*****DO NOT EDIT*** OLD VERSION, SEE LINK BELOW**

Physical Samples and Sample Related Data Sharing Guidance for Authors Submitting to AGU journals

We are reorganizing this document, and are now working on a new version, see link here.

What this document is:

Basic guidance on how to make physical samples and associated data FAIR

Guidance on how to describe and cite samples in AGU journal publications, using the currently available infrastructure for assigning sample identifiers and tracing citations.

Guidance for how to list identifiers consistently in papers, datasets, and other files

This includes basic recommendations for obtaining/assigning sample identifiers, describing sample metadata, publishing associated sample datasets, and referencing/citing samples.

Links to useful resources with more information on domain-specific metadata templates and requirements

Address need for physical sample repositories to get credit when samples are used, and provide information needed to access physical samples

Recommendations that include foundational elements to enable developing more sophisticated complex citations, and associated usage tracking, that include samples in the future

- Recommendations for sample identifiers (PIDs)
- Cross-domain metadata needed for a “sample citation” - info needed to find, access, give credit (e.g. iSamples, DataCite, IGSN registration metadata)

Intimated above and below, it is a living document. We recognize that there are emerging trends, techniques, and technologies. However, this document gives the status quo, and then will be reviewed on regular cycles and updated as necessary.

Following from the previous point, the document is here to raise awareness and promote excellent/good solutions even if they are not yet the norm or are somehow deemed ‘hard’. The document is progressive and trying to move the community forward. If a standard, system, tool, or otherwise is in place and established enough (whatever that means), then it is in consideration for inclusion.

What this document isn’t

Speculation or guidance on the kind of infrastructure needed to better track citations

A technical specification for how to make specimen citations machine readable

Directions on what identifier to use

Guidelines for how to create a physical sample repository [something to connect with the Sampling Nature RCN about, maybe -- working group on uncurated collections]

Guidance for publishers, repository managers, editors etc [we need a separate document for that]

Reinventing the wheel

It is not and should not be bound by AGUs current guidance for Data and Software (or other, similar guidance). Material samples are very different to other research outputs and guidance should be geared towards their unique specificities. The document will not get bogged down with trying to come up with equivalencies that don’t make sense for samples.

A journal article to be submitted for peer review. A companion article will be developed alongside this document, but the priority here is on the guidance itself.

Disclaimer: This is a draft of a document created by the ESIP Physical Samples Curation Cluster. With permission, it contains base text from AGU’s Data and Software for Authors*. This is the internal working copy. For the public copy, please see this document.*

Insert authors

Preamble: why describing and citing samples matters

Physical samples[1] and their associated data are some of the primary building blocks of research published in AGU journals. By “samples” we mean physical subsets or extracts from some feature of scientific interest -- for example, a chip from a geologic outcrop, a soil core section, or

[1] The terms sample and specimen are roughly equivalent, although some practitioners make a distinction between them. We use samples here.

a taxidermied animal taken as a representative of a species, or the residue of some material created through a laboratory experiment.[2]

These objects and their metadata need to be accessible, understandable, and open as possible for reuse to support transparency and replicability. Additionally, sample repositories need to be able to track the demonstrated use of samples to better manage and show the value of their collections. By describing and citing samples used in scientific analyses and papers, we make science more equitable, reliable and cost effective (and therefore, more environmentally-friendly).

These guidelines are intended as a complement to AGU's existing Data and Software for Authors requirements before publication.Here, we offer best practices for describing and citing physical samples that are used in scientific publications, with links to additional information and examples to assist authors in meeting these requirements and recommendations.

Summary of recommendations

- Provide a list of samples used in the text of your paper (e.g. within the methods section, or the data availability statement), or in an associated dataset. Ideally, this should include a structured table containing: the sample IDs, their physical location, and other necessary metadata as dictated by your field of study.
    - If your study uses many samples and would result in a prohibitively long list: provide information on your samples as a separate sample metadata file, published as part of a dataset, which may also include any associated sample data. You can create and publish a dataset for your samples and associated data through an appropriate data repository (such as Figshare, Zenodo, Dryad; or through institution- or domain-specific repositories), and cite that dataset in the references section of your paper.
    - See Table X for links to additional guidance on sample metadata, and metadata file templates. As samples are referenced in the text, cite them inline using a standard PID (e.g. IGSN:IEGRW002B), or in a table. In-line citations should be in addition to (not instead of) a sample availability statement and metadata file.
- Obtain globally unique and persistent identifiers (GUID), such as a DOI or IGSN, for samples used in scientific analyses and published papers. These should be associated

[2] We base our definition of sample on the SOSA ontology: Samples are typically subsets or extracts from the feature of interest of an observation. They are used in situations where observations cannot be made directly on the ultimate feature of interest, either because the entire feature cannot be observed, or because it is more convenient to use a proxy. Samples are thus artifacts of an observational strategy, and usually have no significant function outside of their role in the observation process. The characteristics of the samples themselves are generally of little interest, except to the manager of a sampling campaign, or sample curator.
A Sample is intended to sample some FeatureOfInterest, so there is an expectation of at least one isSampleOf property. However, in some cases the identity, and even the exact type, of the sampled feature may not be known when observations are made using the sampling features.
Physical samples are sometimes known as 'specimens'.
https://www.w3.org/TR/vocab-ssn/#SOSASample

with standard metadata that is openly available via a persistent web landing page (e.g. IGSN:IEGRW002B). GUIDs are particularly important to enable tracking use of physical samples that are stored long-term in physical repositories or for samples that are not stored long-term but may be included in multiple datasets or papers.
- Samples should be associated with standard metadata appropriate to describe collection details of the specific sample type and scientific domain. You may obtain sample PIDs, like IGSN (International Generic Sample Numbers) through services such as SESAR , DataCite, or institution-specific allocating agents…
- If samples are stored long-term in a physical repository, provide the name and location of the physical repository. If the repository requests acknowledgement or citation in some way, please follow those individual guidelines.
- Provide links to related or derived data from samples, and include persistent identifiers for samples in their data.

Note that practices and infrastructure available for sample identifiers, citation, and usage tracking is an area of active work, and will likely improve in the near future (e.g. through efforts such as the new RDA Complex Citations Working group, the ESIP Physical Sample Curation Cluster, and the Internet of Samples project).

The final determination of whether a manuscript meets these requirements is made by the journal editors. Author feedback is appreciated to help ensure that the process remains efficient, feasible, and meaningful. AGU recognizes that the infrastructure for storing and citing metadata about samples is still being developed. We expect these guidelines to change over time.

AGU also recognizes that not all samples and sample data can be fully open. Samples that are sensitive or restricted must be protected through appropriate access controls (i.e permits; access moratoriums). Samples should be as open as possible, as closed as necessary. For samples related to Indigenous Peoples, authors should consult the CARE Principles for Indigenous Data Governance.

Guidelines for Publishing Research that Includes Physical Samples

INSERT RELEVANT HEADINGS
Why use identifiers (the goals - we can get to those points that exist for other materials like knowledge graphs for samples)
Provenance
How to handle subsamples (do they get their own number?)
Destructive Sample and identifiers (why do I need a PID for something that no longer exists)
Where to include the identifier (on the item, in the paper, etc.)

How to choose and how to get an ID would also relate to sample ownership and who has the right to get the identifier if one doesn't exist.

Identify Samples

Samples are often assigned locally-unique identifiers or sample names, with abbreviated codes (S01) or meaningful information (e.g. PineCr_Sed_20220621) when they are collected or split into subsamples. These identifiers are not globally unique and do not allow tracking use of samples across datasets and journal publications.

Persistent, resolvable identifiers can supplement human-readable sample names, and enable tracking provenance and use of physical samples, which may be stored long-term and used over time for various purposes or sent to numerous labs for interdisciplinary analyses. Persistent identifiers for samples, including ARKs, DOIs, and IGSNs, are globally unique and can link directly to metadata describing the sample, and to associated data, papers, and more. They can also be used to explicitly connect a subsample to its parent, making it clear which analyses were performed on which samples.

Researchers can obtain persistent identifiers:

- **Through their organization.** Some organizations, especially those that manage sample repositories, have processes in place to assign persistent identifiers.
- **Through an external organization**, like the System for Earth Sample Registration (SESAR). These organizations provide tools to allow users, including both individual researchers and organizations, to register for, resolve, and manage sample identifiers and metadata (see the next section for guidance about sample metadata).

Note that the persistent identifier must be assigned by the sample's owner. If you are not the owner (for example, if you are working with samples on loan from a repository), contact the owner to request a persistent identifier.

- BEST: Assign all samples and subsamples a persistent identifier. Persistent identifiers for samples are then included in the text or within one or more table(s) of your papers (often in the methods section) in a standard format with hyperlink (i.e. IGSN:IEGRW002B). Published datasets should also include a list of sample PIDs in the dataset metadata. All dataset files involving samples should include the associated sample PID, so that all data is clearly associated with the appropriate sample.
- ACCEPTABLE: Assign samples reasonably distinct sample identifiers (e.g. Darwin Core Triplet, UUID?) . All dataset files involving samples should include the sample identifier, so that all related data is associated with the appropriate sample.

Sample identifiers and destructive analyses: It is useful to assign persistent identifiers to samples even if they are intended for (or have already been used up in) destructive analyses,

particularly for interdisciplinary source material samples. Interdisciplinary source samples may be analyzed for a variety of analyses (e.g. physical, chemical, biological), and published in multiple online data systems, or publications. Assigning a persistent identifier to a sample that is consumed provides the following benefits:

- Enables linking related interdisciplinary sample metadata and data across numerous records or publications online. This in turn enables you and others to efficiently find, access, integrate, and use this sample data in the future.
- Can enable automated sample provenance tracking and use metrics. There is great potential to use PIDs for automated tools that precisely link and exchange related information in the future.
- Allows researchers to record metadata about the exact sample that was analyzed, including that it is no longer available for analysis.
-

Describe Samples - Metadata

Sample metadata includes basic information about the sample--what it is, when and where it was collected, who collected it, who currently owns it. Simple measurements, like the size or weight of the sample, also fall into this category. Detailed metadata can help other researchers find and assess samples for their own research.

One key benefit of using persistent identifiers is that they provide a way to link a sample citation back to information about that sample, thereby allowing researchers to access relevant details that exceed the scope of peer-reviewed literature.

Resources that provide more information on standard metadata and templates used in different scientific disciplines are provided in Table X.

- BEST: Samples have been assigned persistent identifiers that resolve to a web page with detailed standardized metadata, which could be used to link and exchange information about related samples and track use of samples. Published datasets should also include a standard sample metadata file that fully describes samples used in the dataset.
- ACCEPTABLE: Samples are described in the text of associated papers, and any associated datasets include basic information needed to find, access, integrate, reuse samples within dataset metadata (e.g. sample material, date collected, locations), and one or more standard sample metadata file(s).

Archive Physical Samples and Data

Physical sample repositories

Publish Sample Dataset(s)

Reference Samples Used - Current Options

Cite samples where they are first mentioned? Provide a list of samples at the end of the paper, similar to a bibliography? In the availability statement? Links will be disruptive if they are spelled out in the text–can specimens be linked to the PID without showing the PID itself? If so, the links wouldn't show up in print (meh) and might not show up in tools like xDD or publisher APIs (bad). Can put list of samples in methods, supplementary material.

Cite Sample Dataset in Reference Section

Data Availability Statement

Institutions and individuals that contribute samples for a research project are often thanked in an acknowledgement, and may not include that can obscure important details about how other researchers can access the samples used. A separate sample availability statement should be used to ensure that future researchers understand which samples are available and how they can be requested.

For specimens from an institutional collection: Institutional collections differ from personal collections in that they typically have dedicated staff and formal accession, loan, and management policies. For each collection that provided samples, include the name, list of samples, sample availability, and contact information.

For samples from a personal collection: This category includes any collection maintained solely by an individual, even if that collection is stored at their workplace. Provide contact information and a list of samples from the collection. Note whether samples are available for use by other researchers.

Articles based on samples from multiple collections should list each one separately.

Keeping samples accessible: Samples may be lost, consumed, or transferred to new owners over time. To keep samples findable, we recommend providing ownership and contact information in the sample metadata record associated with a sample's persistent identifier. Many systems, including SESAR, allow users to include this information.

Examples

For samples from a personal collection:

A. Smith's personal collection (AS-1, AS-2, AS4). Samples available upon request. Contact A. Smith (asmith@college.edu) to request samples.

For samples of uncertain availability:

A. Smith's personal collection (AS-1, AS-2, AS4). Contact A. Smith (asmith@college.edu) for information about sample availability.

For samples from an institutional collection:

Smithsonian National Museum of Natural History, Department of Mineral Sciences (NMNH 1234, NMNH 2345). Samples available upon request. Contact Leslie Hale (halel@si.edu) to request samples

When to Make Your Samples and Associated Metadata and Data Available

At the time your paper is submitted, you must include a list of samples analyzed for the study. Your sample related data must also be available to the editors and reviewers, following AGU data citation guidelines. Datasets that involve physical samples should include unique identifiers (ideally persistent identifiers) for each sample along with basic metadata describing sample type, material, collector, environmental context, location, and other collection details using a suitable standard format (see Table X for links to domain-specific templates for sample metadata).

At the time your paper is accepted, your samples and associated data availability statements must be clearly stated, with information on the institutions where physical samples are held. Authors must still provide preliminary access to reviewers at the time of submission. Please ensure that your samples and sample-related data are available with details of the online access location(s), data product names, variable names, time ranges, spatial locations, or any other search criteria to allow a reader and reviewer to Find and Access the data used and/or generated for the paper (including those represented in figures and tables). In summary, any data and software utilized in the work contained in the manuscript must be documented for free and open availability. Data or software that are sensitive and require restrictions on access (e.g., personal data, medical information, fossil locations, strategic models) must be preserved in a repository with appropriate access controls.

Finding repositories

Selecting a Repository for your Physical Samples

Selecting a sample repository (physical storage) and determining sample management best practices begins when you propose and fund your research project, through your Data (and Sample) Management Plan. The repository you select for your samples (physical and metadata) may be the same.

For Samples, you need to locate a repository that provides services, such that:

1. Samples will be made accessible according to applicable moratorium policies (publically accessible, access restricted for a period of time).
2. Samples will be preserved and managed with appropriate collection management strategies
3. Curation policies adherent to relevant community standards (nationally, globally, domain specific) inclusive of
    a. Metadata requirements
    b. Providing and tracking identifiers
    c. Resampling procedures
    d. How to cite samples and acknowledge the repository
4. Provide access for discovery and reuse of samples. This may exist within the repository physically curating the samples or may be an external system. *See next section on metadata repositories.*

Resources

- The Geologic Materials Repository Working Group, 2015, The U.S. Geological Survey Geologic Collections Management System (GCMS)—A master catalog and collections management plan for U.S. Geological Survey geologic samples and sample collections: U.S. Geological Survey Circular 1410, 108 p., https://dx.doi.org/10.3133/cir1410.
- Repository Curators - Reach out to physical sample repositories early in your project as they can help when writing your data management plan and throughout the duration of your sample related activities.

Types of Physical Sample Repositories

Your funding may dictate where your samples may reside. Check the requirements of both your funding agency and the repository before making any decisions. We recommend the first three, in an ideal situation, but 4 is sufficient.

Discipline-Specific Repositories: These are specialized repositories which typically provide support and information on required standards for metadata and more.

Institutional Repository: Many universities are supporting research data and software management and compliance requirements for their researchers, and such services are often provided through the library. Librarians can be an excellent source of research data management support, including repository selection, and can help you comply with funder, publisher, and university requirements.

National/State Repositories: Some countries require authors to use their National Repositories for data and/or software preservation.

Personal Collection: May not have formal support and will have to provide access and sustainability through your own methods. *This is an acceptable solution…. To avoid being put in a draw or garage. To avoid sending them to the wrong place or not finding a home.*

Selecting a Repository for your Sample Metadata

Selecting a sample metadata repository (digital storage) and determining sample metadata management best practices begins when you propose and fund your research project, through your Data (and Sample) Management Plan. General considerations when selecting a metadata repository include considering reaching out early assistance and recognizing that some physical sample repositories can not manage their own metadata repository (but this is not always the case). In addition, some physical sample repositories may have internal catalogs (paper records or local database/spreadsheets) but can not make sample metadata accessible externally and so you may need to consider a separate metadata repository to increase discoverability.

For samples metadata, you need to locate a repository that provides services, such that:

1. They provide accession policies and other registration guidelines so that you might assess the repository for appropriate fit. This includes policies on topics such as:
    a. if they support metadata records for samples that no longer exist,
    b. the time frame for accessioning metadata into their system, and
    c. their capacity to register large scale submissions.
2. Have clear metadata requirements including minimum fields required for registration.
3. The repository registers your sample with a persistent identifier that is globally unique such as an IGSN.
4. The data are freely accessible from a landing page that provides information (e.g., metadata) about your data, and preferably version controlled.
5. Trustworthiness of the metadata repository (CoreTrustSeal, etc.)

Optionally, you may also consider if the metadata repository supplies metrics and analysis on your samples over time (e.g. views of sample metadata profiles, tracking samples over time).

Resource

- Klump, J., Lehnert, K., Ulbricht, D., Devaraju, A., Elger, K., Fleischer, D., … Wyborn, L. (2021). Towards Globally Unique Identification of Physical Samples: Governance and

Technical Implementation of the IGSN Global Sample Number. Data Science Journal, 20(1), 33. DOI: http://doi.org/10.5334/dsj-2021-033
- Damerow, J. E., Varadharajan, C., Boye, K., Brodie, E. L., Burrus, M., Chadwick, K. D., … Agarwal, D. (2021). Sample Identifiers and Metadata to Support Data Management and Reuse in Multidisciplinary Ecosystem Sciences. Data Science Journal, 20(1), 11. DOI: http://doi.org/10.5334/dsj-2021-011

Types of Sample Metadata Repositories

Your funding may dictate where your sample metadata may reside. Check the requirements of both your funding agency and the repository before making any decisions. We recommend the first three, in an ideal situation, but 4 is sufficient.

Discipline-Specific Repositories: These are specialized repositories which typically provide support and information on required standards for metadata and more.

Institutional Repository: Many universities are supporting research data and software management and compliance requirements for their researchers, and such services are often provided through the library. Librarians can be an excellent source of research data management support, including repository selection, and can help you comply with funder, publisher, and university requirements.

National Repositories: Some countries require authors to use their National Repositories for data and/or software preservation.

Personal Collection: May not have formal support and will have to provide access and sustainability through your own methods. *This is an acceptable solution with these conditions (do you align with the data management set out with your project?, LOCKSS ("Lots of Copies Keep Stuff Safe"), cloudbased backup, clear and articulate metadata which can be connected back to the samples and their references in publication, https://connectingtocollections.org/upcoming-webinar-dealing-with-digital-assets/ etc.)….*

Selecting a Repository for Derived or Related Data

- AGU has recommendations for data repositories (and we may need to point to those plus augment with sample specific examples)
    - Direct users to domain-specific data repositories (https://data.agu.org/resources/useful-domain-repositories)

Reach out to these data repositories early and can help you with requirements!

Consideration - you may have samples that are not in a repository (maybe the sample no longer exists) and that sample data should still go into a repository.

References

These guidelines were modeled on the following:

Fox, Peter, Erdmann, Chris, Stall, Shelley, Griffies, Stephen M., Beal, Lisa M., Pinardi, Nadia, Hanson, Brooks, Friedrichs, Marjorie A. M., Feakins, Sarah, Bracco, Annalisa, Pirenne, Benoî, & Legg, Sonya. (2021). Data and Software Sharing Guidance for Authors Submitting to AGU Journals. Zenodo. https://doi.org/10.5281/zenodo.5124741

Further reading

- *Science* Guidelines for Specific Types of Studies (see **Studies involving the field collection of samples, specimens, or fossils** and **Paleontological and archaeological studies** subheadings)
- SVP Best Practice Guidelines for Repositing and Disseminating Contextual Data Associated with Vertebrate Fossils (addresses permits and redacting sensitive contextual data)

==========

Not sure where to put this:

Considerations for publications involving physical samples and associated data

.

1. **Sample Citation (required)**
   Include sample citations for the primary and processed research data in the References section of your paper. Doing so ensures proper credit is given for the sample. Note,

English-language (or English translation) for any cited sources is required. The AGU's Data and Software for Authors guide provides information on how to cite your samples.

1. Primary Data and Processed Data: Some repositories "reserve a DOI" before publishing your data. Use this DOI in your data citation. Once your data are published, the DOI will resolve properly. Some repositories use other persistent identifiers or URLs which are permissible (e.g, GenBank). Some repositories will only provide a DOI close to or at the time of paper acceptance; in this case the DOI will need to be added during the final revision.
2. Simulated Data / Model Output Data: See the guidance for Numerical Models.
3. Large Data (>1TB): Preservation of large data may be possible in some repositories for a fee. Authors should account for this in their research budget and Data Management Plan. Also, check with your university or institution and their repository.
4. Data Used from Another Source (e.g., data created by others): Cite the specific source to ensure proper credit and allow readers to access the same version of the data. If the data are also associated with a publication, especially a data publication, cite both the paper and the repository.

During Peer Review

Data: Your data must be available for peer review of your manuscript. Here are options to ensure confidential access to your data.

1. Preserve your data in a repository and make it available for peer review. Depending on the repository, this process can be done in a couple of ways:

    1. Provide a temporary private link ("share link") in the last sentence of the Open Research section of your paper. This link will not be present in your published paper as it is not a persistent link. This approach allows your data to remain private until acceptance of the manuscript for publication.

    2. Provide the persistent identifier (e.g., DOI) for your data; i.e., when your dataset has completed the repository submission process and is now publicly available.

2. 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 prior to paper

acceptance.

Software: For papers where software is central to the research, your software, or workflow, must be available, if the peer reviewer needs access. The options for providing or limiting access to your software are the same as for data.

Peer reviewers, with support from AGU staff, will ensure the link to the data and software resolves properly to a community-accepted, trusted repository and includes the data and software necessary to evaluate your research.

Paper Acceptance

1. Data: Relevant data and model results should be accessible at the time your paper is accepted. Note the possibility that in unusual cases the repository policy may not allow your data to be pul ablished until your paper is published. If that is the case, AGU will accept that your data will be made available after your paper is published. Please coordinate with the repository to ensure the availability of your data.
2. 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 Checklist Examples and Templates

The Availability Statement is a narrative that indicates to the reader where and how to directly access your physical samples and associated metadata and provides any information on licenses and restrictions. An Availability Statement should contain an in-text citation, licensing information (e.g., CC-BY 4.0, MIT) and access restrictions (e.g., authentication required) (here is an example from JAMES). Statements to the effect of “samples available from authors” are not acceptable. Also, statements to the effect of “samples and associated metadata available from http://nasa.gov” are not acceptable since high-level website references do not meet the AGU’s Data and Software for Authors requirements. Provide details of the specific locations, data product names, variable names, time ranges, spatial locations, and any other search criteria needed to Find and Access the data used/ generated in the paper (including those represented in figures and tables).

Sample Availability Statement:

1. **Data archived in a repository:** Samples and associated metadata 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]. Such datasets must be findable and accessible (e.g. via

URLs).

2. **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. Such datasets *must be findable and accessible* (e.g. via URLs). For example:
   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*
   Citations for data in regular publications where data are not findable or accessible, i.e. not available are NOT acceptable. For a made-up example:
   Cline, D., R. Armstrong, R. Davis, K. Elder, and G. Liston. 2002. CLPX-Ground: ISA snow depth transects and related measurements, J. Ice., vol 1 (2), pp. 3-9. (journal is subscription only, and data are not available in the article or supplement information or is in a proprietary format that is no longer readable).

3. **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]. Such datasets must be findable and accessible (e.g. via URLs).

4. **Theoretical papers, or most review papers:** Data were not used, nor created for this research.

5. **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. **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 a 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.

Checklist for authors

Describe your samples. What are they? Where are they from? Who owns them, and where are they kept? Is the sample available to another researcher looking to build upon your

work? Most disciplines have at least basic metadata standards that can be used to help make sure samples are described well enough to remain scientifically useful.

PERMITS

Assign PIDs to samples and subsamples. Track parent-child relationships so that other researchers can build upon and evaluate your work even if a single subsample is exhausted.

What about samples on loan from a repository?

Cite samples by their identifiers in the text of the paper

BEST: Link samples with clickable PIDs when they first appear in the text

What should the link look like?

Should PIDs be legible in a printed copy of a paper?

GOOD: Provide a list of all samples, including their PIDs, as an appendix

Appendices not always available and generally more difficult to use for data mining applications

Provide an easily accessible copy of all sample-derived data for other researchers to use. Tables in papers are difficult to parse and make data inaccessible to the machines. At best, they mean other researchers will have to type out the results of your data. Providing a tidy copy of your data makes it easier for other researchers to find and re-use it.

BEST: Upload your data to a repository specific to the type of analysis (genomics, etc.)

GOOD: Provide a tidy spreadsheet of sample-derived data to an appropriate data repository. Include the DOI for the resulting resource(s) in your paper.

Or just follow AGU data guidance?

Identify a permanent physical repository for significant samples. This is way outside the norm for petrology at least and I don't think these repositories exist.

BEST: Deposit samples in a managed repository prior to publication

ACCEPTABLE: Maintain samples in personal collection and provide readers with instructions for how to contact you about access to samples

What to include in sample availability statement

Institutions and individuals that contribute samples for a research project are often thanked in an acknowledgement, and may not include that can obscure important details about how other researchers can access the samples used. A separate sample availability statement should be used to ensure that future researchers understand which samples are available and how they can be requested.

For specimens from an institutional collection: Institutional collections differ from personal collections in that they typically have dedicated staff and formal accession, loan, and management policies. For each collection that provided samples, include the name, list of samples, sample availability, and contact information.

For samples from a personal collection: This category includes any collection maintained solely by an individual, even if that collection is stored at their workplace. Provide contact information and a list of samples from the collection. Note whether samples are available for use by other researchers.

Articles based on samples from multiple collections should list each one separately.

Keeping samples accessible: Samples may be lost, consumed, or transferred to new owners over time. To keep samples findable, we recommend providing ownership and contact information in the sample metadata record associated with a sample's persistent identifier. Many systems, including SESAR, allow users to include this information.

Examples

For samples from a personal collection:

A. Smith's personal collection (AS-1, AS-2, AS4). Samples available upon request. Contact A. Smith (asmith@college.edu) to request samples.

For samples of uncertain availability:

A. Smith's personal collection (AS-1, AS-2, AS4). Contact A. Smith (asmith@college.edu) for information about sample availability.

For samples from an institutional collection:

Smithsonian National Museum of Natural History, Department of Mineral Sciences (NMNH 1234, NMNH 2345). Samples available upon request. Contact Leslie Hale (halel@si.edu) to request samples