AGU Paper Number: IN12B -12

NASA EOSDIS Data Usage Metrics- Insight and Assessment

Durga N. Kafle[1,2] (durga.n.kafle@nasa.gov), Lalit Wanchoo[1,2] (lalit.wanchoo@nasa.gov), Young-In Won[1,2] (youngin.won@nasa.gov), Jeanne Behnke[2] (Jeanne.behnke@nasa.gov)

[1] ADNET Systems, Inc., 7515 Mission Drive, Suite A100, Lanham, MD 20706, [2] NASA GSFC, Code 423, Greenbelt, MD 20771

Introduction

NASA's Earth Science Data and Information System (ESDIS) Project collects Earth science data usage metrics on a routine basis through the ESDIS Metrics System (EMS), which includes metrics on data archive, distribution of data products, users, data volumes, and number of files from 12 Distributed Active Archive Centers (DAACs) (Table 1), with each DAAC usually focusing on specific disciplines within the Earth Science community. These metrics are key parameters in evaluating system-level performance of any of the 12 DAACs encompassed by NASA's Earth Observing System (EOS) Data and Information System (EOSDIS). EOSDIS data usage metrics illustrate the benefits of making NASA data openly available to the public and show a rapid growth in data distribution to a worldwide user community. In fact, each year since 2014 the EOSDIS has distributed over one billion data files of products from EOS satellite, airborne, and *in situ* observations. An evaluation of the data product archive process and the assessment of the long-term trends of data usage based on various product metadata can provide insights into data accessibility, usability, and manageability, the key factors for any DAACs and the management.

Purpose

The purposes of this study are 1) to analyze long-term archive of data products by various DAACs, 2) to perform a comprehensive analysis of EOSDIS data usage metrics over the last 10 years, and 3) to characterize the distribution of the data products by reviewing key metrics parameters.

Table 1: EOSDIS Distributed Active Archive Centers (DAACs)

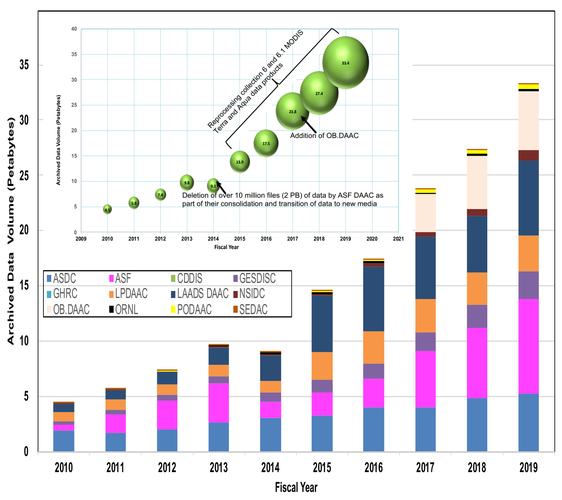
ASDC	Atmospheric Science Data Center
ASF DAAC	Alaska Satellite Facility DAAC
CDDIS	Crustal Dynamics Data Information System
NSIDC DAAC	National Snow and Ice Data Center DAAC
GHRC DAAC	Global Hydrology Resource Center DAAC
ORNL DAAC	Oak Ridge National Laboratory DAAC
GES DISC	Goddard Earth Sciences Data and Information Services Center
OB.DAAC	Ocean Biology DAAC
LP DAAC	Land Processes DAAC
PO DAAC	Physical Oceanography DAAC
LAADS DAAC	Level 1 Atmosphere Archive and Distribution System DAAC
SEDAC	Socioeconomic Data and Applications Data Center

Quality metadata is essential for effective metrics that enhances data accessibility, usability and manageability. Moreover, accurate metadata elements (product level, study area, service or protocol, country etc.) are critical to make easier for users to discover the data and hence thereby create higher data demands.

Accessibility

NASA's EOSDIS began archiving EOS data in FY 2000, and since then has stored a current total of over 33 Pbytes, which does not include near real-time (NRT) data products and the products that have been deleted over time. Figure 1 shows a steady growth of data archive over last 20 years. However, out of that 20 years total, over 86% (29 Pbytes) of the data have been archived in last 10 years (FY10-FY19) only. These data have been accessible to the users.

Figure 1. EOSDIS Total Data Archive Volume as of 30 September 2019



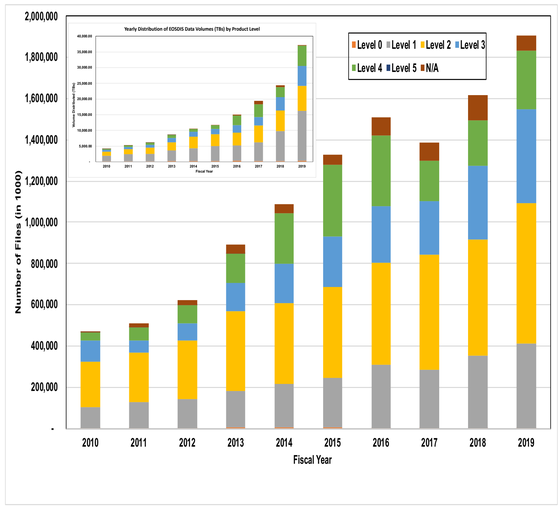
Usability

Distribution metrics provides an insight into data usage and user characterization, and as such becomes one of the key performance indicators for the DAACs. Since FY 2010, a total of 139.44 Pbytes of volume and 11.33 billions of data files have been distributed to over 11 millions distinct public users. Figures 2-11 show metrics on distribution of users, data volume, and number of files in various ways.

Manageability

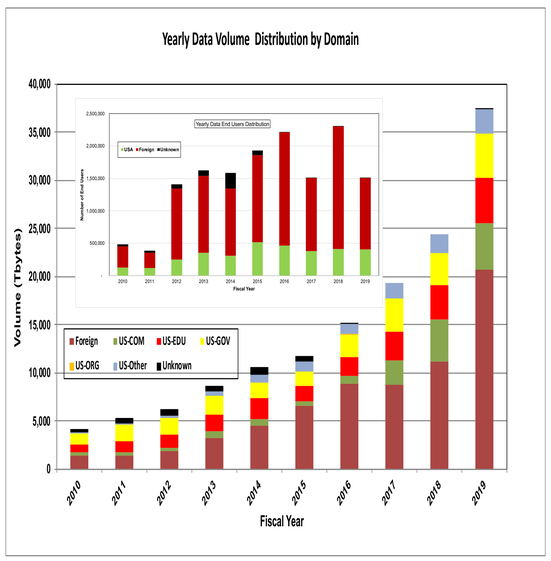
Efficiency in data accessibility and usability depends on the data management mechanisms such as metadata, tools, services and protocols. Improvement of metadata, a better identification of product of interest makes easier for users to find the data. Metrics based on study area improves user characterization. Similarly, the metrics information based on service or protocol demonstrate how the usage of each protocol has been evolved, which may help DAACs as well as NASA management team in deciding where to focus their efforts to better serve users' needs. Figures below illustrate the evolution of such distribution metrics over the last 10 years.

Figure 2. EOSDIS Yearly Data Distribution by Product Level



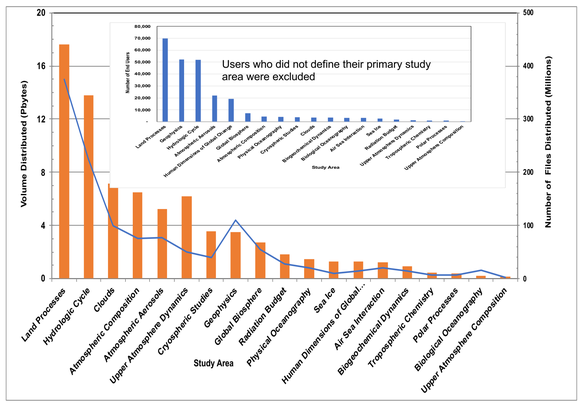
As shown in Figure 2, the most downloaded data products are shifting from levels 0 and 1 to level 2 that means more usage of derived geophysical quantities than lower level radiance. The over all demand for higher level data products has been increasing steadily year after year, which may indicate improved data quality with time.

Figure 3. Yearly EOSDIS Data Download by Internet Domains



The distribution metrics (distinct users, distributed volume and files) by domain (Figures 3) shows that the demand for NASA EOSDIS products in foreign countries has increased over last 10 years.

Figure 4. EOSDIS Data Distribution by Study Area from FY 2010 to FY 2019



Study area was added to the ESDIS user registration system (Earthdata Login), thereby improving user characterization. Figure 4 shows the distribution metrics based on study area for last 10 years. However, there are significant number of users who did not define their study area. For such users "science discipline" may be substituted as shown in Figure 5.

Figure 5. EOSDIS Data Distribution by Science Discipline (FY 2010 to FY 2019)

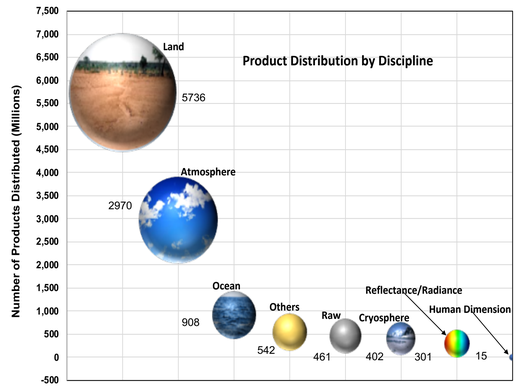


Figure 6. Yearly Number of Users downloading EOSDIS Data Using Various Data Access Methods

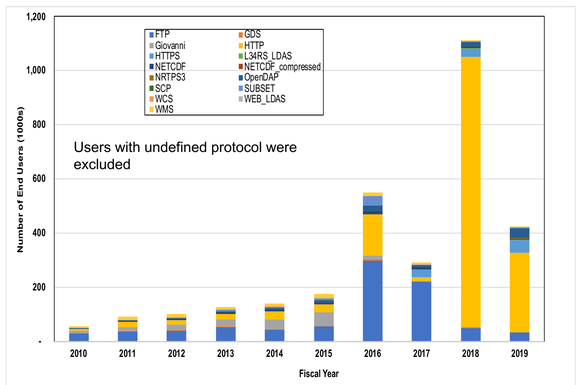
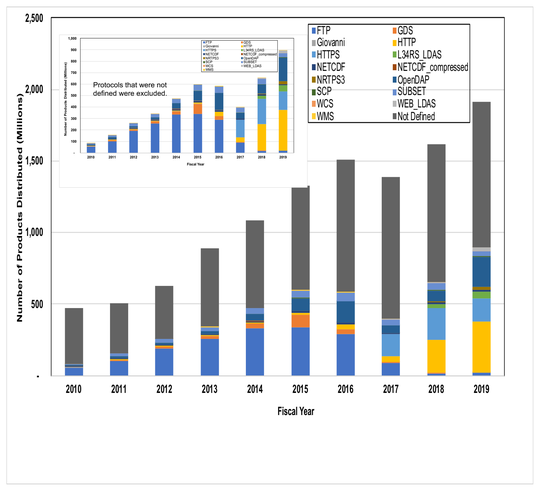
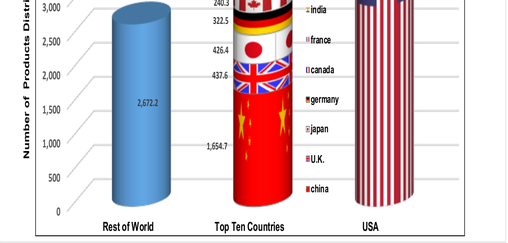


Figure 7. Yearly EOSDIS Data Distribution to End Users Using Various Data Access Methods



Data distribution metrics by protocol is shown in Figures 6-7. FTP, HTTP, HTTPS, OpenDAP and some subset services became more popular year after year. This should be of great help for data providers in deciding where to focus their efforts in maintaining various services and/or deciding which service to terminate.

Figure 8. Number of EOSDIS Data Products Downloaded by Users from Various Countries (FY2010 to FY 2019)



The data and the user community is globally dispersed and the demand from foreign users has increased steadily over time, indicating, in part that the access to the data has been made easier (Figure 8).

Summary

- Total archive size and the distribution of EOSDIS data has been growing steadily year after year.
- Increase in data usage can be attributed to several factors such as better identification of products with improved data quality, data centers provided more special services and protocols via web-based applications making data access easier, and implementation of NASA's Earth Science data open data policy.
- Metrics reports are still limited with our existing systems due to lack of availability of additional metadata information and data accessibility.
- ESDIS project and its constituent DAACs are moving EOSDIS data to the cloud, which will provide greater efficiency for storing and disseminating these data. Users will have improved access to the data sets for a broader range of research.
- There will be interface between EMS and Common Metadata Repository (CMR), which will allow us to get value added metadata such as information about latitude, longitude etc. With that information we will be able to get the metrics over the certain area of investigation we choose. Data usage metrics will be changed, accordingly.

Acknowledgments

This study was funded through the science and Exploration Data Analysis (SESDA IV) GSFC NASA contract No: 80GSFC17C0003