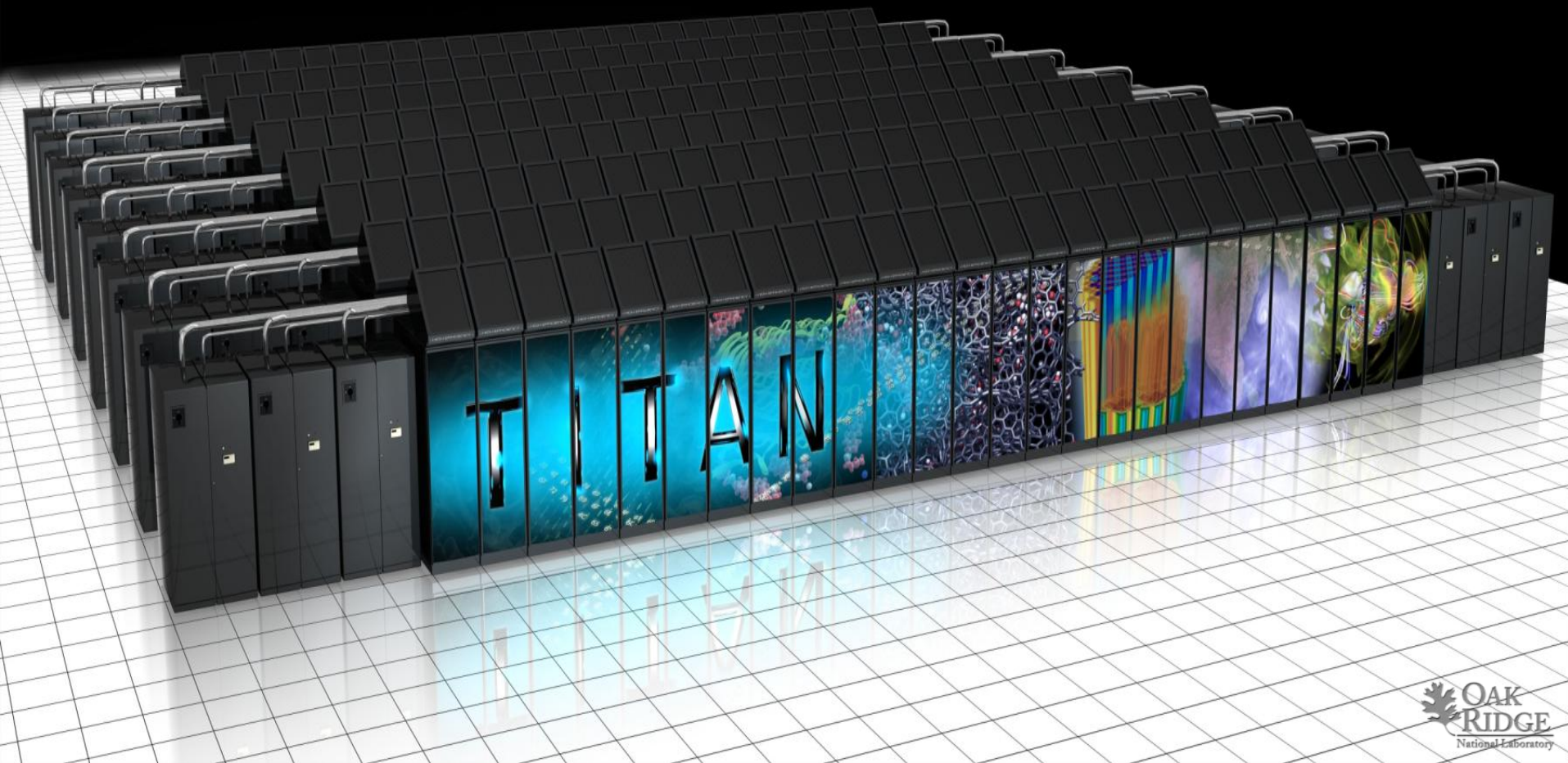


Constellation: Digital Object Identifiers for OLCF Data

Dale Stansberry,
Advanced Data and Workflows Group,
National Center for Computational Sciences



What is a DOI?

- An ISO standard for unique, persistent, and resolvable identifiers for data “objects”
 - Similar to ISBN for books
 - Makes data citable like publications
- Identifier format: 10.<registrant ID>/<data ID>
 - For OLCF: 10.13139/OLCF/12345
- Registration authorities administer ID ranges and store and index DOI numbers and associated metadata
- Includes metadata plus references
- Data can be any digital format
- Resolved by URL that directs access to DOI “Landing Page”
 - Includes access instructions / facilities

Why Do We Need DOIs?

- To improve repeatability and understanding of scientific research
 - Access to data used to support published results
- Office of Science and Technology Policy (OSTP) mandated planning for increased public access to research results.
- Required for any result published in peer-reviewed scholarly publications (supported by digital data) based on federally funded research.
- Specified in the “Holdren memo”, February 2013
- Bottom line: tax payers want more openness / visibility into funded research

Who Provides and Resolves DOIs?

- Within DOE, Office of Scientific and Technical Information (OSTI) is an RA with DataCite
 - OSTI provides DOI service freely for DOE projects
- OSTI registers DOIs with www.doi.org
 - Resolution redirects to landing page of DOI
 - Shows metadata
 - Data access instruction or download (if supported)
 - Example URL: <http://dx.doi.org/10.13139/OLCF/12345>
- OSTI pushes metadata to DataCite
 - Can be searched from DataCite
 - Also gets indexed by external search engines (Google, Bing)
- OSTI does NOT curate data

DOI System Requirements

- Data should be available permanently *
- Provide core metadata (Dublin core)
- Provide landing pages
 - URL that links to the DOI provides necessary context for accessing the data
 - May provide authenticated or open access to data
- Data should have some QA and integrity
- DOI implies data management planning
- * Tombstone page
 - Coordinate with DOI RA (OSTI) to maintain a persistent page when data must become unavailable

Constellation is OLCF DOI System

- Constellation is an NCCS/OLCF Data Portal Project
 - DOI access and creation is key feature
 - Access to OLCF functions under development
- Available to all active OLCF users
- Allows anonymous downloads of DOI data
- URL: <https://doi.ccs.ornl.gov>
 - Requires Globus Account
 - Free registration
- DOI Creation involves internal review process
 - Policy is work-in-progress

Constellation – DOI Download

- All (anonymous) users have ability to browse, search, and download DOIs via Constellation
- Main window shows “All Published DOIs” (from OLCF)
 - Can organize by date, title, or DOI number (toolbar)
 - Search terms match against title, description, and keywords
- View and/or download DOI by expanding DOI in list
- Download initiates staging of data
 - Email address is required for notifications
 - User will be notified once dataset is staged from HPSS into download cache, download link will be included (limited time)
 - Future version will support direct transfer from HPSS via Globus

Constellation – DOI Landing Page

- Alternatively users may arrive at DOI landing page via external reference (search engine, citation)
- Landing page URL is <https://doi.ccs.ornl.gov/ui/doi/<req-no>>
 - Request number is visible to creator/authors via Constellation application
- Shows same information as in Constellation application
- Download functionality from landing page is identical

Constellation – DOI Creation

- User logs-in and selects DOI create option from main menu
- Fill-in form with at least required metadata:
 - Title, description, author(s), dataset type, project ID, originating and sponsoring orgs, DOE contract no.
- Read and accept terms and conditions, then click “Create” button.
- A “DOI Request” is generated and can be found in “DOIs as Creator” panel.
 - Can view/edit metadata as needed
- Next: Upload data files

Constellation – DOI Creation, Cont.

- Upload, view, and manage DOI data files by selecting “files” button
- Opens Globus Online File Transfer window
 - Source endpoint is pre-filled as “OLCF ATLAS”, path /~/
 - Destination endpoint is “OLCF DOI-UPLOADS”, path /<req no.>/
 - May require activating Atlas endpoint
 - Destination endpoint is private
 - Multiple DOIs will have different request numbers (see DOI info)
- Select and transfer files to “uploads”
 - Can delete or overwrite files as needed
 - No time limit for organizing DOI files
- When ready, submit DOI
 - Review process begins

Constellation – DOI Review Process

- OLCF DOI admin is notified of new DOI requests
- Admin may choose to review directly, or assign one or more reviewers
 - Reviewers are granted read access to DOI data via Globus
 - Reviewers provide feedback to DOI admin with accept/reject recommendations and rationale.
- If accepted, DOI number is acquired from OSTI, data is archived into HPSS, and user is notified via email.
 - DOI will not be available for download until archiving completes
- If rejected, submitting user is notified via email
 - User may create duplicate DOI request, with corrective actions

Constellation Demo

- <https://doi.ccs.ornl.gov>
- On-line help describes registration, DOI tasks, and general application use.

Acknowledgements

- Contributors to the feasibility study
 - Doug Fuller
 - Terry Jones
- OSTI Support
 - Mark Martin
 - Jannean Elliott
- Constellation Development Team
 - Sudharshan Vazhkudai – Technology Integration Group Lead
 - Ross Miller
 - Mitch Griffith
 - Matt Li
 - Raghul Gunasekaran

Questions?