

OLCF's DOI Framework

<https://doi.ccs.ornl.gov>

Contact

Raghul Gunasekaran (r2y@ornl.gov)

Sudharshan Vazhkudai (vazhkudaiss@ornl.gov)

OLCF's DOI Web Portal

- <https://doi.ccs.ornl.gov/>
- The portal lists all DOI's created by OLCF users
- Open access to metadata and data
- For downloading DOI data
 - Click on the download link and provide your email address.
 - Data will be staged for open access and you will receive an email with download instructions.
 - Currently downloads are via http and Globus.
- Only active OLCF users can login to the web portal
 - First time user, please register with NCCS credentials, ccs/<username> and passcode (pin + token)
 - Then login with ccs/<username> and passcode.
 - User can view all published DOIs, and new request's created and authored.
- The "HELP" link on the web portal has more detailed instructions.

Request for DOI

Request for DOI is a two step process

- Login to web portal and provide metadata information
- Data upload/transfer to a designated directory
- **First Step: Registration via web portal**
 - Using the web portal submit all the metadata descriptors
 - Metadata descriptors: title, description, keywords, authors, dataset type, subject categories, contributing & sponsoring organizations.
 - User will receive an email with instructions for data upload.
- **Second Step: Upload data**
 - Data should reside on Spider or NFS
 - Login to a dtn node, and run these commands
 - `module load doisub`
 - `doisub -r <request_id> <datafile(s)>`
 - Will request for passcode, please use ccs credentials.
 - User can upload data multiple times and it overwrites the last data upload.
- Once you are done with data upload, please go back to the web portal and click submit. This is to confirm data upload is complete and the user can no longer overwrite the data.

DOI Registration Process

- On receiving a request for DOI
 - Internal approval/review process
 - If request rejected, user notified and data is deleted.
 - If request accepted
 - Request submitted to OSTI, with metadata descriptors
 - OSTI issues DOI, registers with DataCite
 - Data archived in HPSS
- On getting a DOI
 - Every request gets a DOI and a landing page (URL)
 - Example: <https://doi.ccs.ornl.gov/doi/27>
 - Open access to DOI metadata and data
 - OLCF tracks data download frequency

FAQs

- What is the review process for getting a DOI ?

Currently, one or more domain experts will review to provide feedback on the DOI request. A more formal review process on the metadata and data organization is yet to be formulated.

- What is the maximum data size ?

Currently, the service creates a single tar of all the data files associated with a DOI. As downloads are via Globus we are limited by what Globus can support. Though, our upload tool can support only few 10's of GB, it is not a limitation we can work around it to support your case. Also, we are working on supporting data uploads via Globus replacing our upload tool. Globus has supported a 1 PB single file transfer.

In the near future, the service will support a select and download feature to choose from a collection of file's within a large data set identified by a single DOI.

- How long will the DOI data be available?

The general requirement is data be made available for at least 5 years. Beyond 5 years is an OLCF policy decision, and also we track DOI data downloads which should provide a perspective on the DOI's popularity.

- I want to update my DOI data ?

A DOI dataset once published cannot be updated. A new DOI needs to be obtained for the newer dataset, and a metadata descriptor field will identify that the new DOI supersedes the old DOI. All requests for the old DOI will be redirected to the new DOI.

Next Steps / Work in Progress

- Globus integration for data uploads
- Metadata descriptors to make DOIs searchable
 - We request feedback from users on how to best describe scientific datasets across science domains.
- Integration with workflow systems
 - The user will still need to submit a request via the DOI web portal
 - For data upload, a dedicated scratch space will be made available for the user to append data (files).
 - Once all the data files are generated, user can go back to the web portal and click submit, confirming data upload is complete.