DOI Initiative for Heliophysics Data   October 2020

D. Aaron Roberts, NASA GSFC, aaron.roberts@nasa.gov
James Weygand, UCLA, jweygand@igpp.ucla.edu
Todd King, UCLA, tking@igpp.ucla.edu

The recognition and documentation of data use in publications has become more formalized, with many journals adopting increasingly specific requirements for citation and referencing of datasets. Most people agree that these changes are beneficial to all concerned, including readers who want to reproduce research and data providers who want an easy way to keep track of and get credit for data use. **To help NASA's Heliophysics (HP) division with adapting to this change, the SPASE group (http://spase-group.org) is offering a service that makes acquiring a "Digital Object Identifier" easy for data providers.** DOIs have long been routinely used for assigning unique identifiers to journal articles or books, but they are now the *de facto* standard for registering datasets.

One of the requirements of such registration is to have a "landing page" that is a web page associated with the DOI that provides access and other information; this is the equivalent of bringing up a direct reference to a paper. The SPASE product registrations provide all the required information (plus much more), and we have developed a standard format for landing page display.

An example of a data product for which we already have "minted" a DOI would be cited as Russell (1987), with the bibliography entry (the specific format for the data reference will depend on the journal):

Russell, Christopher T. (1987). ISEE 2 Magnetometer 1-min Data at CDAWeb. *University of California at Los Angeles* **and** *NASA's Space Physics Data Facility***. https://doi.org/10.21978/P8T923. Accessed from CDAWeb, 15 March, 2018.**

Entering the DOI URL in a web browser will bring up the product description. (Just click above.) Note that we have already compiled product descriptions for nearly all current and most past NASA missions, so **all we need in order to mint DOIs for datasets is to get agreement with the missions on the publication information, namely, the** *title, creator(s), publisher(s), and publication year***.** (The "creator" is the "author", i.e, the person or organization that made the dataset.)

For a typical resource with a SPASE record, we would retrieve the *title* from the ResourceName as it is in the description. The title could be different from this name, or the

---

Need to find out what Products you have registered? Need to find the DOI to refer to data in a paper that is registered?
The Heliophysics Data Portal (HDP) at https://heliophysicsdata.gsfc.nasa.gov is a "card catalogue" that provides many search mechanisms for finding products. It also provides direct access to plots and data downloads for many products. *To find the example here*, choose "observatory" in the list at the left of the HDP, pick ISEE 2, and click the "Add" button. Entering "magnetic" in the text box and adding it will make the request more precise. The descriptions are always subject to improvement; let us know!

name can also be changed, but consistency is useful. Note that entering any unique piece of the SPASE ID in the text box of the HDP (https://heliophysicsdata.gsfc.nasa.gov) will find the resource, and clicking on its name will deliver the landing page.

The *creator(s)* would typically be an instrument PI and perhaps other major contributors, but it is the decision of the product provider who should be included in the list. Teams or organizations (e.g., "the WIND SWE team") could be used as well, although they can also be taken as implicit. Our initial assumption is that all products from a given instrument would have the same creator(s), but this is not necessarily true. Creators are the people most directly responsible for the creation of the data product; they should agree with the list and the other aspects of the DOI information.

Our plan is to work with the Project Scientists of each mission to move this forward. We hope that the idea of instituting DOIs and our approach to doing it will require little effort from the missions, and that this will be a useful exercise for all involved. *Contact James Weygand or Aaron Roberts for further information or to obtain DOIs*.

The *publisher* is the source of the data. Often data are served from a PI site, and more than one site can be noted if appropriate. The original data site may be included.   If the data are in a major archive (e.g., NASA's Planetary Data System, Space Physics Data Facility, or Solar Data Analysis Center), that archive should be included as the publisher or co-publisher of the data; this will help to assure longevity.

The *publication year* could be inferred from the ReleaseDate of the SPASE descriptor, but this would be much later than the first public release of the data in most cases. For active datasets, the publication date is, in effect, continually changing, so the *date of the most recent data version* could be used. **Note that "dynamic" datasets that are accumulating or undergoing modest upgrades do not need to be registered as new products. In this instance, however, citations/references *must always* include the time of product access.** Changes that are likely to affect scientific results or that change the set of variables in the files should trigger a new DOI. Older, static, datasets can use the date that the product was released in final form as the publication date.

Creators, publishers, and publication year will be in the SPASE entry PublicationInfo. However, nearly all of the resources descriptions currently do not include this information. As we collect the necessary information for minting a DOI we will update the resource description as well, and the latter will then *provide the information needed by users of the DOIs for citations*.

Minting a DOI also requires a "location URL" and "type of resource," but these are taken care of by the SPASE team. The location URL is the SPASE metadata landing page at https://hpde.io, which is what the doi link actually references. The type of resource is "Dataset" for NumericalData and "Collection" for DisplayData (there's an enumerated list of type choices). In the current repository of SPASE metadata there are 2,355 NumericalData and 144 DisplayData resource descriptions. Our goal is to assign a DOI to each described resource.

More information about DOIs can be found at the DataCite pages, https://datacite.org, and the current DataCite Metadata Schema is here:
https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf