DOI Next Steps Telecon - Meeting Notes - 03 Feb 2022

Nathan James

Lalit Wanchoo

Sarah Voorhies

Meeting Recording

03 Feb 2022

Meeting Recording

1. **Agenda**

    a. **DOI System Update (1:05)**

        i. EOSDIS DOI Account with DataCite being moved under NASA Consortiuum managed by NASA OCIO office.
        ii. Personnel Changes - Elli moving on - Replaced by Sarah Voorhies

    b. **DOI Usage: Looking Ahead   (1:15)**

        - Usage-Based Discovery Initiative

            ○ Introduce Vinny Inverso
            ○ UBD Knowledge Graph Follow-up
            ○ Lot of DAAC support so far. Over 10,000 datasets and articles have been collected from the DAACs
            ○ Exciting ideas moving forward on how we might use this data.

                ▪ DOIs and collected data enable us to observe trends about where data is being used and their impacts
                ▪ UBD Graph gives potential to power semantic search moving forward combining ML with Graph can do a lot with natural language processing and advanced search and discovery in general.
            ○ ORNL comment: Our experience is that it takes about minimum of 12 mos after a dataset is published before seeing the 1st citation in public literature.
            ○ UBD Category Titles derived from Air borne data Campaign titles with some augmentation
            ○ See the Chat for links to Presentation and UBD Knowledge Graph
            ○ Dataset Search on Google - Low hanging fruit DOI exercise

                ▪ https://datasetsearch.research.google.com/search?query=10.5067
                ▪ Google Scholar API at https://serpapi.com/google-scholar-api
        - GES DISC Citation Search using DOIs - Irina

            ○ Uses SCOPUS from Elsevoir, Web of Science, Google Scholar and DataCite APIs
            ○ Hard to create automated tools to search articles for citations via DOIs
            ○ NOAA Connection - Discussions about Citations are on-going

                ▪ Coleridge initiative - a suite of machine learning models that identify datasets in paper texts. Inter-agency cooperation with NYU - agencies of NOAA, and 2 other agencies - but not NASA yet. GES-DISC is submitting their datasets to this effort.
            ○ **Issue with CMR/DOI database.** When we search for our datasets via DOIs in the paper. We have retired DOIs in our system if datasets are retired in CMR. Problem with publications that reference those retired DOIs. Those DOIs should say that a dataset is retired on the DOI LP or have a reference in the new DOI that points to the latest version.
                ▪ Q: Does CMR delete the metadata for Retired Collections?
                ▪ A: Lalit: Early discussions with CMR, it was said that old info (metadata) would eventually be purged from CMR. DAACs confirm that deletions are happening. DAAC Mgr disagrees with this approach of deleting Datasets in CMR especially if there is a DOI attached. Take this up with Valerie
                ▪ Smaller DAACs felt different about this initially. They thought it made sense NOT to keep those retired products in CMR. They keep only the active products maintained to minimize labor. But they are reconsidering this as they consider those who reference older products from publications. Other DAACs recommend the ESDIS practice of using DOIs landing page to reference and point back to older versions.
                ▪ Recommendation: There should be a product description in CMR for retired products for identification purposes.
                ▪ ORNL: Curation case to discuss with CMR team - publications using CMR concept ID instead of DOI so now the publication reference does not resolve because the product was retired in CMR.
        - GraphDB Update
        - CMR DOI Developments

            ○ Handling multiple DOIs in a collection - Erich Reiter

                ▪ UMM-C handles DOI's in 2 ways.
                ▪ First is the main DOI field which is the DOI of the collection.
                ▪ Second is the AssociatedDOIs field which holds any number of DOIs of data that is associated with this collection.
                ▪ MMT can be populated with DOIs now

                    - You can enter AssociatedDOI information using MMT.

- EDSC displays the main DOI field at the top called DOI
- EDSC displays the Associated DOIs in a field called Secondary DOIs.
- ASDC Comment: Possible to overload this field with many associated DOIs. Could confuse the users
- SEDAC Bob Downs says to follow the guideline of DataCite use of related DOIs - They have different fields that help specify the type of association by relationship. That could eliminate confusion. **Request for examples of this.**
- **CMR Q: Does an Associated DOI have a description field that allows one to say what the relationship is?**
  - Things We'd like to see in CMR...

    - Group DOIs for single campaigns
    - DOIs from other agencies
    - Products with related datasets
- Open Science Initiative - Nate
c. **Collection of Collection DOIs - Guidelines and Requirements (1:45)**

   i. Data Product Group DOIs - Airborne Campaigns - Debra Smith

      1. Community-defined groupings of data collections with a descriptive group landing-page
   ii. Group Campaign DOIs - ASDC/ORNL/GHRC
   iii. Document Group DOIs - GES - Lena, Thomas
   iv. Group DOI Requirments WG - Lalit