# Permanent Identifier FAQs

## Frequently Asked Questions about permanent identifiers

For full instructions on requesting a DOI, please see https://cyverse-doi-request-quickstart.readthedocs-hosted.com/en/latest/.

### Why should I publish my data in CyVerse Curated Data?

CyVerse Curated Data is the ideal platform for ease of data reuse. Because it is assigned a permanent identifier (DOI or ARK), it is stable and unchangeable, making it ideal for data citation. Because the data is stored in large-scale storage resources that are monitored 24/7, it is secure. Because it allows transfer, upload, and download across different computers and platforms, it can store very large datasets. And because its data is accessible to CyVerse's suite of large-scale computational analysis resources, users can seamlessly analyze, manage, and publish new results. For more information, see Is CyVerse Curated Data Right for My Data?

### What are the conditions for data to be published through the CyVerse Curated Data site?

Several conditions must exist in your data before it can be published in CyVerse Curated Data:

- You must be a registered CyVerse account holder. To register for an account, see Create Account on the CyVerse website.
- A dataset may be up to 100GB in size. If you are interested in depositing a larger dataset, please request an increased data allocation before requesting a permanent identifier.
- Data must be both curated and static. Once the data is published, it cannot be amended (although newer versions can be published).
- Data must be organized to identify the different components (raw, preprocessed, analysis, etc.).
- Compressed files must be in LASzip (http://www.laszip.org/) and the open-source Gzip (http://en.wikipedia.org/wiki/Gzip) family of compression formats including zip, tar, tar.gz, or tgz.
- At minimum, the dataset must include a complete description according to the DataCite standard. Domain-specific schemas, however, and the addition of ReadMe files, publications, or help notes that explain the data as well as how they were obtained and can be used, are welcome. In organizing and documenting the data, users should ask themselves, “What would someone need to reuse this data?”

### Can I publish to the Data Commons if my data is not static and curated by CyVerse?

Yes, you can make data available to the public via Community Released Data. See Publishing Data through the Data Commons.

### What is a DOI ?

A DOI is a Digital Object Identifier. It serves as a permanent, redirectable identifier and URL for your dataset, so that even if the location of your dataset changes, it can still be found with the same ID. DOIs are issued by CyVerse through the DataCite service.

### Do I need to contact CyVerse before requesting a DOI?

The process of requesting a DOI is automated through the DE, but some process must be handled manually, such as DOIs for datasets with more than 1000 files or DOIs for datasets that are stored somewhere other than /iplant/home/share/commons_repo/curated. If you match either of those cases, please contact us a doi@cyverse.org.

Also contact us if you have questions about how to organize your data or what scientific metadata to include.

## How much does a CyVerse permanent identifier cost?

At this point in time, CyVerse does not charge for DOIs or ARKs. However, the dataset must meet the requirements on the page, Is CyVerse Curated Data Right for My Data? In the future, there may be a charge for issuing permanent identifiers in the CyVerse Curated Data site.

## How long will it take to obtain a permanent identifier and publish my data?

Providing that your dataset is in good order and ready to be published, the process may take up to one week, as it may involve a dialogue with the CyVerse Curated Data curators. If your data is well organized and the metadata is complete and accurate, the process will be much faster (1-2 days). It is best to submit your request at least one week before you need the identifier (e.g., for a manuscript submission).

## Can I publish different versions of my data?

Yes. Each new version must be documented, and will be assigned a new permanent identifier that references the original dataset. For new versions, contact us at doi@cyverse.org.

## How small or big should my data be to be published?

The size of the dataset is less important than its utility to the scientific community. Although there is no lower size limit for requesting a DOI or ARK, the default upper size limit for data allocations on CyVerse is 100GB. If you are interested in depositing a larger dataset, please request an increased data allocation before requesting a permanent identifier.

## What is the policy for submitting compressed data to CyVerse Curated Data?

Certain file types are regularly transferred, stored, and used in applications in a compressed form, such as FASTQ for genomic data and LAZ for LIDAR data. Curated Data supports the deposition of files in the following open compressed formats: LASzip (http://www.laszip.org/) and the open source Gzip (http://en.wikipedia.org/wiki/Gzip) family of compression formats including zip, tar, tar.gz, or tgz.

## Can I publish data in CyVerse if I am not a CyVerse user?

You need to have a CyVerse account to publish your data in the Data Commons repositories (Community Contributed or Curated Data). You do not have to be a user of the entire platform, but at minimum you must be able to upload data, add metadata, and use the Discovery Environment. If you have not used the DE's metadata features before, start with Using Metadata in the DE, and read the section on metadata templates.

## How secure is the data in the Curated Data site?

Data in our platform is stored in large-scale storage resources that are monitored 24/7. Data is authenticated through checksum analysis at ingest, and is locally and geographically replicated so that if any one system fails there will always be a safe copy of your data.

## What is CyVerse Data Commons’ long-term commitment to hosting public data?

If and when the Data Commons cannot host your data in CyVerse Curated Data, it will transfer custody of the data to another repository and will change the target URL to which the identifier points.

## What if in the future I want to move my data to another repository?

If you want to move your data to another repository, please send a ticket with the new URL location and we will change the DOI target. You may leave a copy of the dataset in the CyVerse Curated Data site for ease of reuse within the computational environment. We will update the metadata to reflect the relationship between the two identical datasets.

## How can I make it easier for people to give me (and my co-creators) credit for using my dataset?

One way is to provide a short ReadMe file with the data and which provides specific instructions on how to cite the dataset. This is information that you obtain automatically once you are assigned an identifier.

Another way is to link your data to your ORCID (see http://orcid.org/). ORCID provides a persistent digital identifier that distinguishes you from every other researcher and supports automated linkages between you and your professional activities, ensuring that your work is recognized. The DataCite metadata template includes places to list ORCIDs of the creator. ORCIDs of contributors can be added as user metadata.

## Whom do I choose for the creator versus contributor?

A dataset can have only one creator but may have multiple (or no) contributors.

The *creator* is the single person or group with primary responsibility for the dataset. The creator is either the lead author of the dataset, the senior author of the dataset, or a consortium of authors, and should be the same as the name used for the folder containing the data. The creator does not need to be the person who is submitting the identifier request.

You also can add the role of *contributors*, which can include anyone who made a significant contribution to the creation of the dataset.

## Which license can I use to publish my data?

You can choose one of two licenses, depending on the materials you will be publishing:

- ODC PDDL for non-copyrightable materials (i.e., data only).
- CC0 for copyrightable material (Workflows, White Papers, Project Documents).

If you have special circumstances that require a different license (e.g., your dataset is aggregated from previously published data that already has another license), please contact us at doi@cyverse.org.

## Can I make changes to the metadata record?

Once the data is public it will not change, and thus metadata should be stable as well. However, it is possible that you may need to do minor edits or enhance the metadata. Version changes of the metadata will be recorded. To request changes to metadata data, contact us at doi@cyverse.org.

## What metadata standards does CyVerse support

## for data publication?

All data will have a DataCite metadata entry. However, take into consideration that the DataCite metadata is a citation metadata that does not represent the complexity of the research that went behind creating your data, and we encourage you to include additional metadata. We suggest that you include the metadata records and other help documents in your publication package within a folder labeled as “metadata” so it is easily identifiable for other users.

## How can I organize and package my dataset for publication in the CyVerse Curated Data site?

See Guidelines for Organizing Your CyVerse Curated Data for more details on how to prepare your data for publication.

## What if I want to change or add metadata to my public data?

If you need to make changes to the metadata of a dataset with a DOI or ARK, contact us at doi@cyverse.org.

## Where can I go for help on permanent identifiers?

Email the CyVerse DOI team or ask a question on our forum.

**On this page:**

**Related pages:**

- Requesting a Permanent Identifier in the Data Commons
- Email the CyVerse DOI team
- DataCite standard
- Digital Object Identifier System website
- Understanding Identifiers
- Requesting an increase to your Data Store allocation
- Using Metadata in the DE
- Guidelines for Organizing Your CyVerse Curated Data
- Is CyVerse Curated Data right for my data?