AGU Paper Number: IN13C-1854

# Developing Data Citations from Digital Object Identifier Metadata

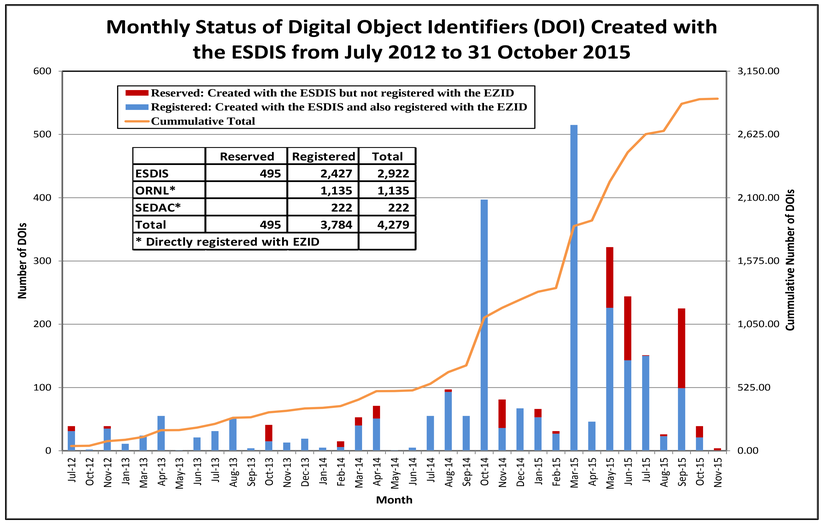
Lalit Wanchoo[1] (lalit.wanchoo@nasa.gov) and Nathan James[2] (nathan.l.james@nasa.gov),

[1] ADNET Systems, Inc., 7515 Mission Drive, Suite A100, Lanham, MD 20706 , [2] Earth Science Data and Information System Project (Code 423), NASA Goddard Space Flight Center, Greenbelt, MD 20771

## NASA Digital Object Identifiers

Digital Object Identifiers (DOIs) are permanent actionable identifiers that enable researchers to find, access, and use data products referenced in publications. NASA's Earth Science and Data Information System (ESDIS ) Project assigns DOIs to its science datasets to track the use of its products and to assign proper credit for the products through data citations. Since its inception 5 years ago, the ESDIS DOI registration system has registered nearly 3000 DOIs with nearly 500 DOIs held in reserve awaiting compliance.

When registering DOIs, ESDIS requires certain DOI metadata elements be collected for the DOI landing page as recommended by NASA's Earth Science Data System Working Group (ESDSWG). The landing page provides sufficient information to 1) identify NASA data as referenced in a science publication, 2) credit data creators and distributors, and 3) access the data itself enabling the trace-ability and reproducibility of the results. However, the required elements for this DOI landing page are also the core required elements for forming an Earth science data citation.




| | Reserved | Registered | Total |
|---|---|---|---|
| ESDIS | 495 | 2,427 | 2,922 |
| ORNL* | | 1,135 | 1,135 |
| SEDAC* | | 222 | 222 |
| Total | 495 | 3,784 | 4,279 |
| * Directly registered with EZID | | | |

## Data Citations

Data citations are gaining significant attention within the scientific community and amongst data centers. By identifying specific data used in published research, the goal is to make it easier for scientists to discover the data and validate the findings as reported in the publication. In addition, data citations credit both the data creator and the data distributors for their roles in the data product lifecycle. Data citations also promote the reproducibility and transparency of research data, which is becoming increasingly more critical for the advancement of the scientific analysis.

Though there are some challenges to creating a properly formatted data citation, they are far out-weighed by the benefits to both the scientific community and the data center. By making data products more easily discoverable, the use of data citations can also result in increased product distribution through the data centers. In addition, when DOIs are incorporated, data citations can also direct researchers to the stewards of the data and other research activities using the same data product.

To facilitate the use of data citations, the ESDIS Project has established citation guidelines and samples that appear on the landing pages of all data products being distributed from its science archives.

**ESDIS DOI WIKI URL: https://wiki.earthdata.nasa.gov/display/DOIsforEOSDIS**

## Citation of Data Products

**1. ESIP recommendations:**

*Creator, Year, Product Title, Version, Distributor, Date of Data Access, DOI Name*

*Example: AIRS Science Team/Joao Texeira; (2013): Aqua AIRS Level 3 Standard Daily Product using AIRS and AMSU with HSB V6; NASA Goddard Earth Sciences Data and Information Services Center. May 2014 http://dx.doi.org/10.5067/AQUA/AIRS/DATA302*

**2. DataCite Citation tool results:**

*Creator, Year, Product Title, Distributor, DOI Name*

*Example: AIRS Science Team/Joao Texeira; (2013): Aqua AIRS Level 3 Standard Daily Product using AIRS and AMSU with HSB V6; NASA Goddard Earth Sciences Data and Information Services Center. http://dx.doi.org/10.5067/AQUA/AIRS/DATA302*

**3. American Meteorological Society:**

*Creator, Year, Product Title, Version, Distributor, Date of Data Access, DOI Name*

*Example: AIRS Science Team/Joao Texeira; (2013): Aqua AIRS Level 3 Standard Daily Product using AIRS and AMSU with HSB V6; NASA Goddard Earth Sciences Data and Information Services Center. May 2014 http://dx.doi.org/10.5067/AQUA/AIRS/DATA302*

**4. The Economic and Social Research Council (ESRC) - United Kingdom Data Services:**

*Identifier, Creator, Title, Publisher, and Publication year, Resource Type, Version*

*Example: University of Essex. Institute for Social and Economic Research and National Centre for Social Research, Understanding Society: Wave 1, 2009-2010 and Wave 2, Year 1 (Interim Release), 2010 [computer file]. 3rd Edition. Colchester, Essex: UK Data Archive [distributor], February 2012. SN: 6614, http://dx.doi.org/10.5255/UKDA-SN-6614-3.*

*Version – not currently in ESDIS DOI model but most of the data providers include it in the Title.*
*Date of Data Access: to be provided by the authors*
*Resources Type, Version: These are optional*

## Citation of Services

To give the Earth science data users the capability of creating data products with user-specified parameters, data centers are developing tools and technologies, such as Geospatial Interactive Online Visualization ANd aNalysis Infrastructure (GIOVANNI), Global Imagery Browse Services (GIBS), and the Land and Atmospheric Near Real-Time Capability for EOS (LANCE). Some of these products are archived only for a very short time period; e.g., LANCE products are archived for two weeks, and GIOVANNI and GIBS products are both created on-the-fly without being archived. This provides a challenge in assigning DOIs and developing data citations for the data products that are created through such tools. For these special service tools, a DOI would be assigned to the service, and the user would supply the parameters selected for the generation of the product. This would ensure the reproduction of the exact same data product that was used in the research. For example, the Worldview service is an interactive global image browse tool that interfaces with GIBS for visualization data. At the end of a data session, Worldview generates a URL containing the parameters used to generate the visualization product. That URL could be quoted in the data citation with the service tool DOI, making it a complete citation.

**Citation Example:**

Global Imagery Browse Service, Worldview Earthdata, 2015, 10.5067/services/vis_tool100, https://worldview.earthdata.nasa.gov/?p=geographic&l=MODIS_Terra_SurfaceReflectance_Bands143,MODIS_Aqua_CorrectedReflectance_TrueColor%28hidden%29,MODIS_Terra_CorrectedReflectance_TrueColor,MODIS_Terra_Brightness_Temp_Band31_Day,MODIS_Terra_Aerosol,Reference_Labels%28hidden%29,Reference_Features%28hidden%29,Coastlines&t=2015-11-20&v=-177.890625,-48.796875,47.109375,55.265625

## Citation of Special Application Products Generated by Services

Another use case would be the researcher requesting a DOI for a visualization product to be referenced in his science paper. For the DOI Landing page, ESDIS would require 1) a link to the service application pre-loaded with the parameters needed to regenerate the visualization, and 2) a brief description of each of those parameters. The DOI would resolve to the landing page where the visualization would be described, parameters listed, and data product reproduced with one-click. The DOI would be managed by the data center providing the service application.

## Recommended DOI Elements for Data Citations

The metadata attributes required for DOI registration are the same attributes required for data citations. This allows for the automatic generation of data citations using DOI metadata elements. Also, having well-formed data citations on the DOI Landing page enables science data users to easily copy and paste citations into their publications.

### ESDIS DOI Metadata Requirements

**DOI Registration**

- Identifier
- Creator Name
- Title
- Publisher
- Publication Year
- URL

**DOI Landing Page**

- Identifier
- Data Set Long Name
- Version No.
- Data Set Description
- Data Access
- Data Citation
- Associated Landing Pages

### Data Citation Requirements

**Identifier, Creator Name, Title, Publisher, Publication Year, URL**

## Acknowledgments

Lalit Wanchoo's contributions to this study were funded through the Science and Exploration Data Analysis (SESDA III) GSFC NASA Contract No: NNG12PL17C and Nathan James's contributions to this study were made as a part of his employment by NASA.