

PRACTICE PAPER

NASA's Earth Observing Data and Information System – Near-Term Challenges

Jeanne Behnke¹, Andrew Mitchell¹ and Hampapuram Ramapriyan^{1,2}¹ NASA Goddard Space Flight Center, US² Science Systems and Applications, Inc., USCorresponding author: Hampapuram Ramapriyan (Hampapuram.Ramapriyan@ssaihq.com)

NASA's Earth Observing System Data and Information System (EOSDIS) has been a central component of the NASA Earth observation program since the 1990's. EOSDIS manages data covering a wide range of Earth science disciplines including cryosphere, land cover change, polar processes, field campaigns, ocean surface, digital elevation, atmospheric dynamics and composition, and inter-disciplinary research, and many others. One of the key components of EOSDIS is a set of twelve discipline-based Distributed Active Archive Centers (DAACs) distributed across the United States. Managed by NASA's Earth Science Data and Information System (ESDIS) Project at the Goddard Space Flight Center, these DAACs serve over 4 million users globally. The ESDIS Project provides the infrastructure support for EOSDIS, which includes other components such as common metadata and metrics management systems, specialized network systems, standards management, and centralized support for use of commercial cloud capabilities. Given the long-term requirements, and the rapid pace of information technology and changing expectations of the user community, EOSDIS has evolved continually over the past three decades. However, many challenges remain. Challenges in three key areas are addressed in this paper: managing volume and variety, enabling data discovery and access, and incorporating user feedback and concerns.

Keywords: Data systems; Earth science; Remote sensing; Big data; Data discovery; Data access; Data preservation; Metadata

Introduction

NASA's Earth Observing System (EOS) Data and Information System (EOSDIS) has been a central component of the NASA Earth observation program since the 1990's. The data collected by NASA represent a significant public investment in research. During the 1990s, a common set of data exchange and access principles was created by the Canadian, Japanese, European and U.S. International Earth Observing System (IEOS) partners (IEOS 1996). Consistent with these international principles, NASA developed its free, open and non-discriminatory data and information policy. NASA's policy aims at maximizing access to data. "In this context, the term 'data' includes observation data, metadata, products, algorithms including scientific source code, documentation, models, images, and research results" (NASA 1999). An updated version of the policy can be found at (NASA 2019a). EOSDIS manages data covering a wide range of Earth science disciplines. The data managed by EOSDIS include observations from instruments on board satellites and aircraft, and field campaigns, as well as derived products. EOSDIS is comprised of partnerships among NASA Centers, other US agencies and academia that process and disseminate remote sensing and in situ Earth science data. One of the key components of EOSDIS is a set of twelve discipline-based Distributed Active Archive Centers (DAACs). Because of their active role in NASA mission science and with the science community, they perform many tasks beyond basic data stewardship, representing a distinct departure from typical data archives. For example, they interact closely with the flight projects in support of pre-launch end-to-end testing of satellite data flows, work with science teams responsible for data product generation in defining metadata, and establish levels of service for the diverse sets of data that they archive and distribute. They are

collocated with scientific expertise in their respective Earth science disciplines. Managed by NASA's Earth Science Data and Information System (ESDIS) Project at the Goddard Space Flight Center and distributed across the United States, these DAACs serve over 4 million users globally. The ESDIS Project provides the infrastructure for EOSDIS including other components such as common metadata and metrics management systems, specialized network and security systems, standards management, and centralized support for use of commercial cloud capabilities. A distributed set of Science Investigator-led Processing Systems produce derived digital data products and deliver them to the DAACs for archival and distribution. During the nearly 30 years since the beginning of the design of EOSDIS, there have been significant developments in information technology and consequently in the expectations of the user community. As a result, the ESDIS Project has had to evolve EOSDIS continually as a system of systems to address many challenges. There are several other systems available to the community such as the Australian National Data Service (ANDS 2019), European Space Agency (Albani 2018, ESA 2019), US National Oceanic and Atmospheric Administration's (NOAA) National Centers for Environmental Information (NCEI 2019), and the US Geological Survey (USGS 2019) that provide significant volumes and variety of Earth observation to data and face similar challenges. The purpose of this paper is to describe some of the key challenges that the ESDIS project has observed during its nearly three decades of experience and the approaches being taken to address them. In the era of big data, it is important to consider emergent issues in light of the ongoing challenges that science archives like EOSDIS have addressed and are continuing to address. The details of the overall architecture of EOSDIS from the enterprise, information and service viewpoints can be found at ESDIS (2017a).

Challenges

As a long-lived system that manages data from many diverse sources and serves a multi-disciplinary user community, EOSDIS faces many challenges. These challenges can be grouped into three main categories: 1. Managing volume and variety; 2. Enabling data discovery and access; and 3. Incorporating user feedback and concerns. These challenges are discussed in the three subsections below.

Managing volume and variety

When EOSDIS was conceived in the 1990s, it was understood that it would always be a growing collection of Earth science datasets. It started with very small collections that NASA had funded at various locations, which became the DAACs of the EOSDIS. The NASA EOS program was planned to consist of several multi-instrument platforms that would collect data continuously. From the management and funding perspective, it made sense to have a single system to manage multi-mission operations, as opposed to the old model of creating a new processing/archiving system with each mission. Since its inception, EOSDIS has added new missions to the Earth science collection expanding the variety and volume every year. With each orbit, instruments continue to acquire sensor data adding to the collection. In addition, the data grow as scientists improve the measurements from the instruments deriving new parameters and products. Data formats that were chosen at launch must adapt to meet new standards and feed new software applications reliant on improved metadata. In the 1980s, the Global Change Master Directory (GCMD) worked on metadata issues at the dataset level, eventually developing a rich directory of Earth Science data from all over the world. Starting in the 1990s, staff at the ESDIS Project had a difficult time convincing scientist data providers of the value of metadata at the file level – assuring them that the most metadata they would have to insert into the complex Hierarchical Data Format (HDF) would be no more than 18 individual fields. Today, metadata is a ubiquitous word – everyone understands the value of it and the EOSDIS metadata model has grown to cover not only data, but services as well. Over the years, various standards have been used by data providers to supply metadata to the EOSDIS metadata repository. The standards include: The GCMD Directory Interchange Format (DIF) for collection level metadata, the EOS Clearing House (ECHO) format for collection and granule (i.e., file) level metadata, GCMD's Service Entry Resource Format (SERF) for service metadata, ISO 19115-1 (ISO 2014) and ISO 19115-2 (ISO 2019a). Recently, a Unified Metadata Model (UMM) has been developed (ESDIS 2015). Rather than adopting a single metadata standard, the UMM permits the continued use of those used previously while providing for easily translating from one supported standard to another. The ESDIS Standards Office (ESO 2019) provides a list of currently accepted standards for data and metadata for NASA's Earth science data. It also provides a mechanism for reviewing, approving and adopting new standards. The standards ISO 19115-1 and ISO 19115-2 are now placed as a requirement by NASA for metadata, and guidance is provided for their implementation including other relevant international standards (ESDIS 2018). In addition, the Dataset Interoperability Working Group, one of NASA's Earth Science Data System Working Groups (ESDSWG), provides recommendations (e.g., compliance with the Climate and

Forecast (CF) Conventions) (Eaton *et al.* 2017) for data producers to follow to improve interoperability of NASA datasets among themselves and with those from other organizations, thus promoting data use by a broad user community (NASA 2019b).

At the end of 2018, EOSDIS had over 420 million granules identified in its repository from over 7,000 data collections. The access to the archive and the metadata repository for loading and deleting data is strictly controlled. The EOSDIS Common Metadata Repository software carefully manages and maintains the repository, but open source versions of the software are available along with programming interfaces that allow anyone to access the repository. The ESDIS Project provides an infrastructural software system, Earthdata Search, as a user interface to the repository and reference client that can help users build their own interfaces. Because of the diversification by discipline, the workload in maintaining the EOSDIS collection is shared by the DAACs. Vigilance is still needed as inconsistencies in the collected information become more readily apparent in Earthdata Search and other user interface applications. For example, the use of keywords to describe rain (rainfall, precipitation, etc.) must either be uniform or must have associated terms in our metadata thesaurus to allow complete searching of the archive. The GCMD Keyword Management System is used for managing keywords, and uses the Simple Knowledge Organization System (SKOS) concepts (GCMD 2017). One way the ESDIS Project manages inconsistencies is to establish an independent review committee composed of metadata professionals. These professionals have recently completed a two-year task of scrubbing through the DAAC collections, focusing on metadata, to create targeted reports of errors, misspellings, inconsistencies, etc. (IMPACT 2019) This task has brought many improvements to the user experience of searching through the EOSDIS data collection. Because of its success, the ESDIS Project has extended the task to include new and continuing efforts to improve the EOSDIS metadata.

As has been the case over the history of EOSDIS, the diversity of science data producers contributing to the variety and volume of the collection continues to present challenges. Physical storage and hardware challenges are always expected as the collection grows. However, the challenge presented by the diversity of data producers is inescapable. Although we have required standards for data and metadata, like the HDF and ISO19115 (ESDIS 2017b), we do not precisely control the way the standard is implemented by a particular science instrument team or SIPS. This challenge means that with each organization-wide system change, whether for new versions or transferring to new technology, each dataset must be handled individually. In the early days of EOSDIS, the DAACs were able to develop very close relationships with their data providers and would be able to help in the development of datasets. With the addition of new missions, the DAACs are more thinly spread and individualized support is more difficult. The ESDIS Project is working toward developing software and guides to help data providers choose keywords, select more common quality flags, and other variables. Today, such software and guides are still very discipline specific, which is useful for the individual disciplines, but need some modifications to support multi-disciplinary studies.

An additional recent challenge is that several of the original EOS instruments are at the end of their life. Prior to the EOSDIS, NASA scientists would write data directly to tape media and the tapes would be racked and users could request the data from the archive. The archive might make a copy or send the original but it would be up to the users to determine how to read the tapes. Today all data are online and always available, and the data are easier to access, maintain and re-version. During the active life of an instrument, the various versions of data resulting from initial processing and reprocessing are kept track of rigorously by assigning new Digital Object Identifiers (DOI) to each new version. A consequence of easy access is that it enables data providers to continue to adjust and improve the datasets even after the instrument has been decommissioned. Thus, EOSDIS may no longer have a "final" version of data as was done in the past. Heritage datasets will always need to be maintained at the DAACs and the ESDIS Project plans ahead to keep them updated to the latest data and metadata standards. However, in the case of missions completed several years ago for which the planned support from NASA has stopped (other than for updates for standard compliance), if members of the scientific community generate new versions of an entire mission's dataset with improved algorithms, the challenge arises as to how to accommodate them at the DAACs. Such datasets may not have undergone the review processes as the original mission datasets had, and no funds would have been allocated to their management. An example of this is the ICESat (2003–2010) mission data where researchers have reprocessed the entire mission dataset with better calibration data. To accommodate such cases, the ESDIS Project has developed review procedures for deciding whether the older versions should be replaced or augmented and be distributed by the archive to users.

EOSDIS, like many other space data archives, is looking at the use of the commercial cloud as the next avenue for data storage and services. Using cloud systems, instead of managing in-house hardware systems at the DAACs, is very appealing because of the potential for improved services to users without a

costly infrastructure improvement. Several prototyping tasks have been undertaken to gauge the effort of managing data in commercial cloud structures (McInerney 2017). The chief effort is to build an infrastructure that allows all of the EOSDIS components to work in a controlled fashion on the various cloud platforms so that resources are appropriately scaled and costs kept under control. One challenge is the effort to make certain that we understand the security aspects associated with the use of a commercial system and to meet this challenge. A commercial system, such as the Amazon Web Services cloud, allows its customers to perform all types of computing working within its structures. Given this paradigm, we are carefully considering how to secure the EOSDIS systems from vulnerabilities, threats and attacks that focus on our cloud structures along with honest mistakes made as we move into this new, complex environment. As part of cloud system security planning, we are examining how to manage analytics, the use of configuration management, and how to incorporate tools like GuardDuty and CloudTrail in our development and operations environment (Amazon 2019a, b). The attention to security culminates in implementation of tools and processes that will secure the EOSDIS environment in the AWS Cloud. We also keep a risk register to identify risks and mitigation approaches. Risks include vendor lock-in, demand for Cloud skills in the workforce, complex financial reporting requirements, operational uncertainties and so on. By examining the risks weekly, we insure that we are focusing our development efforts in the right direction. Another issue is the use of various networks to access the cloud systems. Improper use of the networks could increase the cost of cloud use significantly, especially for somewhat unpredictable volumes of data downloaded by the user community. Throttling mechanisms need to be used to control these costs. Several test efforts are ongoing to document various traffic patterns and usage. In addition, we have prototyped the development of a system, based on existing processes, for ingest and archive of data within the cloud environment. This system is undergoing functional and performance tests to work out the many issues that have been encountered. However, one of the greatest challenges will be managing the overall cost of using the cloud by the various components of EOSDIS. Developing the processes for such management is ongoing and proving to be problematic but not insurmountable. We expect that by using the cloud as a platform, the advantages to the user community will be myriad. Examples of advantages are: use of only as much as needed of compute resources while the load fluctuates – as in the case of reprocessing of data products over an entire mission from its beginning to the current date; access to computing on petabytes of data close to where the data are stored obviating the need for large data transfers; and ease of preparation for high-data-rate missions. Researchers will be able to gain new insights into the data and users will enable new applications, which indeed is the ultimate goal for the big data era.

Enabling Data Discovery and Access

A continuing challenge is to provide users with just the data they need. Typically users search for data using keywords as well as spatial and temporal constraints. In EOSDIS, with thousands of datasets, typical queries from users may result in hundreds of hits meeting their criteria. Ensuring that the most relevant of the datasets appear first in the results list is crucial to users. The obvious steps one can take towards increasing search relevance are ranking the datasets based on spatial and temporal criteria. Also, ranking newer versions higher, and applying information about community usage of datasets (e.g., through automated analyses of scientific literature) for ranking are useful steps. Observation of the real usage of the Earthdata search capability in EOSDIS and characterizing the search and access will also help in continuous improvements to data discovery.

In the case of data that can be represented as images, it is beneficial for users to be able to visualize them and select the data that they want to download and analyze. Enabling this for large volume datasets is a challenge that the ESDIS Project has successfully addressed through its Global Imagery Browse System (GIBS) and the WorldView client software (Murphy *et al.* 2015). The GIBS consists of a database of images stored in a hierarchical manner to enable rapid access to data at multiple resolutions. The WorldView client takes advantage of this data structure and enables users converge within a few seconds on their area of interest at the highest resolution offered by the dataset.

The access to data by users has changed significantly over the last two decades. In the 2000's, EOSDIS data were stored in robotic tape silos. Users would discover what they needed and place online orders for data from the respective DAACs. The DAACs would copy data to media and mail them to the users or stage the data on disk and email users so that they could download the data. With the move starting in 2006 to online storage, users now select data granules (files) that meet their search criteria and are provided with URLs, which they can use to download the granules. The online storage also has enabled the users to request services such as subsetting, reformatting and reprojection conveniently prior to downloading the

data. However, as the volume of data is expected to increase significantly in the near future, new challenges arise for the user. In the 2020s, the new missions such as Surface Water and Ocean Topography (SWOT) and NASA-ISRO Synthetic Aperture Radar (NISAR) will make individual files in the multi-gigabyte range. SWOT is planned to be launched in 2021 and is expected to generate nearly 20 terabytes per day of data. NISAR is planned to be launched in 2022 and is expected to generate over 72 terabytes per day of data. While files can presently be downloaded efficiently through the network, the new missions will make this impossible.

Providing to users data that are ready for ingest into algorithms and for analysis saves them considerable amount of traditional preparatory work, such as downloading large amounts of data, subsetting, reprojection, mosaicking, etc. This idea of “analysis-ready data” is becoming more popular recently, especially with respect to Landsat data (Dwyer *et al.* 2018). There are several “Data Cube” projects, either implemented or in progress (e.g., Reid 2018; Adrup 2018; Giuliani *et al.* 2018; Lewis *et al.* 2017). Of course, to extend this idea to all the Earth science disciplines is a challenge due to the differences in the way different science disciplines deal with data. We have only started examination of analysis-ready data concepts in the context of EOSDIS through the ESDSWG (NASA 2019b). A step in this direction is the preparation of visualization-ready data in the EOSDIS Global Image Browse System (GIBS) mentioned above (Murphy *et al.* 2015). While not quite analysis-ready, nearly 800 different types of data that can be represented as images are available for full-resolution visualization. Defining analysis-ready data for different disciplines and preparing the data to meet their diverse needs would take significant effort, especially in a well-established system such as EOSDIS with hundreds of millions of data files requiring reorganization. The next step is to carefully evaluate typical use cases in different disciplines and prioritize implementation efforts. Also, the large and increasing volumes of data make it impractical for users to download them into their own systems for analysis. Near-archive analysis capabilities, as in the case of archiving data in a commercial cloud environment, will alleviate this problem significantly. The challenges of security and managing costs in the cloud environment are real, and are being addressed as described earlier.

Another challenge in this area is ensuring access to data decades into the future. Increased attention to data preservation and future access is evidenced in a series of conferences on “adding value and preserving data” (see Conway and Winfield 2018 for the latest). The data and derived products from NASA's missions are a valuable asset resulting in many important scientific discoveries and influential findings. Therefore, they need to be preserved so that future users are able to discover, access, read, understand and reuse them. Future users should be able to verify, reproduce or question the science as necessary without having access to the science teams that produced the products. The contents needed to be preserved with the data can be referred to as associated knowledge. The ESDIS Project has developed a “Preservation Content Specification (PCS)” that identifies the classes of content that need to be preserved (NASA, 2011). Similar efforts have been documented by the European Space Agency (Albani *et al.* 2018) and the Committee on Earth Observing Satellites (CEOS) Working Group on Information systems and Services (WGISS) (CEOS, 2015). Since then, NASA has adopted PCS as a requirement for its missions, and we have educated our components, DAACs and SIPS, on preservation of content, especially with regard to instruments at end-of-life. It is important that the broader community also consider long-term preservation and accessibility. Therefore, we have worked to encourage and participate in the development of international standards for preservation of data and metadata (ISO 2019b).

Incorporating User Feedback and Concerns

As a system that serves a diverse global community of over 4 million users, EOSDIS receives feedback from them in several different ways. Responding to the diversity of the feedback is a challenge. Users can provide direct feedback including suggestions, problem reports or questions on the webpage <http://earthdata.nasa.gov>. Each of the DAACs has a user services team responsible for analyzing applicable user feedback and responding to requests for help. Also, each of the DAACs has a user working group (UWG) consisting of science and applications users representing the DAAC's specific discipline(s). The UWGs meet periodically with the ESDIS Project and DAAC staff members to review the data holdings, tools and services offered by the DAACs and provide advice on priorities and future plans. The ESDIS Project employs an independent company to conduct an annual survey of users to derive the “American Customer Satisfaction Index (ACSI)”. While the ACSI is a number indicating how satisfied the users are, the survey also includes several questions for which users provide free-form answers. The ESDIS Project and DAACs analyze these answers for suggestions for system improvement. In addition, focused efforts have been made within the ESDSWG for user needs assessment. More details on the influence of user feedback on EOSDIS and how it affects implementation of its capabilities can be found in Ramapriyan and Behnke (2019).

A related challenge is a concern by users regarding privacy while the system requires them to be registered in order to obtain most of the data and services. As a system that manages data from NASA as well as other non-U.S. partners, EOSDIS must comply with different rules regarding access restrictions and privacy policies regarding collection of information about data users. NASA has had a free and open data policy for Earth science data since the beginning of the EOS Program in 1990. However, working under agreements for archiving and distributing data from international partners, NASA complied with their more restrictive policies regarding charging for data and requiring users to be registered and authorized to obtain the partners' data. For example, until April 1, 2016, NASA had to charge for data from Japan's ASTER instrument that was on NASA's Terra satellite. The distribution of data from JERS-1 requires a restricted data use agreement (ASF DAAC 2019). Until 2012 NASA did not have a registration system for users to access the data from NASA missions, because it was felt unnecessary given the free and open data policy. However, NASA's "earthdata login" is now used for registering users, but with minimal information needed for registration so that more accurate metrics are collected about numbers and organizations of users (ESDIS 2019), and users can be contacted about new datasets and features offered by EOSDIS. The need for better metrics and services to users is balanced relative to privacy rules.

Conclusions

As a long-lived data system, EOSDIS has faced a number of challenges over the past three decades since its development was started. In this paper, we have grouped these into three categories: 1. Managing volume and variety; 2. Enabling data discovery and access; and 3. Incorporating user feedback and concerns. The following three paragraphs provide a summary and conclusions about each of these categories of challenges.

In the early days of EOSDIS, the scientific community responsible for the data products was reluctant to accept the need for comprehensive metadata to ensure the discovery, access and understanding of their products by users. Also, given the diversity of disciplines dealt with by the EOSDIS DAACs there were several "standards" used for providing metadata. Today the importance of metadata is well understood. The challenge of multiple standards has been dealt with by developing a Unified Metadata Model (UMM) providing for easy mapping from one standard to another without needing to go through an expensive conversion of older metadata into the recent ISO standards. A Common Metadata Repository (CMR) has been developed to manage a database of over 420 million individually addressed granules (files). Open source versions of the CMR software are available along with programming interfaces that allow anyone to access the repository. Keywords, important for proper discovery of data, are managed by the GCMD Keyword Management System, which uses the Simple Knowledge Organization System (SKOS) concepts. The ESDIS Project has used an independent review committee composed of metadata professionals to ensure that inconsistencies in metadata are detected and corrected. Like many other space data archives, EOSDIS is proceeding towards the use of the commercial cloud for data storage and services. Examples of advantages of the cloud are: use of only as much as needed of compute resources while the load fluctuates – as in the case of reprocessing of data products over an entire mission from its beginning to the current date; access to computing on petabytes of data close to where the data are stored obviating the need for large data transfers; and ease of preparation for high-data-rate missions such as SWOT and NISAR expected to be launched in the early 2020s.

Migration of most of the EOSDIS archives to online storage since 2006 has provided easier and faster access to data and also has enabled the users to request services such as subsetting, reformatting and reprojection conveniently prior to downloading. However, with very high data rates (and volumes) expected from the upcoming missions, new challenges arise, which might be best addressed by providing near-archive computation resulting in reduction in network loading. We have started examination of analysis-ready data concepts. Several efforts are in progress by other groups in this area, especially with respect to remotely sensed land imaging. EOSDIS provides "visualization-ready" data through its Global Image Browse System (GIBS) for nearly 800 different types of data that can be represented as images. However, defining analysis-ready data for different disciplines and preparing the data to meet their diverse needs would take significant effort. Ensuring access to data decades into the future is a challenge that is being met by developing a preservation content specification for use by NASA's earth observing missions, as well as participating in the development of international standards for preservation of data and metadata.

There are several ways the ESDIS Project receives feedback on EOSDIS from the user community. These are discussed in detail in Ramapriyan and Behnke (2019). One of the mechanisms for feedback is an annual survey of users to derive the "American Customer Satisfaction Index (ACSI)". While the ACSI is a number indicating how satisfied the users are, the survey also includes several questions for which users provide

free-form answers. The free-form answers provide important information about user needs and assist in planning on improvements to the system. Given NASA's free and open data policy, until 2012, it was felt that there was no need for keeping track of users through a registration process. This led to not being able to obtain accurate metrics and the inability to contact users to new datasets or services offered by EOSDIS that were relevant to specific user groups. A registration system has been in use since 2012 with minimal information being collected from users. The need for better metrics and services to users is balanced relative to privacy rules.

Acknowledgements

Jeanne Behnke and Andrew Mitchell contributed to this paper as a part of their duties as employees of NASA. Hampapuram Ramapriyan was supported by NASA contract NNG15HQ01C with Science Systems and Applications, Inc. Mention of names of companies or software packages does not imply their endorsement by NASA or the authors. This work has been placed in the public domain, via the CC BY 4.0 license. A NASA approval process permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Jeanne Behnke improved on the initial draft of the paper developed by Hampapuram Ramapriyan. Both of them identified the challenges documented in this paper, based on their many years of experience in NASA's Earth Science Data and Information System (ESDIS) Project. Andrew Mitchell, the ESDIS Project Manager, reviewed the drafts and suggested improvements to some of the sections.

References


- Adrup, J.** 2018. Pioneering Steps towards Use of Data-cubes in the Global Earth Observation System of Systems. In Conway and Winfield (eds.), *Proceedings of the 2018 conference on adding value and preserving data (PV2018)*, 48–50. ISSN 1362-0231.
- Albani, M,** et al. 2018. ESA Data Preservation System. In Conway and Winfield (eds.), *Proceedings of the 2018 conference on adding value and preserving data (PV2018)*, 48–50. ISSN 1362-0231.
- Amazon.** 2019a. Amazon GuardDuty. Available at <https://aws.amazon.com/guardduty/> (last accessed July 3, 2019).
- Amazon.** 2019b. AWS CloudTrail. Available at <https://aws.amazon.com/cloudtrail/> (last accessed July 3, 2019).
- ANDS.** 2019. Australian National Data Service. Available at <https://www.ands.org.au/about-us> (last accessed July 1, 2019).
- ASF DAAC.** 2019. JERS-1 Data Access Request. Available at <https://www.asf.alaska.edu/get-data/alaska-satellite-facility-daac-restricted-data-access-request/> (last accessed June 26, 2019).
- CEOS.** 2015. Earth Observation Preserved Data Set Content (PDSC). Available at http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content_v1.0.pdf (last accessed April 26, 2018).
- Conway, E and Winfield, K** (eds.). 2018. PV2018: Proceedings of the 2018 conference on adding value and preserving data, Harwell, UK, 15th–17th May, 2018. ISSN 2362-0231.
- Dwyer, JL,** et al. 2018. Analysis ready data—Enabling analysis of the Landsat archive. *Remote Sensing*, 10(9): 1363. DOI: <https://doi.org/10.20944/preprints201808.0029.v1>
- Eaton, B,** et al. 2017. NetCDF Climate and Forecast (CF) Metadata Conventions. Available at <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html> (last accessed July 3, 2019).
- ESA.** 2019. ESA Earth Online. Available at <https://earth.esa.int/web/guest/home> (last accessed July 1, 2019).
- ESDIS.** 2015. Unified Metadata Model (UMM). Available at <https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr/umm>, latest update June 5, 2019 (last accessed June 26, 2019).
- ESDIS.** 2017a. EOSDIS Handbook, March 2017. Available at <https://earthdata.nasa.gov/files/EOSDIS%20Handbook%201.3.pdf> (last accessed July 8, 2019).
- ESDIS.** 2017b. Standards, Requirements and References. Available at <https://earthdata.nasa.gov/esdis/eso/standards-and-references> (last accessed April 26, 2018).

- ESDIS.** 2018. ISO 19115 Geographic Metadata Standard – Implementation Requirement and Guidance. Available at https://earthdata.nasa.gov/files/423-ESO-035_ISO_19115_Implementation_Guidance_Original.pdf (last accessed June 26, 2019).
- ESDIS.** 2019. ESDIS Metrics System (EMS). Available at <https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/ems>, last updated June 5, 2019 (last accessed June 27, 2019).
- ESO.** 2019. ESDIS Standards Office (ESO). Available at <https://earthdata.nasa.gov/esdis/eso>, last updated June 5, 2019 (last accessed June 26, 2019).
- GCMD.** 2017. Keyword Management Service Application Program Interface. Available at <https://wiki.earthdata.nasa.gov/display/gcmdkey/Keyword+Management+Service+Application+Program+Interface>, last updated December 14, 2017 (last accessed June 27, 2019).
- Giuliani, G,** et al. 2018. Towards Sentinel-2 Analysis Ready Data: a Swiss Data Cube Perspective. *IGARSS 2018 – 2018 IEEE International Geoscience and Remote Sensing Symposium*. DOI: <https://doi.org/10.1109/IGARSS.2018.8517954>
- IEOS.** 1996. IEOS Data Exchange Principles. Available at http://www.codata.info/data_access/policies.html#IEOS (last accessed June 26, 2019).
- IMPACT.** 2019. Analysis and Review of CMR (ARC). Available at <https://earthdata.nasa.gov/esds/impact/arc>, last updated June 6, 2019 (last accessed June 27, 2019).
- ISO.** 2014. 19115-1:2014 Geographic information – Metadata – Part 1: Fundamentals. Geneva: International Organization for Standardization (ISO), 2014.
- ISO.** 2019a. 19115-2:2019 Geographic information – Metadata – Part 2: Extensions for imagery and gridded data. Geneva: ISO, 2009.
- ISO.** 2019b. 19165-2 Geographic information – Preservation of digital data and metadata – Part 2: Content specifications for earth observation data and derived digital products (under development). <https://www.iso.org/standard/73810.html> (last accessed June 28, 2019).
- Lewis, A,** et al. 2017. The Australian Geoscience Data Cube – Foundations and lessons learned. *Remote Sensing of Environment*, 202: 276–292. 1 December 2017. DOI: <https://doi.org/10.1016/j.rse.2017.03.015>
- McInerney, M.** 2017. EOSDIS Cloud Evolution. Available at <https://earthdata.nasa.gov/about/eosdis-cloud-evolution>, last updated August 2, 2019 (last accessed August 5, 2019).
- Murphy, KJ,** et al. 2015. LANCE, NASA's Land, Atmosphere Near Real-Time Capability for EOS. *Time-Sensitive Remote Sensing*. New York, NY: Springer, 2015. DOI: https://doi.org/10.1007/978-1-4939-2602-2_8
- NASA.** 1999. Data and Information Policy: NASA Earth Science Enterprise Statement on Data Management. *1999 EOS Reference Handbook*. Available at <https://eosps.nasa.gov/sites/default/files/publications/1999ReferenceHandbook.pdf> (last accessed June 26, 2019).
- NASA.** 2011. NASA ES Data Preservation Content Spec (423-SPEC-001, Nov 2011). Available at <https://earthdata.nasa.gov/esdis/eso/standards-and-references/preservation-content-spec> (last accessed August 5, 2019).
- NASA.** 2019a. Data and Information Policy, last updated June 18, 2019. Available at <https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy> (last accessed June 26, 2019).
- NASA.** 2019b. Earth Science Data System Working Groups. Available at <https://earthdata.nasa.gov/collaborate/esdswg>, last updated May 5, 2019 (last accessed June 26, 2019).
- NCEI.** 2019. National Centers for Environmental Information. Available at <https://www.ncei.noaa.gov/> (last accessed July 1, 2019).
- Ramapriyan, H** and **Behnke, J.** 2019. Importance and Incorporation of User Feedback in Earth Science Data Stewardship. *Data Science Journal*, 18(24): 1–9. DOI: <https://doi.org/10.5334/dsj-2019-024>
- Reid, S.** 2018. Data Cube as a National Geo-spatial Information System. In Conway and Winfield (eds.), *Proceedings of the 2018 conference on adding value and preserving data (PV2018)*, 51–54. ISSN 1362-0231.
- USGS.** 2019. Data Management. Available at <https://www.usgs.gov/products/data-and-tools/data-management> (last accessed July 1, 2019).

How to cite this article: Behnke, J, Mitchell, A and Ramapriyan, H. 2019. NASA's Earth Observing Data and Information System – Near-Term Challenges. *Data Science Journal*, 18: 40, pp.1–9. DOI: <https://doi.org/10.5334/dsj-2019-040>

Submitted: 24 January 2019 **Accepted:** 01 August 2019 **Published:** 21 August 2019

Copyright: © 2019 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.

OPEN ACCESS 