Topics

- Ideal State
- Preservation Challenge
- Preservation Content
- DOI Implementation
Ideal State

- Traceability of everything related to a dataset to be able to answer all possible questions that user may raise, e.g.,
  - What were the inputs?
  - How was the dataset generated – software, algorithm, computer, operating system, etc.?
  - Who were the authors of algorithm?
  - What instrument(s) did data come from?
  - What satellite did it (they) fly on?
  - Who funded the development?
  - What is the quality of data? What are the limitations?
  - What can the dataset be used for?
  - What publications have used the dataset?

- Unambiguous references important for scientific understanding and reproducibility

- Long-term assumption – authors of datasets and related items not available for answering questions
Preservation Challenge

Instrument Teams / PI’s → Science Data Product Documentation

Product Generation Support Teams (SIPSs) → Science Data Product Software

Science Data Products → DAACs

DAACs → Mission Data Calibration

Mission Data Calibration → Calibration Team

Mission Logs → Mission Operations

Level 0 Data → Science Data Software Tools

Ancillary data sources (e.g., NOAA) → Science Data Product Algorithm Input

Science Data Product Algorithm Input → Preflight/Pre-Operations

Preflight/Pre-Operations → Data gathering project (e.g., flight project)

Data gathering project (e.g., flight project) → Validation Team

Validation Team → Science Data Product Validation

Science Data Product Validation → Major production data

Other artifacts
Preservation Content Categories

1. **Preflight/Pre-Operations**: Instrument/Sensor characteristics including pre-flight/pre-operations performance measurements; calibration method; radiometric and spectral response; noise characteristics; detector offsets

2. **Science Data Products**: Raw instrument data, Level 0 through Level 4 data products and associated metadata

3. **Science Data Product Documentation**: Structure and format with definitions of all parameters and metadata fields; algorithm theoretical basis; processing history and product version history; quality assessment information

4. **Mission Data Calibration**: Instrument/sensor calibration method (in operation) and data; calibration software used to generate lookup tables; instrument and platform events and maneuvers

5. **Science Data Product Software**: Product generation software and software documentation

6. **Science Data Product Algorithm Input**: Any ancillary data or other data sets used in generation or calibration of the data or derived product; ancillary data description and documentation

7. **Science Data Product Validation**: Records, publications and data sets

8. **Science Data Software Tools**: product access (reader) tools.

9. **Checklist**: “metadata” about the above 8 categories showing how and where items in each category are preserved

ESDIS Project started implementing DOIs in 2011 for datasets held in EOSDIS – goal is to assign DOIs to all datasets

- ESDIS is DOI issuing authority for datasets at most DAACs – prefix 10.5067
- Exceptions are ORNL and SEDAC - preceded ESDIS in DOI implementation
DOI for datasets - Implementation

Step 1: Visit ESDIS DOI wiki Website
(https://wiki.earthdata.nasa.gov/display/DOIsforEOSDIS/)

Step 2: Download DOI Submission Form
(https://wiki.earthdata.nasa.gov/display/DOIsforEOSDIS/DOI+Submission+Form/)

Step 3: Fill the New/Update Form
(check examples on the wiki)

Step 4: Submit Form to ESDIS
(ESDIS Contact Team)

Step 5: Review of DOI Information

Step 6: Process DOI Information

Step 7: Reserve/Register/Update DOI Information

Step 8: Post Information on the ESDIS wiki website

Credit: Lalit Wanchoo, ADNet/ESDIS
Status of Digital Object Identifiers (DOI) Created with the ESDIS by the Data Providers as of 16 June 2016

- Registered: Created with the ESDIS and also registered with EZID
- Reserved: Created with the ESDIS but not registered with EZID

<table>
<thead>
<tr>
<th>Distributor</th>
<th>Registered</th>
<th>Reserved</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESDIS</td>
<td>3,070</td>
<td>374</td>
<td>3,444</td>
</tr>
<tr>
<td>ORNL*</td>
<td>1,151</td>
<td>0</td>
<td>1,157</td>
</tr>
<tr>
<td>SEDAC*</td>
<td>226</td>
<td>0</td>
<td>226</td>
</tr>
<tr>
<td>Total</td>
<td>4,447</td>
<td>0</td>
<td>4,827</td>
</tr>
</tbody>
</table>

* Directly registered with EZID
Under the umbrella of NASA’s ESDSWG, identifiers and citations have been themes of WG’s since 2012

- Data Stewardship WG
- DOI WG
- Citations and Identifiers WG

Topics considered:

- Past: DOI Syntax, Assignment Process, Landing Page Contents, DOI Field Formatting, Citations Reformatting, Citation Policy
- Current: Identifiers for Non-Dataset Objects (focus on software right now)
Citations and Acknowledgements

Open Data and the Importance of Data Citations: The NASA EOSDIS Perspective

Data Citations and Acknowledgements

- ASDC - https://eosweb.larc.nasa.gov/citing-asdc-data
- ASF DAAC - https://www.asf.alaska.edu/about/how-to-cite-data/
- CDDIS - http://cddis.gsfc.nasa.gov/About/Citing_our_data.html
- LAADS - http://modaps.nascom.nasa.gov/services/faq/LAADS_Data-Use_Citation_Policies.pdf
- LP DAAC - https://lpdaac.usgs.gov/citing_our_data
- OB.DAAC - http://oceancolor.gsfc.nasa.gov/cms/citations
- PO.DAAC - http://podaac.jpl.nasa.gov/CitingPODAAC
Challenges

- Takes time to implement DOI’s for all datasets
- Identifiers for non-dataset objects still being considered
- Provenance implementation – establishing links among different objects - still in experimental stage
Acknowledgments

- Thanks to Nate James (ESDIS Project) and Amanda Leon (NSIDC DAAC) for their comments on a draft of this presentation.

- H. K. Ramapriyan's work was supported by NASA’s contract NNG15HQ01C with SSAI.