

Response to RFI regarding NASA's move to Open Science.

Peter J Minnett. Department of Ocean Sciences, Rosenstiel School of Marine, Atmospheric, and Earth Science, University of Miami, Miami. (pminnett@earth.miami.edu)

Background

I am now Professor Emeritus at the Rosenstiel School of the University of Miami. I have been part of the MODIS Science Team since before the launches of Terra and Aqua, and similarly for VIIRS on the S-NPP satellite. My specialty is the derivation of the skin sea-surface temperature (SST_{skin}) from the infrared measurements of MODIS and VIIRS. I have also received funding for other research from NASA and other agencies. The comments below were developed with colleagues at RSMAS.

I attended the on-line Webinar held on July 17 hosted by Dr. Louis Barbier.

Timeliness of data release

Part of my NASA-funded activity involves shipboard installation of instruments, including those to measure the spectra of infrared emission from the sea surface and atmosphere, to provide data for the refinement of the algorithms used to derive SST_{skin} from the satellite measurements, and to assess the accuracy of the derived values. Other instruments installed include a weather station and an all-sky camera to record the conditions in which the radiometric measurements were taken. The instruments run autonomously and continuously with daily connection over the ships' internet links to our lab to allow monitoring of the status of the instruments. The 24/7 operations are required to ensure appropriate measurements are made at the time of every satellite overpass. The bandwidths of the internet links are not sufficient to allow the real-time transmission of the measurements to the lab. Instead, the data disks are swapped out when the ships are in US ports, which can be after intervals of many months. It is only when the datasets are available in the lab can they go through quality control to remove unusable data, such as those contaminated by precipitation or sea-spray. The instruments are robust allowing operation at sea for over a year, sometimes several years, and when needed they are brought back to the lab for recalibration and refurbishment. Only after the instruments have had post-cruise calibration can we assess the accuracy of the measured and derived variables and record their calibration history, which is a requirement for open access to data.

In the Introduction of NASA's Public Access Plan, it is stated that it responds to the new 2022 OSTP guidance "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research." I am sure I am not alone amongst those who take field data in finding the "Immediate" requirement a cause of great concern. In my case it is not feasible to distribute field data immediately. Even if it were, it is illogical and most unwise to disseminate to all and sundry data that have not undergone quality assurance and are without an appropriate calibration history or assessment of accuracy. The wording of the timeliness of data release in the final version of the NASA instructions to PIs must be very carefully thought through and very clearly written, ideally by people with first-hand knowledge of taking data in the field.

Preliminary data.

The OSTP requirement excludes several categories of preliminary material, but “preliminary data” is missing. On p8 of the NASA Data Plan, preliminary data are included in the list of excluded items. It is important that preliminary data be excluded from the final NASA requirements. If this is to be the case, the timeliness issue is less of a concern.

Withheld data.

It is good practice to withhold a subset of data to be used to validate the accuracy of the retrievals of algorithms or models derived from similar data sets. For example, we have used withheld shipboard data to assess the accuracy of atmospheric profiles of temperature and humidity and SST_{skin} represented in re-analysis fields (MERRA-2 and ERA5). This would not be possible if all data were made available and thereby probably be used in a way that prevents their use as independent, withheld data.

Graduate students.

It is stated that recipients of “Educational grants and grants to individual students” are exempt from these open science requirements. I urge you to extend this exemption to other students using measurements taken with NASA funding, but who are not directly supported by NASA. A recent example is a graduate student of mine who was supported by a university fellowship, but worked with data taken in a NASA-funded project. Some graduate students from overseas are supported by grants or fellowships from their home country, and some may even be self-funded.

Promotion and award of tenure.

In response to a question during Webinar, Dr Barbier stated that it was acknowledged that time spent preparing data and software for release, including documentation, takes away from time that would otherwise be spent on research that would lead to a publication, and from preparing that publication. Dr Barbier tried to reassure us that NASA program managers had been instructed to accept release of data sets and software as equivalent to reviewed publications in terms of satisfying the requirements of a NASA grant. However, junior faculty in universities face many hurdles in achieving promotion and the award of tenure. Major factors considered include the number of publications, the rate of publication, and the number of citations. In most universities, external review letters, including from abroad, are sought as part of the file that will be judged by a review panel that makes a recommendation to the faculty, dean, provost, and president of a university. It is clear to me that junior faculty who are successful in securing research funding but who are consequently sidetracked into preparing open data and open software will be disadvantaged in developing their careers.

Software.

The requirement to make code openly available and sufficiently well documented to be usable easily by others will impose a huge burden on researchers. A comment was made during the webinar that it takes ten times longer to write well-documented code for others to understand than developing code for one’s own use. This is, I think, greatly underestimated. There is also a

fundamental risk that code a researcher releases contains one or more undetected errors that will be propagated into future analyses by others, whereas code written by the second researcher to replicate the results of the first will likely reveal such mistakes. In my experience, coding errors can occur even when the software is developed by a professional programmer. The aspiration of open software is laudable in terms of saving time and effort across the community and society, but it is not without risk of negative consequences.