

Blog post: <https://www.esipfed.org/esip-interviews/making-data-matter-with-rama>

Blog title: Making Data Matter with Rama

Interviewee: Hampapuram (Rama) Ramapriyan

Affiliation: Science Systems and Applications, Inc. and NASA Goddard Space Flight Center

Date: April 4, 2019

Location: Remote (GTM)

Interviewer: Arika Virapongse

Blog highlight: The possibility of being able to implement things that you thought about 20 years ago because the computational capability is available now is quite exciting. -- Rama

Arika: Could you tell me a little bit about your work history and what you are doing now?

Rama: For the past 4.5 years (since October 2014), I've been working as a contractor for Science Systems and Applications Inc. I used to be the Assistant Project Manager for the Earth Science Data and Information System (ESDIS) project at NASA (National Aeronautics and Space Administration) Goddard Space Flight Center (GSFC) in Greenbelt (Maryland). I was with the project since it got started in 1990. I worked at GSFC for close to 38 years--the first two years as a contractor and 36 years as a civil servant.

What types of projects are you working on now?

With the Assistant Project Manager role, I focused on science data production, managed Science Investigator-led Processing Systems (SIPS) and also worked on data stewardship and preservation activities. I supported the MEaSUREs (Making Earth System Data Records for Use in Research Environments) Program as a liaison between the NASA Headquarters Program Manager, the [DAACs](#) (Distributed Active Archive Centers), and the PIs who produced data. I continued that liaison activity for a couple of years after I retired from the government.

My current focus is primarily on data stewardship, data quality, and information quality. I'm co-chairing the Data Quality Working Group (DQWG) under NASA's Earth Science Data System Working Groups ([ESDSWG](#)). I also chair the ESDSWG Data Product Developer's Guide working group, where we are developing a guide document for data producers so they can generate products that are more easily usable by end users.

When and how did you get started working in the field of Earth Science Data?

I have a PhD in electrical engineering from University of Minnesota (1970). My PhD focused on control theory. In those days, they did not offer degrees in computer science. If they did, I probably would have gone for that. Control theory is essentially analyzing signals in one dimension--looking at how systems behave with respect to time. My thesis was about the stability and optimal control of large scale systems.

I joined Computer Sciences Corporation (CSC) in 1971 as a contractor working for the Marshall Space Flight Center (MSFC) in Huntsville, Alabama. They wanted me to develop algorithms to process x-ray images of rocket engines that showed cracks. It was my first exposure to image processing. Some of the material that I had learned in control theory could be applied to image processing, because I could extend what I knew about one dimension to two dimensions. (Gray scale in an image is a function of two independent variables--the X and Y dimensions.) I found the work on image processing quite exciting because we could write an algorithm, display the results, and find the effect of what we did almost immediately. It was instant gratification. I ended up developing many programs for image analysis.

I gradually migrated into remote sensing when LandSat-1 (then called ERTS--Earth Resources Technology Satellite) was launched in 1972. Using the image processing work that we had done in the previous year, we could apply and develop new algorithms and software for image processing and radiometric/geometric correction of LandSat data. It was a matter of understanding the documentation that was written for LandSat, so we could make sense of the images being collected and do the processing. We did some collaborative studies with the local area governments (TARCOG--Top of Alabama Regional Council of Governments), and produced a number of images for them. For several years I was primarily a user of remote sensing data.

In 1976, the CSC contract at MSFC ended and several people in my group were transferred to different places by CSC, while others chose to stay in Huntsville and join other companies. I transferred to NASA GSFC, where I joined a program called IntraLab (later called ERRSAC--Eastern Regional Remote Sensing Applications Center) as manager of a section in CSC supporting IntraLab/ERRSAC. The mission of ERRSAC was to promote the use of LandSat data in state and local governments. In 1978, I switched from being a contractor to being a civil servant. My job then was to teach people from state and local governments how to use image processing techniques on LandSat.

In preparation for the launch of LandSat 4 (in 1982), I helped manage development of an internal system in GSFC to process one LandSat image per day; they needed someone who understood processing software. This was done as a back-up, while GE (General Electric), who was the contractor at the time for LandSat, was getting their larger system to process data on a production basis. Then I became a Principal Investigator on the Shuttle Imaging Radar Program (SIR-B, launch date October 5, 1984), performing automated stereo-image matching to derive elevation maps using the [Massively Parallel Processor](#) (developed in the mid-1980s) at GSFC.

Towards the end of the 1980s, the EOSDIS (Earth Observation System Data and Information System) was getting started. I got tapped to join the Information Processing Division (Code 560) to help monitor the Phase B studies that were being conducted by TRW (TRW Inc.) and Hughes (Hughes Aircraft Company). Once these studies finished, I was involved in generating and reviewing the presentation material for the science data processing parts of the data system for the Non-Advocacy Review (NAR) conducted by GSFC. The NAR reviewed the entire EOS

Program and included presentation of costs of the data system. It was an essential step before approval of the Program.

Starting with fiscal year 1991, the EOS Program was approved by Congress as a part of the “Mission to Planet Earth”. The EOS Program had several flight missions and the data system (EOSDIS). The flight mission budget was intentionally kept separate from the data system budget. The ESDIS Project (Code 423) at GSFC was started in early 1991 and I joined the Project as its Deputy Project Manager.

Between 1991 and 1995, there were many reviews about the cost of the program because the budget for the EOS Program was considered to be very large. The program was re-structured, re-scoped, re-baselined, and re-shaped through these reviews (see [Ward, 2008, on page 5](#)). The Payload Panel, one of 12 panels of the EOS Investigators Working Group (IWG) played a very important role in advising NASA on how to maximize the science return while constraining the costs. Another of the IWG panels, the EOSDIS Advisory Panel (a.k.a. Data Panel), was responsible for advising the EOSDIS development.

Around 1995, the National Research Council (NRC) reviewed the EOS Program and had a subcommittee review the EOSDIS. I supported that review by providing details about EOSDIS costs. The review called for performing the EOSDIS functions through a federation of competitively selected information partners ([Dutton et al, 1995 in appendix F](#)). That’s how the term “ESIP” (Earth Science Information Partners) came into being.

To respond to the NRC review recommendations, NASA set up a “Response Task Force (RTF)” led by Dixon Butler (Program Manager and head of the division at NASA Headquarters responsible for EOSDIS) and consisting of experts from the science and technology communities. Instead of using the federation to take on almost all of the functions of the EOSDIS, as was suggested in the review, the RTF recognized that many users depended on the observational data and their being processed on a regular operationally robust manner. So EOSDIS remained, and ESIP came into being with “Type I”, “Type II”, and “Type III” members. Operational robustness was represented in Type I. Research and innovation was Type II. Applications, which eventually became independent of NASA funding, was Type III. I helped Dixon Butler as his deputy on the RTF.

In summary, I gradually moved in my career from the user end of remotely sensed data to its production end. As a LandSat user, I was using data. When LandSat 4 went up, I was closer to the production end by producing products for other users. EOSDIS is the other extreme, I suppose, where we were building a data system that produces, archives and distributes data that the whole world-wide community can use. That was my career path.

It seems to me that as we’ve learned about the changing climate, it has stimulated interest in collecting even more information, as well as making more resources available

for Earth observation. Do you agree with that? Can you speak to how and why that support has changed over the years, as well as the public perception around that?

The initial definition of the EOS Program was based on concern about climate change. In those days, they referred to it as global warming. They changed the term to climate change over time because it is the change that people are concerned about.

I was not involved in the definition of the program, but people like Dixon Butler and Shelby Tilford at NASA Headquarters were forefathers of the program, and there was a large scientific community involved. There are many [reports](#) you can refer to for more information about how the program got started.

I'm not a climate scientist. Instead, I see that our job as data providers is to provide data that intelligent climate scientists can take, interpret, and produce products that are useful and provide enough information. Even the scientists are not policy makers. They need to provide information to policy makers, so they can take appropriate actions.

I couldn't go into the political aspects of things over time because politics happens and science goes on. Rather than comment on that, I should say how reasonably steadily we have had support for the EOS program over the years. The Earth science program has been well supported since 1990 and even before. We have had a large and diverse set of data collected from Earth observing satellites since the old times of the Nimbus satellites and even earlier.

I wrote a book chapter (Ramapriyan, 2002) that covers very briefly the history of remote sensing. The chapter primarily focuses on describing the processing steps and techniques used for satellite data, as well as some history of the satellites for remote sensing.

What do you think are some of the major challenges and issues facing Earth Science and Earth Science data and informatics today?

From an external observer's point of view, I can see that the political climate is such that there is a reasonably large set of people who are skeptical about the conclusions drawn by the climate scientists.

To address that, it's critically important that data are accompanied by full provenance and openness. When I say provenance, I mean traceability of lineage. If someone is skeptical and wants to find out how a scientific conclusion was drawn, you need to provide links to everything that went into it. That includes the papers and data that conclusions are based on, and descriptions of the multiple low-level products, such as those coming from instruments that flew on different satellites, that higher level data products depend on. From a data science or data system point of view, I think we need to enable the availability of that full lineage.

The kinds of things that we are doing to support that goal are: citations of data, assigning digital object identifiers, and preserving the provenance and context type of information. For example, the NASA Data Quality Working Group makes recommendations along these lines, so that

people document, capture, and make it easy to discover and use the quality-type of information that is associated with the data. So if you are predicting changes expected 100 years from now, the range and uncertainty of the predicted parameters should be clearly defined, and the data systems should make such information available in a consistent manner.

Aside from the challenges that you've already mentioned, are there any others that you'd like to add?

When we got started with our satellite missions, our main challenge was: How are we going to manage all of the data that are coming? Our concern was: We are going to have a few petabytes of data. How do we keep track of them? How do we store them? It is going to be very expensive to store. That part of the challenge got addressed by advances in technology.

Data volume requirements and storage capabilities form a "feedback loop", so to speak. People want to do what is possible at a given time. People plan to do a highly challenging thing, and technology advances to accommodate challenges by meeting the requirements. Then, the requirements grow based on the advancement of technology. It is a two-way street.

Now we have a challenge that the upcoming missions are going to be producing data 10 to 15 times faster than before. NASA and other organizations are addressing this challenge by moving data to the cloud. This also makes it easier for people to do their analysis where the data are located (i.e., moving the analysis to the data, rather than moving the data to the analyst's desktop).

What do you think drives this technology development? I'm thinking about the role of the federal government and private industry, as well as public needs.

I think it happens in both private industry and in federal government-supported activities. Internet, for example, evolved from the ARPANET - Advanced Research Projects Agency Network, funded by the Department of Defense. So it started with government seeding of the technology but then it grew wild with private industry. A similar thing happens with NASA--they produce a bunch of technology with federal government funding, and then it gets taken over and multiplied by private companies.

Right--since one of the government's goals for supporting these types of activities is to stimulate the economy. Where do you think Earth Science or Earth Science data and informatics is going in the future? You've spoken about the cloud already, is there anything else you would like to add?

I think that ESIP is a great place where informatics discussions are taking place. There are so many different clusters and committees that it is hard to keep up with all of them. Each of them has interesting discussions that happen every month during the respective telecons as well as during the two annual meetings. We are in very good times for Earth Science informatics.

From my own observations, 2003 was roughly when AGU (American Geophysical Union) was starting to talk about Earth and Space Science informatics. During the first year, there were 1 or

2 oral sessions, and a couple of poster sessions. Nowadays there are several parallel oral sessions going on and hundreds of posters, if not thousands. A lot of progress is being made and a lot of material is being presented in Earth Science informatics. That's gone on in IEEE (Institute of Electrical and Electronics Engineers) Geoscience and Remote Sensing Society as well, where there is an Earth Science Informatics committee. I chaired the Data Archiving and Distribution Technical Committee during 2009-2013. Towards the end of my term, we proposed that they change the name to Earth Science Informatics because it more accurately represented the breadth of interests of the members. In addition, the EGU (European Geosciences Union) holds ESSI (Earth and Space Science Informatics) sessions. Of course, RDA (Research Data Alliance) also does a lot of work with Earth Science informatics. It's becoming a fast growing field.

It also seems to me that the field is growing exponentially. Why now? Why not in the past?

There is so much information that is openly and freely available that the interest in analyzing and making use of the data is quite high. The volume, velocity, variety, veracity, value--all the big data terms that start with the letter "V"--pose interesting problems and challenges for computer scientists to solve. It is another reason for the growth in Earth Science informatics.

Deep learning is something that has taken off because the high performance computing capabilities that are available enable people to do things that they could not do before. It's not that these or similar techniques have not been talked about or thought of 20 or 30 years ago. It's just that they have become more practical because of the advances in high performance computing. The possibility of being able to implement things that you thought about 20 years ago because the computational capability is available now is quite exciting.

References

Dutton, JA, FP Bretherton, RL Jenne, S Karin, S Volansky, F Webster, C Zraket (1995) "The Earth Sciences Information System" (Appendix F). In National Research Council, *A Review of the U.S. Global Change Research Program and NASA's Mission to Planet Earth/Earth Observing System*. National Academy Press, Washington DC, pp. 75-87.

Ramapriyan, HK (2002) "Satellite Imagery in Earth Science Applications", Chapter 3. In L. D. Bergman and V. Castelli (editors), *Image Databases, Search and Retrieval of Digital Imagery*, Wiley, Inc., pp. 35-82.

Ward, AB (2008) "The Earth Observer: 20 Years Chronicling the History of the EOS Program". In, Ward, AB (editor), Perspectives on EOS [Special Issue], *The Earth Observer* 20 (2): 2-6.

[Disclaimer: Any opinions or recommendations expressed in this interview are those of the interviewee and do not necessarily reflect the views of Science Systems and Applications Inc, NASA, or any other organizations listed. This interview also represents an “oral history” (a recollection of history), so its value is in the personal perspectives and insights of the interviewee, rather than specific dates, years, and titles for reference.]