

NASA EOSDIS 20 Years of Data Usage and User Assessment in Support of Open Science Initiative

Lalit Wanchoo^{1,2} (lalit.wanchoo@nasa.gov), Francis Lindsay² (francis.lindsay-1@nasa.gov), Dany Turcios^{1,2} (dany.x.turcios@nasa.gov), Yan Hao^{1,2} (yan.hao@nasa.gov), Heather Weir^{1,2} (heather.weir@nasa.gov)

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

Introduction

NASA EOS Data and Information System (EOSDIS) will be celebrating a milestone in 2024 of 30 years of distributing data to world-wide users free and with open access. Since the launch of NASA's Terra satellite in 1999, more than 17,000 data products have been archived and distributed by NASA-funded Distributed Active Archive Centers (DAACs) that are part of NASA's EOSDIS. As of September 30, 2023, more than 102 PB of data archived by EOSDIS have been made available to public users and during 2023 over 112 PB have been distributed to public users worldwide. Over the past 23 years, data have shown significant increase in the distribution of various data products. EOSDIS has collected this data over these years from 106s of missions and research programs that include missions that are currently in-orbit. In addition there are new missions planned that will provide additional data to EOSDIS in coming years. Figure 1 shows the list of the Earth science missions that are in-orbit and future planned missions that have been provided or will provide data to EOSDIS. Such a large amount of data made available through multiple Distributed Active Archive Centers (DAACs) as listed in Table 1 to worldwide users unrestricted under free and NASA's Open Data Policy, have been an important step towards open science initiative.

The purposes of this study are 1) to perform a comprehensive investigation of the archive and distribution patterns of EOSDIS data products for last 23 years, 2) to identify and characterize the global user community for those data, 3) analyze the increased demand for data products, 4) evaluate distribution of higher level products because those are the ones most frequently used in the studies of natural disasters by public users (those data requestors not involved directly in the production or validation of the data products) and contribute globally to the advance scientific understanding of the Earth-Atmosphere Systems.

Approach

This study examines yearly data archived and amount of data distributed to public users for the key missions, that are AQUA, AURA, GPM, ICESAT-2, International Space Station instruments, JPSS, S-NPP, SENTINEL, SMAP, and TERRA. All other missions included non-operational and legacy missions that are combined together as OTHER category. This data is available from NASA's EOSDIS Project's Metrics System (EMS), which collects archive, distribution, and user information from EOSDIS data provider on daily basis. EMS information is stored in a relational database from which quantitative metrics of NASA Earth Science data were retrieved and analyzed for this study. In addition, to understand the data usage, user characterization was performed for the user data download and their affiliations.

The product levels, as described in Table 2, indicate various levels of processing of the data products that are made available to users. Higher level data products, such as levels 2, 3, and 4, have geophysical parameters with various formats. Level "Other" is for those products that either do not have a product level, such as ancillary data or metadata was not available. This study will also examine the distribution of various data products based upon their product level. As such, distribution is an indicator of the data usage by the science community.

Archive Data

There are 12 NASA Distributed Active Archive Centers (DAACs) located throughout the United States as listed in Table 1. More than 50 percent of the data archived comes from top 10 missions as shown in Figure 2 which Aqua and Terra, the two flagship missions contributing majority of the data archive. Sentinel 1A/1B, European missions, whose data is also made available through EOSDIS DAACs has contributed more than 24 PBs of data in the archive at one of the DAACs.

In comparing the 10 targeted Earth observing missions to the legacy missions, the number of data files archived in the targeted missions are approximately 72% of the total number of files in the archive. The data volume archived from the Sentinel missions is approximately 24% of the entire archive volume, even though the number of files in the archive is 1.3%. Aqua and Terra consist of more than 50% of the entire data volume archived, including more than 100 missions.

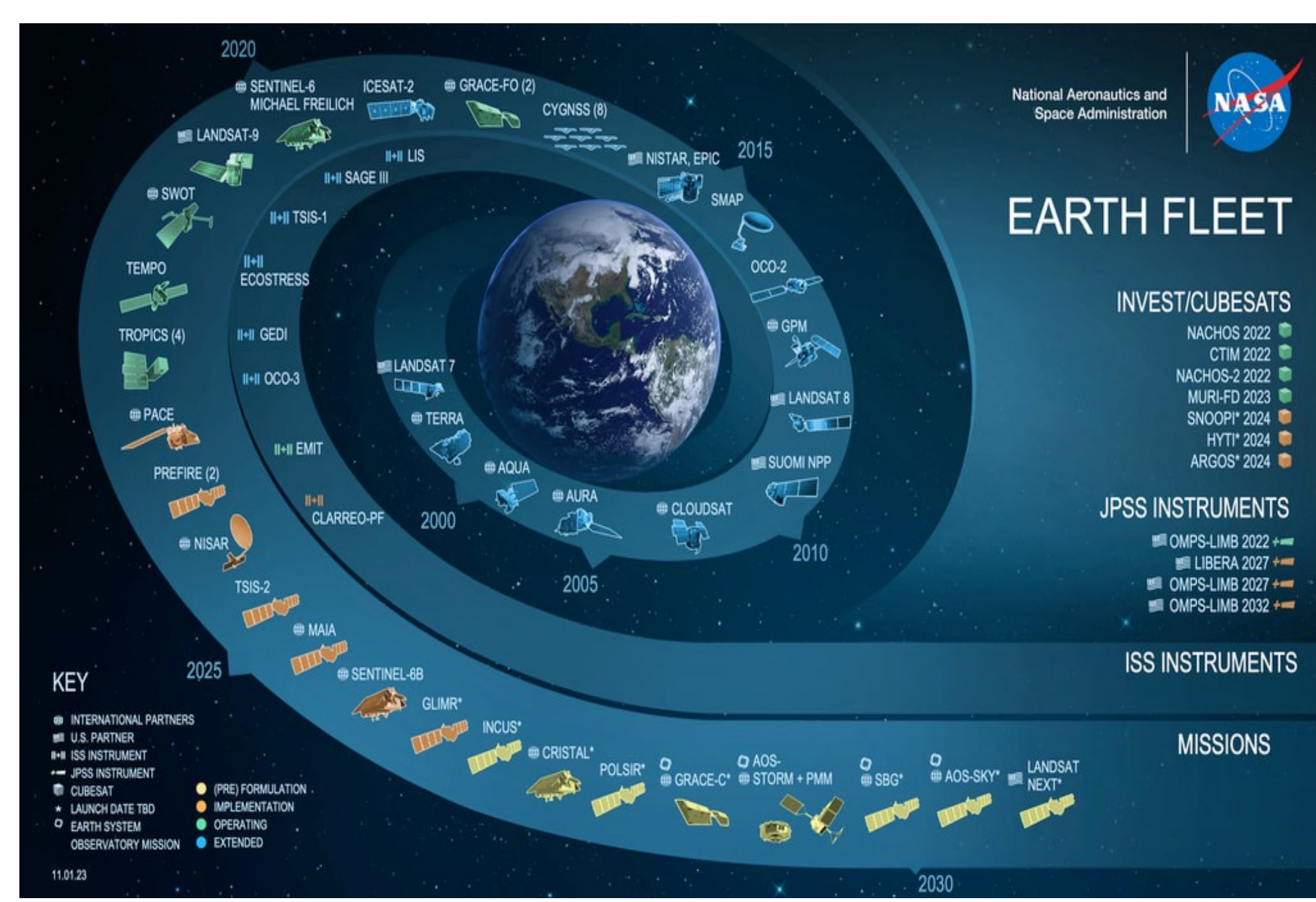


Figure 1: Current Earth Fleet Chart

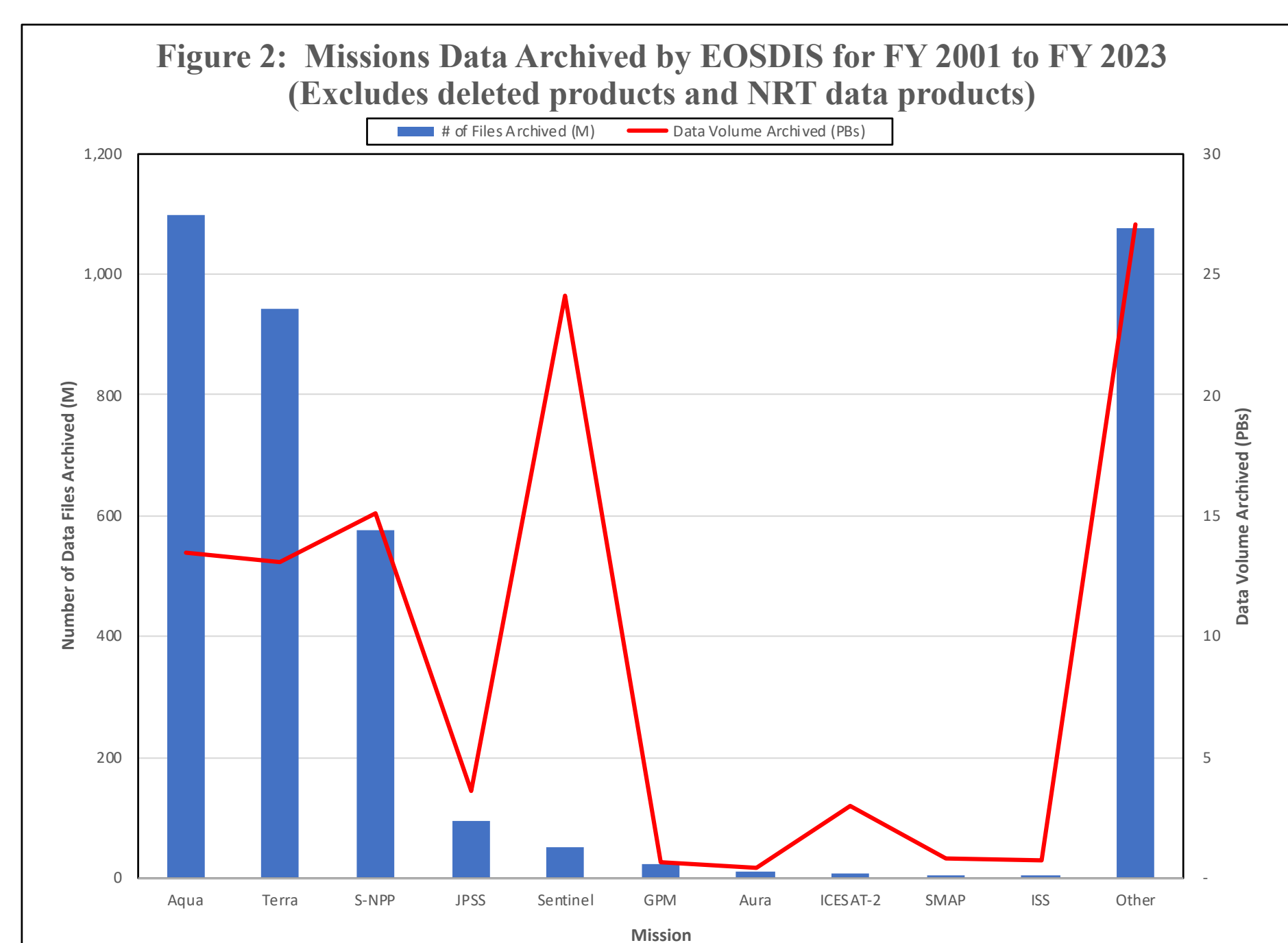


Figure 2: Missions Data Archived by EOSDIS for FY 2001 to FY 2023 (Excludes deleted products and NRT data products)

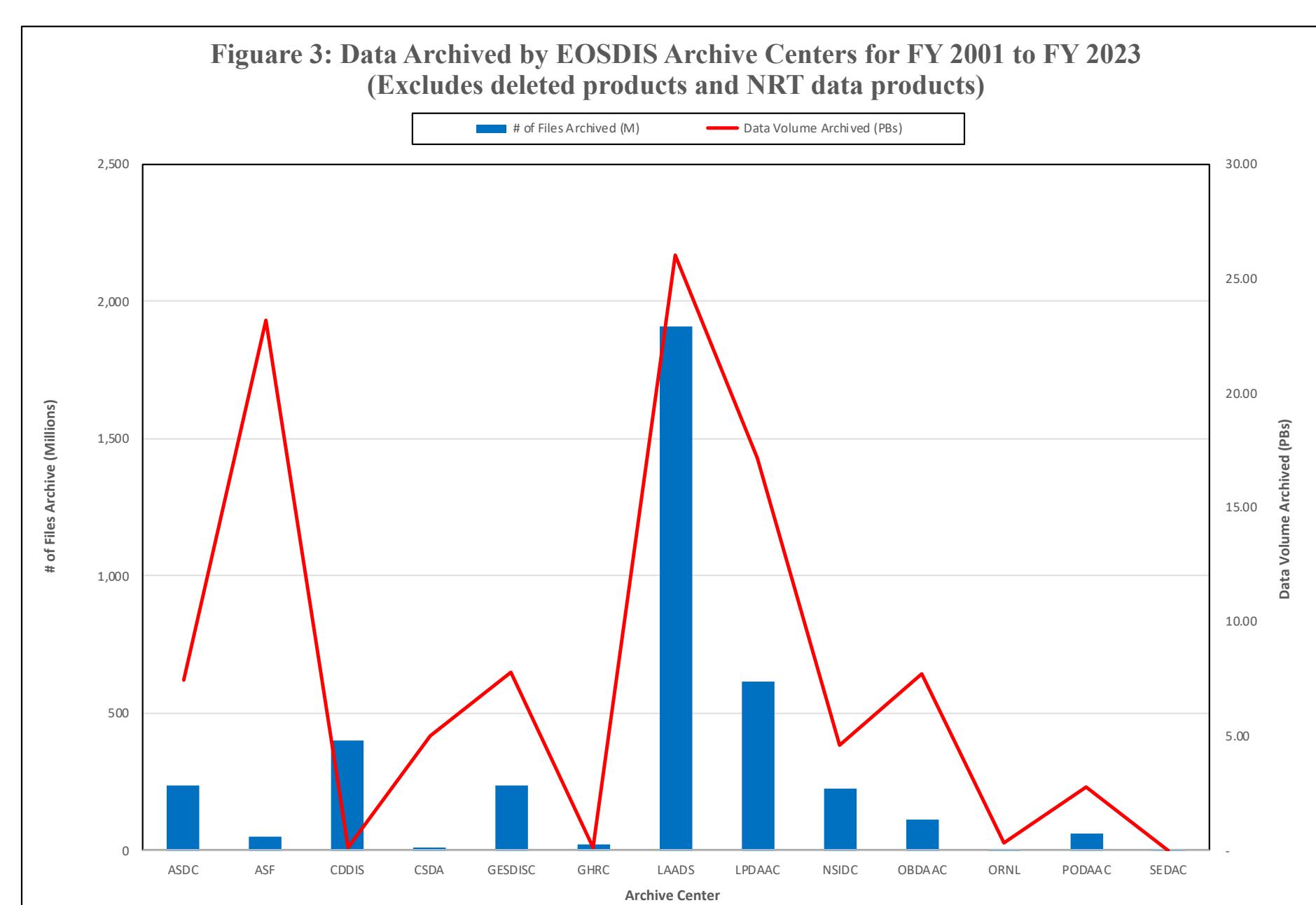


Figure 3: Data Archived by EOSDIS Archive Centers for FY 2001 to FY 2023 (Excludes deleted products and NRT data products)

Table 1: EOSDIS Distributed Active Archive Centers

DAAC	DESCRIPTION
ASDC	Atmospheric Science Data Center
ASF	Alaska Satellite Facility
CDDIS	Crustal Dynamics Data Information System
CSDA	Commercial Smallsat Data Acquisition
GES DISC	Goddard Earth Sciences Data and Information Services Center
GHRC	Global Hydrometeorology Resource Center
LAADS	Level 1 and Atmosphere Archive and Distribution System DAAC
LPDAAC	Land Processes DAAC
NSIDC	National Snow and Ice Data Center DAAC
OBDAAC	Ocean Biology DAAC
ORNL	Oak Ridge National Laboratory DAAC
PODAAC	Physical Oceanography DAAC
SEDAC	Socioeconomic Data and Applications Center

* CSDA is a program established by NASA's Earth Science Division (ESD) to identify, evaluate, and acquire commercial small-satellite (smallsat) data that support NASA's Earth science research and application goals.

Table 2: Data Product Level Definitions

Data Product Level	Description
Level 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 (EDOS) provides these data to the data centers as production data sets for processing by the Science Data Processing Segment (SDPS) or a SIPS to produce higher-level products.)
Level 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 ephemeris) computed and appended but not applied to Level 0 data.
Level 1B*	Level 1A data that have been processed to sensor units (not all instruments have Level 1B source data).
Level 2	Derived geophysical variables at the same resolution and location as Level 1 source data.
Level 3	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.
Level 4	Model output or results from analyses of lower-level data (e.g., variables derived from multiple measurements).

* Level 1A and 1B have been combine for reporting purposes.

Reference: EarthData Website, <https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels>

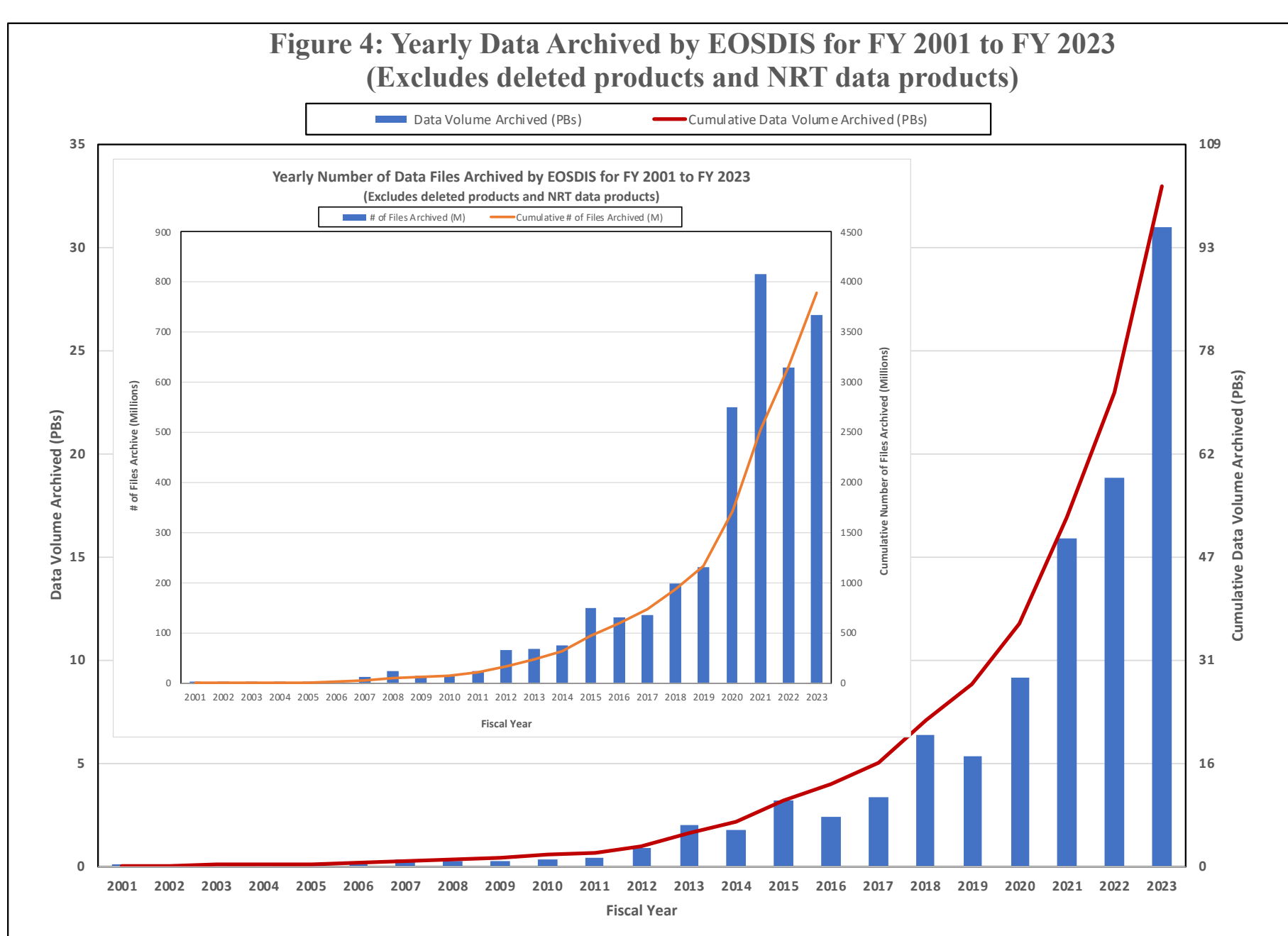


Figure 4: Yearly Data Archived by EOSDIS for FY 2001 to FY 2023 (Excludes deleted products and NRT data products)

LAADS and ASF have the majority of the data volume archived (PB), respectively, than the rest of the DAACs combine. Figure 3. LAADS also archives more data files than any other DAAC. LAADS archive MODIS Aqua and MODIS Terra data while as ASF archive Sentinel data.

The DAACs process, archive, and distribute data from NASA's past and current Earth-observing satellites and field measurement programs, providing easy accessibility to end users. This means that the total EOSDIS data archive volume is distributed over multiple DAACs with each DAAC having a specialized discipline. As of September 30, 2023 over 102 PBs of the data has been archived by all DAACs with significant yearly increase as shown in Figure 4. Figure 4 shows drop in the FY 2022 number of files archived and it was due to the deletion of the older version data by one of the DAACs.

There has been significant, steady growth. Over the past four years, the cumulated data volume archive (PBs) has more than tripled in size. Figure 4. The cumulated number of files archived (Millions) has more than doubled in the same time frame. This has been due to the addition of new mission data to the archives.

Acknowledgments

Lalit Wanchoo, Dany Turcios, Yan Hao, and Heather Weir contributions to this study were funded through the Science and Exploration Data Analysis (SEDA V) GSFC NASA contract No: 80GSFC23CA040 and Francis Lindsay's contributions to this study were made as a part of his employment by NASA.

Data Distribution

EOSDIS DAACs provide free access to all data in the EOSDIS collection to worldwide data users and is constantly evolving services to make discovering and accessing these data as simple and efficient as possible. In FY 2022 and FY 2023, Sentinel data products were the most distributed mission data in terms of volume, approximately 51 PBs and 66 PBs respectively. The decrease in the number of files for FY 2023 is due to LAADS and LPDAAC. Some of the data products are not available as older versions of their data products were deleted. The increase in data volume is due to the addition of JPSS and Sentinel data. The later is available through ASF.

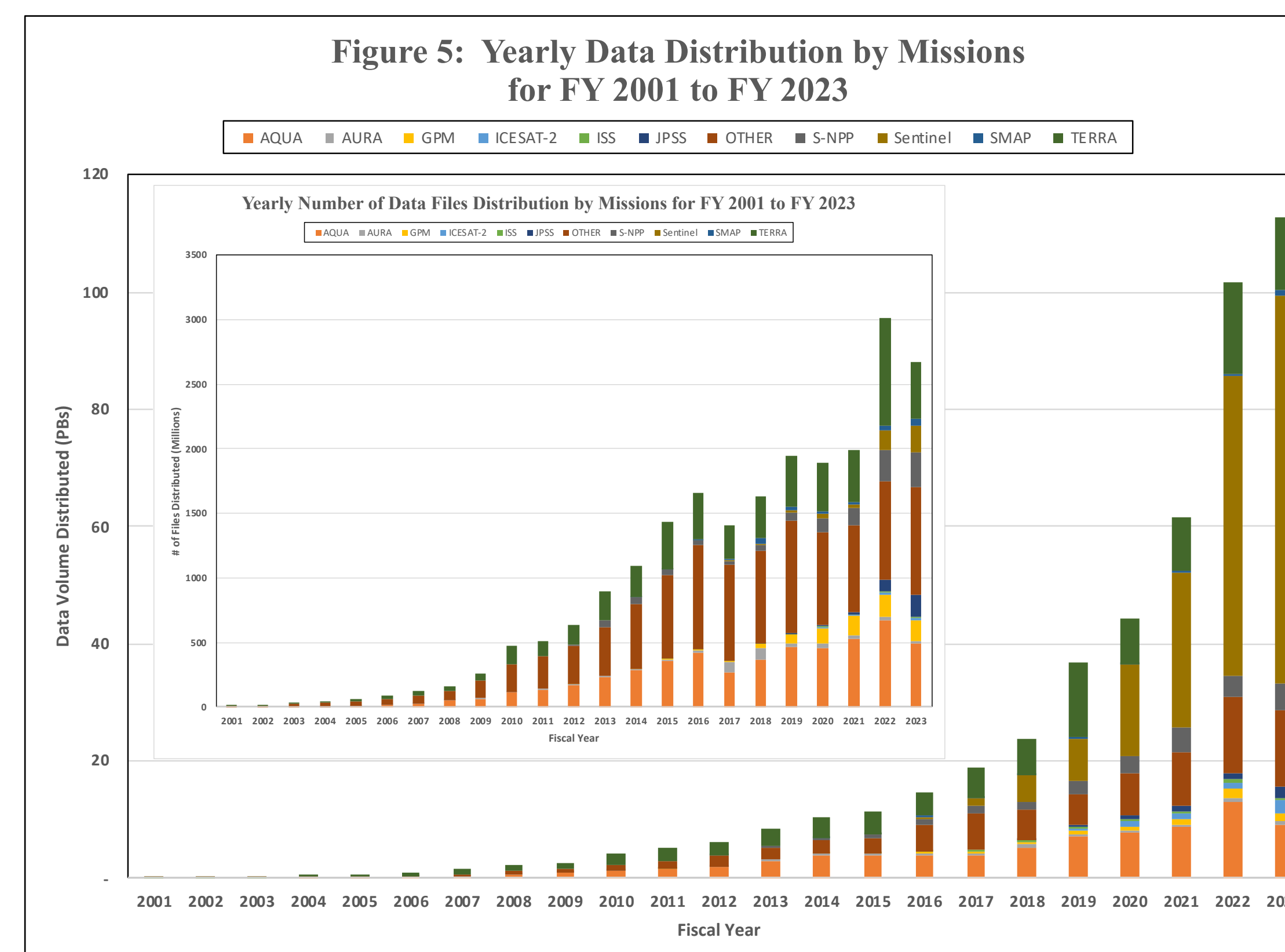


Figure 5: Yearly Data Distribution by Missions for FY 2001 to FY 2023

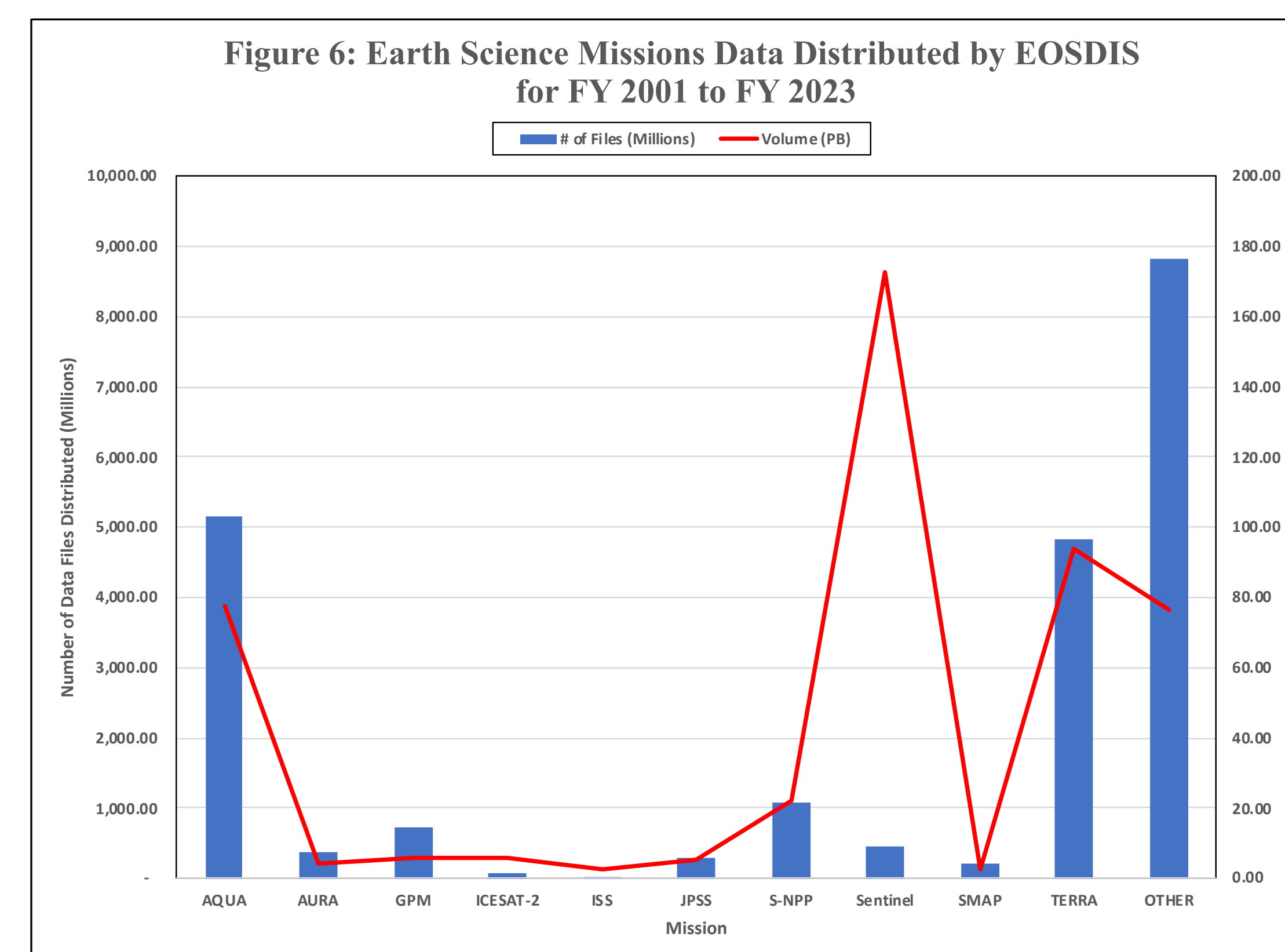


Figure 6: Earth Science Missions Data Distributed by EOSDIS for FY 2001 to FY 2023

AQUA and TERRA, Figure 6, comprise of 45% of the total number of files distributed compared to OTHER which was only 40%. OTHER represents the rest of the missions including aircraft missions, ground observations, and other programs.

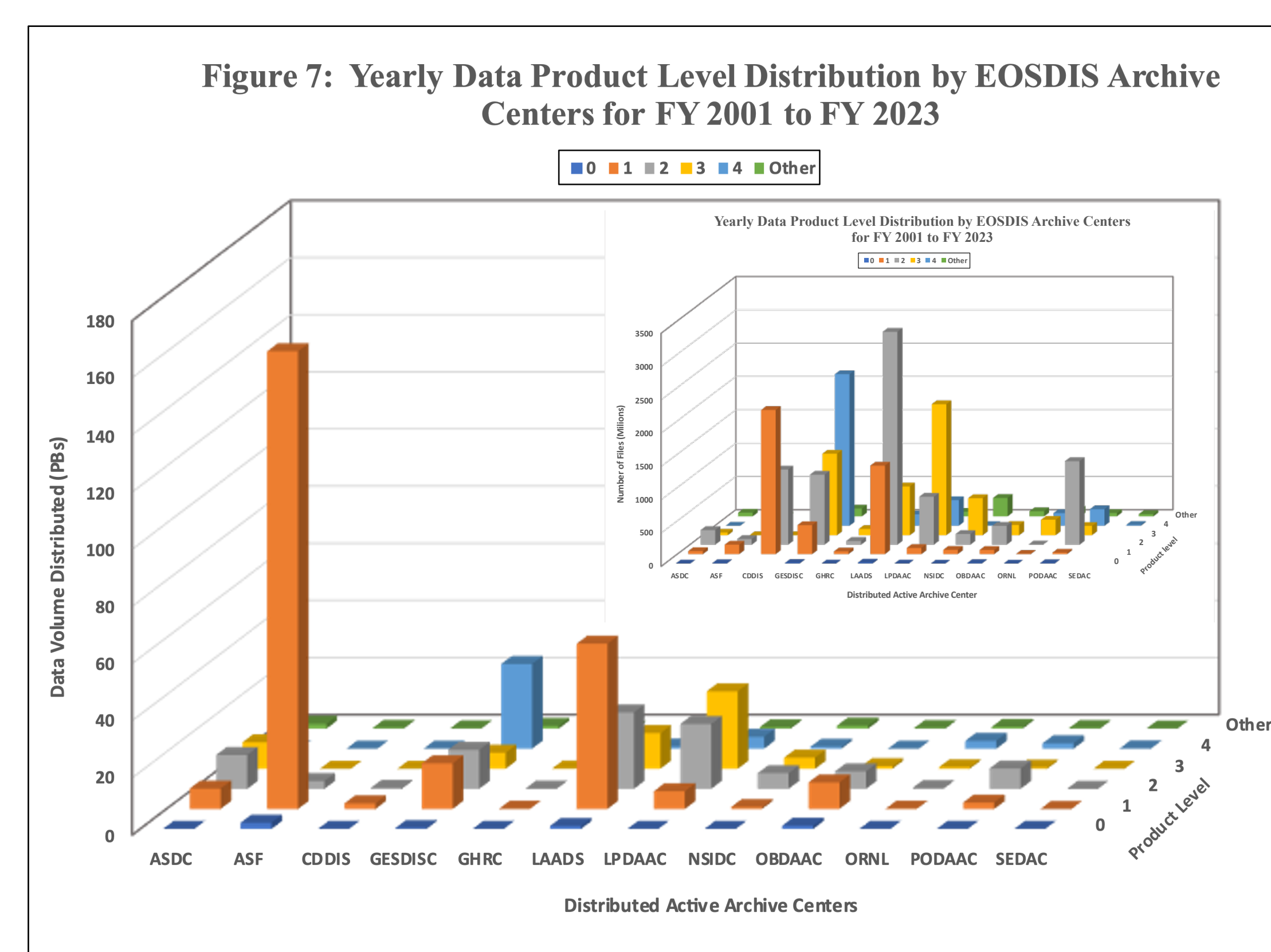


Figure 7: Yearly Data Product Level Distribution by EOSDIS Archive Centers for FY 2001 to FY 2023

ASF, Figure 7, provides the majority of the data volume distributed (PB) at the Level 1 product level, specifically from Sentinel. LAADS has a large number of files at Products Level 2, which is more than any other DAAC.

Data User Characterization

The EOSDIS data user community has evolved significantly over the past 23 years. As the number of mission has increased, so has the number of users downloading Earth science data. Users downloading data has grown from more than 140 in FY 2001 to more than 2.3 million in FY 2023. Figure 6 illustrates the distribution of EOSDIS data to more than 22.7 million users from the United States and foreign countries over the last 23 years. Half of EOSDIS data was downloaded by U.S. users. Figure 10 shows that majority of data of higher product level were downloaded by the users.

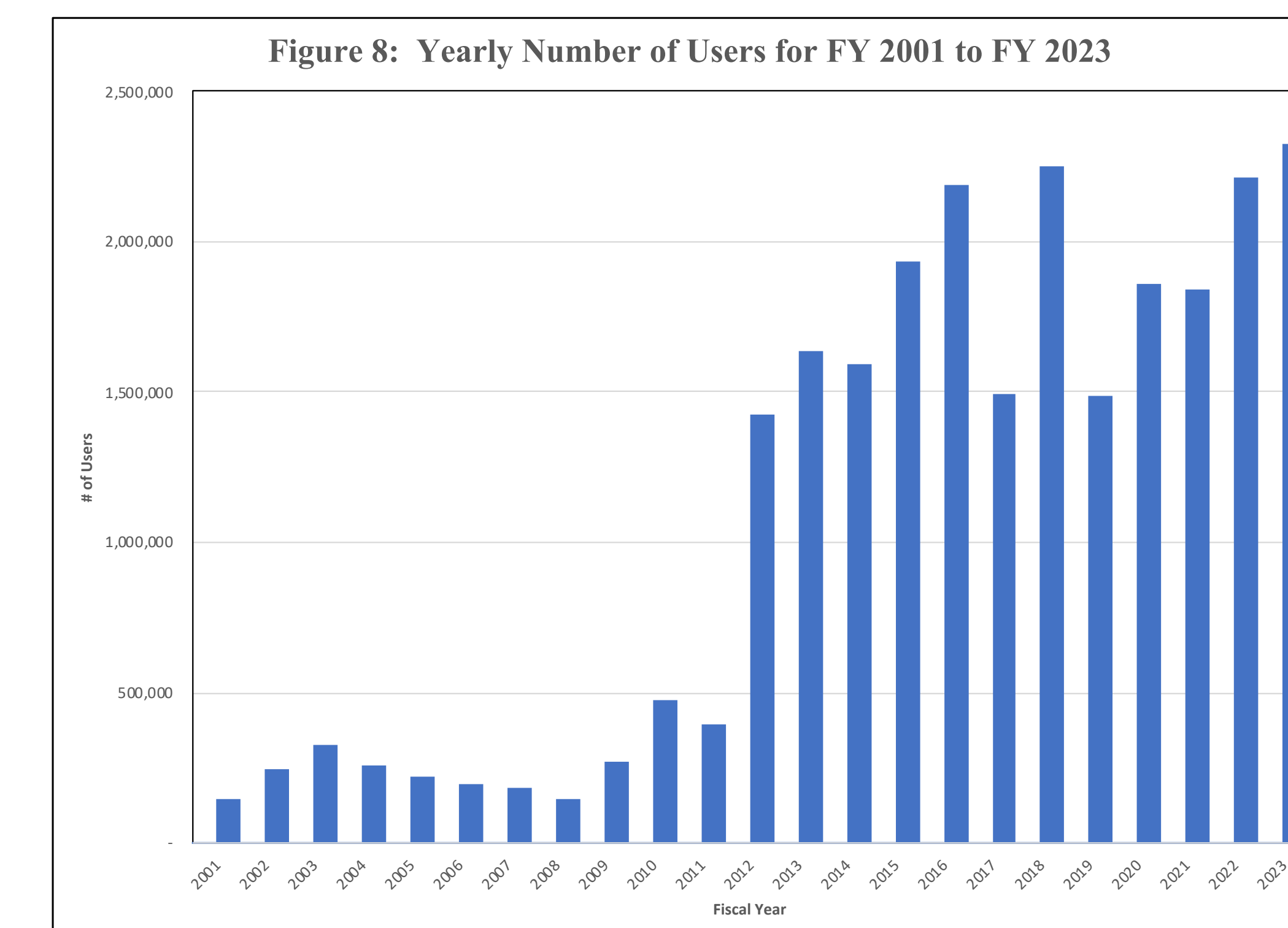


Figure 8: Yearly Number of Users for FY 2001 to FY 2023

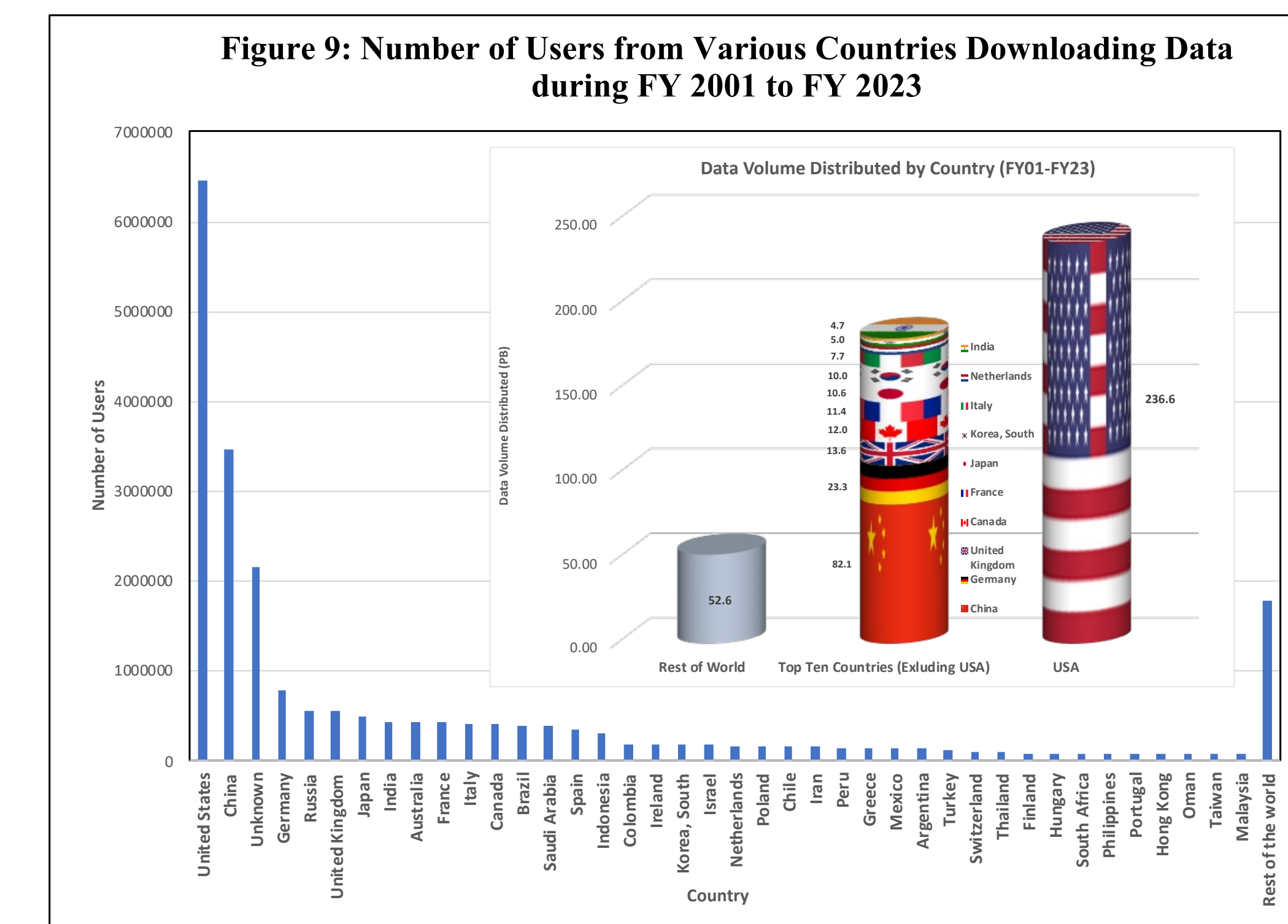


Figure 9: Number of Users from Various Countries Downloading Data during FY 2001 to FY 2023

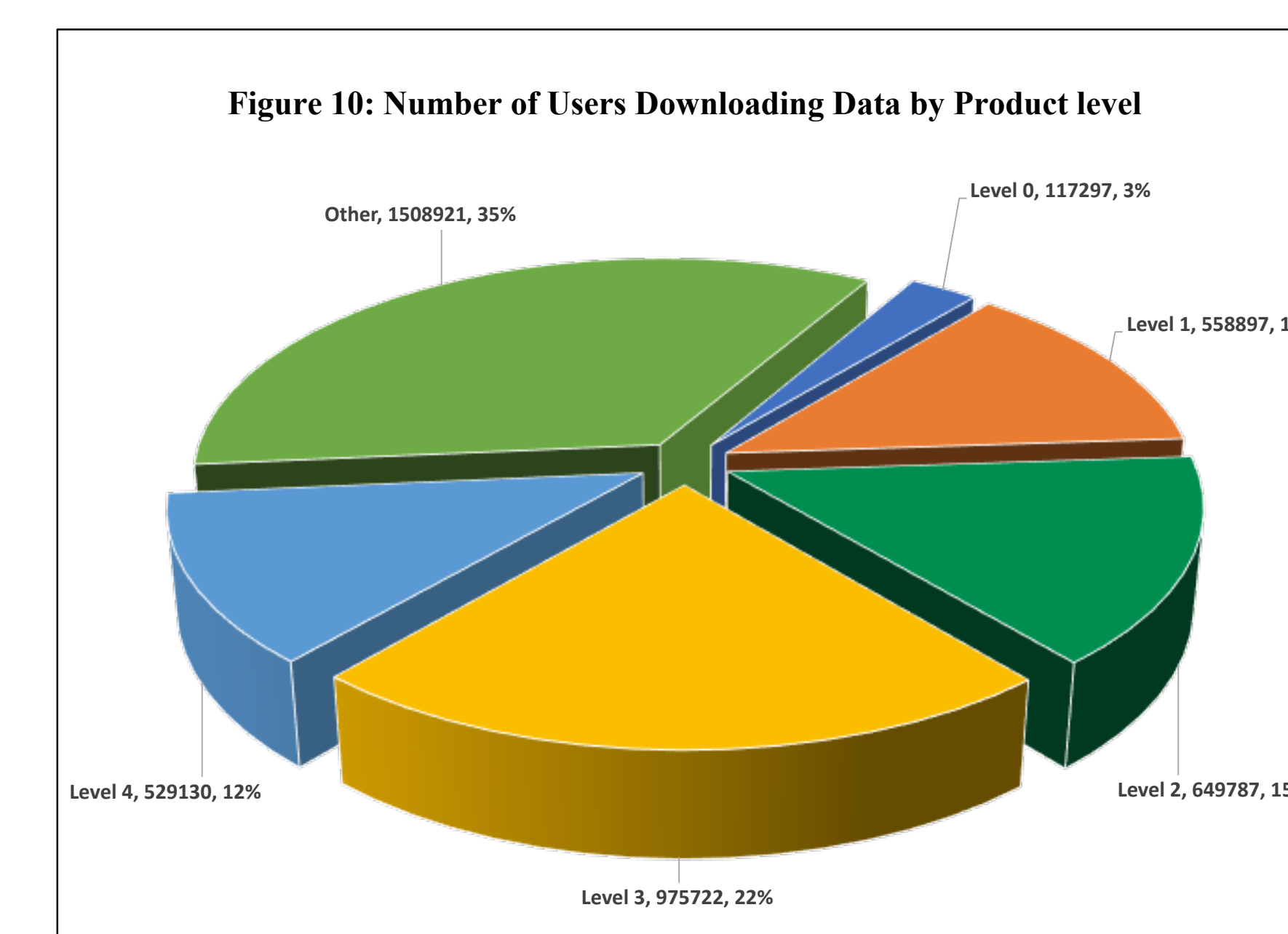


Figure 10: Number of Users Downloading Data by Product level

Summary

Figure 11 shows how the archiving and distribution of EOSDIS data have evolved over last 23 years. A comparison of FY 2001 and FY 2023 data metrics shows the following: 1) From only a few hundred data products, the EOSDIS archive has grown to more than 10,000 data products, 2) The distribution of EOSDIS data has grown from 105 Terabytes (TB) to 112 PB per FY, and 3) The number of users downloading data has grown from about 143 to 2.3 million, with data being distributed to world-wide users free and with open access. These metrics for EOSDIS archive and distribution data provide insight to the innovation and work of EOSDIS DAACs in managing the growth of the Earth science data archive and facilitating the efficient distribution of these data. This may have been possible largely due to free and NASA's Open Data Policy, thereby an important step towards open science initiative. EOSDIS DAACs continue to provide guidance for the requirements of future data systems that will be needed to effectively and efficiently handle increasing amounts of data produced by on-going and future Earth science missions. This process will continue providing unrestricted access to the data to world-wide users thereby accomplish the first and most important step for the US Open Science Initiative directive.

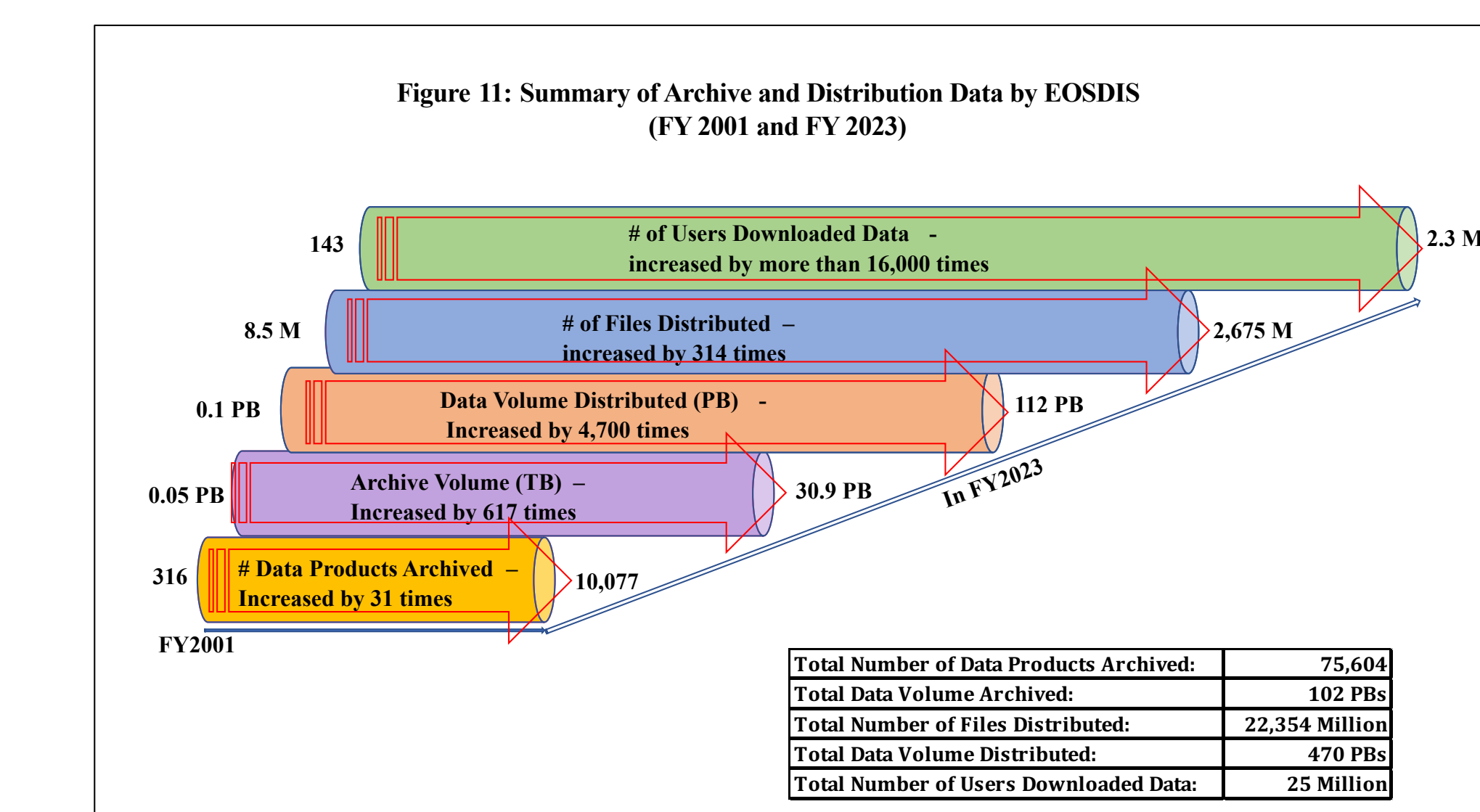


Figure 11: Summary of Archive and Distribution Data by EOSDIS (FY 2001 and FY 2023)

Total Number of Data Products Archived:	75,604
Total Data Volume Archived:	102 PBs
Total Number of Files Distributed:	22,354 Million
Total Data Volume Distributed:	470 PBs
Total Number of Users Downloaded Data:	25 Million