



March 04, 2026 | DOE Data Days

Constellation 2.0: Data Curation and Publication Support at OLCF

Meghan Berry

National Center for Computational Sciences



U.S. DEPARTMENT
of **ENERGY**

ORNL IS MANAGED BY UT-BATTELLE LLC
FOR THE US DEPARTMENT OF ENERGY

FRONTIER



Constellation supports OLCF open data

220+
datasets

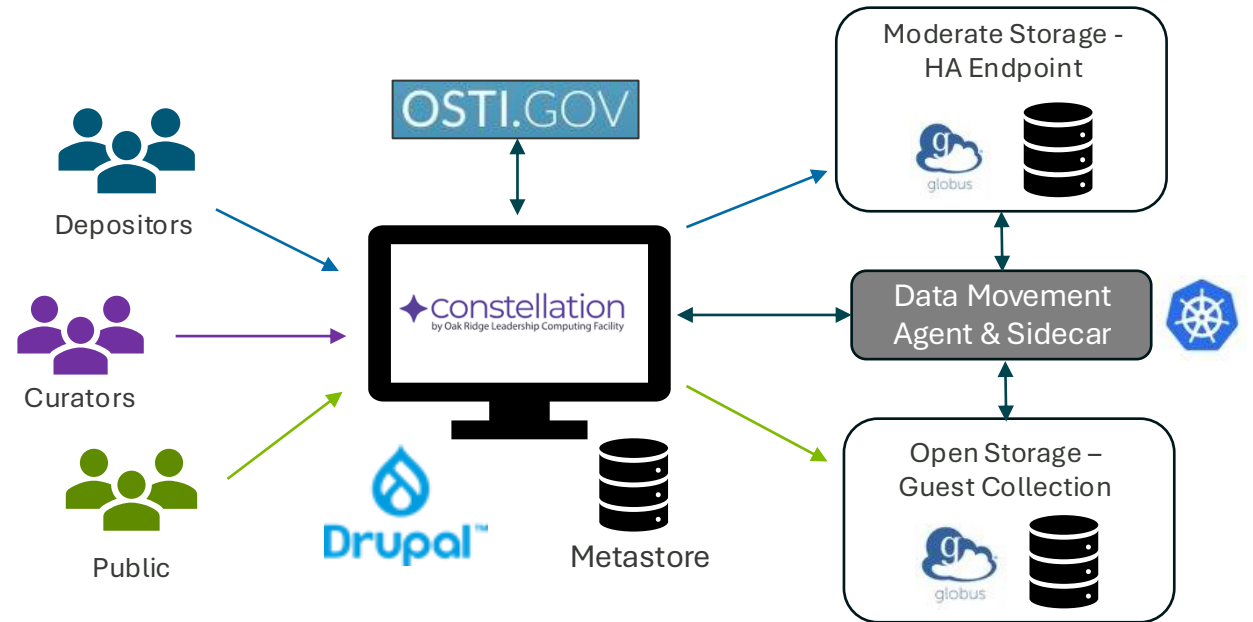
4.5 PB
of published
data

2 PB
largest
dataset

- Generalist data repository hosted by the Oak Ridge Leadership Computing Facility
 - Exascale era = large scale scientific data
 - All open data
 - Free publishing for OLCF users
 - Available to ORNL researchers across domains
 - Automates registry for DataCite DOIs via OSTI

Constellation 2.0 launched end of FY25

- Homegrown software was developed in 2016
- Motivations for redesign:
 - Service growth
 - Modernize architecture and interface
 - Following FAIR principles
- Key features of Constellation 2.0:
 - DKAN data portal (Digital Public Good)
 - Added support for OSTI E-Link 2



Migration to Constellation 2.0

Software migration

- Sidecar implemented for data movement
- Drupal front end development

Challenges

- Meeting robust security standards
- Testing and deployment pipelines

Metadata migration

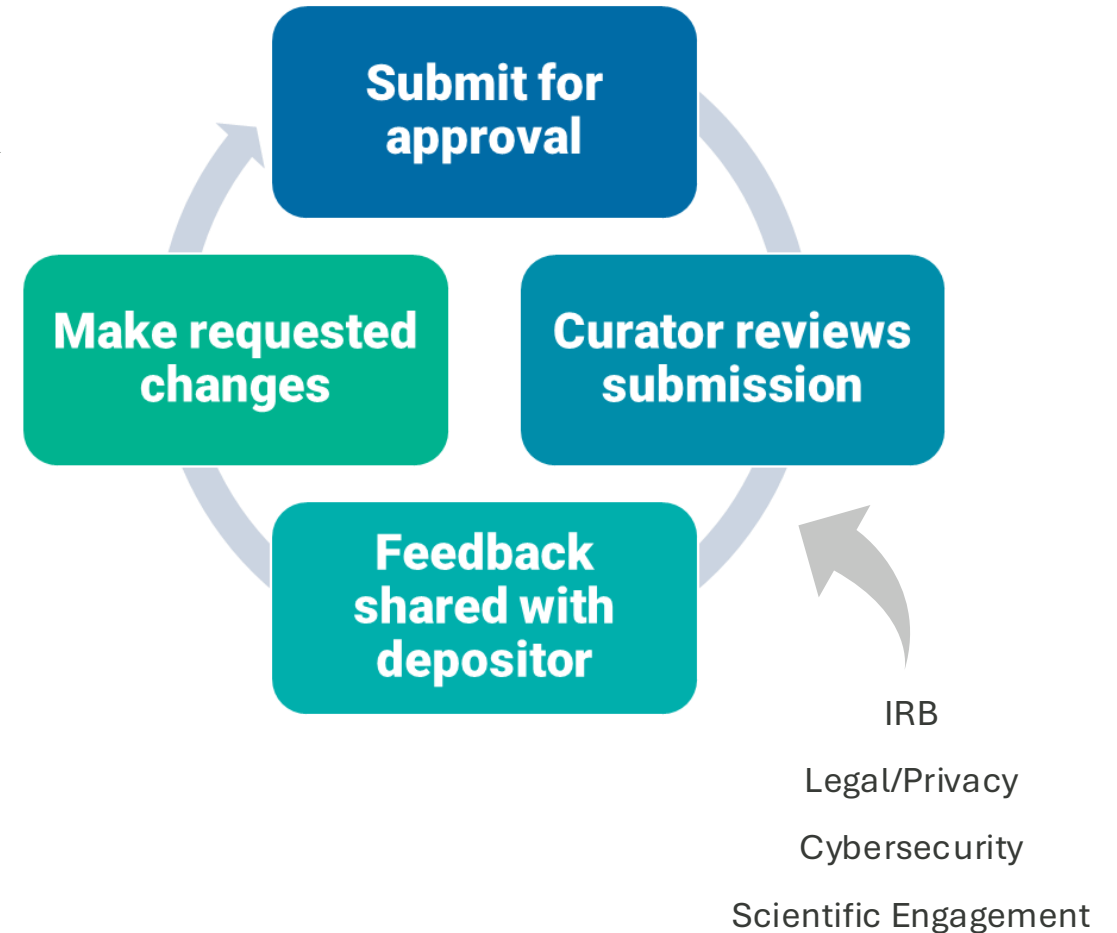
- New schema to meet DKAN, OLCF, and OSTI requirements
- Record cleanup and conversion pipeline

Challenges

- Collecting and validating funding IDs
- Ongoing metadata cleanup

Enhancing data curation workflows

- What data curation is and isn't
- Adoption of the Data Curation Network CURATE(D) framework
 - Record of work in curation logs
- Benefits of a structured curation process:
 - Metadata quality checks and enhancement
 - Supports DOE review and release alignment
 - Common issues are identified
- Curation at scale requires new tools



Fostering a culture of data stewardship

- Customer service focus and proactive communication
- Encouraging high quality data publication
 - Support for creating documentation
 - Identify high impact modifications
- Curation increases data reusability
- Furthering awareness of FAIR principles



Growing user outreach and engagement

Strategies to build awareness, trust, and engagement:

- Documentation for different audiences
- User group presentations
- Building relationships with research teams
- Communication mirrors other OLCF services

Outcomes:

- Every new publication includes a README file
- 47% increase in new DOIs FY24-25

The image shows a screenshot of a web application. The top section is titled "Data Depositor Guide" and contains a welcome message and instructions. Below this is a list of steps to publish data. The bottom section is titled "Data Publication (Constellation)" and contains a list of links to various pages. A navigation menu is visible in the center, listing various services and applications.

Data Depositor Guide

Welcome! Constellation is a research data repository service of the Oak Ridge Leadership Computing Facility (OLCF), available for OLCF users as well as Oak Ridge National Laboratory (ORNL) staff. Datasets published in Constellation are assigned a Digital Object Identifier (DOI) by the DOE Office of Scientific and Technical Information (OSTI) and reported for inclusion in the [DOE Data Explorer](#). Constellation can assist projects in meeting the objectives of the 2023 DOE Public Access Plan for federally funded research.

Data submitted to this repository may be the results of research:

- 1) conducted using a project allocation on OLCF resources including Frontier, Summit, Slate, Andes, etc., or
- 2) funded wholly or in part by the US Department of Energy under its contract with UT-Battelle, LLC (DE-AC05-00OR22725) for the management of ORNL.

For questions and help, contact [Data Publication \(Constellation\)](#) or [Data Publication \(Constellation\)](#).

Steps to publish a dataset

There are seven steps to publish a dataset:

1. Obtain a Globus endpoint
2. Create your Constellation account
3. Reserve a DOI for your dataset
4. Add metadata and documentation
5. Add data files and upload
6. Address any questions

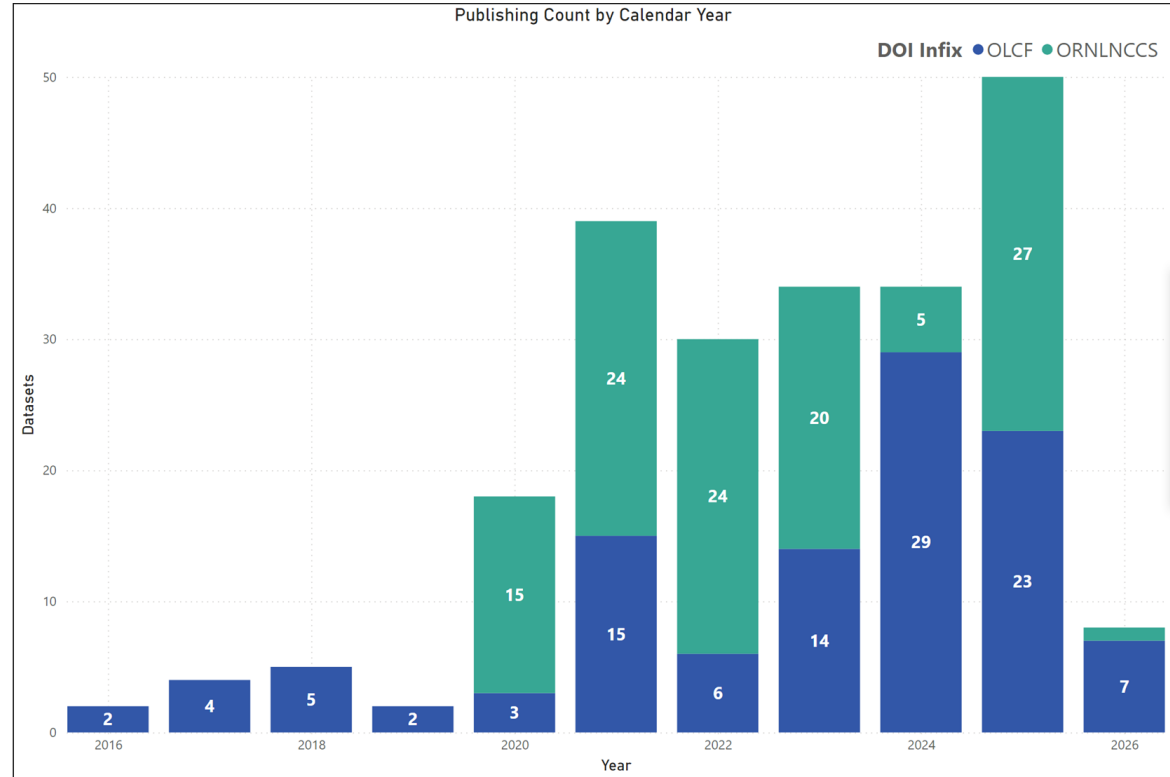
Data Publication (Constellation)

- Constellation Overview
- Infrastructure
 - Preparing Your Dataset
 - Data Review and Release
 - Reserving a Draft DOI
 - Removal of Dataset Drafts
 - Adding Data and Documentation
 - Review States
 - Data Curation
- Reusing Constellation Data

[Previous](#)

Measuring repository usage and impact

- Google Analytics and Umami for web traffic and user behavior
- Planned user dashboards – downloads and citations
- User feedback drives development



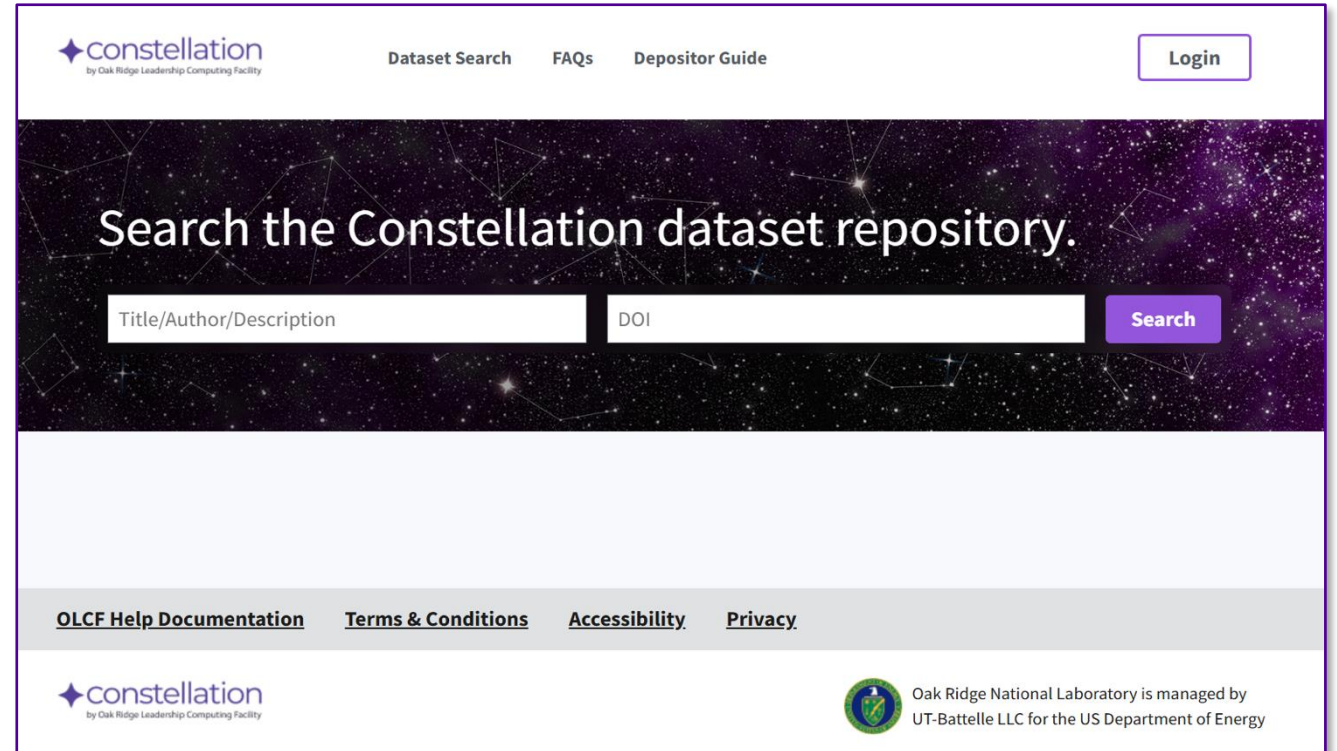
Lessons learned

-  Document and prepare for edge cases
-  Balance needs of various user groups
-  Look for insight into end data users
-  Cross-team collaboration is essential
-  Plan ahead to collect metrics and feedback

Future directions

- Constellation is ready for the future with upgraded infrastructure and curation
- Planning for new features – automated curation tools, additional metrics for users
- AI data readiness

Questions?



This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Advanced Scientific Computing Research programs in the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.



ORNL IS MANAGED BY UT-BATTELLE LLC
FOR THE US DEPARTMENT OF ENERGY