



Developing Metrics for NASA Earth Science Interdisciplinary Data Products and Services

RESEARCH PAPER

ZHONG LIU

CHUNG-LIN SHIE

ANTHONY J. RITRIVI

GUANG-DIH LEI

GARY T. ALCOTT

MARY GREENE

JAMES ACKER

JENNIFER C. WEI

DAVID J. MEYER

ANGELA LI

ATHEER F. AL-JAZRAWI

][ubiquity press

**Author affiliations can be found in the back matter of this article*

ABSTRACT

Metrics are measures that are able to produce quantifiable information. There are many applications of metrics in Earth science data and services; for example, metrics are frequently used to track service performance and progress. In short, developing, collecting and analyzing metrics are essential activities to better support Earth science research, applications, and education.

As one of the largest repositories of Earth science data in the world, NASA's Earth Science Data and Information System (ESDIS) Project supports twelve Distributed Active Archive Centers (DAACs). Standard metrics have been developed by the ESDIS Metrics System (EMS). These metrics are collected and analyzed routinely at each DAAC. As it is expected that the total data volume will continue to grow rapidly, and the timely developed technologies (e.g., cloud computing, AI/ML) will continue to improve data discovery and accessibility, opportunities for developing new data services for the Earth science community will also arise, especially in interdisciplinary research and applications. However, developing such metrics has become a challenge because multiple datasets are often needed. Current metrics are designed for a single predefined dataset or service, a disadvantage for collecting metrics for interdisciplinary data services.

In this paper, we assess current metrics using one of the NASA DAACs, the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), as an example, to discuss challenges and opportunities, along with recommendations for developing metrics addressing interdisciplinary satellite data products and services.

Highlights

Overview of NASA GES DISC Earth science datasets and services

- Overview of existing metrics collection methods and analysis tools with examples
- Discuss challenges and opportunities in collecting metrics for Earth science interdisciplinary data and services

CORRESPONDING AUTHOR:

Zhong Liu

NASA Goddard Earth Sciences
Data and Information Services
Center (GES DISC), US; George
Mason University, US

zhong.liu@nasa.gov

KEYWORDS:

Metrics; NASA; Earth
Science; Remote Sensing;
Interdisciplinary; Data and
Services

TO CITE THIS ARTICLE:

Liu, Z, Shie, C-L, Ritrivi, AJ, Lei,
G-D, Alcott, GT, Greene, M,
Acker, J, Wei, JC, Meyer, DJ,
Li, A, Al-Jazrawi, AF. 2022.
Developing Metrics for NASA
Earth Science Interdisciplinary
Data Products and Services.
Data Science Journal, 21: 5,
pp. 1–15. DOI: [https://doi.
org/10.5334/dsj-2022-005](https://doi.org/10.5334/dsj-2022-005)

Metrics are measures that are able to produce quantifiable information. There are many applications of metrics, ranging from monitoring data system performance to benchmarking a project or mission success. Metrics (e.g., [Table 1](#)) are routinely collected in data repositories and provided to data providers, mission or project management, and scientists and software engineers, to analyze (e.g., compare, benchmark) and track a range of performance-related activities, such as system performance, data access and usage in research, applications, allocation of information technology resources, and benchmarking a mission or project success (Cousijn et al. 2019; Par et al. 2019; Behnke et al. 2019; Shie et al. 2019; Kafle et al. 2019; Cruse et al. 2019; Su et al. 2019; O'Brien et al. 2020). In short, developing, collecting, and analyzing metrics is essential to better support Earth science research, applications, and education (e.g., NASA EOSDIS 2021a, 2021b, 2021c, 2021d, 2021e).

TYPE OF METRICS	METRICS
Key Metrics	The operational distribution metrics recording overall user data/service access and download activities for the following three major groupings: <ul style="list-style-type: none"> i) Number of Distinct or Registered Users ii) Number of Distributed Data Files iii) Size of Distributed Data Volume in four categories: Country (e.g., United States, Canada), Protocol (e.g., HTTPS, OPeNDAP), Project (e.g., TRMM/GPM, MERRA-2) and Domain (e.g., '.edu', '.gov').
Bugzilla Ticket Metrics	Collecting and retrieving significant and useful info from user questions or feedback mentioned in User Assistance Tickets: <ul style="list-style-type: none"> i) User Background: <ul style="list-style-type: none"> a) What they are: Researcher; Professor/Graduate Students; Industry; etc. b) Where they come from: USA; Africa; Asia; Australia; Europe; and the Middle East; etc. ii) Number of User Assistance Tickets: Monthly; Seasonal; Yearly distributions (per routine Daily collections) iii) Application/Study: Hydrology; Atmospheric Chemistry; Oceanography; etc. iv) Portal: Giovanni; MERRA-2; TRMM/GPM; etc. v) Data Variable: Air Temperature; Wind Fields; Precipitation; Aerosol; etc.
Giovanni Publication Metrics	Collecting/gleaning significant and useful info from our Giovanni user journal publications in regard to: <ul style="list-style-type: none"> i) Applied Variable: Atmos. Aerosol; Precipitation; Air Temperature; etc. ii) Product Source: TRMM/GPM; MODIS; MERRA-2; etc. iii) Studied Subject: Hurricane; Aerosol/Dust; Rain/Water Vapor; etc. iv) Studied Temporal Period: Long-term; Mid-term; Short-term v) Studied Spatial Domain: Global; Regional; Local vi) Studied Region: Continents; Oceans; Countries; Lakes; etc. vii) Journal Origins: America; Europe; Asia; Middle East; International/Open Access; etc.
Website Metrics (via Google Analytics)	Collecting useful info on User Website Access via utilization of the Google Analytics Tool. <ul style="list-style-type: none"> i) Dataset Keyword search: "rainfall"; "TRMM"; "merra-2"; "trmm"; "GPM"; etc. ii) Information Keyword search: "precipitation"; "trmm"; "rainfall"; "merra-2"; "Giovanni measurements"; etc. iii) Content Type search: "Data Collection"; etc. iv) Traffic and/or Referral Sources: Direct Access; Google; etc. v) Datasets subsetted/downloaded directly from search results page: "trmm_3b42_v7"; "m2t1nxslv_v5.12.4"; "m2i3npasm_v5.12.4"; etc. vi) Most Sorted Columns: "begin date"; "time res."; "end date" vii) Most Browsed Categories: "subject"; "measurement"; "source"; "project"; "spatial resolution"; "temporal resolution" viii) Most searched Content Type: "data collections"; "data documentation"; "image gallery"; "how-to's"; "tools"; "faqs"

Table 1 Four major types of metrics collected at GES DISC.

As one of the largest repositories of Earth science data in the world, NASA's Earth Science Data and Information System (ESDIS) Project (NASA ESDIS 2021a) supports twelve Distributed Active Archive Centers (DAACs) (NASA DAACs 2021). Standard metrics have been developed and collected by the ESDIS Metrics System (EMS) (NASA ESDIS 2021b) for routine analysis at each DAAC. Other metrics for both ESDIS and each DAAC are also collected, such as user satisfaction.

As the total data volume is expected to grow rapidly and technologies (e.g., cloud computing, AI/ML) continue to improve data discovery and accessibility, opportunities for developing new data services for the Earth science community will also arise. However, developing such metrics has become a challenge, because in this era multiple datasets are often needed. Current metrics are designed for a single predefined dataset or service, a disadvantage for collecting metrics for interdisciplinary data services.

In this paper, we use one of the DAACs, the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) (NASA GES DISC 2021), as an example to assess the current status of metrics. We will discuss challenges and opportunities, with recommendations in developing metrics for interdisciplinary data and services in conjunction with the FAIR guiding principles (Wilkinson et al. 2016) and community recommendations.

The structure of the paper is as follows: Section 2 overviews existing datasets and services for collecting metrics; Section 3 lists current metrics, collection methods and operations, along with examples; Section 4 discusses current issues and future needs for new metrics; and Section 5 provides the summary and recommendations.

2. DATASETS AND SERVICES AT GES DISC

The GES DISC provides a large number of NASA Earth science multidisciplinary datasets (i.e., atmospheric composition; water and energy cycles; climate variability; carbon cycle and ecosystems) to research, application, and education communities across the globe. Datasets from several well-known NASA satellite missions and projects are included, such as the NASA-JAXA Tropical Rainfall Measuring Mission (TRMM), and the Modern-Era Retrospective-analysis for Research and Applications (MERRA). As of this writing, more than 1200 datasets are archived at GES DISC.

The GES DISC homepage ([Figure 1](#)) is the primary gateway for accessing datasets and relevant information (e.g., technical documents). The homepage ([Figure 1](#)) contains Web components, which include: 1) dataset search from data collections; 2) links to dataset-related publications; 3) access to data tools; 4) archives of dataset-related information (news, alerts, service releases, glossary); and 5) libraries of supporting information (FAQs, Data in Action articles, How-To's). There are several search functions developed to help refine and sort search results (e.g., refining by subject, measurement, source, data processing level, project, and temporal and spatial resolutions).

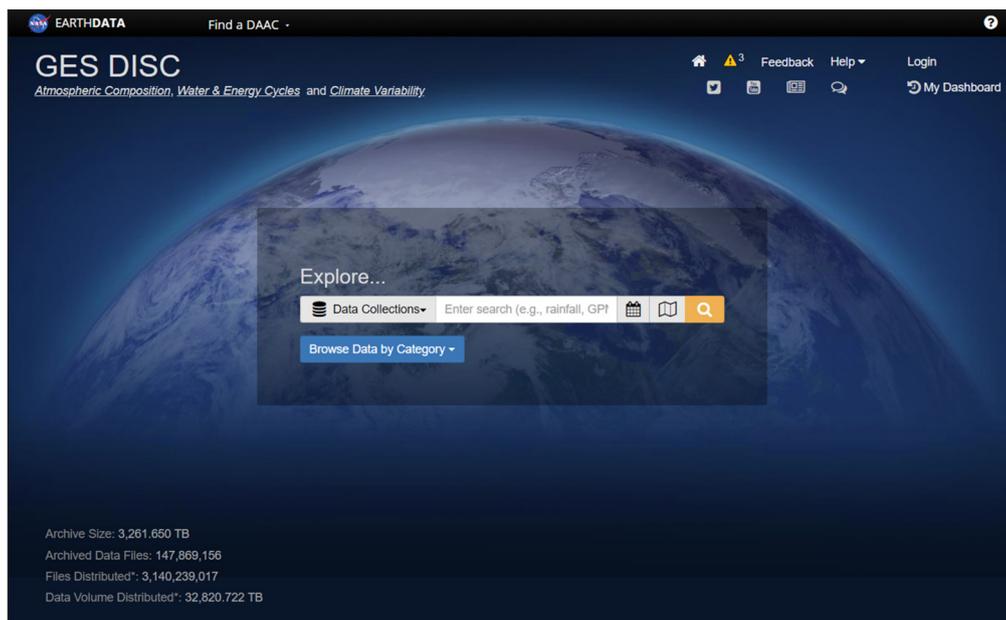


Figure 1 The homepage of GES DISC with search capabilities for datasets, tools, documentation, alerts, data releases, news, FAQs, publications and more.

Each dataset at GES DISC has its own landing page (e.g., [Figure 2](#)) which serves a one-stop 'shop' for data services and information. The dataset landing page is still evolving; at present, it

provides: 1) data access methods; 2) a brief dataset summary; 3) dataset citation (e.g., Digital Object Identifier, DOI); and 4) key dataset documentation (e.g., the Algorithm Theoretical Basis Document, also known as ATBD).

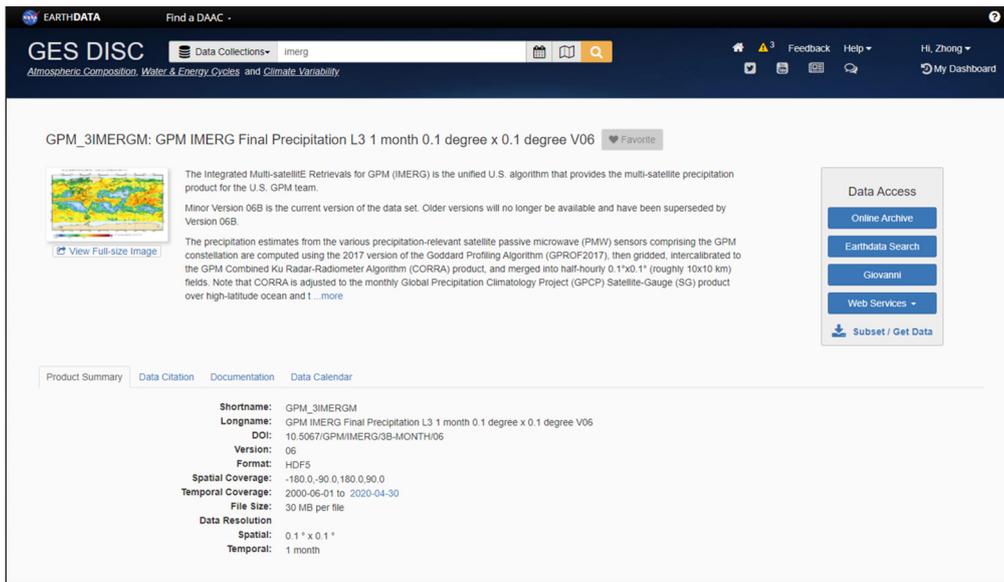


Figure 2 An example of the dataset landing page for the popular NASA Integrated Multi-satellite Retrievals for GPM (IMERG) monthly dataset. A one-stop ‘shop’ design allowing easy access to data and dataset related information.

3. EXISTING SYSTEMS AND SOFTWARE FOR COLLECTING METRICS

3.1. FRAMEWORK OF COLLECTING METRICS AT GES DISC

A framework of collecting metrics has currently been developed at GES DISC (Figure 3). Four major/different types of metrics (Table 1) are collected and analyzed at GES DISC (Shie et al. 2019) including: 1) key metrics (Figure 4) extracted from the operational distribution metrics, recording overall user data/service access and download activities and submitted to EMS; 2) Bugzilla (Bugzilla 2021) user ticket metrics (Figure 5) from User Assistance tickets; 3) Giovanni (Liu and Acker 2017; NASA Giovanni 2021) publication metrics (Figure 6) from journal publications by users acknowledging their Giovanni usage; and 4) website metrics (via Google Analytics) (Figure 7) for information on user website access.

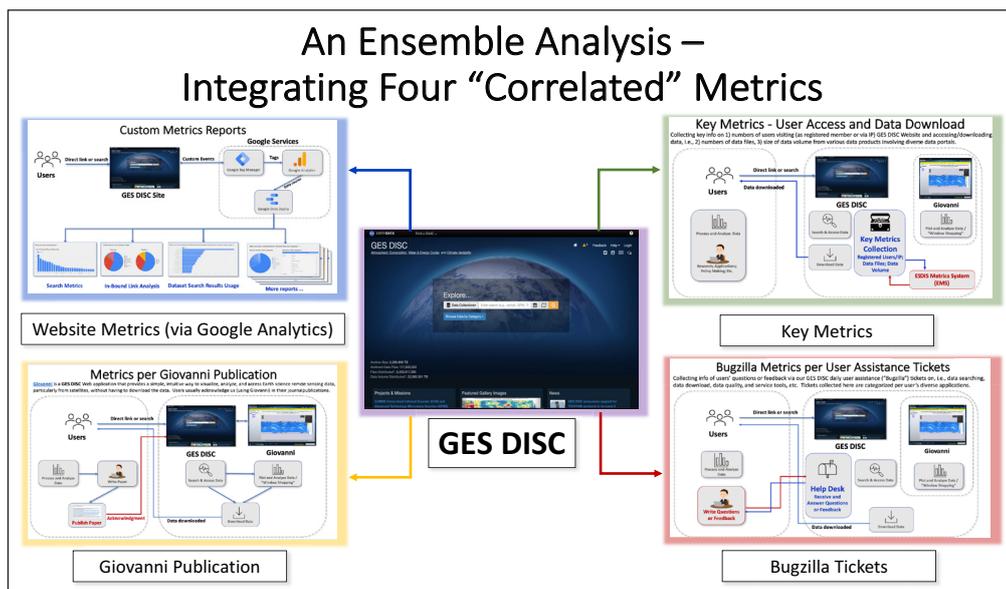


Figure 3 A schematic of four “correlated” metrics at GES DISC. More details with examples are shown in Figures 4, 5, 6 and 7.

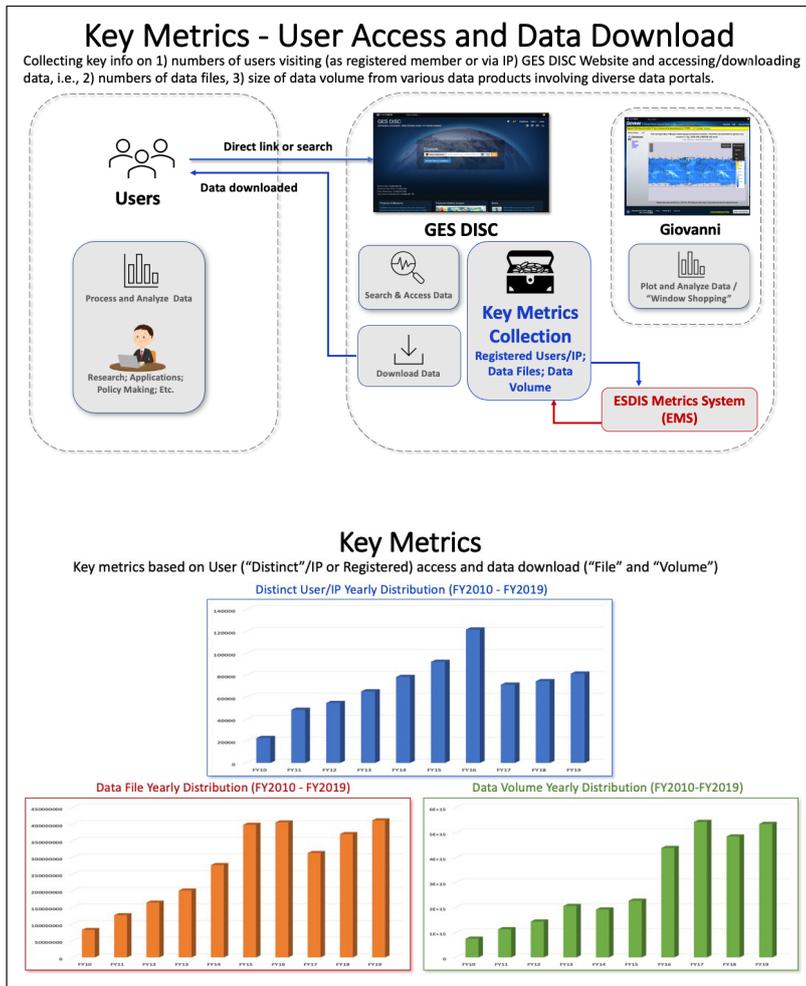


Figure 4 The schematic of the collection workflow (top), and the yearly, i.e., FY2010 – FY2019, distributions of distinct user/IP (middle), data file (bottom left), and data volume (bottom right).

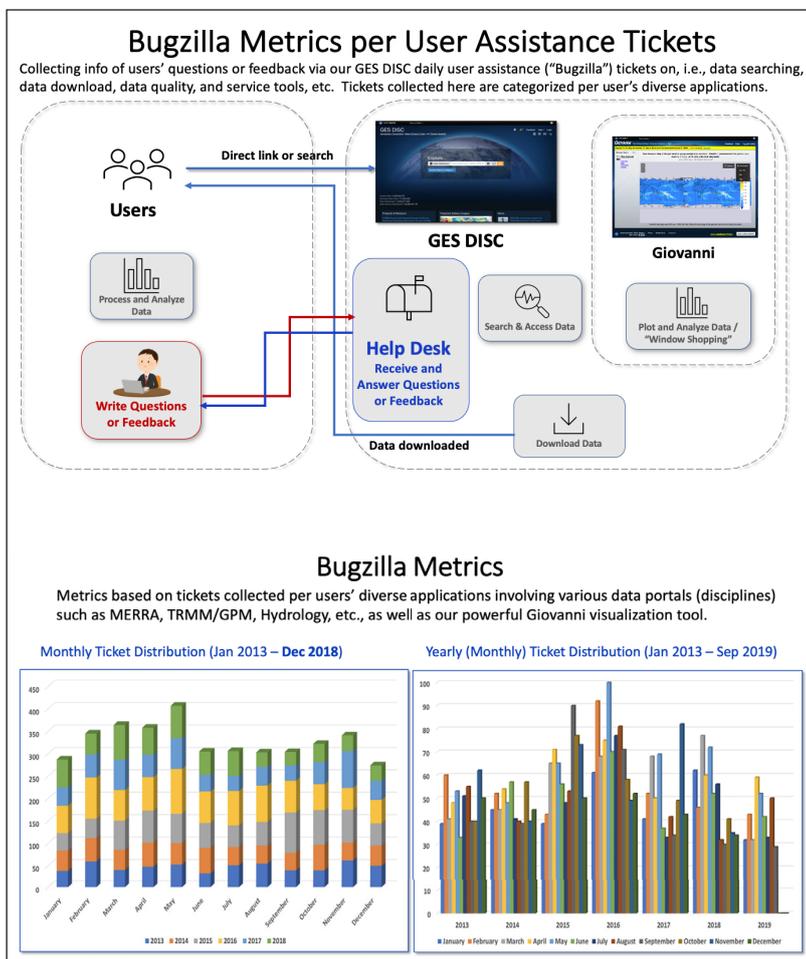


Figure 5 Bugzilla metrics – user assistance tickets. Top: a schematic of the collection workflow. Bottom: Monthly (2013–2018, left) and yearly (201301–201909, right) ticket distributions presented in two different perspectives.

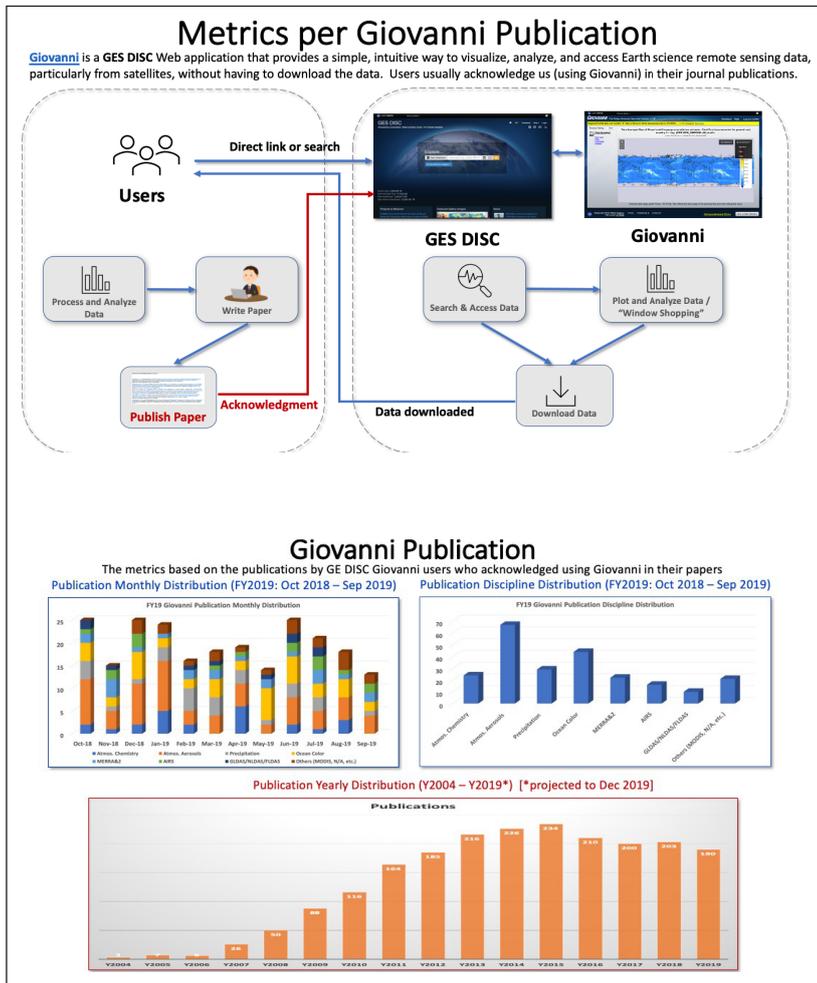


Figure 6 Giovanni publication metrics. Top: a schematic of the collection workflow. Middle: Monthly publication distributions of diverse disciplines (left) and the respective distributions of individual disciplines (right) for FY2019. Bottom: Yearly publication distributions for Y2004-Y2019* [*projected to Dec 2019].

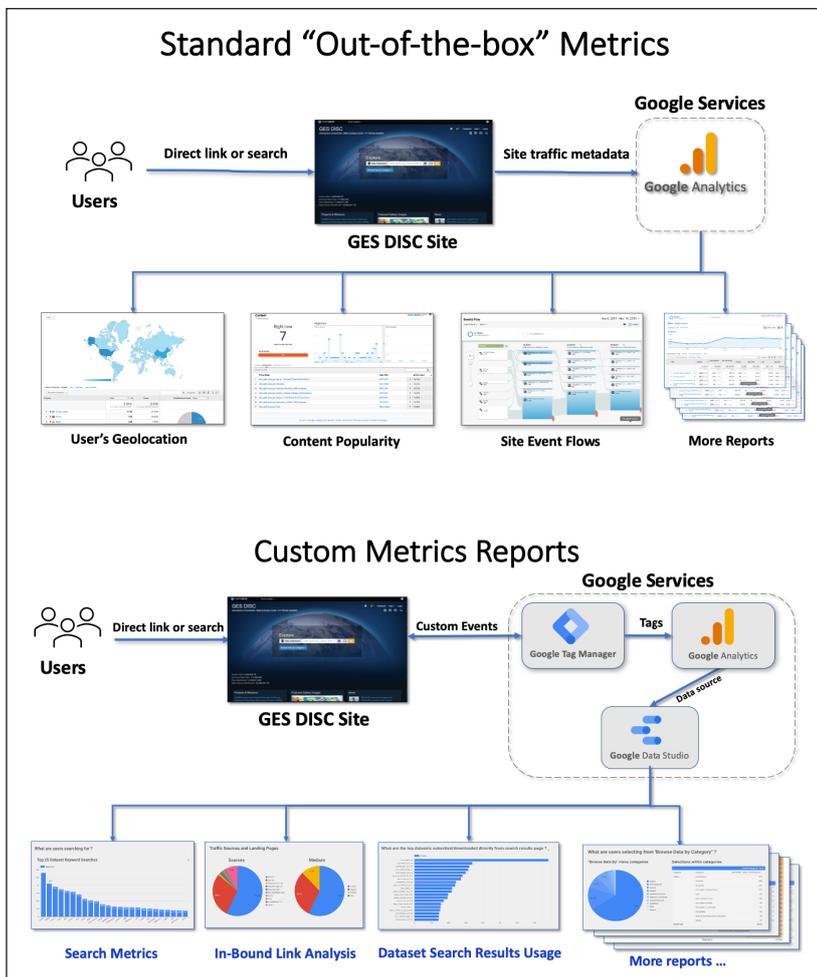


Figure 7 Standard “out-of-the-box” metrics. Top: a schematic of the collection workflow. Bottom: a workflow to generate GES DISC website custom metrics reports.

3.2. NASA ESDIS METRICS SYSTEM (EMS)

ESDIS EMS (NASA ESDIS 2021b) establishes requirements and methods for each DAAC to collect data activity and usage metrics. The metrics, analysis, and reports (NASA ESDIS 2021c) are generated and provided to NASA management to inform the best allocation of resources for the scientific user community (NASA ESDIS 2021b). Metrics are also analyzed at DAACs for better understanding the usage of their data products by end users, which can potentially or genuinely guide data centers to improve data services.

Digital analytics products (e.g., Google Analytics 360) not only include the traditional website analytics (e.g., the IBM NetInsight), but are also capable of acquiring and analyzing metrics from all possible sources, including again, traditional website analytics, and social media, mobile devices, etc. (NASA ESDIS 2021b). At present, Google Analytics 360 is used at EMS for digital analytics (NASA ESDIS 2021b). In short, digital analytics products provide additional information for data service operations and decision makers.

3.3. COLLECTING METRICS FOR EMS AT GES DISC

Data product collection metadata serve as a key element in the GES DISC reporting capabilities. Product attributes consist of metadata such as instrument, mission, product level, and discipline, all describing the characteristics of a data product. This metadata information, together with the product search term information, will be linked or “mapped” to each record within a distribution file because of unique patterns contained within a record. Accuracy of EMS reports is highly dependent on comprehensive, consistent, and timely updates to the product attributes. The required fields of the collection metadata and its search terms are listed in [Table 2](#). They are used to collect metrics information from all data ingest, archive, and distribution interfaces throughout EOSDIS for analysis and reporting.

Table 2 Required fields of the EMS collection metadata and its search terms.

FIELD NAME	DESCRIPTION	MAX LENGTH	EXAMPLE
product	This is a product identifier or the short name of the dataset...	80	AIRIBRAD
metaDataLongName	Identification of the long name associated with the collection or granule.	1024	AIRS/Aqua infrared geolocated radiances
productLevel	NASA data processing levels (i.e., 0, 1, 1A, 1B, 2, 3, 4).	10	1B
discipline	Designates the scientific area of application (i.e., Ocean, Atmosphere, Land, Cryosphere, Volcanic, Solar, Raw data, Radiance).	500	Atmosphere
mission	An operation to provide scientific measurements with space-based and/or ground-based measurement systems (i.e., platforms, satellites, field experiments, and aerial measurements, etc.). For a multi-mission product, list all missions separated by a semi-colon (;). The primary mission should be listed first. Each mission should have 1 or more instruments associated with it. If there are multiple missions and multiple instruments, then the relationships between the missions and instruments should be defined.	80	Aqua
instrument	Consisting of a collection of one or multiple sensor instruments to provide scientific measurements. For a multi-instrument product from one mission, all instruments are listed and separated by a comma (,). If the product (e.g., a combined product) involves multiple missions and multiple instruments, the instruments from each mission are separated by a semi-colon (;). The order of instruments should be in the same sequence as the mission field. If not applicable, enter: “N/A”. (NOTE: the number of missions entered must pair evenly to number of instruments delimited by “;” i.e., if two missions entered: “mission1;mission2” then at least two instruments: “instrument1;instrument2” or “N/A;N/A” or “instrument1a,instrument1b”; “instrument2a,instrument2b” etc.)	80	AIRS

(Contd.)

FIELD NAME	DESCRIPTION	MAX LENGTH	EXAMPLE
processingCenter	Data center where this product was generated.	80	GESDISC
archiveCenter	Data center where the data product is archived. This value is usually 'GESDISC'.	50	GESDISC
eosFlag	Flag to indicate whether the data product is an EOS (NASA EOS 2021) or Non-EOS product. Values: E for EOS and N for Non-EOS.	1	E
productFlag	Flag denotes the type of product. Values: 1 = Data Product, 2 = Instrument Ancillary, 3 = System/Spacecraft and 4 = External. For a non-ECS product, use the value 1.	1	1
publishFlag	Flag to indicate whether the product and its associated granules be published to EMS or not. This value is usually 'Y'.	1	Y
searchTerm	File name, directory, path, ESDT, Data Provider internal product IDs or other information that uniquely identifies a data product as it appears in an EMS Data file. The searchTerm should not include URL query strings and associated name value pairs. searchTerms can include full strings or substrings. Values within this field are always treated as regular expressions (e.g., '+MOD1[1-9].+'). Therefore, reserved grep/egrep characters should only be used when they are needed. By default, we will use the product Shortname. Let OPS staff know if any specific pattern needs to be added to a product.	200	AIRIBRAD
dataSource	Assigns the data source (e.g., the system, subsystem, file, table or other identifying information) where the logs/flat files/metadata are generated (e.g., airs, aura, disc, reason, urs). Currently, GES-DISC has identified five data source groups. Each group and its associated hosts are listed as follows: <ul style="list-style-type: none"> • 'airs' – airscal1u, airscal2u, airspar1u, airspro2u, airspro3u, airspro5u, airsraw1u, airsraw2u, airsraw3u, rep2u, rep1 • 'aura' – acdisc, aurapar1u, aurapar2u, auraraw1u, goldsfs1u, goldsmr1, goldsmr2, goldsmr3, rep5u • 'reason' – reason, neespi, atrain, agdisc, hydro1 • 'disc' – disc1, disc2, disc3, tads1u, gdata1, gdata2, rep3, rep4 • 'urs' – discrnt1, discrnt2 	50	airs

3.4. AMERICAN CUSTOMER SATISFACTION INDEX (ACSI) REPORTS

The metrics for user activities described so far are passively collected from the servers at GES DISC. By contrast, initiated by ESDIS (NASA EOSDIS 2021d) and tasked to the Claes Fornell International (CFI) Group (CFI 2021), the ACSI survey (ACSI 2021) has been conducted annually in the user community of NASA DAACs in a proactive manner since 2004. The ACSI survey measures user satisfaction with NASA EOSDIS data services to identify key areas for continuously improving data services to users, and to track trends in user satisfaction with each DAAC.

4. CHALLENGES AND OPPORTUNITIES

From [Table 1](#), it is seen that key and website metrics collected at GES DISC are likely to be found in other data repositories as well. The other two metrics, namely Bugzilla ticket and Giovanni publication metrics, may depend on whether the similar services are available. Nonetheless, current key metrics at GES DISC, in general, are designed for a single dataset or service. Interdisciplinary research and applications, necessitating the use of multiple datasets, have increasingly relied on data from multiple sources (e.g., satellites, models, *in situ* observations). For example, multiple datasets from different disciplines (e.g., meteorology, hydrology, oceanography) are often needed or demanded in tropical weather and climate research and applications (Liu et al. 2020). For interdisciplinary research, datasets (involving multiple satellite

missions or projects) are often acquired from multiple DAACs or even multiple domestic and/or international organizations (e.g., NOAA, NSF, ESA). Accordingly, adequate and meaningful correlated metrics associated with multiple datasets and collecting methods should also be further considered and developed.

4.1. INTERDISCIPLINARY AND INTEGRATION CHALLENGES

As technologies evolve and scientific requirements change with time, new Earth science data services have correspondingly been developed and improved. For example, the cloud computing environment (NASA EOSDIS 2021f) has aimed to improve capability and potential to develop and provide numerous new data services, which may be otherwise difficult to implement 'on premises', for handling large volume remote sensing or model datasets. Eventually all datasets at the twelve DAACs will be made available in the cloud environment (NASA EOSDIS 2021f). Such a cloud environment offers new opportunities to develop customized dataset services capable of generating on-the-fly datasets from one dataset and/or merging several datasets from one DAAC or even multiple DAACs. A new design for such customized datasets is possible by organizing multiple datasets into different groups (e.g., netCDF-4) to form a new dataset. However, existing single-dataset-based data product development guides (e.g., Ramapriyan and Leonard 2020) need to be expanded for this new dataset design.

In current EMS metrics, metadata are required to be defined in advance in order to be collected for metrics from server logs. With more on-the-fly datasets expected to be available in future data services, it would likely become quite a challenge to develop this kind of metrics, as those datasets lack full definitions prior to usage. Rauber et al. (2015) proposed a scalable dynamic data citation methodology with three core recommendations for data (versioning, timestamping, and identification), which can be considered for use with on-the-fly datasets.

The four major metrics groups at GES DISC have been found to be correlated to each other to a certain degree (Shie et al., 2019). It is necessary to integrate these metrics groups for a holistic analysis. Additional dataset-related metrics are needed, in particular science-related metrics (Parr et al. 2019). For example, dataset citations are a key measurement for impact in the scientific community (e.g., Cousijn et al. 2019). Initial efforts have been carried out at GES DISC to add dataset-related publications (NASA GES DISC 2021) to the individual dataset landing page; however, it remains a challenge to include and sort all related refereed and non-refereed publications, as well as publications in which multiple datasets (not only including the mentioned individual dataset) are involved. For very popular datasets, such as Global Precipitation Climatology Project (GPCP), which can have well over 1000 citations, certain proper management capabilities (e.g., for sorting and filtering) may also need to be developed.

4.2. DATA QUALITY ISSUES

Attention to and studies on data and information quality have significantly grown, especially during the past two decades (e.g., Shankaranarayanan and Blake 2017; Ramapriyan et al. 2017; Azeroual et al. 2018). Without data quality being genuinely assessed and/or its information being timely facilitated to the user community, it would be difficult for users to select and use the right dataset in research and applications (Ramapriyan et al. 2017). This should be another important area for metrics to address or tackle. Data quality consists of two main components: 1) quality information itself generated by either dataset or service providers, and 2) dissemination of the respective quality information to the users. Ramapriyan et al. (2017) introduced and defined four components of information quality (scientific, product, stewardship, and service) aiming to promote consistent quality assessment and quality information on data products and services for the Earth science community. However, concerning how to timely provide and effectively facilitate such comprehensive information to the user community, there still remain many obstacles to overcome.

For example, to obtain reliable scientific data quality information for satellite-based global datasets could be difficult. Researchers or application workers (e.g., flood forecast, crop monitoring) around the world often seek and acquire existing data quality information relying on available (limited) ground validation results from previous data quality assessment studies, such as reports or publications. Conducting ground validation activities on a global scale is a challenge because *in situ* observations are very limited, especially in remote regions and

over vast oceans. For datasets relying on multiple satellite observations, to gain and provide data quality information is even more challenging. Data quality for derived datasets, e.g., from 3-hourly to daily, is not easy to define either.

There are also challenges in developing, implementing, and presenting data quality information in (or along with) datasets. At present, there is no widely implemented community standard for data quality parameters or variables, along with metadata in a dataset. Over the years, the Data Quality Working Group (DQWG), one of NASA's Earth Science Data System Working Groups (ESDSWG), has developed several documents for data quality recommendations addressing different need areas and offering useful information or suggested guidance (Wei et al. 2019; Liu et al. 2019; Ramapriyan et al. 2019a; Ramapriyan et al. 2019b). More efforts (e.g., via education, training, project management, etc.) are needed to implement and continuously evaluate and optimize, as well as eventually disseminate, these recommendations.

4.3. APPLICATION AND RESEARCH METRICS

Data (and/or service) application metrics collected from users' data and/or service usage feedback are useful for data providers and project management. For new or novice users, existing applications can act as examples of using the datasets. Application metrics might be difficult to obtain, however, because they would require users to actively report their activities with some timely detail. Most of the NASA Earth observation datasets are global and many applications (e.g., monitoring crop conditions, landslide prediction, environmental hazard, and disaster management) around the world rely on NASA satellite data for development and operations. However, to collect such application metrics has been a challenge, especially from private industries where such information can be proprietary.

Nonetheless, a new design for comprehensive metrics, especially those being scientific-activity-related (e.g., citation, data quality, application), is needed and it requires all involved parties (e.g., DAACs, stakeholders) to participate and set up the requirements (Bell et al. 2011; National Research Council 2005; Evely et al, 2010; Fazey et al. 2014; Ferguson et al. 2016; Wall et al. 2017). The National Research Council (2005) has developed principles for developing metrics in the climate change science program. One simple and crucial principle (National Research Council 2005, 2014) is that data metrics need to be uncomplicated and easy to understand, along with a few other principles involving funding, leadership, planning, or implementation, respectively. Providing straightforward data metrics is a challenge, which requires many reiterations and collaborations among software developers and stakeholders (e.g., those using the metrics for various purposes) to ensure that data metrics are easy to use and benefit stakeholders' activities.

4.4. METRICS INTEROPERABILITY

Metrics interoperability is an important area, especially when datasets come from various sources or repositories. Without interoperability and the thus-to-be-generated (new) metrics, more time will be spent on data processing (e.g., format conversion), which is less efficient, and it will remain difficult to evaluate the experience and success of multidisciplinary researchers and/or their research collaboration. The EMS standards have been a good example to ensure the interoperability in metrics among DAACs, which is a model for integrating, enhancing, and developing new metrics. On a larger scale (e.g., interagency and international), standardization of metrics is necessary to ensure interoperability. For example, the project, COUNTER (COUNTER 2021), enables publishers and vendors to report standardized and consistent usage of their electronic resources (e.g., data usage).

4.5. USER SURVEY

At present, it is a common practice to passively collect metrics by recording user activities (e.g., website visits, data download, and user inquiry), which may not sufficiently or often reflect user's opinions or provided feedback, e.g., when a user's choices are limited to an individual DAAC that may not provide a vitally needed dataset. GES DISC occasionally receives user feedback or suggestions through the user support service or Bugzilla tickets, but they may contain bias due to limited sample size, essentially the viewpoint of individuals. On the other hand, conducting active user surveys often receives low response rates, which is a major challenge. Short and

focused surveys in different forms may be more effective and receiving better response rates and thus should be considered as alternatives.

4.6. DISSEMINATION OF METRICS

Nowadays, it is imperative to develop and provide an all-in-one dataset landing page (e.g., [Figure 2](#)) to users who look for related services and information in one place. Over the years, especially in recent years, software engineers and scientists at GES DISC have been working to integrate dataset information (e.g., DOI, documents) and services in one place for easy discovery and access, compared to several years ago when information and services were more scattered in different places and difficult to find or remember. More dataset-related items (e.g., FAQs, How-To's, Data-in-Action articles) have been added to the dataset landing page. Likewise, metrics need to be integrated together from different sources and made available in the dataset landing page. Currently, some of these metrics are only available internally and submitted to EMS where metrics from other DAACs are integrated. NASA EMS provides an annual report for the entire data system and each individual DAAC (NASA ESDIS 2021c). However, detailed information about an individual dataset is not available in this annual report. A data provider or a project principal investigator (PI) who develops the dataset algorithm will have to make a special request for the associated dataset metrics to the DAAC where the dataset is archived. Since mid-2016, user registration has been routinely required to download NASA Earth science data, so it should now be more feasible to produce more accurate user usage metrics and make them available in the dataset landing page (e.g., [Figure 2](#)) for the respective data providers or PIs, or even future users, to fetch.

As metrics increase in volume and variety, a Big Data approach that capably integrates different types of metrics from different sources for ensemble analysis is needed to help better understand these metrics and disclose their possible interwoven correlations. Han et al. (2016) conducted a case study to investigate metrics collected from the CEOS (Committee on Earth Observation Satellites) federated catalog service with emphasis on catalog service integration. Their integrated and deployed metrics reporting system provides insightful information for different users such as stakeholders and developers. Google Analytics is another example of providing on-the-fly data analysis and visualization for website metrics, as mentioned earlier. Similarly, newly improved or developed tools should and will be able to integrate metrics from different sources (e.g., servers, data quality, and citation).

5. SUMMARY AND KEY RECOMMENDATIONS

Using NASA GES DISC, a multidisciplinary data center, as an example, we have described current metrics for satellite data products and services. The EMS standards are used at each DAAC to collect dataset metrics, requiring registration of dataset attributes in advance. Each DAAC also participates in an annual ACSI survey of its user community.

The pace of scientific progress and objectives is dynamic and rapid. Meanwhile, technologies continue to evolve. Metrics activities, including definitions, collection, analysis, and visualization, need to keep up with the changes accordingly.

Key recommendations from the discussion are:

1. Current predefined metrics are collected mainly for a single dataset or service. Expansion of metrics is needed to support interdisciplinary activity and on-the-fly data services. Recommendations from community efforts can be leveraged.
2. Data quality metrics are important for research and applications. However, very few metrics or limited quality information is currently available, especially for satellite products with global coverage. More efforts are needed from data product developers and providers.
3. Application and research metrics are an important part of the information for users, data product developers and providers, management, etc. Collecting such information is a challenge, requiring broad participation from product users and publishers as well as a system to report and collect such information.

4. Interoperability for metrics is a challenge. The NASA EMS standards may not be interoperable with other data repositories. When developing metrics, it is important to have the FAIR guiding principles as a guideline to optimize or maximize the use of metrics. Existing efforts (e.g., COUNTER) can also be leveraged.
5. Metrics need to be integrated for forming a holistic view. User-friendly tools for analysis and visualization are also needed.
6. For the Earth science data community as a whole, the challenges are even bigger, requiring all associated stakeholders to collaboratively identify the problems inherent to this scientific endeavor and to work together on common solutions and standards.

COMPUTER CODE AVAILABILITY

No code or software has been developed for this research. Google Analytics 360 was used in collecting and generating GES DISC website custom metrics in [Figure 7](#).

ACKNOWLEDGEMENTS

Special thanks to Elaine Owens and Jerry Shiles at GES DISC for managing metrics data and routine Bugzilla tickets. Thanks are extended to the NASA EMS staff (in particular, Jianfu Pan, Lalit Wanchoo, and Nelson Casiano) for providing comments and clarification about EMS during the manuscript preparation. We thank two reviewers who provided constructive and thought-provoking comments. We acknowledge the staff at GES DISC for development and maintenance of data services. This work and GES DISC were funded by NASA's Science Mission Directorate (SMD).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Z. Liu (ideas, writing, editing, visualization), C.-L. Shie (ideas, co-writing, metrics analysis, editing, visualization), A.J. Ritrivi (Google Analytics, retrieving GES DISC website metrics and producing the web-metrics schematic chart), G.-D. Lei (providing original metrics data for analysis and plots, editing), G.T. Alcott (editing, technical management), Mary Greene (providing Bugzilla metrics data), J. Acker (editing, providing Giovanni publication lists), J.C. Wei, D. Meyer, A. Li and A.F. Al-Jazrawi (technical management).

AUTHOR AFFILIATIONS

Zhong Liu  orcid.org/0000-0001-8150-7556

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), US; George Mason University, US

Chung-Lin Shie  orcid.org/0000-0002-1115-1029

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), US; University of Maryland Baltimore County, US

Anthony J. Ritrivi

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), US; Adnet Systems, Inc., US

Guang-Dih Lei

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), US; Adnet Systems, Inc., US

Gary T. Alcott

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), US

Mary Greene

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), US; Telophase Corporation, US

James Acker

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), US; Adnet Systems, Inc., US

REFERENCES

- ACSI. 2021. The American Customer Satisfaction Index (ACSI). Available online: <https://www.theacsi.org/>. Last accessed, September 29, 2021.
- Azeroual, O, Saake, G and Wastl, J. 2018. Data measurement in research information systems: metrics for the evaluation of data quality. *Scientometrics*, 115: 1271–1290. DOI: <https://doi.org/10.1007/s11192-018-2735-5>
- Behnke, J, Mitchell, A and Ramapriyan, H. 2019. NASA's Earth Observing Data and Information System – Near-Term Challenges. *Data Science Journal*, 18(1): 40. DOI: <https://doi.org/10.5334/dsj-2019-040>
- Bell, S, Shaw, B and Boaz, A. 2011. Real-world approaches to assessing the impact of environmental research on policy. *Res. Eval*, 20: 227–237. DOI: <https://doi.org/10.3152/095820211X13118583635792>
- Bugzilla. 2021. Bugzilla, available online: <https://www.bugzilla.org/>. Last accessed, September 29, 2021.
- CFI Group. 2021. *Claes Fornell International Group*. Available online: <https://cfigroup.com/>. Last accessed, September 29, 2021.
- COUNTER. 2021. *The Code of Practice*. Available online: <https://www.projectcounter.org/>. Last accessed, September 29, 2021.
- Cousijn, H, Feeney, P, Lowenberg, D, Presani, E and Simons, N. 2019. Bringing Citations and Usage Metrics Together to Make Data Count. *Data Science Journal*, 18(1): 9. DOI: <https://doi.org/10.5334/dsj-2019-009>
- Cruse, P, Garza, K, Budden, AE, Chodacki, J, Fenner, M, Jones, MB, Lowenberg, D, Stall, S and Vieglais, D. 2019. *Make Data Count and PARSEC: Two efforts Towards Data Usage Metrics Standardization, the 2019 AGU Fall Meeting*, 9–13 December 2019, San Francisco, CA.
- Evely, AC, Fazey, I, Lambin, X, Lambert, E, Allen, S and Pinard, M. 2010. Defining and evaluating the impact of cross-disciplinary conservation research. *Environ. Conserv.*, 37: 442–450. DOI: <https://doi.org/10.1017/S0376892910000792>
- Fazey, I and Coauthors. 2014. Evaluating knowledge exchange in interdisciplinary and multi-stakeholder research. *Global Environ. Change*, 25: 204–220. DOI: <https://doi.org/10.1016/j.gloenvcha.2013.12.012>
- Ferguson, DB, Finucane, ML, Keener, VW and Owen, G. 2016. Evaluation to advance science policy: Lessons from Pacific RISA and CLIMAS. In: Parris, AS, et al. (eds.), *Climate in Context: Science and Society Partnering for Adaptation*. Wiley, 215–234. DOI: <https://doi.org/10.1002/9781118474785.ch10>
- Han, W, Di, L, Yu, G, Shao, Y and Kang, L. 2016. Investigating metrics of geospatial web services: The case of a CEOS federated catalog service for earth observation data. *Computers & Geosciences*. 92: 1–8. July 2016. DOI: <https://doi.org/10.1016/j.cageo.2016.04.005>
- Kafle, D, Wanchoo, L, Won, Y-I, and Behnke, J. 2019. *NASA EOSDIS Data Usage Metrics – Insight and Assessment, the 2019 AGU Fall Meeting*, 9–13 December 2019, San Francisco, CA. DOI: <https://doi.org/10.1002/essoar.10501904.1>
- Liu, Z, Ramapriyan, HK, Wei, Y, Shie, C-L, Moroni, D, Downs, RR, Habermann, T, Scott, D and Huffman, G. 2019. “High-Priority Data Quality Recommendations for Data Producers and Distributors, Technical Note ESDS-RFC-034.” *NASA ESDIS Standard Office (ESO)*, April 19, 1–17. <https://cdn.earthdata.nasa.gov/conduit/upload/11247/ESDS-RFC-034.pdf>
- Liu, Z, Shie, C-L, Li, A and Meyer, D. 2020. “NASA Global Satellite and Model Data Products and Services for Tropical Meteorology and Climatology.” *Remote Sensing*, 12(17): 2821. DOI: <https://doi.org/10.3390/rs12172821>
- NASA DAACs. 2021. *EOSDIS Distributed Active Archive Centers (DAACs)*. Available online: <https://earthdata.nasa.gov/eosdis/daacs>. Last accessed, September 29, 2021.
- NASA EOS. 2021. *NASA's Earth Observing System Project*. Available online: <https://eosps.nasa.gov/content/nasas-earth-observing-system-project-science-office>. Last accessed, September 29, 2021.
- NASA EOSDIS. 2021a. *Earth Observing System Data and Information System (EOSDIS)*. Available online: <https://earthdata.nasa.gov/eosdis>. Last accessed, September 29, 2021.
- NASA EOSDIS. 2021b. *System Performance and Metrics*. Available online: <https://earthdata.nasa.gov/eosdis/system-performance>. Last accessed, September 29, 2021.
- NASA EOSDIS. 2021c. *EOSDIS Annual Metrics Reports*. Available online: <https://earthdata.nasa.gov/eosdis/system-performance/eosdis-annual-metrics-reports>. Last accessed, September 29, 2021.

- NASA EOSDIS.** 2021d. *American Customer Satisfaction Index (ACSI) Reports*. Available online: <https://earthdata.nasa.gov/eosdis/system-performance/acsi-reports>. Last accessed, September 29, 2021.
- NASA EOSDIS.** 2021e. *NASA Earthdata*. available online: <https://earthdata.nasa.gov/>. Last accessed, September 29, 2021.
- NASA EOSDIS.** 2021f. *Earthdata Cloud Evolution*. Available online: <https://earthdata.nasa.gov/eosdis/cloud-evolution>. Last accessed, September 29, 2021.
- NASA ESDIS.** 2021a. *Earth Science Data and Information System (ESDIS) Project*. Available online: <https://earthdata.nasa.gov/esdis>. Last accessed, September 29, 2021.
- NASA ESDIS.** 2021b. *ESIDS Metrics System (EMS)*. Available online: <https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/ems>. Last accessed, September 29, 2021.
- NASA ESDIS.** 2021c. *Metrics Planning Group (MPG)*. Available online: <https://earthdata.nasa.gov/esdis/mpg>. Last accessed, September 29, 2021.
- NASA GES DISC.** 2021. *NASA Goddard Earth Sciences Data and Information Services Center (GES DISC)*. Available online: <https://disc.gsfc.nasa.gov>. Last accessed, September 29, 2021.
- NASA Giovanni.** 2021. *NASA Giovanni*. Available online: <https://giovanni.gsfc.nasa.gov>. Last accessed, September 29, 2021.
- National Research Council.** 2005. *Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program*. Washington, DC: The National Academies Press. DOI: <https://doi.org/10.17226/11292>
- National Research Council.** 2014. *Enhancing the Value and Sustainability of Field Stations and Marine Laboratories in the 21st Century*. Washington, DC: The National Academies Press. DOI: <https://doi.org/10.17226/18806>
- O'Brien, M, Parr, C and Gries, C.** 2020. *Value Metrics for Data Repositories in Earth and Environmental Sciences, the 2020 AGU Fall Meeting, 1–17 December 2020, Online Everywhere*.
- Parr, C, Gries, C, O'Brien, M, Downs, RR, Duerr, R, Koskela, R, Tarrant, P, Maull, KE, Hoebelheinrich, N and Stall, S.** 2019. A Discussion of Value Metrics for Data Repositories in Earth and Environmental Sciences. *Data Science Journal*, 18(1): 58. DOI: <https://doi.org/10.5334/dsj-2019-058>
- Ramapriyan, HK, Peng, G, Moroni, D and Shie, C-L.** 2017. "Ensuring and Improving Information Quality for Earth Science Data and Products". *D-Lib Magazine*, July/August 2017. DOI: <https://doi.org/10.1045/july2017-ramapriyan>
- Ramapriyan, HK, "Rama", Scott, D, Armstrong, E, DiMiceli, C, Downs, RR, Gacke, C, Gluck, S, Huffman, G, Liu, Z, Moroni, D, Shie, C-L, Smith, D and Wei, Y.** 2019b. "Data Quality Working Group Recommendations for the Data Management Plan Template for Data Producers, Technical Note ESDS-RFC-032. *NASA ESDIS Standard Office (ESO)*, February 2019, 1–7. <https://cdn.earthdata.nasa.gov/conduit/upload/10744/ESDS-RFC-032v1.pdf>.
- Ramapriyan, HK, "Rama", Scott, D, Armstrong, E, DiMiceli, C, Downs, RR, Gacke, C, Gluck, S, Huffman, G, Liu, Z, Moroni, D, Shie, C-L and Wei, Y.** 2019a. "Data Management Plan Template for DAACs, Technical Note ESDS-RFC-031." *NASA ESDIS Standard Office (ESO)*, February 2019, 1–14. <https://cdn.earthdata.nasa.gov/conduit/upload/10743/ESDS-RFC-031v1.pdf>.
- Ramapriyan, HK, "Rama" and Leonard, PJT.** 2020. *Data Product Development Guide (DPDG) for Data Producers version1*. NASA Earth Science Data and Information System Standards Office, 9 July 2020. DOI: <https://doi.org/10.5067/DOC/ESO/RFC-041VERSION1>
- Rauber, A, Asmi, A, van Uytvanck, D and Proell, S.** 2015. *Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC)*. DOI: <https://doi.org/10.15497/RDA00016>
- Shankaranarayanan, G and Blake, R.** 2017. From content to context: The evolution and growth of data quality research. *Journal of Data and Information Quality (JDIQ)* 8(2), January 2017. DOI: <https://doi.org/10.1145/2996198>
- Shie, C-L, Ritrivi, AJ, Lei, G-D, Greene, M, Acker, J, Alcott, GT, Li, A, Wei, JC, F, Al-Jazrawi, A and Meyer, DJ.** 2019: *Integrated Analysis of Multiple User Metrics – A "Sequel"; and Introducing the Google Analytic (eLightning), the 2019 AGU Fall Meeting, 9–13 December 2019, San Francisco, CA*. eLightning presentation available at <https://agu2019fallmeeting-agu.ipostersessions.com/default.aspx?s=73-69-FA-06-81-63-5A-47-D7-5C-4E-60-C3-14-4F-62>.
- Su, J, KC, B, Loeser, C, Rui, H, Shen, S, Lei, G and Ostrenga, D.** 2019. *Metrics Learning at NASA GES DISC, the 2019 AGU Fall Meeting, 9–13 December 2019, San Francisco, CA*.
- Wall, TU, Meadow, AM and Horganic, A.** 2017. Developing Evaluation Indicators to Improve the Process of Coproducing Usable Climate Science. *Wea. Climate Soc.* 9: 95–107. DOI: <https://doi.org/10.1175/WCAS-D-16-0008.1>
- Wei, Y, Ramapriyan, HK, Downs, RR, Shie, C-L, Liu, Z, Moroni, D, Habermann, T, Khalsa, SJ and Peters, B.** 2019. "Data Quality Working Group's Comprehensive Recommendations for Data Producers and Distributors, Technical Note ESDS-RFC-033." *NASA ESDIS Standard Office (ESO)*, August 27, 1–81. <https://cdn.earthdata.nasa.gov/conduit/upload/12101/ESDS-RFC-033.pdf>.
- Wilkinson, M, Dumontier, M and Aalbersberg, I, et al.** 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3: 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>

TO CITE THIS ARTICLE:

Liu, Z, Shie, C-L, Ritrivi, AJ, Lei, G-D, Alcott, GT, Greene, M, Acker, J, Wei, JC, Meyer, DJ, Li, A, Al-Jazrawi, AF. 2022. Developing Metrics for NASA Earth Science Interdisciplinary Data Products and Services . *Data Science Journal*, 21: 5, pp. 1–15. DOI: <https://doi.org/10.5334/dsj-2022-005>

Submitted: 27 August 2021

Accepted: 24 January 2022

Published: 11 February 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.