Drivers and Challenges for NASA’s Earth Science Data Holdings

Kathleen Fontaine[1], Francis Lindsay[2], and Martha Maiden[2]

[1]NASA Goddard Space Flight Center, Code 610.2, Greenbelt, MD 20771
kathy.fontaine@nasa.gov

[2]NASA Headquarters, 300 E Street SW, Washington, DC 20546
flindsay@hq.nasa.gov, martha.e.maiden@nasa.gov

Abstract. NASA’s Earth-Sun System Division has the responsibility for managing the distributed data system assets that contain over four petabytes of Earth science data. The issue of long-term data stewardship is an important piece of the data and information management for the Agency. This paper explores and discusses specific science drivers that influence the planning for on-going data access by science communities and strategies for the eventual long-term archival needs for these data. NASA is guided by a number of policies whose requirements help shape how the Agency faces the many challenges to supporting Earth science research and applied uses of Earth science data and models. We discuss the lessons learned from the development and operation of this very large data system and present the emerging strategies to assure data and data products generated from NASA’s numerous Earth focused assets will remain available for future research.

1 Introduction

Of all national assets, archives are the most precious; they are the gift of one generation to another and the extent of our care of them marks the extent of our civilization.

Sir Arthur G. Doughty
Dominion Archivist, Canada, 1904-1935

The National Aeronautics and Space Administration (NASA) is the US space and aeronautics research agency. Part of its mission is to conduct research into the workings of the Earth and its multiple complex systems, gain insight from these learn enough from that research to develop predictive capabilities in a variety of areas, and then iterate the newest data sets through to fine-tune those capabilities. Inherent in this process is the need for the preservation of not only the well-calibrated and well-validated data sets along with all associated processing algorithms and documentation, but also to capture the Agency's policies and practices to ensure that all the necessary data and information is available when needed, as well as removed when superseded. Balanced against this need is a set of science and policy drivers which, to a large extent, determine both the short and long term direction of NASA’s Earth science program, including how and where data are archived, by whom, and for how long.

There are many drivers influencing NASA, from both internal and external sources. Internal in this context refers to the agency itself, its mission and vision, and its organizational implementation of these objectives. External refers to those drivers originating outside the agency, but which NASA must, nonetheless, be aware of and responsive to, while pursuing its own goals. The challenge for NASA is balancing these competing influences while maintaining the integrity of its science data and the support of its science communities.

2 External Drivers

NASA, as one of many federal agencies, is cognizant of and responsive to a variety of legislation and related national efforts. Some of the more prominent of those to NASA are discussed below.

2.1 Space Act of 1958

Many people associate NASA with human space flight – the Apollo program and lunar landings, the space shuttle and space station, and the future exploration of Mars. But NASA is not solely in the business of human space flight and the exploration of space. The Space Act of 1958 established that among NASA's many objectives is "The expansion of human knowledge of the Earth and of phenomena in the atmosphere and space...". This manifested itself early on in the Earth science realm as a series of land remote sensing spacecraft, beginning with instrumentation on board the Television Infrared Observation Satellite (TIROS) program, and progressing through the initial Landsat period through to later Landsats and on to the EOS era. Each of these spacecraft collected and transmitted back to Earth a variety of data in multiple formats from several sensors, and so began a long journey from these early systems of transmission and storage to the advanced capabilities of present today.

2.2 OMB Policies and Circulars

There were other organizations in the Federal government that are focused on the subject of data collection and dissemination. The Office of Management and Budget (OMB) primarily monitors how other federal agencies spend their money. OMB has in the course of these duties assembled policies on the collecting, retaining, and disseminating data and information gleaned from both public and private sources. OMB determined that Federal agencies must only collect, retain, or disseminate data that is part of their mission; that is available at no more than the cost of reproduction; and that does not violate the privacy of any individual. These limitations are of significance when data providers, such as NASA, attempt to collect metrics on their system usage and performance, and when all US agencies try to form both international and interagency agreements relating to data sharing and/or ownership.

2.3 USGS and NOAA

In addition to OMB policies, practices within other agencies affect NASA. The US Geological Survey (USGS), for example, has had the responsibility for mapping the US and its territories since 1879, and has gained permission to keep receipts (funds) from the sale of its map products. This policy does not mean that the agency can charge consumer excessive sums, rather it means that any money they collect in map sales, they can maintain as a portion of their operating budget. NASA, in contrast, does not have this ability. In putting in place agreements for Landsat data, for example, at one point Landsat 4 and 5 data was available for sale from the USGS, and Landsat 7 data was available for no charge from NASA. USGS is now responsible for distribution of all Landsat data.

The National Oceanic and Atmospheric Administration (NOAA) was formed more recently in 1970 by combining the US Coast and Geodetic Survey, the Weather Bureau, and the Bureau of Commercial Fisheries. These agencies jointly had the responsibility of collecting information that protected US maritime and commerce interests, and so needed to gather real-time data for weather prediction, water temperatures, shipping lane conditions, and the like. These agencies did perform some trend analysis, but this was focused on local conditions (confined to, for example, when the St. Lawrence Seaway tends to ice over each year, etc.), as opposed to regional or global phenomena. NOAA is considered an ‘operational’ agency rather than a ‘research’ agency. This distinction can result in the need for extensive negotiation from time to time, such as the recent talks about putting a Landsat follow-on instrument on an operational NOAA spacecraft. This plan has not gained wide acceptance primarily because the research sensor required more pointing accuracy than was designed into the spacecraft. This sort of conflict impacts the continuity of trend data sets such as Landsat, and illustrates the conflicting goals and pressures evident between two agencies with similar capabilities.

2.4 Other Legislation

With the Land Remote Sensing Act of 1972, ownership of Landsat was moved from the Department of Commerce (NOAA) to NASA and the Department of Defense, jointly. In 1992, Congress directed the Department of the Interior to establish a permanent government archive containing satellite remote sensing data of the Earth’s land surface, and to make these data easily accessible and readily available for study. Residing in the USGS Earth Resources Observation System (EROS) Data Center near Sioux Falls, South Dakota, this collection of information is known as the National Satellite Land Remote Sensing Data Archive (NSLRSDA). It is a comprehensive, permanent, and impartial record of the planet's land surface derived from more than 40 years of satellite remote sensing. The legislation establishing NSLRSDA directs that all land remote sensing data to be permanently archived here.

More recently, Congress has established NOAA as the permanent archive for Earth Observing System (EOS) data from NASA, which will result in a wholesale movement of data sets from NASA active archives to NOAA. The Comprehensive Large Array-data Stewardship System (CLASS) is a web-based data archive and distribution system for NOAA’s environmental data. NASA and NOAA are currently in discussions about which EOS data sets will ultimately reside within CLASS. CLASS is an extension of the Satellite Active Archive (SAA), which was part of the EOS Data and Information System (EOSDIS) in the early 1990s.

2.5 Other National Programs

The U.S. Global Change Research Program (USGCRP) supports research on the interactions of natural and human-induced changes in the global environment and their implications for society. The USGCRP began as a Presidential initiative in 1989 and was codified by Congress in the Global Change Research Act of 1990 (P.L. 101-606), which mandates development of a coordinated interagency research program. Participants in the USGCRP include NASA, the Agency for International Development, Department of Agriculture, Department of Commerce, NOAA, Department of Defense, Department of Energy, Department of Health and Human Services, National Institutes of Health, Department of State, Department of Transportation, Department of the Interior, US Geological Survey, Environmental Protection Agency, National Science Foundation, and the Smithsonian Institution.

The Climate Change Science Program integrates the activities of the USGCRP and the Climate Change Research Initiative activities.

And most recently, the Earth Observing Summits, resulting in the Global Earth Observation (GEO) System of Systems (GEOSS) concept may exert an influence on NASA activities. The US response to GEO is the US GEO, currently comprised of 18 agencies including NASA. US GEO is positioned as a sub committee to the Office of Science and Technology Policy within the White House, and, as such, will offer guidance and recommendations on the best way to integrate existing US Earth science assets into observing systems that address the national, and possibly international, components of each societal benefit area.

2.6 Impacts of External Drivers

For the most part, all of the external drivers above define either what NASA does or with whom it does it. What is important to note here is that NASA does not work in a vacuum, so to speak, with respect to the research data it collects for its own reasons (to be covered in the next section). NASA must participate in data collection, research, and stewardship in an era of dwindling resources, and must take the time to make the necessary agreements with the necessary parties to ensure that the relationship it has established with its science community over the years remains productive. In the near term, the impacts to NASA's systems and decision processes may not be noticeable, but the longer term shift in emphasis and funding will become apparent over time.

3 Internal Drivers

While navigating the changing regulatory and legislative environment, NASA must stay true to its mission and vision, even as the emphasis changes.

3.1 NASA Vision and Mission

Chief among the internal drivers are the NASA mission and vision, and the Earth science research goals. The NASA Vision – "To improve life here, To extend life to there, To find life beyond"; and Mission – "To understand and protect our home planet, To explore the universe and search for life, To inspire the next generation of explorers . . . as only NASA can", both set the parameters for NASA activities. The Science Mission Directorate's Earth-Sun System Division Research and Analysis Program has identified six science focus areas (climate variability and change, Earth surface and interior, weather, carbon cycle and ecosystems, atmospheric composition, and water and energy cycle.) that determine the missions to be flown and the measurements to be collected. Collectively, the data from these six focus areas will help enable reliable prediction of:

- Climate variability and change, and scientific assessment of the impacts of changes in global sea level and ocean circulation, regional temperature, precipitation, and soil moisture
- Recovery of Earth's atmospheric ozone shield and assessment of the quality of the air we breathe
- Global terrestrial and ocean biological productivity, ecosystem health, and interactions with the climate system, and the implications for food and fiber production

- Extended weather patterns, and early formation and the probable pathway of severe storms and hurricanes for planning evacuations and protecting life and property
- Seasonal flooding, droughts and water supply by region globally, and their impact on agriculture and fire hazards
- Volcanic eruptions on monthly time scales and estimation of earthquake probabilities for selected tectonic zones for the protection of life and property.

3.2 Research Needs

In each of the six focus areas for Earth science research, NASA seeks the input of the Earth science community in universities and elsewhere to identify the scientific questions to be addressed and to define effective strategies to pursue the answers to those questions. The following is a list of the 23 science questions that guide NASA's Earth science research.

How is the global Earth system changing?

1. How are global precipitation, evaporation, and the cycling of water changing?
2. How is the global ocean circulation varying on interannual, decadal, and longer time scales?
3. How are global ecosystems changing?
4. How is stratospheric ozone changing, as the abundance of ozone-destroying chemicals decreases and new substitutes increases?
5. What changes are occurring in the mass of the earth's ice cover?
6. What are the motions of the earth and the earth's interior, and what information can be inferred about earth's internal processes?

What are the primary causes of the Earth system variability?

7. What trends in atmospheric constituents and solar radiation are driving global climate?
8. What changes are occurring in global land cover and land use, and what are their causes?
9. How is the earth's surface being transformed and how can such information be used to predict future changes?

How does the earth system respond to natural and human-induced changes?

10. What are the effects of clouds and surface hydrologic processes on earth's climate?
11. How do ecosystems respond to and affect global environmental change and the carbon cycle?
12. How can climate variations induce changes in the global ocean circulation?
13. How do stratospheric trace constituents respond to change in climate and atmospheric composition?
14. How is global sea level affected by climate change?
15. What are the effects of regional pollution on the global atmosphere, and the effects of global chemical and climate changes on regional air quality?

What are the consequences of change in the Earth system for human civilization?

16. How are variations in local weather, precipitation and water resources related to global climate variation?

17. What are the consequences of land cover and land use change for the sustainability of ecosystems and economic productivity?
18. What are the consequences of climate and sea level changes and increased human activities on coastal regions?

How well can we predict future changes in the Earth system?

19. How can weather forecast duration and reliability be improved by new space-based observations, data assimilation, and modeling?
20. How well can transient climate variations be understood and predicted?
21. How well can long-term climate trends be assessed or predicted?
22. How well can future atmospheric chemical impacts on ozone and climate be predicted?
23. How well can cycling of carbon through the earth system be modeled, and how reliable are predictions of future atmospheric concentrations of carbon dioxide and methane by these models?

Given this extensive list of research questions, the Earth science research program at NASA is structured to choose specific scientific questions for which the Agency's technology and remote sensing expertise can make a defining contribution; pursue answers to these questions using an "end-to-end" systems approach that includes observation, research and data analysis, modeling, and scientific assessment in collaboration with our partners; engage the broader Earth science community throughout the process, from question formulation to the final release of findings to decision makers and the public; identify and generate a specific set of validated climate data records in collaboration with the science community and our domestic and international partners; create data assimilation capabilities for available diverse data types; and develop computational modeling capabilities for research focus areas

3.3 Advisory Committees

NASA also has several external advisory committees formed under the Federal Advisory Committee Act, which allow independent parties to review and provide advice on NASA's overall direction. These committees hold open meetings on a regular basis to hear status on NASA programs, and invite comment both from the public and from the committee members. These committees generally concentrate on a specific aspect of NASA's program, for example, looking at Earth science technologies, and render advice and opinions on the subject.

3.4 Earth Science Communities

And finally, the Earth science communities, some of whom get NASA funding on a periodic basis, express their views concerning the direction of NASA's Earth science research endeavors. Several scientific organizations such as the American Geophysical Union, for example, have written editorials, testified before Congress, and met with senior NASA officials on occasion to express their support or concerns.

There are also grassroots organizations within the Earth science communities that began as funded research projects, and now have matured into organized bodies that provide recommendations to NASA and enable linkages to Earth science professionals across the spectrum of the science community. As one example, the Earth Science Information Partnerships Federation (ESIPs) began as a collection of data providers funded by initially by NASA to prototype the viability of a more distributed data and information system. The ESIPs have since formed a formal federation, and continue to involve other US agencies and commercial entities in their efforts to bring Earth science to wider audiences. There is also a set of 'community-based' working groups (discussed in more detail below), which provide thoughtful recommendations to NASA's Earth science management.

3.5 Impacts of Internal Drivers

The above listed highlights of the structured and inclusive approach NASA's Earth science management has adopted to enable Earth science research. What this means for NASA is that, in reality, just as the agency cannot work in a vacuum among other federal entities, it also cannot work wholly independently from these broader science communities. Because NASA's main science goals are to perform research, quite a lot of discussion occurs with research communities, and, by extension, these community members take some ownership for and responsibility to the outcome of NASA Earth science endeavors. Shorter-term needs may be met through the judicious use of research solicitations, but as with the external drivers, longer-term fixes must be in synch with the overall direction of the Earth science research program, as well as the national space priority. The tension created between NASA's federal responsibilities and its constituent's need is nowhere more evident that in the data systems themselves.

4 Effects of External and Internal Influence on Managing Earth Science Data

Specifically for data systems, the collective data collection, access, and archive responsibilities of USGS, NASA, and NOAA have created a reality where data once 'owned' by one agency will ultimately be the responsibility of another. Agencies that may have operated within their own 'stovepipes' may now find the need to cooperate and strategize with other agencies. The following are a few examples of the impact of these forces on NASA data management.

During the late 1980s to early 90s, NASA's Earth Observation System Data and Information System was designed to accommodate data from a series of NASA spacecraft (originally called EOS-AM, EOS-PM, and EOS-CHEM), each containing from five to seven instruments, operating over a decade or longer timeframe, and estimated to collect one terabyte of data per day. Over time, the EOS program has changed, the data identified as being 'within the EOS program' has expanded, and thus the data requirements have changed. Early requirements (agreed to by the science community) stated a turn-around of 24 hours for data, from sensor to desktop. It is recognized today that it is not wise to simply stream raw data in this manner. Science teams now work very carefully and deliberately to produce the best quality, most well-calibrated, well-validated data sets they possibly can. But with Congressional direction to move EOS data from its current home (NASA's Distributed Active Archive Centers (DAACs) to a NOAA facility, NASA and NOAA have initiated a pilot project to move the Moderate Resolution Imaging Spectroradiometer (MODIS) Level 0 data set. from the Goddard DAAC into NOAA's CLASS facility.

Prototyping this sort of migration raises issues around the difference between the needs of an operational agency and the needs of a research agency. For instance, NOAA, as an operational agency, has many real-time weather-related responsibilities that require transmission of large amounts of data during stress periods such as hurricanes. If such a stress period occurred during the transmission of an entire collection of EOS instrument data, and some data were lost, what would the responsibility be of each party (would NASA simply resend the entire data set, or would NOAA check what they got and the ask for the rest, for example?). Additionally, as the largest EOS data set, the MODIS Level 0 data prototype would prove the concept and pave the way for all other small data sets to be transferred, once identified.

Now that the final EOS mission, Aura, is on orbit, NASA is looking to the future of its data systems, and exploring several avenues for restructuring its assets to place the data where it belongs, while continuing its positive relationship with the research community. In 1998, NASA initiated a study entitled NewDISS (New Data and Information Systems and Services), which called for NASA to modify its data system configuration to a more distributed, heterogeneous architecture. This led to the Strategic Evolution of Earth Science Data Systems (SEEDS) study. The SEEDS Study established a framework for distributed data management to maximize availability and utility of NASA's Earth science products; leverage community expertise, ideas, and capabilities; and improve overall effectiveness of NASA-funded Earth science systems and services. SEEDS recommended the creation of community-based groups that would give inputs and recommendations to NASA management, to which NASA would respond. Following the release of the study, four of the working groups were formed, and are addressing issues related to technology transfer, standards, software reuse, and metrics planning and reporting. Membership to these working groups is open, and most entities volunteer some or all of their time and effort.

At the completion of this study, NASA created a tiger team effort specifically to evolve the most visible and successful of its Earth science data systems, EOSDIS. While evolving EOSDIS, the team has taken into account a new direction in data, services and information systems at NASA, namely a move toward 'measurement-based systems'. Measurement-based systems have the driving characteristic that the measurement teams are responsible for ingest, production, dissemination, and archive of the data sets. Some of the long-term goals for NASA's future data systems include:

- NASA's research communities have access to all EOS data through services at least as rich as any contemporary science information system, for example:
  - Data access latency is no longer an impediment
  - The physical location of data storage is irrelevant
  - Finding data is based on common search engines (e.g., Google2015)
  - Services are primarily invoked by machine-to-machine interfaces
  - Multiple data and metadata streams can be seamlessly combined
  - Custom processing (e.g., subsetting, averaging, reprojection) provides only the data needed, the way they are needed
  - Open interfaces and best practice standard protocols are universally employed

- The research and value-added provider communities use EOS data interoperably with any other relevant data sources (e.g., NPOESS, METOP, GPM, numerical models, in situ systems) and systems (e.g., Global Earth Observation System of Systems).

- The EOS archive holdings are regularly peer reviewed for scientific merit:
  - Procedures for such reviews have been developed and tested over a decade

- Derived products that are not deemed scientifically useful are phased out.

One major consequence of these internal and external influences is the need to critically assess the science data holdings and their future. With over 1500 data sets within EOSDIS alone, it is clear that a structured approach to identifying data sets for deletion or long-term preservation is vital. Two efforts underway now will address this need. First, NASA management is establishing a peer review procedure for identifying candidate data sets for deletion or preservation. Working with key NASA science leaders and the appropriate DAACs, data products will be analyzed for their access history, from which an independent panel will determine the fate of the product. Second, NASA management is considering the creation of an internal 'advisory' entity that would concentrate solely on data life cycle issues for Earth science data, including long-term archive. Within the next year, NASA hopes to have both processes identified and tested.

It is evident that the changing data system environment will require some form of adjustment on the part of all parties involved. The hope is that careful planning and prototyping will enable better access to the needed data and services required by NASA's science communities.

5 Conclusion

The drivers and influences discussed in this paper provide insight into the challenges faced by only one agency within the US government that makes use of space-based remote sensing capabilities. Increasing emphasis on national and international activities such as US GEO and GEOSS will require all those agencies using remote sensing to understand the needs of its primary stakeholders (in NASA's case, the research community) and other stakeholders (such as the applications and decision support communities).

NASA's data systems evolution is at a critical stage, having to adapt to the end of the EOS program, tight federal funding, an increased sharing among other agencies, and newly established international programs and initiatives. The way forward clearly includes cooperation with many other agencies and entities, and will certainly entail some adjustments for all involved. If, however, all the stakeholders, from federal agencies to individual researchers keep the greater goal in mind, as quoted at the start of this paper, the whole nation will benefit.