

## **Data Citation breakout session notes**

About 40 attendees in the room.

Online Participants: Pam Michael, Peter Grimm, Subhash Kuvelker

### **Intro - Dan Kowal, NGDC**

Need IDs to ensure, in part, that datasets are not lost or orphaned over time.  
Nice if ID scheme was not developed in-house but is more broadly applicable.

### **ESIP Data Citation Guidelines - Ruth Duerr, NSIDC**

Citation provides scientific reproducibility, credit, transparency and tracks data ; tracks impact

ESIP Citation work started with a review of members' citation guidelines

Mandatory parts of citation

- \* Author - Credit/Responsibility/accountability could also include editors or compilers
- \* Release dates - depends on regularity of updates
- \* Title
- \* Version
- \* Archive or distributor - data stewards/publishers
- \* Locators - URL/DOIs/ARKs or handles - if there is one fixed medium included it. point to a landing page for dataset. don't include org name in case it changes.

DOI can be assigned before dataset is "complete." Should assign at first citable moment.

Pick the right moment, because there is a maintenance cost.

Major version change of dataset => new DOI assigned.

[http://wiki.esipfed.org/index.php/Interagency\\_Data\\_Stewardship/Citations/provider\\_guidelines](http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations/provider_guidelines)

Alternatives to DOI: Handle or ARK or PURL.(?) DOI favored by publishers.

DOI = Digital Object Identifier.

ARK = Archival Resource Key.

Citation may include date of access. Should it also include the subset used (e.g., time range of data actually used in the research)? => There exists optional Subset field in citation guideline.

Referencing individuals (as creator or editor of dataset, for example): should everybody have an identifier such as a Researcher ID? <http://www.researcherid.com/>

Data sets can be a compilation of smaller data sets.

Citation can list if it is an updated data set.

### **NASA EOSDIS Citation Guidelines - John Moses, NASA**

So far ORNL DAAC is the only one using DOIs, but these will become more prevalent in NASA.

NASA has a subscription for DOIs (provider: California Digital Library (CDL) EZID). John is subscription holder.

Assign DOI to each EOSDIS Standard Data Product. (Not to each granule, for example.)

NASA has suffix syntax including possible information like mission, instrument, processing level, etc. ESIP & NWS & NGDC argue against adding semantics to the identifier.

Can you reserve a series of DOIs?

Not best practice to have any identifying info

2012 NOAA EDM Conference  
Tuesday, May 15-17

**Data Citation breakout session notes**

NOAA Data Citation Procedural Directive - Jeff de La Beaujardière

Looking to have draft in 2012 Q4 catching up to NOAA

Working with journals to mandate data citation

NOAA would like to subscribe to NOAA-wide license for minting DOIs

Would like for people to get credit - i.e. this person has 6 citable datasets that they curated... Hoping that unique ids to find duplicates.

How would you deal with alternative titles?

What about other digital products - tools, algorithms,

ESIP is thinking of other types of identifiers

Also links to other processing steps, etc.

Proposed by someone after the session: There should be a checksum or something similar in the identifier. If the text of the DOI is accidentally mangled, there is no way to recover automatically.

<end session notes>