PSD's Approach to Data Management Plans (DMPs)

Michael H. New, PhD
NASA Headquarters
Science Mission Directorate
Planetary Science Division

Why the change?

- Bureaucratic answer: Because NASA has been directed to improve public access to the results of NASA-funded research by the White House's Office of Science and Technology Policy (OSTP).
- Philosophical answer: Because the American taxpayer provides large amounts of money for us to go off and study things that *we* think are important and the least we can do is give everyone access to the knowledge and data we produce.

Conan, what is “data”?

- According to OMB:
  Research data *are defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings.*

- “Data” does not include:
  - physical objects such as
    - astromaterials,
    - analog specimens,
    - experimental run products;
  - preliminary and other unpublished data;
  - plans for future research;
  - peer reviews;
  - trade secrets, commercial information, or materials necessary to be held confidential by a researcher until they are published;
  - personal and medical information;
  - data in prepublication documents; and,
  - private communications.

How should I manage my data?

PSD's guiding philosophy is that all relevant data should be made publicly available (*i.e.,* without fee or restriction of use) at the time of publication, or at the earliest practical time thereafter, through a stable and long-term supported data repository.

“all relevant data”

- From ROSES-2016, Appendix C.1, “all relevant data” means:
  - Any data needed to validate the scientific conclusions of peer-reviewed publications, particularly data underlying figures, maps, and tables.
  - Also, any other data and software that would enable future research or the replication/reproduction of published results.
- The precise nature of the data will vary from discipline community to discipline community.
  - This is why Data Management Plans are reviewed by peer review panels.

“publicly available”

- This means that the data are accessed without fee or restriction of use.
  - So putting data behind a “paywall” is not acceptable.
  - Putting data into “Supplementary Material” may be acceptable if the “Supplementary Material” is not behind a “paywall”.
  - NASA recognizes that this may cost money for page or open-access charges. NASA is willing to pay for this. Adjust the budgets of new proposals accordingly.

“a stable and long-term supported data repository”

- Putting your data on your personal or your research group’s website is not sufficient.
  - “long-term” may be thought of as meaning “longer than one person’s career”.
- Many university libraries are creating data repositories for faculty. These are acceptable.
- Except for DDAP and PDART, archiving results in the PDS is not required, although it is clearly acceptable.
- Other acceptable respositories include but are not limited to:
  - IEDA (*www.iedata.org*)
  - EarthChem (*earthchem.org*)
  - HITRAN (*www.cfa.harvard.edu/hitran/*)
  - MRCTR (astrogeology.usgs.gov/facilities/mrctr)

How do I communicate my plans?

- Every proposal to PSD is allowed two extra pages for a Data Management Plan (DMP).
- This plan shall contain (*cf:* ROSES C.1):
  - A description of data types, volume, formats, and (where relevant) standards;
  - A description of the schedule for data archiving and sharing;
  - A description of the intended repositories for archived data, including mechanisms for public access and distribution;
  - A discussion of how the plan enables long-term preservation of data; and,
  - A discussion of roles and responsibilities of team members in accomplishing this plan.
- Any funds needed to implement the DMP should be included in the usual budget and budget justification sections. Don't be shy about asking for necessary funds for this!

But I don’t produce any “data”

- Ok, that’s possible and perfectly acceptable based on the definition of “research data”.
- BUT, then the DMP should state that no data preservation or data sharing is needed, and why that is the case.
- In a case where no appropriate archive exists for a particular data set, the DMP should discuss alternative methods for making the data publicly available.

Did I see “software”?

- You read that correctly! For many of us, software is as important a tool as a geologist’s hand-lens or electron microprobe and out results can’t be divorced from the software we use to generate or process data. But we’re trying to be realistic.

- *Software…created as part of a NASA award should be made publicly available when it is practical and feasible to do so and when there is scientific utility in doing so. Stand-alone code that is straightforward to implement or whose utility is significantly outweighed by the costs to share it is not expected to be made available.* (ROSES-2016, C.1)
- Otherwise, NASA expects that the source code, with associated documentation sufficient to enable the code’s use, will be made publicly available via
  - GitHub (github.com/NASA-Planetary-Science),
  - the PDS (for mission-specific code, when appropriate), or
  - an appropriate community-recognized depository (for instance, the homepage of the code base for which a module was developed).
- Archiving software in a public repository does not require the proposer to maintain the code

Will this be on the test?

- DMP's will be reviewed as a part of a proposal's peer review.
- For the time being, most (but not all) programs in PSD are not including the evaluation of the DMP as a part of the Merit score.
  - Inevitably, this will change as we all become used to writing DMP's.
- Awards will not be issued, however, until an acceptable DMP is in place.

QUESTIONS?