feature articles

Digital Object Identifiers for NASA's Earth Observing System

John Moses, NASA's Goddard Space Flight Center, Earth Science Data and Information System Project, john.f.moses@nasa.gov
Jeanne Behnke, NASA's Goddard Space Flight Center, Earth Science Data and Information System Project, jeanne.behnke-1@nasa.gov

This article will briefly describe the function and "fit and finish" of DOIs and their utility in the Earth Observing System (EOS) framework. Members of NASA's Earth Science Data and Information System (ESDIS) Project are working with several EOS science and instrument teams and principal investigators to develop methods for assigning DOIs to EOS products.

Introduction

The science community has long recognized the importance of citing data in published literature to encourage replication of experiments and verification of results. Authors who try to cite their data often find that publishers do not accept Internet addresses. These addresses are viewed as transient references, frequently changed by the data provider after the paper is published. *Digital Object Identifiers* (DOIs) were created to avoid this problem by providing a unique and persistent reference to online data. An indicator of the value of DOIs is the fact that they have emerged as the most accepted data identifier in the publishing community.

This article will briefly describe the function and "fit and finish" of DOIs and their utility in the Earth Observing System (EOS) framework. Members of NASA's Earth Science Data and Information System (ESDIS) Project are working with several EOS science and instrument teams and principal investigators to develop methods for assigning DOIs to EOS products. By assigning DOIs to EOS products, authors and publishers should find it easier and more compelling to cite EOS data products in their research and publications.

What Is a DOI?

A DOI is a unique alphanumeric character string used to identify any entity, i.e., a physical or digital object. The identification occurs via the DOI System[1]—a *resolution service* that works much the same way a uniform resource locator (URL) is resolved to a specific Web site by the Internet's *domain name system resolver.* The DOI System was developed by the International DOI Foundation (IDF) to support the publishing industry, and first placed into operation in 2000.

A DOI is permanent, such that when it is assigned and registered, it can always be used to locate the data object it refers to. In published citations, DOIs are used in place of Internet addresses (or URLs) along with other information arranged in accordance with the publisher's requirements. For example, a typical data citation would include the author (creator), publication year, (data product) title, publisher (distributer), DOI, and Internet access date.

The alphanumeric strings that comprise a DOI have two components, which together form the *DOI name*. The components are separated by the "/" character and take the form doi: [prefix]/[suffix]. The prefix consists of "10.[number]" where "10" is the assigned value for the DOIs in the Handle System®[2]. The prefix "number" value is assigned by a registration agency for use by the organization that wishes to register DOI names (e.g., the publisher; in our case, NASA). The "/" delimiter identifies the data item as decided by the registrant or agent. There are no significant restrictions on the suffix string; however, the guidance is to keep it simple and short for ease of use; more on this later.

In practice, a DOI name is typed or pasted into the text box of the DOI system's permanent online resolver service at *dx.doi.org*. The browser returns the specific Web page that provides access to the data object. The relationship between the DOI name and Internet location is maintained in the DOI system, usually by the publisher through services

---

[1] To learn more about the DOI® System, visit: *www.doi.org*.
[2] The Handle System® provides efficient, extensible, and secure resolution services for unique and persistent identifiers of digital objects. See *www.doi.org/doi_handbook/3_Resolution.html* for details of the Handle System.

offered by a DOI registration agency. The registration agency works under agreement with the International DOI Foundation and interfaces with the DOI system.

One of the key values of defining a permanent identifier is that if the Internet location of the object changes, the DOI is still valid. This means that an archive can change the URL of a data product without destroying previous references in already published literature. In addition, the publisher (or archive-distributor) of the object can change, and the new owner/publisher can still support the same DOI as the object's permanent locator. To make this work however, it is important that DOI names be universally acceptable among the research and publication community. For example, the DOI name should not contain information that may change in the future, such as references to the publisher or distributer of the data object.

Implementing DOIs for EOSDIS

ESDIS is developing an operations concept and scope for ESDIS and Distributed Active Archive Center (DAAC) roles and responsibilities. Specifically, ESDIS is working to:

- Prepare guidelines for DOI suffix profile, citation, and location information, drawing on the experience of others;

- make DOIs attractive to users, soliciting feedback from EOS science teams, DAAC user working groups, and the Earth science research community;

- assign DOI and maintain citation, and location information in the DOI system;

- add DOIs to DAAC product citation Web pages;

- add DOIs to DAAC product databases, the Global Change Master Directory (GCMD), and the EOS Clearinghouse (ECHO) through metadata updates;

- embed DOIs into existing product metadata at next reprocessing;

- add DOI metadata to the NASA Technical Report Server (NTRS) for searchable documentation; and

- set up metrics collection based on journal citation reports.

The ESDIS DOI-Assignment Infrastructure

To provide the best possible platform for creating EOS-related DOIs, ESDIS joined the California Digital Library (CDL)'s EZID—pronounced *easy-eye-dee*—service in February 2012 to facilitate creating and managing unique, long-term identifiers. The EZID Web service supports assigning an unlimited number of DOIs, and provides an application programming interface (API) for developing custom DOI management tools. CDL is a member of the DataCite Registration Agency[3], and as such requires compliance with the relatively simple DataCite metadata standard. The DataCite citation profile requires at minimum a URL, creator (organization or person), title, publisher, publication year, and resource type (e.g.,"Dataset"). ESDIS assigns DOI values and provides corresponding metadata in accordance with DataCite mandatory requirements using the EZID service. Upon subscribing to EZID, the unique prefix number *10.5067* was assigned for NASA ESDIS use.

Using this infrastructure, ESDIS is responsible for managing the "uniqueness" of DOI suffixes when creating new DOIs; they are also responsible for maintaining the DataCite-mandatory metadata. The value of DOIs assigned by ESDIS through EZID will always begin with the prefix 10.5067. This unique prefix assures that whatever alphanumeric string follows in the suffix portion can co-exist with all other DOIs in the

One of the key values of defining a permanent identifier is that if the Internet location of the object changes, the DOI is still valid. This means that an archive can change the URL of a data product without destroying previous references in already published literature.

feature articles

---

[3] DataCite is an international entity organized to facilitate access to research data. For more detail, see *datacite.org*.

ESDIS is exploring a strategy for determining DOI suffix names that will correspond to each EOS product. Since the development of the DOI System, DOIs have often been used to locate research papers in journal publications.

DOI system, as well as with all other identifiers in the Handle System®. *DOIs are case-insensitive,* so letters coded in upper or lower case will refer to the same DOI.

ESDIS is exploring a strategy for determining DOI suffix names that will correspond to each EOS product. Since the development of the DOI System, DOIs have often been used to locate research papers in journal publications. In these cases, the suffix has often been a combination of a few letters from the initials of the journal, followed by a sequence number incrementally assigned for each new article. In other cases, internal data center identifiers have been combined with prefixes to quickly establish a DOI structure for scientific data products. The German World Data Center for Climate has used this approach to assign DOIs to output from climate model simulations for reference by published literature. A legacy model data reference syntax is used for the suffix, based on codes for various components such as activity, product, institute, model, experiment, frequency, modeling realm, variable name, version number, and more—which tends to be quite long and complex. CDL EZID recommends random generation of an alphanumeric number. Systems have been devised that range from random to formal controlled vocabularies.

DOIs Are Valuable for EOS Product Users

Currently there are approximately 1850 EOS products archived and distributed by Earth Observing System Data and Information System (EOSDIS) DAACs. EOS products can be composed of thousands of files that contain observational data covering orbits and suborbital scenes. New versions of products can replace older versions. A DOI could be assigned to an aggregate of products (i.e, for an EOS mission or instrument), to all versions of a product within the instrument/mission, to each individual version, or to each digital file. To this end, ESDIS has not only been focusing their efforts on assigning DOIs to EOS products, but is also working to ensure that identifiers at multiple levels can co-exist and complement each other where appropriate, as individual projects will want to explore and develop applications.

In addition, for active on-orbit satellite missions, observations are continuously being added to each product. DOIs assigned to products from active missions cannot alone determine an accurate data citation because the data products will continue to grow as new observations are added. For an accurate and complete data citation—i.e., to uniquely determine all of the observations, or files, in a product—DOIs that were used in the publication citing the data must be accompanied with a product *access* date-time.

An overarching goal is for DOI names to be attractive for researchers to use in publication applications: Keeping their structure and content short and simple is, therefore, paramount. On the other hand, since DOIs are to be used with corresponding citation information, the suffix should have enough recognizable meaning so that a user will have confidence in knowing that they are correctly paired with the other citation information, without referring to any information that may not be permanent—e.g., publication information. The primary purpose is to add just enough meaning so that the reader knows the DOI value is likely for the cited data (e.g., an approximate match with the dataset title).

DOIs in the Product Metadata

Most users of EOS data download files to their local facilities for further analysis or for use in applications. Local software extracts attributes from the file metadata for visual examination or further processing. By embedding DOIs in the file metadata, users have access to the DOI name, which provides the potential for users to find documentation about the product in the future—long after it has left the contextual environment of the data provider. ESDIS is cognizant of this reality and has promoted and encouraged standard ways in which DOIs should be embedded in science product metadata.
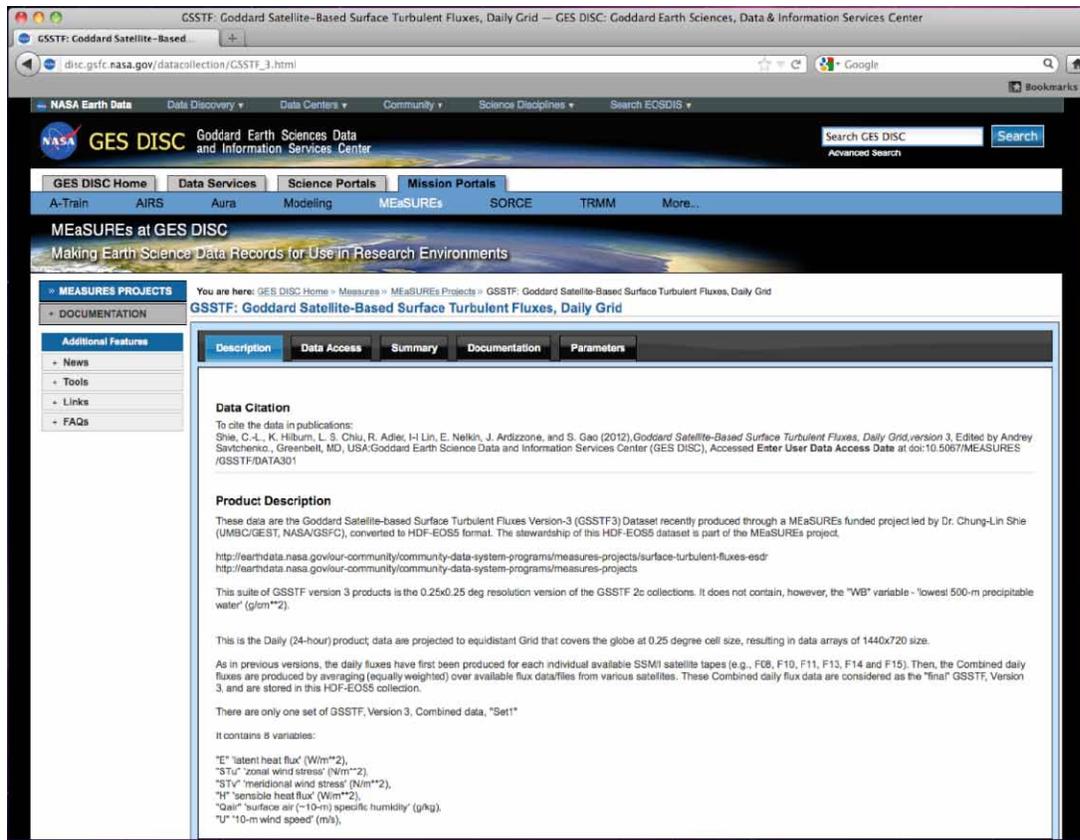
Landing at the Publisher

The next step is for the NASA DAACs to prepare the Internet location for access to the data product. The concept of a *landing page* refers to the Internet web page that the DOI service takes the DOI user to—i.e., it is the Internet address part of the DOI metadata. Since the DOI user has a specific product reference, the landing page accesses a product specific location. This requires a landing page (or section of a page) for each product.

The page should have enough information such that upon arrival, users know they have reached the source of the product and can gain access to it. A summary description of the product, as well as all of the other citation information about the product would satisfy the first requirement. Direct access to the product can be tricky because in many cases there are multiple versions of the product and each version consists of thousands of granules. As mentioned earlier, new versions of products will receive new DOIs. Of particular interest is how we handle the "old" DOIs that point to the previous product versions that are replaced with new versions. Since the old DOI is permanent, the landing page information will need to be updated or the corresponding URL in the DOI system will need to point to a new landing page. In either case, the DAACs will need to provide a landing page with explanation (i.e., disposition of the old version, availability of the new version) so that the old DOI will still work.

The DAACs have search tools—or advocate use of the ESDIS *Reverb*[4] search tool—as their primary means for searching through all products within geospatial and temporal constraints. If a search tool is used as the primary access method for a product-specific landing page, it should be prepopulated with a keyword (e.g., product shortname and version) that will uniquely identify the product within the search database. The search tool uses the keyword to find the product version and then allows the user to define additional spatial and temporal qualifications as desired to retrieve all or some of the product granules. **Figure 1** shows an example of the landing page at the God-

*The concept of a **landing page** refers to the Internet web page that the DOI service takes the DOI user to—i.e., it is the Internet address part of the DOI metadata.*

[4] For more information about Reverb, read *Reverb—The Next Generation Earth Science Discovery Tool* in the September-October 2011 issue of *The Earth Observer* [**Volume 23**, **Issue 5**, pp. 24-25].

feature articles

dard Earth Sciences Data and Information Services Center (GES DISC) Website for the Making Earth System data records for Use in Research for Earth Science (MEaSUREs) Program[5] product titled *Goddard Satellite-Based Surface Turbulent Fluxes, 0.25x0.25 deg, Daily Grid, V3, (GSSTF)*. Included on this page is descriptive information about how to format the data citation in publications, and a text-based description of the data product. Other tabs accessed from this page show how to access the data; details of the data product's short name, version format, coverage, and data resolution; an extensive set of references about the data, its provenance, and its utility; and a list of the parameters in the data file, with descriptions and units.

Pilot Studies

ESDIS is collaborating with algorithm implementers of the EOS Science Investigator-led Processing Systems (SIPS), and with MEaSUREs data providers to investigate and test various implementation options in *pilot study*[6] efforts. Pilot studies for products from the High Resolution Dynamics Limb Sounder (HIRDLS), the Atmospheric Infrared Sounder (AIRS), the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), the Moderate Resolution Imaging Spectroradiometer (MODIS), the Geoscience Laser Altimeter System (GLAS), and four MEaSUREs Projects[7] have already begun. In addition, two new missions—the Soil Moisture Active Passive (SMAP) and the second Ice, Clouds, and land Elevation Satellite (ICESat-2)—are also involved in the collaboration, planning early for implementation of DOIs in their products.

At this time ESDIS has worked through the entire process with two MEaSUREs projects and two DAACs. Two examples of the DOIs from each project are offered below. They can be examined using the DOI resources mentioned previously. The products can be retrieved and product metadata can be examined using standard tools to see the embedded DOI attribute names and DOI names.

doi: 10.5067/MEASURES/DMSP-F17/SSMIS/DATA301
doi: 10.5067/MEASURES/DMSP-F17/SSMIS/DATA304

doi: 10.5067/MEASURES/GSSTF/DATA301
doi: 10.5067/MEASURES/GSSTF/DATA302

Future Steps

The official archive for NASA scientific and technical documentation is the NASA Scientific and Technical Report Server (NTRS). NTRS—managed by the NASA Scientific and Technical Information (STI) Program—holds most of the definitive documents and research publications about EOS data products. ESDIS has discussed a concept and prototype for linking EOSDIS with the NTRS. In this concept, EOS would provide to NTRS product DOIs and lists of the related STI documentation. NTRS would add the DOI to their document's metadata database. A new NTRS search service would enable users to search for all the definitive documentation associated with a particular EOS product using the product DOI.

---

[5] MEASURES was funded under the Research Opportunities in Space and Earth Sciences (ROSES) 2006 Announcement of Opportunity.
[6] A *pilot study* is a small-scale preliminary study conducted in an attempt to improve the methods used in a larger study or full-scale research project.
[7] Including principal investigators **F. J. Wentz** [Remote Sensing Systems], **C. L. Shie** [Joint Center for Earth Systems Technology, University of Maryland Baltimore County, and NASA's Goddard Space Flight Center (GSFC)], **J. R. Herman**, [GSFC], and **P. K. Bhartia** [GSFC].

Shown in **Figure 2** is an overall view of how DOIs will flow through the EOSDIS architecture. As mentioned earlier, the product DOI names will be embedded in the EOS data products at the time of product generation, as depicted in the diagram; DOIs are also extracted from the products and added to the NASA DAAC's metadata repository. For missions that have ended and the resulting observational datasets finalized, DOIs will be assigned and entered separately into the DAAC's product metadata repository, since they are not available in the products.

Conclusion

ESDIS plans to continue developing DOI governance policies and distributed maintenance functions through integration with existing metadata infrastructures. By assigning DOIs, DataCite citation metadata, and data center landing pages for EOS products, ESDIS will make consistent citation information easily accessible for use by the research community. ■

By assigning DOIs, DataCite citation metadata, and data center landing pages for EOS products, ESDIS will make consistent citation information easily assessable for use by the research community.
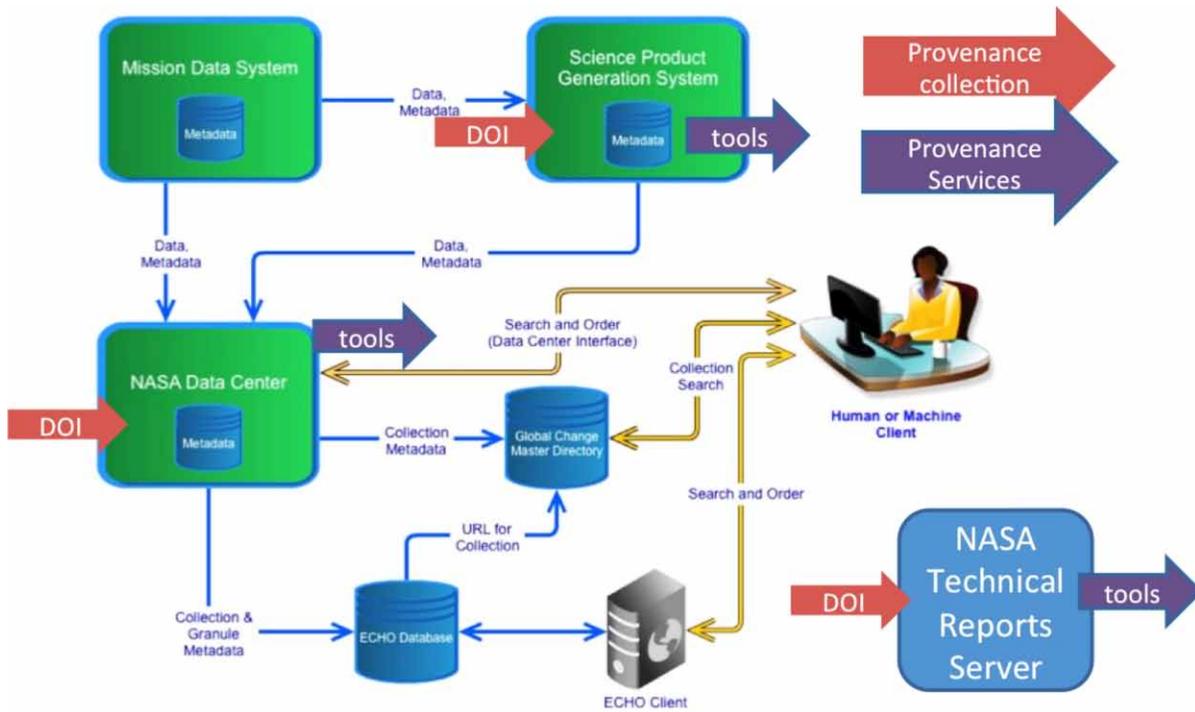

Figure 2. DOIs and provenance services in the EOSDIS architecture. **Image credit**: ESDIS Project