With the current intensive use of software including specialized tools and models for scientific research problems, software has evolved to become a key research resource requiring transparency and disclosure to the readers of any scientific article. With work by the FORCE11 Software Citation Implementation Working Group, citation of software can now be consistently implemented by publishers who follow the guidelines in this paper, establishing both data citation and software citation as important elements in peer-reviewed papers.

Long-term preservation and sharing of software and data that support scientific results in trustworthy scientific repositories provide for discovery, transparency, validation, and re-use by other researchers. Publishers should provide the means to associate these products to research papers through citation in the references section, using the structured implementation guidance for software and data citation in publishing workflows, provided in this article. Software citation, like data citation, provides a direct path to FAIRness for these essential research components (Groth et al., 2020). The updated guidance provided here includes recent improvements in the practices, challenges publishers are encountering, and recommendations for addressing those challenges.

Software and Data Citations: Checklist of Best Practices for Journal Publishers
Journal publishers should make best efforts to undertake the following practices, on which we provide further detailed guidance in this article.

1. **Instructions to Authors:** Provide author instructions that define which software and datasets should be cited in their paper, how to structure these citations, information on the best possible scientific repositories to use for software and data, and what information to put in an Availability Statement.
2. **Publication Policies:** Update publication policies to include the importance of citing the software and datasets that support the transparency, integrity, and reproducibility of the research.
3. **Education and Outreach:** Educate editors, reviewers, and staff on the policy, requirements, areas of flexibility, and unique aspects of software and dataset citations.

Journal Production Guidance for Software and Data Citations

27 December 2022

4. **Technical Updates:** Put into place the necessary technical updates, as defined in this document, to ensure that the machine-readable representation of the software and dataset citations is sustained throughout the publication workflow and properly formatted when registered to Crossref.
5. **Production Teams:** Define for publication production team members the unique aspects of software and dataset citations; work together to implement necessary changes.
6. **Metrics:** Establish quality metrics and measures to ensure consistent, accurate results for software and dataset citations.

This document is the product of working sessions of the FORCE11 Software Citation Implementation Working Group's Journals Task Force, conducted over a period of 2 years. It reflects lessons learned from a project supported by an NSF Grant (2025364) to the American Geophysical Union to ensure that software and data citations were transferred reliably from publications to NSF's Public Access Repository. It provides best practices to journals such that the machine-readable representation of the software and dataset citations is sustained throughout the publication workflow and properly formatted when deposited to Crossref. This optimizes their ability to be machine actionable and linked as research objects in support of services provided by DataCite, Scholix, and others.

The guidance here is primarily intended for those who support the journal production process, including those engaged in copy editing, markup, and preparing the paper for publication on various types of digital and print platforms. To implement these recommendations, coordination with journal submission systems, authors guidelines, editors, reviewers, authors, and others involved with the front-end process of publications will also need to be included. Since journal production depends on a number of activities that take place both prior and subsequent to production in the publication workflow, we include brief descriptions of what is expected in those activities. The FORCE11 Journals Task Force includes both society and commercial publishers of all sizes, research infrastructure providers, and others interested in implementation challenges for software and data citations.

In this guide we describe a set of necessary journal-related activities, organized by role, along with what is needed for datasets and software that supports the paper to be identified and cited. We provide use cases for journals to consider as they ensure that their production processes are handling data and software citations optimally.

Problem Statement:

Data and software citations have unique use cases and workflows that distinguish them from article or book citations. In particular, software and data are often registered by DOI registry agencies other than Crossref.

We have discovered that at many publishers, when scripts in the journal production process attempt to validate these citations, they do not do so correctly, and as a result, the integrity of the machine-actionable link to the software or dataset is not maintained and not provided to Crossref. As a result, important downstream services are not supported.

Journal Production Guidance for Software and Data Citations

27 December 2022

The global research community is now advancing the importance of citing software and datasets in scholarly literature as a means to be more transparent and support steps towards better reproducibility and reusability of software and datasets. Data and software are research objects separate from the paper with the potential of reuse and citation in later works. Tracking these citations accurately so that authors/creators of these objects receive recognition is important and ensures the scholarly record is more complete.

Journal practices are influenced by other participants in the research ecosystem and the processes and services they provide. For this set of recommendations, these include Crossref (or other Digital Object Identifier Registries), Crossref's Event Data service, DataCite's Commons, and Scholix. These work together to associate software and datasets with publications for many other uses.

Crossref provides Digital Object Identifier (DOI) registration services for publishers of English language papers, their journals, and other publications. Using the Reference Section of each registered paper, Crossref Event Data records the relationship between the software and datasets cited in the paper with the paper itself in a database that other services can use. Event Data can also be populated if a relationship is asserted (with an identifier) to software or datasets in the journal metadata. Crossref has encountered challenges in providing consistent support to Event Data. In 2021, Crossref completed updates that fixed and improved these processes; however, incorrect publisher markup of the software and dataset citations still prevents full implementation.

DataCite, like Crossref, provides Digital Object Identifier (DOI) registration services. DataCite specializes in preserving scientific datasets, some preprint servers (e.g., arXiv.org), physical samples (e.g., IGSNs), and software (e.g., Zenodo). DataCite is the preferred DOI registry agency for datasets and software because of its robust identifier system, metadata, and support services.

DataCite Commons accesses Crossref Event Data as well as the metadata from both Crossref and DataCite's APIs, ORCIDs, and RORs to visualize the connections between works (e.g., paper, data, software), people, and organizations. For example, you can see the works that cite a particular dataset.

Scholix is an interoperability initiative that collects and exchanges links between research data and literature (Cousijn et al., 2019; Khan et al., 2020). Scholix is designed to support multiple types of persistent identifiers, although at the moment only DOIs are included. Both Crossref and DataCite feed DOIs to Scholix. If a data repository includes metadata about related publications that use a dataset, through the metadata included in the registry process DataCite will provide that information to Scholix. Similarly, if a paper includes a dataset in the Reference Section, Crossref will report that to Scholix. Scholix manages multiple entries when the same paper-to-dataset link is provided by two different sources. Software is not supported at this time.

Journal Production Guidance for Software and Data Citations

27 December 2022

Workflow Description

The following activities describe a typical publishing workflow, grouped by role, used to capture software and dataset citations properly in both human- and machine-readable formats to support linking and automated attribution and credit. The order of activities may be slightly different for each publisher's production process.

Author Preparation

1. **Use scientific, trustworthy repositories that register data and software with persistent, globally-resolvable identifiers.** Trustworthy repositories provide long-term curation and preservation data services guided by the TRUST (Transparency, Responsibility, User focus, Sustainability and Technology) principles (Lin et al., 2020). CoreTrustSeal provides a list of certified repositories that follow those principles (L'Hours et al., 2019). Specific dataset curation services may require time to review and prepare the data prior to publication. This is most common with repositories that specialize in a particular scientific domain or data type. The time needed for domain-specific data curation, the value-added process of making the dataset interoperable and more likely to be reused, may be up to several weeks or more. *Generalist repositories such as Dataverse, Dryad and Zenodo are typically far quicker* as they do not undertake domain-specific curation. *However, domain-specific repositories may be the norm for certain disciplines.* Journals should encourage researchers to be aware of this time constraint where applicable, and to plan accordingly. Well-curated data provides transparency and integrity to research. Tools like DataCite Commons or the CoreTrustSeal website support authors in finding a trustworthy discipline-specific repository.
2. **Include in the submitted paper both an Availability Statement for software and datasets and citations in the Reference Section paired with in-text citations in the paper.** Publishers should provide support with author guidance and journal templates. All software and datasets can be described in the Availability Statement, but frequently the information needed for a complete citation is not fully available.
    a. **An Availability Statement** is designed for the reader and states where the supporting datasets and software are located and any information about accessibility that is needed. Include an in-text citation that has a corresponding entry in the Reference Section. This is a statement on availability so that those looking to analyze or reuse datasets or software can easily find these objects. The information provided should lead the reader to the exact location where the software or dataset can be found. See Availability Statement Recommended Elements Section for more information.
    b. **Paired Citations and References**
        i. **In-text citations** in the Methods and/or Data Section that refer to corresponding entries (citations) in the Reference Section for the dataset or software. These in-text citations provide discrete linkage from scientific claims to the specific dataset or software supporting them and / or methods used in analysis.

ii. **Citation (in Reference Section)** for the dataset and software used in this research. These should be listed in the Reference Section. This allows for automated attribution and credit through Crossref's Event Data. See Techniques for identifying a data or software citation in your References Section for a suggestion on using a "bracketed description" in the citation format to make identification of data and software citations easier. Also, see Software and Data Citation Resources Section.

3. **Be aware that reference tools used by many authors may not have an item type for datasets or software yet.** In order to follow the best practices in this guide, further editing of the citation may be required. For example, Zotero does not support datasets, but does support software. Zotero provides a work around to support datasets.

Journal Staff Review

1. **Ensure the presence of a sufficiently descriptive and actionable Availability Statement** for both software and datasets supporting the research.
2. **Ensure that for all software and datasets identified in the Availability Statement there is a supporting citation and the statement is complete.** Journal review guidance is helpful (e.g., AGU's Data and Software Availability and Citation Checklist). The journal should promote the importance of a citation in the Reference Section as much as is feasible to encourage automated attribution and credit. Provide examples relevant to your journal. Citations should be to both newly created software and datasets as well as to software and datasets already deposited in repositories by other researchers and reused or referred to in the research by the authors. See Availability Statement Recommended Elements Section for more information.
3. **Provide guidance to authors recommending that software and datasets in the Availability Statement be preserved for the long term in a trustworthy repository** complying with the TRUST principles (Lin, 2020) including its inherent long-term commitment for sustainable operations. DataCite has enhanced their Commons tool to include repositories; In addition, CoreTrustSeal provides a list of certified repositories. Use keywords to discover potential repositories and review relevant characteristics like certification (Murphy et al., 2021).
4. **Ensure citation follows journal style guides.** Best practice is to include an APA-style "bracketed description" to easily identify the citation as a dataset, software, computational notebook, etc. This is important to inform automated and manual reference section review as well as accurate production markup. See Software and Data Citation Resources Section for more information.

Reviewer/Editor Evaluation

1. **Review the submitted paper to determine if the software and datasets referenced in the Methods and Data or Methods and Materials Section appears to fully support the research**, such that all has been accounted for.

2. **Review the Availability Statement to ensure all the software and datasets referenced are included.** Use the information provided to locate and observe the software and datasets. Ensure that the links resolve and the content is as described. Request clarity as needed. Ensure the Availability Statement is complete. See Availability Statement Recommended Elements Section for more information.

Publisher Responsibility

1. **Ensure the entire production process team, such as copy editors, markup vendors, and staff, are aware of the uniqueness of software and dataset citations.** This includes providing guidance, education, and support for questions. See the Software and Data Citation Resources Section for more information.
2. **Ensure the publication policies are current** with the best practices of including software and dataset citations using the TOP Guidelines (Nosek et al., 2016) and a Research Data Policy Framework for All Journals and Publishers (Hrynaszkiewicz et al., 2020) and extending to software as well.
3. **Implement quality assurance processes** for measuring software and data citation accuracy.
4. **Establish quality controls, measures and metrics** to track consistency and establish thresholds for when to take action when the measures indicate a degradation in quality.

Automated Verification Activity - Reference Section (Publisher or third-party vendor responsibility)

1. **Check software and dataset citations in the Reference Section.** When checking citations, note that software and dataset citations formats can include repository names and version numbers. Guidance from FORCE11 and RDA provides recommended formats. See the Software and Data Citation Resources Section for more information. The format for persistent identifiers for software and datasets can vary. DOIs are commonly registered through DataCite and potentially other DOI Registration Agencies. Consider using content negotiation to validate.
2. **Ensure that software and dataset citations are tagged accurately.** Avoid tagging them as “other”. Refer to JATS4R guidance for data citations (NISO, 2020) and software citations (NISO, 2021).
3. **Avoid removing citation information**, especially the persistent identifier.

Copy Editor Review (re: language and style editing)

Check that the software and datasets mentioned in the Methods and Data Section have corresponding entries in the Availability Statement and citations in the Reference Section. Note: not all data or software can be supported with a citation that includes a persistent identifier.

Journal Production Guidance for Software and Data Citations

27 December 2022

Production Markup Review (supports machine-readable formats)

1. **Methods/Data Section:** Check that in-text citations and/or text in the Availability Statement link correctly to citations.
2. **Availability Statement:** This text should include availability statements for all software and datasets that support the research. Currently most journals do not mark up this text. However, JATS4R now has a specific recommendation to do so.
3. **Citations:** Journals should review and update the markup requirements for dataset citation (NISO, 2020) and software citation (NISO, 2021). The persistent identifier or URL is an active link in the paper to the reference. Some publishers provide visualization tools of the software or dataset. Consult the Software and Data Citation Resources Section of this paper for information on formatting. Dataset and software research outputs should use the same journal citation style and treatment of the persistent identifier with slight adjustment to include software version information and bracketed descriptions.

Content Hosting Provider Activities

1. **Register the paper with Crossref and ensure metadata is sent to Crossref**, including the full reference list (and the Availability Statement when added to the Crossref schema). Ensure all citations are included in the file going to Crossref, and not being removed inadvertently. Use the Crossref Participation Report to check publisher metadata overview of what is provided to Crossref. Consult with I4OC for community recommendations on open citations.
2. **Use the Software and Dataset Citation Use Cases provided below as a guide to ensure coverage of most variations.**
3. **Display the human readable citation correctly.** Consult the Software and Data Citation Resources Section of this paper for information on formatting. Dataset and software research outputs should use the same journal citation style and treatment of the persistent identifier with slight adjustment to include software version information and possibly bracketed descriptions.
4. **Provide the machine-readable citation correctly** to downstream services (e.g., Crossref). It is important to note the guidance provided by Crossref in their blog (Lin, J., 2017) has since been amended. If a journal implemented this guidance previously, there is a high probability that a correction is needed. The Software and Dataset Citation Use Cases table below, specifically the Crossref information, includes those corrections. In short, if there is a DOI, include the DOI XML tag.

Crossref Activities

1. Receive the article XML from the **Content Hosting Provider/Publisher** and create the necessary Event Data entries for software and dataset citations included in the Reference Section.

We include below a list of specific software and dataset use cases with recommended JATS XML and Crossref submission record elements.

Journal Production Guidance for Software and Data Citations

27 December 2022

Following the use cases are recommendations for techniques for identifying a software or dataset citation in the Reference Section.

Software and Dataset Citation Use Cases

The table below lists the common use cases for software and dataset citations, the corresponding JATS XML elements, and Crossref Metadata depository schema elements to assist publishers with creating the necessary adjustments to their production systems that result in proper machine-readable and actionable citations. Use cases not included are listed below the table.

	Use Case Name, Description	**Desired Outcome for JATS XML version 1.3 (ANSI/NISO, 2021)***	**Desired Outcome for Crossref Metadata depository schema 5.3.1 (Crossref, 2021)**	**Notes:**
	Software Citation Use Cases			

A	**Software Citation with registered DOI,**
	
	A registered DOI has the desired metadata to support this citation and allows for content negotiation
		**<ref id="bib11">
		<element-citation publication-type="software">
		<person-group person-group-type="author">
		<name>
		<surname> Wang</surname> <given-names>Yafei</given-names> </name>
		...
		</person-group>
		<year iso-8601-date="2022">2022</year>**
			**Tags:
			//citation/unstructured_citation AND
			//citation/doi**
			
			e.g.
			<citation key="XXX">
			<unstructured_citation>Wang Yafei, …, 2022, pc4covid19//pharmacodynamics2GUI:
				**Pro:
				Crossref can update Event Data.
				At this time software is not included in Scholix.**

		<part-title>pc4covid19/pharmacodynamics2GUI: Version 2.0 </part-title> <source>zenodo</source> <pub-id pub-id-type="doi">10.5281/zenodo.3967327</pub-id> </element-citation> </ref>	Version 2.0, zenodo, 10.5281/zenodo.3967327</**unstructured_citation**> <**doi**>10.5281/zenodo.3967327</**doi**> </**citation**>	
B	**Software Citation with URL (not persistent) e.g., GitHub, GitLab, BitBucket,** **The community accepted location for the software does not use persistent identifiers. A URL is provided instead. No DOI registry services are available.**	**<ref id="bib3"> <element-citation publication-type="software"> <person-group person-group-type="author"> <name> <surname>Kotila</surname> <given-names>M</given-names> </name> … </person-group> <year iso-8601-date="2019">2019</year> <part-title>Numerical computing is fun</part-title> <source>GitHub</source> <ext-link ext-link-type="uri" xlink:href="https://github.com/eka-foundation/numerical-computing-is-fun" >https://github.com/eka-**	**Tags: //citation/unstructured_citation** **e.g. <citation key="XXX"> <unstructured_citation>Kotila M, …, 2019, Numerical computing is fun, GitHub, https://github.com/eka-foundation/numerical-computing-is-fun" >https://github.com/eka-foundation/numerical-**	**Con: Software is <u>not</u> in a trusted preservation repository. The URL is subject to <u>link rot</u> and it is <u>not FAIR</u>.**

		foundation/numerical-computing-is-fun</ext-link> </element-citation> </ref>	**computing-is-fun</unstructured_citation> </citation>**	
C	**Software Citation with Software Heritage PID,** **A registered PID has the desired metadata to support this citation and allows for content negotiation.**	**<ref id="bib2"> <element-citation publication-type="software"> <person-group person-group-type="author"> <name> <surname>Zhang</surname> <given-names>Martin</given-names> </name> </person-group> <year iso-8601-date="2021">2021</year> <part-title>tabula-muris-senis</part-title> <source>Software Heritage</source> <pub-id pub-id-type="swhid">swh:1:dir:7dc782970300a97e9bca9038ba34728c857a0638</pub-id> </element-citation> </ref>**	**Tag: //citation/unstructured_citation** **e.g. <citation key="XXX"> <unstructured_citation>Zhang Martin 2021, tabula-muris-senis, Software Heritage,swh:1:dir:7dc782970300a97e9bca9038ba34728c857a0638</unstructured_citation> </citation>**	**Con: Non-DOI PIDS are typically not converted to clickable links on journal websites and some journals encourage use of a URL as a preference.**

D	**Software Citation with other PID (e.g., RRID, ARKs, Handle, ASCL, SwMathID),** **A PID is used for the software but is not a DOI.**	**<ref id="bib2"> <element-citation publication-type="software"> <person-group person-group-type="author"> <collab>Broad Institute</collab> </person-group> <part-title>CellProfiler Image Analysis Software</part-title> <pub-id pub-id-type="rrid">SCR_007358</pub-id> </element-citation> </ref>**	**Tag:** **//citation/unstructured_citation** e.g. <citation key="XXX"> <unstructured_citation>Broad Institute, CellProfiler Image Analysis Software,SCR_007358</unstructured_citation> </citation>	**Con:** **Non-DOI PIDS are typically not converted to clickable links on journal websites.**
	Data Citation Use Cases			
1	**Data Citation with registered DOI,** **A registered DOI has the desired metadata to support this citation and allows for content negotiation.**	**Example:** **<ref id="d1"> <element-citation publication-type="data" > <person-group person-group-type="author"> ... <name> <surname>van Beethoven</surname> <given-names>Ludwig</given-names> </name> ... </person-group>**	**Tags:** • **//citation/unstructured_citation** **AND** • **//citation/doi** **e.g.** **<citation key="XXX"> <unstructured_citation>van Beethoven Ludwig, 2014,Title of dataset,**	**Pro:** **Crossref can update Event Data and Scholix.**

		<data-title>Title of dataset</data-title> <year iso-8601-date="2014">2014</year> <source>Repository Name</source> <pub-id pub-id-type="doi">10.1234/1234321</pub-id> </element-citation> </ref>	**Repository Name, 10.1234/1234321 </unstructured_citation> <doi>10.1234/1234321</doi> </citation>** • **</citation>**	
2	**Data Citation with unregistered DOI (DOI does not resolve),** **Some data repositories do not complete the DOI registration process until after the paper is published. The DOI is valid but will not resolve to a dataset yet. Likely a temporary "share link" was used for paper peer review. Make sure the temporary link is not accidentally included here, instead of the persistent identifier.**	**Same as above**	**Tags:** • **//citation/unstructured_citation** e.g. <**citation** key="XXX"> <**unstructured_citation**>van Beethoven Ludwig, 2014,Title of dataset, Repository Name, 10.1234/1234321</**unstructured_citation**> </**citation**>	**Con:** **Crossref rejects the submission when a DOI tag is used. Only the "unstructured" tag can be used.** **The DOI is registered after the paper is published, and can not be used to validate the citation.** **Crossref can NOT update Event Data.**

3	**Data Citation with URL (not persistent) (e.g., project website, ftp site),** **The community accepted location for the data does not use persistent identifiers. A URL is provided instead. No DOI registry services are available.**	**Use <ext-link> to provide a link directly to the data citation. If there is no pub-id, ext-link should be included.** **<ref id="bib11">** **<element-citation publication-type="data">** **<person-group person-group-type="author">** **<name>** **<surname>Radoshevich</surname> <given-names>L</given-names>** **</name>** **...** **</person-group>** **<year iso-8601-date="2015">2015</year>** **<data-title>Title of dataset</data-title>** **<source>ArrayExpress</source> <ext-link ext-link-type="uri" xlink:href="https://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-1243">https://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-1243</ext-link> </element-citation>** **</ref>**	**Tags:** ● **//citation/unstructured_citation** **e.g.** **<citation key="XXX">** **<unstructured_citation>Radoshevich L, …,2015,Title of dataset, ArrayExpress, https://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-1243,</unstructured_citation>** ● **</citation>**	**Con:** **Data are <u>not</u> in a trusted preservation repository.** **Scholix is <u>not</u> updated.** **The URL is subject to <u>link rot</u> and it is <u>not FAIR</u>.** **Metadata about the data are difficult to determine.**

Journal Production Guidance for Software and Data Citations

27 December 2022

4	**Data Citation with non-DOI PID (e.g., RRID, ARKs. Handle, Accession) Additional information in the Software and Data Citation Resources section below.** **A PID is used for the data but is not a DOI.**	**<ref id="bib11"> <element-citation publication-type="data"> <person-group person-group-type="author"> <name> <surname>Radoshevich</surname> <given-names>L</given-names> </name> ... </person-group> <year iso-8601-date="2015">2015</year> <data-title>Title of dataset</data-title> <source>ArrayExpress</source> <pub-id pub-id-type="accession" xlink:href="https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3649/">E-MTAB-3649</pub-id> </element-citation> </ref>**	**Tags:** • **//citation/unstructured_citation** e.g. <citation key="XXX"> <unstructured_citation>Radoshevich L, …,2015,Title of dataset, ArrayExpress, E-MTAB-3649, https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3649/</unstructured_citation> </citation>	**Con:** **Non-DOI PIDs are not included in Scholix. Non-DOI PIDS are typically not converted to clickable links on journal websites and some journals encourage use of a URL as a preference.**

*** The example used is <element-citation> but <mixed-citation> is equally acceptable.**

Use cases not included in this guidance:

1. Citation to a physical location such as a museum.
2. Mixed content citation that includes, for example, software, data, and other digital objects where individual attribution to a specific item is difficult or not possible.
3. Physical samples using IGSN, RRID, or other persistent identifiers.
4. Executable environments, for example, MyBinder, ERA, Code Ocean.
5. Datasets associated with different persistent identifiers, for example journal supplemental files for a journal article where the journal article DOI is used for citation.
6. Datasets and software together with an online software implementation associated with a publication are reviewed, published and preserved long-term, used data and software is cited and the provenance documented: IPCC Atlas Chapter with WGI Interactive Atlas and WGI Github Repository.

Journal Production Guidance for Software and Data Citations

27 December 2022

Software and Data Citation - Crossref Metadata Schema examples

Software Citation Example

Published Paper

Shumate A and Salzberg S. LiftoffTools: a toolkit for comparing gene annotations mapped between genome assemblies [version 1; peer review: awaiting peer review]. *F1000Research* 2022, 11:1230 (https://doi.org/10.12688/f1000research.124059.1)

Availability Statement (located in the 'Software Availability' Section in this paper)

Archived source code as at time of publication: https://doi.org/10.5281/zenodo.6967163 (Shumate, 2022)

Software Citation in Reference Section

Shumate, A. (2022). agshumate/LiftoffTools: (v0.4.3.2) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.6967163

Crossref Metadata

```
{
    "key": "ref6",
    "doi-asserted-by": "publisher",
    "unstructured": "Alaina Shumate, 2022, agshumate\/LiftoffTools:
    Version 0.4.3.2, zenodo, 10.5281\/zenodo.6967163",
    "DOI": "10.5281\/zenodo.6967163"
},
```

Data Citation Example

Published Paper

Zhang, Y., Li, X., Liu, Z., Rong, X., Li, J., Zhou, Y., & Chen, S. (2022). Resolution Sensitivity of the GRIST Nonhydrostatic Model From 120 to 5 km (3.75 km) During the DYAMOND Winter. In Earth and Space Science (Vol. 9, Issue 9). American Geophysical Union (AGU). https://doi.org/10.1029/2022ea002401

Availability Statement (located in the Open Research Section in this paper)

Data for supporting this study are available at: https://zenodo.org/record/6497071 (GRIST-Dev, 2022).

Data Citation in Reference Section

GRIST-Dev. (2022). DYAMOND winter of GRIST nonhydrostatic model (A21) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.6497071.

Crossref Metadata


```
{
    "key": "e_1_2_7_7_1",
    "doi-asserted-by": "publisher",
    "unstructured": "GRIST-Dev. (2022).DYAMOND winter of GRIST
    nonhydrostatic model
    (A21)[Dataset].Zenodo.https://doi.org/10.5281/ZENODO.6497071.",
    "DOI": "10.5281/ZENODO.6497071"
},
```


Techniques for identifying a data or software citation to properly tag it in the Reference Section metadata

Citations for data or software can be difficult to discern from journal citations. We offer these techniques for your consideration.

Request Authors to include Bracketed Descriptions in Data and Software Citations: As defined by the Publication Manual of the American Psychological Association, 7th Edition, bracketed descriptions “help identify works outside the peer-reviewed academic literature (i.e., works other than articles, books, reports, etc.), provide a description of the work in square brackets after the title and before the period.”

Benefit: Easier identification of data or software citations in the Reference Section.

Challenge: At this time there is no broadly accepted standard list of terms and adding a bracketed description for data and software citations adds to the burden on authors and likely will not be consistent.

Request Authors to include Availability Statement as a Guide to Reviewing the Reference Section for Data and Software Citations.

Benefit: A complete Availability Statement that has been validated through journal staff and/or peer review can be used to determine which citations in the Reference Section are data or software.

Challenge: Staff, reviewers, and authors need clear guidance on what should be included in the Availability Statement and examples of data and software citations. The publication process should include steps to review and provide guidance to authors on completing or improving the Availability Statement. See the Availability Statement Recommendations Section for a list of elements to include.

Use Content Negotiation on machine-readable metadata from the repository landing page.

Benefit: Crossref and DataCite provide a tool that validates Digital Object Identifiers (https://citation.crosscite.org/docs.html) and returns information that includes the "type" of content registered, such as "data set" or "software." Content Negotiation often also works with references to URLs, especially when requesting Schema.org ("application/ld+json") metadata. Many repositories also have machine-readable metadata in Schema.org format on their landing page, which makes it a universal way to request metadata and type (data / software / creative work).
Challenge: Often, content negotiation only works for DOIs registered with Crossref or DataCite. Additionally, for generalist repositories, the "type" information might not be accurate. Researchers depositing their data without support from a Data Manager will usually select the default type, misidentifying the files. The metadata format of Crossref and Datacite is different, so an implementation for both formats is required. It may also be advisable to request Schema.org metadata.

Conclusion

Publishers today have a responsibility to the scientific community, to establish and provide digital connections between the published paper and the software and datasets that support that paper. This is critical for research transparency, integrity, and reproducibility in today's data- and software-intensive science enterprise. Scientific communications that do not provide these connections can no longer be considered fully robust.

In implementing required changes to support these responsibilities many journals do not yet have policies and practices strong enough to ensure the necessary dataset and software citations are included in the references. Further, if such citations are included, the journal guidance has heretofore often been murky on how the machine-actionable citations should be digitally formatted for downstream services to provide automated attribution and credit.

This paper provides the much needed guidance to help journals review their production processes and work with their authors, editors, reviewers, staff, production teams to ensure high-quality dataset and software citations that support automated attribution.

The authors of this paper are members of the FORCE11 Journals Task force representing many society, institution, and commercial publishers world-wide. They encourage their peers to adopt these practices and help enable proper machine-actionable dataset and software citations.

Journals should use this checklist, duplicated here from the Introduction for convenience, to start their journey of improving machine-actionable dataset and software citations.

Software and Data Citations: Checklist of Leading Practices for Journal Publishers
Journal publishers should make best efforts to undertake the following practices on which we provide further detailed guidance in this article.

Journal Production Guidance for Software and Data Citations

27 December 2022

1. **Instructions to Authors:** Provide author instructions that define what software and datasets should be cited in their paper, information on the best possible scientific repository, and the information to put in an Availability Statement.
2. **Publication Policies:** Update publication policies to include the importance of citing the software and datasets that support the transparency, integrity, and reproducibility of the research.
3. **Education and Outreach:** Educate editors, reviewers, and staff on the policy, requirements, and areas of flexibility.
4. **Technical Updates:** Put into place the necessary technical updates, as defined in this document, to ensure that the machine-readable representation of the software and dataset citations is sustained throughout the publication workflow and properly formatted when registered to Crossref.
5. **Production Teams:** Define for publication production team members the unique aspects of software and dataset citations; work together to implement necessary changes.
6. **Metrics:** Establish quality metrics and measures to ensure consistent, accurate results for software and dataset citations.

We strongly encourage our colleagues in academic publishing to adopt these practices. Colleagues with questions, or wishing to share success stories, should contact the lead author.

Journal Production Guidance for Software and Data Citations

27 December 2022

Glossary

This glossary is provided to help the reader with the many acronyms, concepts, and organizations involved in scientific publishing. The descriptions are short to provide a basic understanding, with the links to the source of the text for more information.

AGU - The American Geophysical Union (AGU), an Earth and space society serving over 130,000 science enthusiasts with 23 journals. (https://www.agu.org/Learn-About-AGU/About-AGU/)

APA - The American Psychological Association (APA) is a scientific and professional organization that represents psychologists in the United States. (https://www.apa.org/)

API - Application Programming Interface, a software interface providing connections between computer programs.

ARK - An Archival Resource Key (ARK) is a Uniform Resource Locator (URL) that is a multi-purpose persistent identifier for information objects of any type. (https://registry.identifiers.org/registry/ark)

ASCL - The Astrophysics Source Code Library (ASCL) is a free online registry for software that have been used in research that has appeared in, or been submitted to, peer-reviewed publications. (https://registry.identifiers.org/registry/ascl)

CODATA-ICSTI - The international and interdisciplinary Committee on Data for Science and Technology (CODATA) and the International Council for Scientific and Technical Information (ICSTI) jointly formed the Task Group on Data Citation Standards and Practices in 2010. (https://codata.org/initiatives/task-groups/previous-tgs/data-citation-standards-and-practices/)

DataCite - DataCite is a Registration Agency of the International DOI Foundation, which issues robust persistent identifiers for datasets and maintains a registry of associated rich metadata. (https://datacite.org/value.html)

Computational Notebook - are laboratory notebooks for scientific computing. They are used to combine software code, computational output, explanatory text and multimedia resources in a single document. (e.g., Jupyter, R Markdown, Quarto). (https://doi.org/10.1038/d41586-018-07196-1, https://quarto.org/docs/reference/formats/ipynb.html)

Content Negotiation (for at DOI) - allows a user to request a particular representation of a web resource. DOI resolvers use content negotiation to provide different representations of metadata associated with DOIs. A content negotiated request to a DOI resolver is much like a standard HTTP request, except server-driven negotiation will take place based on the list of acceptable content types a client provides. (https://citation.crosscite.org/docs.html)

Journal Production Guidance for Software and Data Citations

27 December 2022

CoreTrustSeal - an international, community based, non-governmental, and non-profit organization promoting sustainable and trustworthy data infrastructures. CoreTrustSeal offers to any interested data repository a core level certification based on the Core Trustworthy Data Repositories Requirements. This universal catalogue of requirements reflects the core characteristics of trustworthy data repositories. (https://www.coretrustseal.org/about/)

DOI - Digital Object Identifier (DOI), a persistent identifier based on the Handle system, issued by Registration Agencies of the International DOI Foundation, such as CrossRef and DataCite. (https://www.doi.org/factsheets.html)

ERA - Executable Research Articles (ERAs) are enriched narrative and research articles with code, data and interactive figures that can be executed in the browser, downloaded, and explored. (https://elifesciences.org/for-the-press/eb096af1/elife-launches-executable-research-articles-for-publishing-computationally-reproducible-results) (Lasser, 2020)

Event Data - is a Crossref service to capture online mentions of Crossref records. They monitor data archives, Wikipedia, social media, blogs, news, and other sources. The main focus has been on gathering data from external sources, however there is a great deal of Crossref metadata that is made available as events including relationship metadata and citations between records. (https://www.crossref.org/blog/event-data-now-with-added-references/)

FAIR - Findable, Accessible, Interoperable, and Reusable (FAIR), is a set of architectural patterns originally embodied in the FAIR Data Principles (Wilkinson et al., 2016) for achieving the goals identified in the acronym for data on the Web. FAIR has now been extended to apply to research software and other digital research objects.

FORCE11 - is a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing. (https://force11.org/info/about-force11/)

IGSN - International Generic Sample Number (IGSN), a globally unique and persistent identifier for physical samples. (https://www.igsn.org/)

I4OC - Initiative for Open Citations (I4OC), is a collaboration between scholarly publishers, researchers, and other interested parties to promote the unrestricted availability of scholarly citation data. (https://i4oc.org/)

JATS XML - Journal Article Tag Suite (JATS) Library, a standard publishers use to mark up journal article content in XML format. (https://www.niso.org/standards-committees/jats)

Journal Production Guidance for Software and Data Citations

27 December 2022

JATS4R - JATS for Reuse (JATS4R) are recommendations for tagging content in JATS XML, devoted to optimizing the reusability of scholarly content. (https://jats4r.org/)

MyBinder – is a free Binder service, powered by BinderHub, an open-source tool that runs on Kubernetes. Binder allows users to create custom computing environments that can be shared and used by many remote users. (https://mybinder.readthedocs.io/en/latest/)

NISO - The National Information Standards Organization (NISO) is a non-profit membership organization that identifies, develops, maintains, and publishes technical standards to manage information. (https://www.niso.org/welcome-to-niso)

ORCID - Open Researcher and Contributor ID (ORCID), is a unique identifier for researchers. (https://orcid.org/)

PID – a Persistent Identifier (PID) is a long-lasting reference to web resource. Persistence is achieved technically by separating metadata resolution from data resolution, in a two-stage resolution process, which includes long-term persistence of the metadata hosting activity. (https://support.orcid.org/hc/en-us/articles/360006971013-What-are-persistent-identifiers-PIDs-)

RDA - Research Data Alliance (RDA), was launched as a community-driven initiative in 2013 by the European Commission, the United States Government's National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation with the goal of building the social and technical infrastructure to enable open sharing and re-use of data. (https://www.rd-alliance.org/about-rda)

ROR - Research Organization Registry (ROR) is a community-led project to develop an open, sustainable, usable, and unique identifier for research organizations worldwide. (https://ror.org/)

RRID - Research Resource Identification (RRID) are identification numbers assigned to help researchers cite key resources (antibodies, model organisms and software projects) in the biomedical literature to improve transparency of research methods. (https://www.rrids.org/)

Scholix - Scholarly Link eXchange (Scholix), is an initiative to establish a high level interoperability framework and guidelines for exchanging information about links between scholarly literature and research data. (http://www.scholix.org/about)

TOP - the Transparency and Openness Promotion Guidelines (TOP) provide a suite of tools to guide implementation of better, more transparent research. (https://www.cos.io/initiatives/top-guidelines)

Journal Production Guidance for Software and Data Citations

27 December 2022

TRUST - Transparency, Responsibility, User focus, Sustainability and Technology (TRUST). The TRUST Principles provide a common framework to facilitate discussion and implementation of best practice in digital preservation by all stakeholders. (https://www.nature.com/articles/s41597-020-0486-7)

URL - Uniform Resource Locator (URL), commonly called a web address, is a reference to a web resource that specifies its location on a computer network and a mechanism for retrieving it. (https://csrc.nist.gov/glossary/term/uniform_resource_locator)

Journal Production Guidance for Software and Data Citations

27 December 2022

Software and Data Citation Resources

This software and data citation resource list includes discipline-agnostic, community-developed and vetted guidance for scholarly publishers, scientific repositories, authors, software developers, and researchers.

Software Citation Resources

Chue Hong, N.P., Allen, A., Gonzalez-Beltran, A., de Waard, A., Smith, A.M., Robinson, C., Jones, C., Bouquin, D., Katz, D.S., Kennedy, D., Ryder, G., Hausman, J., Hwang, L., Jones, M.B., Harrison, M., Crosas, M., Wu, M., Löwe, P., Haines, R., Edmunds, S., Stall, S., Swaminathan, S., Druskat, S., Crick, T., Morrell, T., Pollard, T., (2019). Software Citation Checklist for Authors. Zenodo. https://doi.org/10.5281/ZENODO.3479198

Chue Hong, N.P., Allen, A., , Gonzalez-Beltran, de Waard, A., Smith, A.M., Robinson, C., Jones, C., Bouquin, D., Katz, D.S., Kennedy, D., Ryder, G., Hausman, J., Hwang, L., Jones, M.B., Harrison, M., Crosas, M., Wu, M., Löwe, P., Haines, R., Edmunds, S., Stall, S., Swaminathan, S., Druskat, S., Crick, T., Morrell, T., Pollard, T., (2019). Software Citation Checklist for Developers. Zenodo. https://doi.org/10.5281/ZENODO.3482769

Katz, D.S., Chue Hong, N.P., Clark, T., Muench, A., Stall, S., Bouquin, D., Cannon, M., Edmunds, S., Faez, T., Feeney, P., Fenner, M., Friedman, M., Grenier, G., Harrison, M., Heber, J., Leary, A., MacCallum, C., Murray, H., Pastrana, E., Perry, K., Schuster, D., Stockhause, M., Yeston, J. (2021). Recognizing the value of software: a software citation guide. In F1000Research (Vol. 9, p. 1257). F1000 Research Ltd. https://doi.org/10.12688/f1000research.26932.2

Smith AM, Katz DS, Niemeyer KE, FORCE11 Software Citation Working Group. (2016). Software citation principles. PeerJ Computer Science 2:e86 https://doi.org/10.7717/peerj-cs.86

Data Citation Resources

Cousijn, H., Kenall, A., Ganley, E., Harrison, M., Kernohan, D., Lemberger, T., Murphy, F., Polischuk, P., Taylor, S., Martone, M., Clark, T. (2018). A data citation roadmap for scientific publishers. Sci Data. https://doi.org/10.1038/sdata.2018.259

Data Citation Synthesis Group. (2014). Joint Declaration of Data Citation Principles. Force11. https://doi.org/10.25490/A97F-EGYK

Journal Production Guidance for Software and Data Citations

27 December 2022

Availability Statement Recommended Elements

This text is adapted from AGU's Data and Software Availability and Citation Checklist (AGU, 2021).

1. Description of the Type(s) of data and/or software - [Required]
   Examples:
   a. Data - The facilities of IRIS Data Services were used for access to waveforms and related metadata from the International Geodynamics and Earth Tide Service (Network Of Superconducting Gravimeters, 1997)
   b. Software - Figures were made with Matplotlib version 3.2.1 (Caswell et al., 2020; Hunter, 2007)
2. Repository Name(s) where the data/software are deposited - [Best Practice]
3. URL/link to the data/software, preferably Persistent Identifier (e.g., DOI) and resolves - [Required]
   Examples:
   a. Software - https://doi.org/10.5281/zenodo.3714460
   b. Data - https://doi.org/10.7283/633E-1497
4. Access Conditions - [Best Practice]
   Examples:
   a. Registration/fee required
   b. Database where certain functionality, selections, or query need to be made. Provide the details.
5. English Translation - [Required]
   Examples:
   a. Site includes translation functionality
   b. Translation available via browser plug-in
   c. Author guides the readers/makes translation available
6. Licensing - [Best Practice]
   Examples:
   a. Software - MIT License (others)
   b. Data - CC-BY (others)
7. In-text Citation where possible - [Best Practice]
   Examples:
   a. Software - Figures were made with Matplotlib version 3.2.1 (Caswell et al., 2020; Hunter, 2007), available under the Matplotlib license at https://matplotlib.org/.
   b. Data - Data from the KNMI archive with Federation of Digital Seismograph Networks (FDSN) network identifiers NL (KNMI, 1993) and NA (KNMI, 2006) were used in the creation of this manuscript.
8. If software, also include
   a. Version (e.g., Version 3.2.1) - [Best Practice]
   b. Link to publicly accessible Development Platform (e.g., GitHub) - [Best Practice]
   Examples:

i. Part of the software (version 1.0.0) associated with this manuscript for the calculation and storage of PSDs is licensed under MIT and published on GitHub https://github.com/Jollyfant/psd-module/ (Jollyfant, 2021).

c. Author, Project Name(s) instead of username(s) (e.g., username123)
d. Additional Context/Description beyond acronym or code name (e.g., Longhorn pipeline scripts for reducing data vs Longhorn)

Journal Production Guidance for Software and Data Citations

27 December 2022

References

AGU (2021), “Data and Software Availability and Citation Checklist”. https://data.agu.org/resources/availability-citation-checklist-for-authors

ANSI/NISO Z39.96-2021, JATS: Journal Article Tag Suite, version 1.3. NISO. https://doi.org/10.3789/ansi.niso.z39.96-2021

Brase, J. (2009). DataCite - A Global Registration Agency for Research Data. In 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology. 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology (COINFO). IEEE. https://doi.org/10.1109/coinfo.2009.66

Cousijn, H., Kenall, A., Ganley, E., Harrison, M., Kernohan, D., Lemberger, T., Murphy, F., Polischuk, P., Taylor, S., Martone, M., & Clark, T. (2018). A data citation roadmap for scientific publishers. In Scientific Data (Vol. 5, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/sdata.2018.259

Cousijn, H., Feeney, P., Lowenberg, D., Presani, E., & Simons, N. (2019). Bringing Citations and Usage Metrics Together to Make Data Count. In Data Science Journal (Vol. 18). Ubiquity Press, Ltd. https://doi.org/10.5334/dsj-2019-009

Crossref (2021), “Metadata deposit schema 5.3.1 ”. https://www.crossref.org/documentation/schema-library/metadata-deposit-schema-5-3-1/

Data Citation Synthesis Group. (2014). Joint Declaration of Data Citation Principles. Force11. https://doi.org/10.25490/A97F-EGYK

Groth, P., Cousijn, H., Clark, T., & Goble, C. (2020). FAIR Data Reuse – the Path through Data Citation. In Data Intelligence (Vol. 2, Issues 1–2, pp. 78–86). MIT Press. https://doi.org/10.1162/dint_a_00030

Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R., & Goudie, S. (2020). Developing a Research Data Policy Framework for All Journals and Publishers. In Data Science Journal (Vol. 19, Issue 1, p. 5). Ubiquity Press, Ltd. https://doi.org/10.5334/dsj-2020-005

Katz, D. S., Chue Hong, N. P., Clark, T., Muench, A., Stall, S., Bouquin, D., Cannon, M., Edmunds, S., Faez, T., Feeney, P., Fenner, M., Friedman, M., Grenier, G., Harrison, M., Heber, J., Leary, A., MacCallum, C., Murray, H., Pastrana, E., … Yeston, J. (2021). Recognizing the value of software: a software citation guide. In F1000Research (Vol. 9, p. 1257). F1000 Research Ltd. https://doi.org/10.12688/f1000research.26932.2

Khan, N., Pink, C. J., & Thelwall, M. (2020). Identifying Data Sharing and Reuse with Scholix: Potentials and Limitations. In Patterns (Vol. 1, Issue 1, p. 100007). Elsevier BV. https://doi.org/10.1016/j.patter.2020.100007 .

Lasser, J. (2020). Creating an executable paper is a journey through Open Science. In Communications Physics (Vol. 3, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s42005-020-00403-4

L'Hours, H., Kleemola, M., & De Leeuw, L. (2019). CoreTrustSeal: From academic collaboration to sustainable services. In IASSIST Quarterly (Vol. 43, Issue 1, pp. 1–17). University of Alberta Libraries. https://doi.org/10.29173/iq936

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. In Scientific Data (Vol. 7, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41597-020-0486-7

Lin, J. (2017, March 02). How do you deposit data citations? Crossref Blog. https://www.crossref.org/blog/how-do-you-deposit-data-citations/

Journal Production Guidance for Software and Data Citations

27 December 2022

Murphy, F., Bar-Sinai, M., & Martone, M. E. (2021). A tool for assessing alignment of biomedical data repositories with open, FAIR, citation and trustworthy principles. In F. Naudet (Ed.), PLOS ONE (Vol. 16, Issue 7, p. e0253538). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pone.0253538

Nelson, Alondra. OSTP Memo : "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research", August 25, 2022: https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

NISO (2020) RP-36-2020, JATS4R Data Citations, Version 2.0. https://doi.org/10.3789/niso-rp-36-2020

NISO (2021) RP-36-2020, JRP-40-2021, JATS4R Software Citations, Version 1.0. https://doi.org/10.3789/niso-rp-40-2021

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. In Science (Vol. 348, Issue 6242, pp. 1422–1425). American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.aab2374

Pirani, A., Alegria, A., Khourdajie, A. A., Gunawan, W., Gutiérrez, J. M., Holsman, K., Huard, D., Juckes, M., Kawamiya, M., Klutse, N., Krey, V., Matthews, R., Milward, A., Pascoe, C., Van Der Shrier, G., Spinuso, A., Stockhause, M., & Xiaoshi Xing. (2022). The implementation of FAIR data principles in the IPCC AR6 assessment process. https://doi.org/10.5281/ZENODO.6504468

Uhlir P. For Attribution - Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop (2012) [Internet]. The National Academies Press; 2012 p. 220. Available from: http://www.nap.edu/catalog.php?record_id=13564

Journal Production Guidance for Software and Data Citations

27 December 2022

Wellcome (2017), “Data, software and materials management and sharing policy”, https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. In Scientific Data (Vol. 3, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/sdata.2016.18

Funding

Shelley Stall is partially funded by the National Science Foundation (Grant ID 2025364). This paper is a deliverable for that project.