

Evolution of NASA's Earth Science Digital Object Identifier Registration System

Lalit Wanchoo¹ (lalit.wanchoo@nasa.gov) and Nathan James² (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

Introduction

NASA's Earth Science Data and Information System (ESDIS) Project has implemented a fully automated system for assigning Digital Object Identifiers (DOIs) to Earth Science data products being managed by its network of 12 distributed active archive centers (DAACs). A key factor in the successful evolution of the DOI registration system over last 7 years has been the incorporation of community input from three focus groups under the NASA's Earth Science Data System Working Group (ESDSWG). These groups were largely composed of DOI submitters and data curators from the 12 data centers serving the user communities of various science disciplines. The suggestions from these groups were formulated into recommendations for ESDIS consideration and implementation.

The ESDIS DOI registration system has evolved to be fully functional with over 5,000 publicly accessible DOIs and over 200 DOIs being held in reserve status until the information required for registration is obtained. The goal is to assign DOIs to the entire 8000+ data collections under ESDIS management via its network of discipline-oriented data centers.

DOIs make it easier for researchers to discover and use earth science data and they enable users to provide valid citations for the data they use in research. Also for the researcher wishing to reproduce the results presented in science publications, the DOI can be used to locate the exact data or data products being cited.

Objective

ESDIS DOIs also provide data "Provenance" which is information about the creation and history of the data in question. This would include when the data was collected, which instrument was used to collect the data, and the version of the product at the time the DOI was assigned. Also, over the past few years, requests for DOIs have increased significantly as DAACs assign DOIs to both legacy data from earlier missions and new data products from in-orbit missions. Therefore, this study will evaluate the evolution of DOI registration and its usage over those years comparing data products as they are organized by mission, science discipline and data product level.

Discipline-based DAACs

The ESDIS Project since 1994 has been responsible for managing the 12 EODIS DAACs and multiple Science Investigator-led Processing Systems (SIPS). SIPS are responsible for the generation, processing and transfer of science data products to the appropriate DAAC(s) for archiving and distribution. DAACs are responsible for performing all the various data management functions to include product ingest, archive, distribution, and documentation. Each of these facilities has a different discipline focus as shown in Figure 1 with details in Table 1. According to ESDIS policy, DOIs are assigned and managed by the DAAC responsible for the archiving and distribution of the science product requiring a DOI.

ESDIS DOI WIKI URL: <https://wiki.earthdata.nasa.gov/display/DOIsforEODIS>

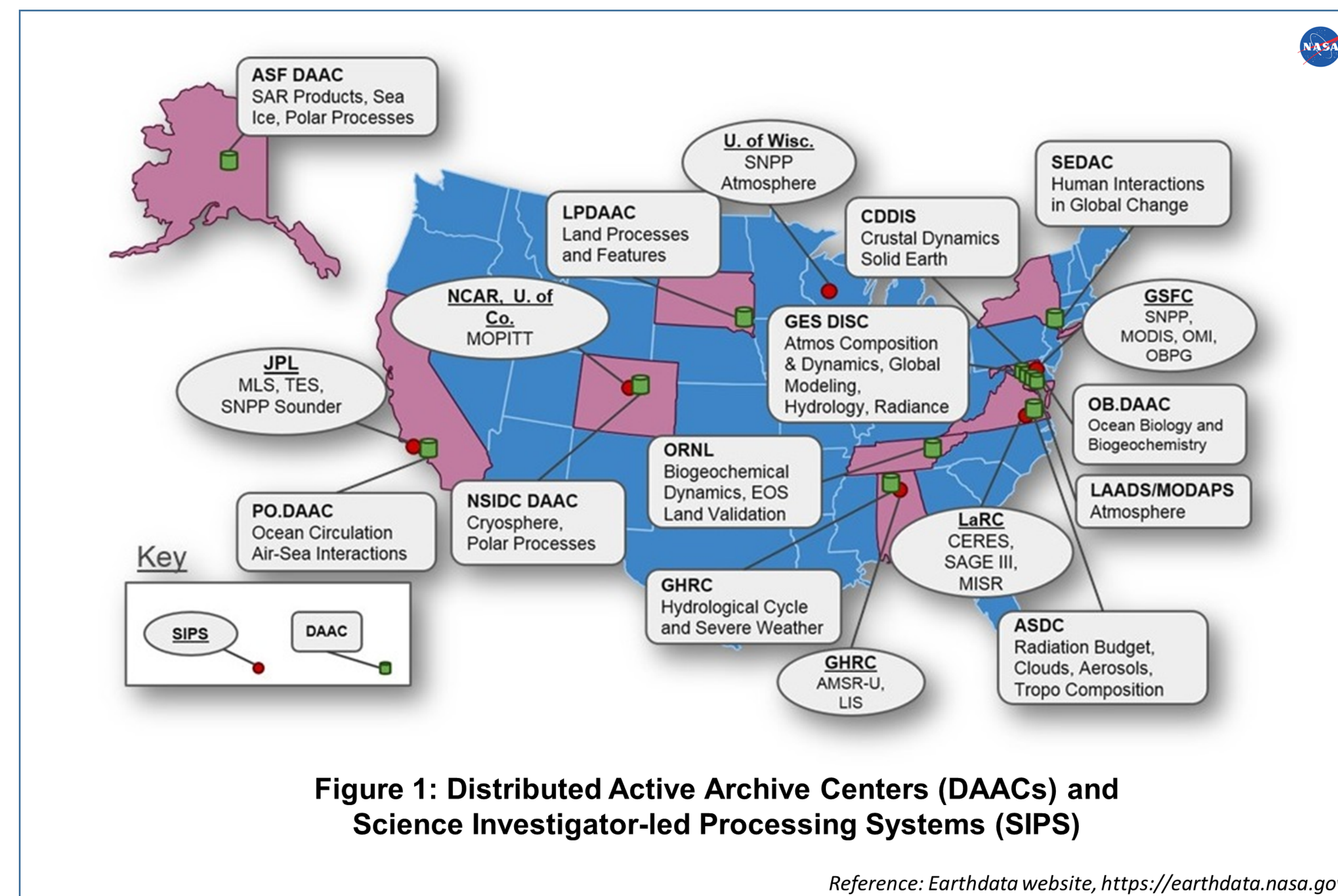
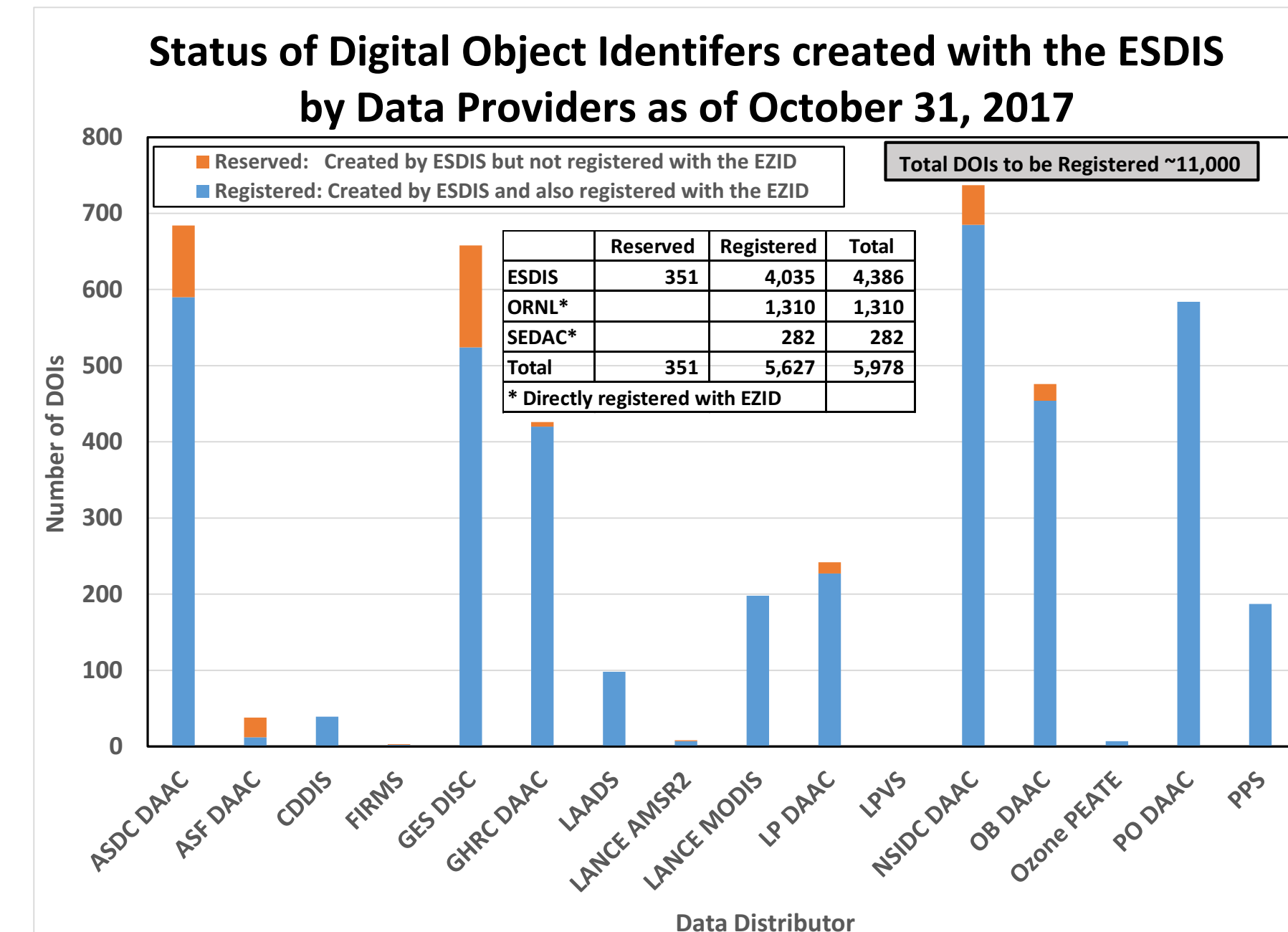


Table 1: Name and Science Discipline of Various Data Distributors

Data Distributor	Data Distributor Name	Science Discipline
ASDC DAAC	Langley Atmospheric Science Data Center DAAC	Radiation Budget, Clouds, Aerosol, Tropospheric Chemistry
ASF DAAC	Alaska Satellite Facility DAAC	Synthetic Aperture Radar Products, Sea Ice, Polar Processes, Geophysics
CDDIS	Crustal Dynamics Data Information System	Space Geodesy, Solid Earth
FIRMS ¹	Land Atmosphere Near real-time Capability for EOS Fire Information for Resource Management System	Near real-time active fire data
GES DISC	Goddard Earth Sciences Data and Information Services Center	Global Precipitation, Solar Irradiance, Atmospheric Composition and Dynamics, Global Modeling
GHRC DAAC	Global Hydrology Resource Center DAAC	Hydrologic Cycle, Severe Weather Interactions, Lightning, Atmospheric Convection
LAADS	Level 1 and Atmosphere Archive and Distribution System	MODIS Level-1 and Atmospheric Products
LANCE AMSR2 ²	Land Atmosphere Near-real-time Capability for EOS (LANCE) AMSR2 at the GHRC DAAC	Near real-time global precipitation and ocean parameters excluding sea surface temperature.
LANCE MODIS ²	LANCE MODIS at the MODAPS	Near real-time Land and Atmospheric data
LPDAAC	Land Processes DAAC	Surface Reflectance, Land Cover, Vegetation Indices
LPVS ²	Land Product Validation Subgroup (Working Group on Calibration and Validation Committee on Earth Observation Satellites)	Land Product Validation
NSIDC DAAC	National Snow and Ice Data Center DAAC	Snow and Ice, Cryosphere, Climate Interactions, Sea Ice
OB DAAC	Ocean Biology DAAC	Ocean Biology, Sea Surface Temperature
Ozone PEATE ²	Ozone Product Evaluation and Test Element	Ozone Measurements
ORNL	Oak Ridge National Laboratory DAAC	Biogeochemical Dynamics Ecological Data, Environmental Processes
PO DAAC	Physical Oceanography DAAC	Gravity, Sea Surface Temperature, Ocean Winds, Topography, Circulation and Currents
PPS ¹	Precipitation Processing System	Global Precipitation Measurement
SEDAC	Socioeconomic Data and Application Data Center	Human Interactions, Land Use, Environmental Sustainability, Geospatial Data

¹Non-DAAC organization associated with a DAAC.

²Non-DAAC organization.



DOIs are also assigned to :

- Near Real-time (NRT) data products generated under the LANCE (FIRMS, LANCE AMSR2, and LANCE MODIS)
- the documents that are related to the data products (LPVS).

DOI Assignments by Product Level

EODIS data products are processed at various levels ranging from Level 0 to Level 4. Level 0 products are raw data at full instrument resolution. At higher levels, the data are converted into more useful parameters and formats. Product level descriptions are given in Table 2.

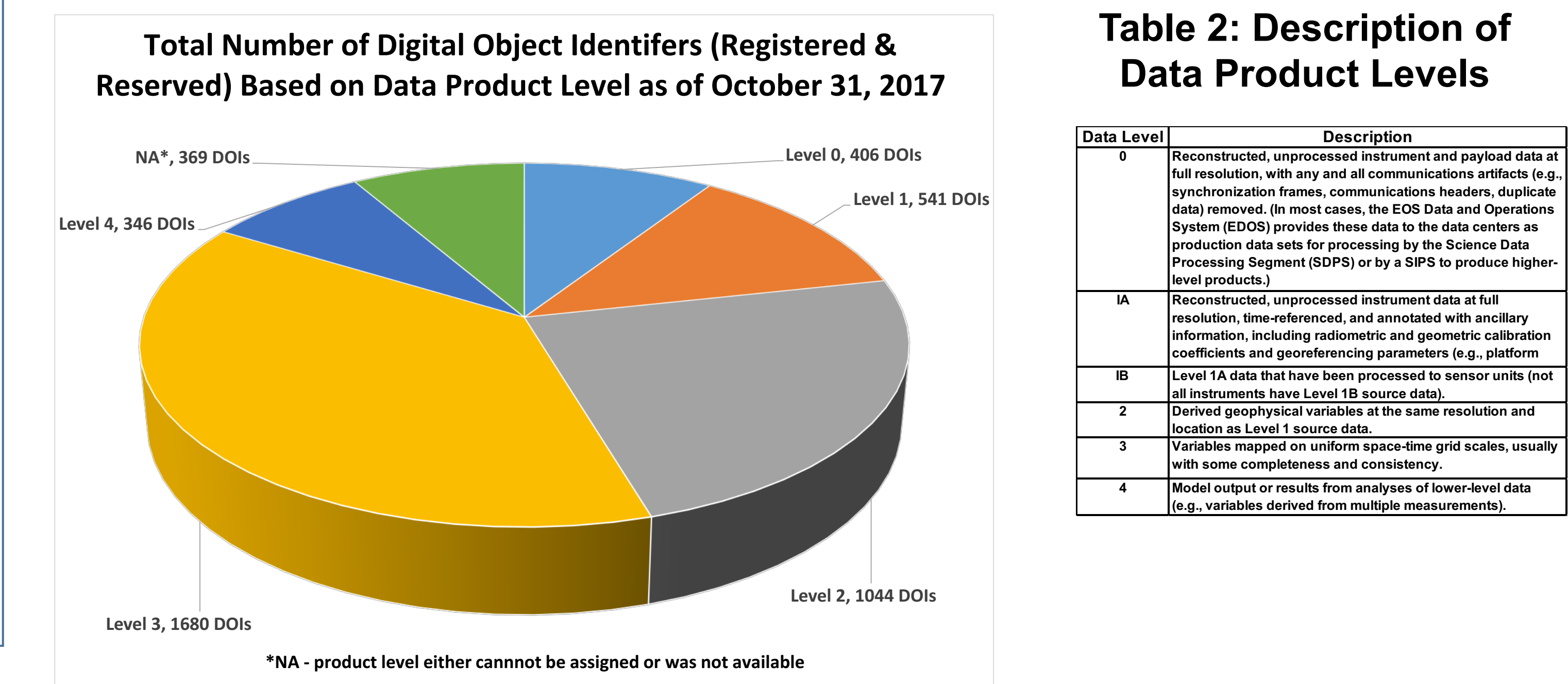
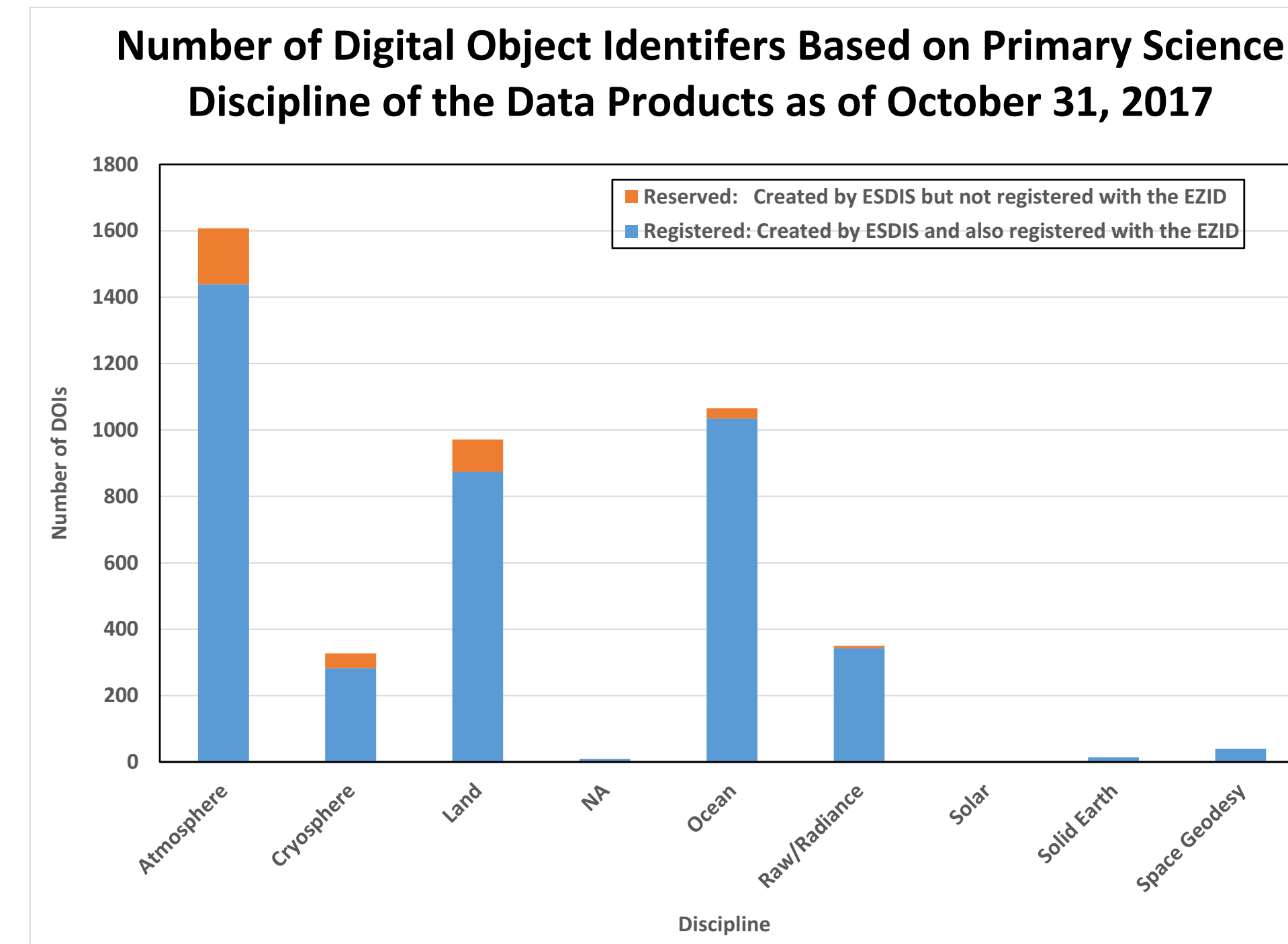


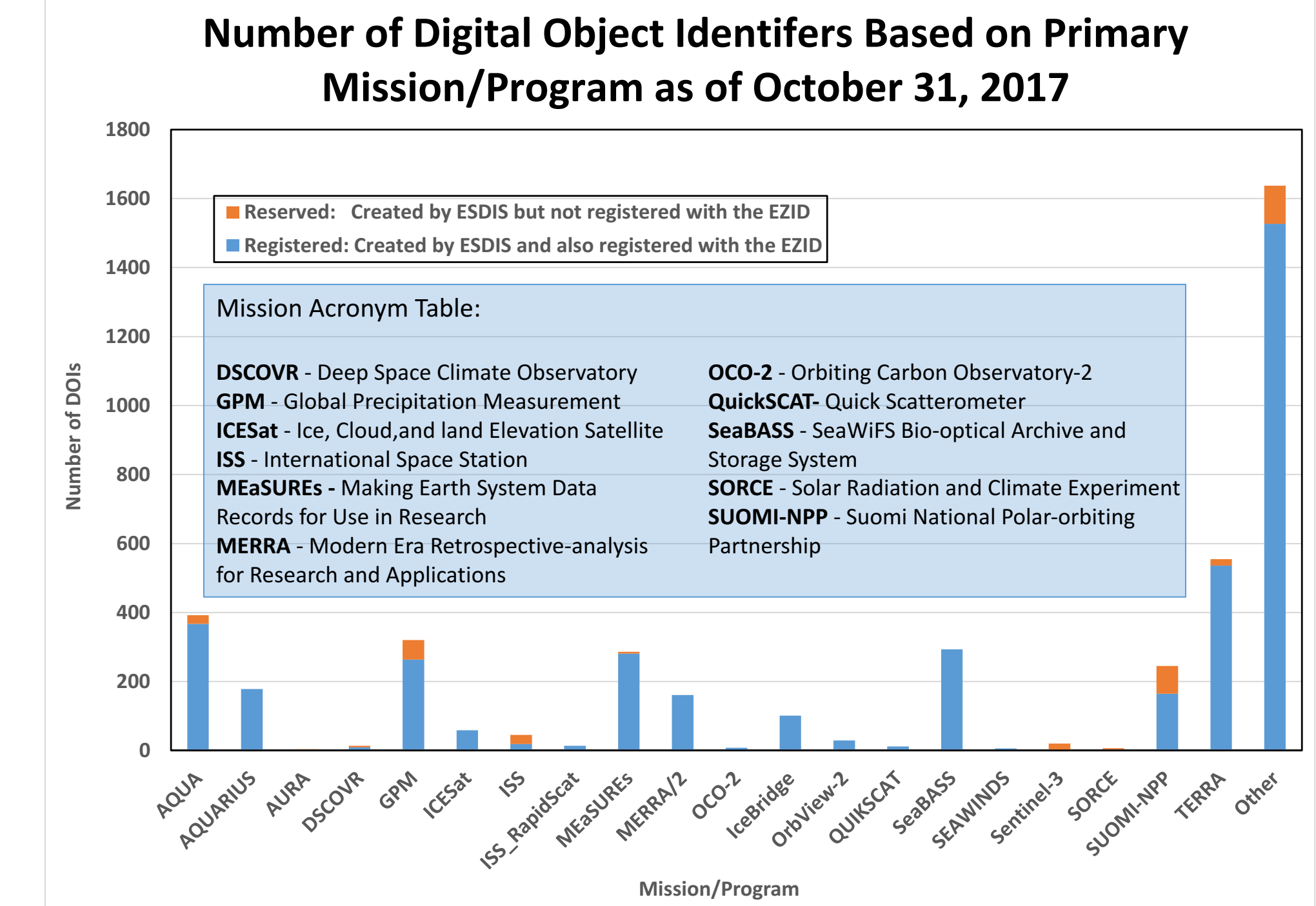
Table 2: Description of Data Product Levels

Data Level	Description
0	Reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e.g., synchronization frames, communications headers, duplicate data) removed. (In most cases, the EOS Data and Operations System (EOS) provides these data to the data centers as production data sets for processing by the Science Data Processing Segment (SDPS) or by a SIPS to produce higher-level products.)
1A	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (e.g., platform location).
1B	Level 1A data that have been processed to sensor units (not all instruments have Level 1B source data).
2	Derived geophysical variables at the same resolution and location as Level 1 source data.
3	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.
4	Model output or results from analyses of lower-level data (e.g., variables derived from multiple measurements).

Science Discipline Assigned DOIs



DOI Assignments by Mission/Program

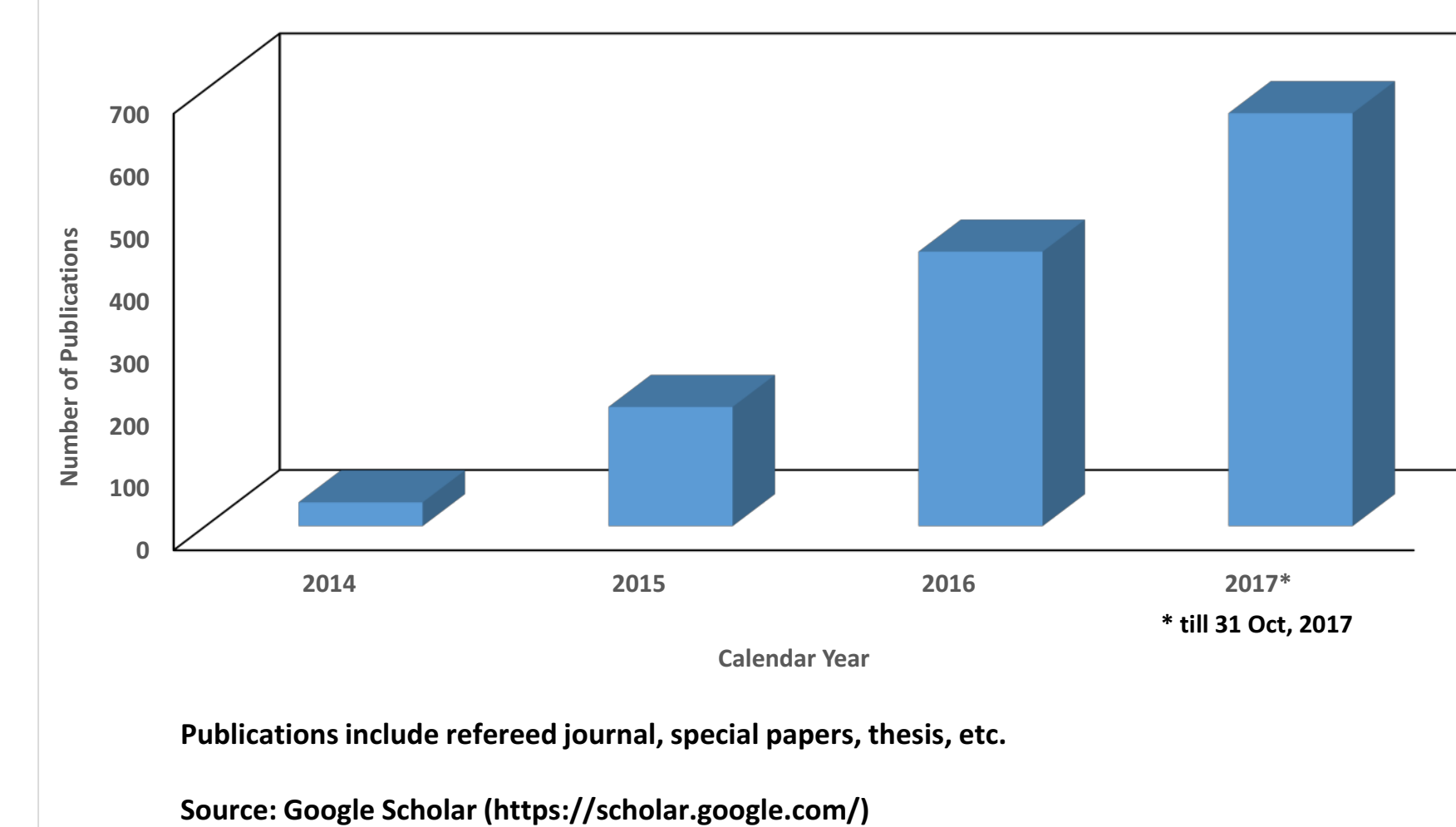


Conclusion

DOI names mapping to the science discipline, mission/program and product level were done based on the product identification provided by the DAACs during the submission of the DOI metadata. In addition, mapping was complemented by matching against the ESDIS Metrics System (EMS) metadata.

Over the last 7 years, requests for DOIs have increased significantly and the above charts show that DAACs assign DOIs to both legacy data from earlier missions, currently in-orbit missions, and all product levels with the majority of DOIs assigned to level 2 and level 3 products. Also highlighted is the significant increase of publications citing ESDIS DOIs over the past few years.

Yearly Number of Publications Citing ESDIS DOIs starting with 10.5067



Acknowledgments

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