Planetary Science Division FAQ for Data Management Plans in Proposals Submitted to ROSES

[Last Updated 07/25/2016]

NASA, and the Planetary Science Division (PSD) specifically, have a long-standing policy of making planetary data products available for use by the scientific community and the public at large. PSD manages mission data via the Planetary Data System (PDS), and relevant non-mission data via its support of Earth-based observatories and community data repositories.

While the sharing of data and tools has always been encouraged, or even expected by certain ROSES program elements, there was not a universal requirement placed on research grants until ROSES 2015. That mandate was introduced by the *NASA Plan for Increasing Access to Results of Federally Funded Research*, which created the data management plan (DMP) requirement for most NASA research proposals. The *NASA Plan* is the basis on which PSD has developed a framework addressing the data needs of the diverse planetary science community.

This FAQ addresses DMPs for proposals submitted to the Planetary Science Division's ROSES program elements. It is intended to complement and clarify Section 3.5 of Appendix C.1 of the ROSES solicitation, which explains the requirements for DMPs and defines important terms. Proposers are strongly advised to read the solicitation and the relevant appendices.

1. I'm not sure whether I have to submit a DMP with my proposal. How do I find out?

Consult the ROSES appendix for the program element to which you are proposing. Unless it explicitly states that you don't need a DMP, you need one. PDART (Appendix C.7) requires a discussion of data management in the 15-page Scientific/Technical/Management section of the proposal, rather than as a separate section of the proposal.

2. Are you telling me that I have to archive everything I do and make it all available to the public?

The requirement is that you provide a plan that describes whether and how you will preserve the data generated by your project, and whether and how you will arrange for public access to the data that are needed for the validation of published results and that are scientifically appropriate for enabling future research by other investigators.

3. What does "needed for the validation of published results" mean?

This refers to the data that support and have led to the scientific understanding and conclusions presented in a scholarly publication. At a minimum, these data will include the "data behind the figures and tables", i.e., the actual quantitative information presented in the publication in graphical or tabular form. They may also include other scientifically appropriate data that have been used in the development of the scientific conclusions. But the term does not necessarily encompass every measurement or calculation that was made in the course of the project (see FAQ #6).

4. What does "scientifically appropriate" mean?

It means there is a reasonable probability that somebody may want to use the data (or software; see FAQ #13) in the future and won't readily be able to re-create it on their own.

How this applies to your products will be driven by the standards and needs of your research communities. For example: perhaps only fully calibrated data from your instrument are of use to others (e.g., if the methods of calibration aren't in question). In that case, archiving and providing access to calibrated data may be sufficient. On the other hand, if there are other potential scientific uses for the raw data, then there should be a plan to make them available, as well as any calibration data which potential users would not be able to find or generate independently.

5. My project won't generate any data. What do I have to put in my DMP?

It is unlikely that your project will generate <u>no</u> data, unless it involves nothing but non-computational theory or instrument development. But if that is really the case, your DMP still needs to explain why no data will be generated.

6. My project will generate data that some people might want, but there are good reasons why they shouldn't be released. What do I have to put in my DMP?

Certain things are explicitly exempt in the *NASA Plan*. These are:

1. Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law.
2. Personal and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.
3. Preliminary data or analyses, laboratory notebooks, drafts of scientific papers, plans for future research, reviews of or comments on the funded research, or communications of the funded researchers with colleagues.
4. Physical objects, such as laboratory samples or specimens.
5. Data resulting from educational grants and grants to individual students.
6. Data that are proprietary, export-controlled, Sensitive But Unclassified (SBU), and National Security classified.
7. Data resulting from Small Business Innovation Research (SBIR) and Small Business Technology Transfer (SBTT) contracts.
8. Research data that are not possible or scientifically appropriate to share or publicly archive.
9. Research data that may be reproduced from other publicly accessible data without an intensive work effort or specialty knowledge and equipment, and with instructions included in publications.

If your data fall into one of these categories, or if there are other compelling reasons why preservation and sharing are inappropriate or impossible, your DMP must explain why this is the case, and how your results could be validated in the absence of preserved data.

7. I understand that I need an actual DMP. What do I have to do?

You should carefully consider the following aspects of data management:

- Data types, volume, formats, and standards;
- Schedule for archiving and sharing;
- Appropriate data repositories, including their mechanisms for public access and long-term data preservation; and
- Roles and responsibilities of team members.

First consider whether you can meet community standards in these areas through publications and supplementary materials in mainstream, widely accessible scientific journals. If this is the case, see FAQ #10 below. If data management for your project requires more than this, your DMP needs to describe how you are handling each of these areas, to the extent they are appropriate to your project, and must explain either (a) how data sharing and preservation will enable validation of published results, or (b) how results could be validated if data are not shared or preserved.

8. How is my DMP going to be evaluated? Where is the "bar" set? Is it going hurt my intrinsic merit score if I don't give you what you want?

Data management is not a part of the intrinsic merit evaluation, except in the PDART program, where the discussion is intrinsically a part of the proposal itself rather than an additional statement. In general, DMPs will be evaluated in the peer-review process according to the prevailing current standards of the research communities. NASA will instruct peer reviewers to consider whether submitted DMPs meet the needs of the relevant communities, but will not attempt to set absolute standards nor impose rubrics for DMP evaluation. A weak DMP will not, by itself, prevent an otherwise strong proposal from being selected for funding (except in PDART, Appendix C.7); however, NASA will require proposers to submit an acceptable DMP before funds are released.

9. Data management sounds like real work. Who's going to pay for this?

We consider data management to be a normal part of the research process, and proper data management practices should happen as a matter of course in any well-managed project. Although this expectation is, arguably, different from what may have been considered standard practice in previous decades, our ability to acquire, manipulate, and store data has grown far beyond what was previously possible. In most cases data management should not require additional funding. However, if you have data or software (see FAQ #13 below) that requires extensive special handling, documentation, or maintenance, you can request support for this aspect of the work in your proposal. Be sure that the request is included in the budget, explained in the budget justification, and referred to in the DMP.

10. My project is small and straightforward; it's totally theoretical in nature or makes use of modest amounts of already public data employing standard methods. The results are going to be published in the traditional way in mainline journals. What do I need to put in my DMP?

It's possible that all you need to worry about is publishing the data behind the figures and tables. You should investigate the options that your favorite journals have for including data in published papers; nearly all of the major publishers have some capability for publishing supplemental data. You may want to consider whether some may be more appropriate to your particular datasets than others. Your DMP should explain your rationale for deciding what and where you will publish, an estimate of the amount and types of data that will be involved, a description (to the extent possible) of the format(s) in which you intend to publish the data, a brief statement of how your plan will enable validation of the results, and an explanation for why no further data management is needed. But you should also consider whether you may have intermediate results that could be scientifically valuable enough to warrant archiving.

11. My project will generate data beyond what can be distributed through publications and supplementary material. How do I figure out where to archive it?

Consider first whether it falls in the purview of any of the PDS nodes. If so, contact the relevant node and inquire about archiving. PDS will ask you to provide information about the nature of your data so they can decide whether they want to host it. The information you provide to PDS will be the same information you need for your DMP, so thinking about these essential items early will help you in the long run. If PDS is amenable, ask for a letter of support and include it in your proposal (see C.1, Section 3.5.2).

If your data are not appropriate for the PDS, then you should investigate whether there may be a research-community-based or university-library-based archive that may be equally or comparably good. Most desirable would be a "PDS-equivalent archive," which is defined in the PDART program as an archive that:

1. is managed by someone other than the major data provider;
2. is managed for the long term (at least 25 years);
3. is accessible to the scientific and lay public without preapproval;
4. ensures that data are searchable;
5. ensures that data are citable;
6. is considered by its user community to be the standard archive for its subfield; and
7. requires data to be submitted in standardized formats and data types.

If there is no PDS-equivalent archive available, you should make a good-faith effort to find a workable alternative having, at minimum, attributes 2, 3, 4, and 5 from the above list. If there is no workable alternative, your DMP should explain the situation so that peer reviewers and NASA can be made aware of significant obstacles to data preservation.

12. Can't I just put the data on my personal web page?

We do not encourage this. Personal web pages tend not to make good provision for long-term data preservation. Active researchers who serve on NASA proposal review panels, as well as NASA program officers, tend to be aware of this and it may affect their evaluations of your DMP.

13. What about software? Are you going to force me to make all of my code public?

PSD's position is that software may be essential for validating the results of NASA-funded projects, and should be made publicly available when it is scientifically appropriate to do so. In the case of PDART, archiving of source code at NASA's Github repository is <u>required</u>; the Github repository is a resource that should make it easier for researchers funded by PDART and other PSD program elements to share code with their communities. But we also recognize that there are multiple layers of complexity in requiring software release, e.g.:

- Some software may have potential commercial value, and intellectual property rights should be protected.
- Software can be brittle, and susceptible to changes in operating systems or compilers, necessitating maintenance which requires effort and funding.
- There may not be a significant user community for highly specialized software, in which case release and maintenance could be a needless effort.
- Software used in a federally funded project may have been developed without federal support, making a presumption of a public right to access questionable.

For these and other reasons we are not mandating any specific course of action at this time with regard to code. If development or use of software is a major element of your proposed work, your DMP should explain the nature of the software, and, as it does for data, explain whether the code will be made public; if so, how, and if not, why not.

14. My project will acquire data from an established observatory that archives every image file. Is that good enough?

Data curation by the observatory could certainly contribute to your DMP and would deserve to be mentioned. But you are probably going to put in a lot of work to get from the raw images to the science results, and the data that can be said to support the conclusion(s) of your paper(s) will probably be the fully reduced, calibrated images and not the raw frames. The common-sense criterion at work here is that someone who might want to use your observations, either to confirm your results or to do new science, ought to be able to work from the calibrated data and should not be duplicating your effort on data reduction.

15. My project involves collaboration with unfunded international partners who will provide data from their own facilities. Do they have to agree to allow their data to be archived in the U.S. for public access? Can I use their data if they don't agree?

We would hope that you would talk with your international collaborators to find out what requirements for data management they are working under, and whether they are amenable to some form of data release. But NASA has no authority over non-U.S. entities, and no interest in limiting the data that researchers can use in their science investigations. If there are factors that make release of certain data impossible, explain why this is the case in your DMP (see FAQ #6).

16. In Section 4 of the *NASA Plan* it says that "DMPs must provide a plan for making research data ... accessible at the time of publication or within a reasonable time period[4] after publication..." and footnote 4 says, "This time period will defined in the final Data Access plan." What is this "reasonable" time period?

For data archived and distributed through publications and supplemental supporting materials (see FAQ #10), release should occur at the time of publication. For other, more extensive data sets, you can propose and justify in your DMP a time period considered reasonable by your research community (for example, 12 months); but generally this should not be later than the end of the award.

17. In Section 7.1 of the *NASA Plan* there is a footnote about an "official approval process (signature process) for data release." What is that?

That footnote is intended to apply only to large data sets that go into the official NASA archives. We delegate that approval responsibility to those responsible for running those archives, each of which has a process for reviewing submitted datasets.

18. The SMD-wide FAQ page for data management plans has examples of a "minimal DMP" and a "real DMP" that are very short and simple. Can I just copy those?

We wouldn't recommend that. The SMD FAQ page reflects the minimum DMP requirements common to all divisions in SMD. PSD is setting its requirements above this minimum, and so we suggest that you craft your DMP following the advice given here.

19. Suppose my proposal gets funded. Do I have to report what I do with regard to data management?

Yes, you will need to do this in your annual progress reports, and summarize in your final report at close-out. Datasets made publicly available separately from journal articles should be listed as separate products in the list of publications.

20. What if I don't do what I said I was going to do in my DMP? Is that bad?

It could be, just like it could be bad if you don't do the science you said you were going to do in your proposal. We understand that projects can take unexpected turns, and that it is possible for things not to end up as intended, through nobody's fault. Nonetheless, funded researchers, research institutions, and NASA centers are responsible for ensuring and demonstrating compliance with the DMPs approved as part of their awards. Principal Investigators are responsible for data management issues on their entire projects, including data from all subawards/subcontracts, facilities, and collaborations. Awardees who do not make a good faith effort to fulfill the intent of their DMPs may have continuing funds withheld, and this may be considered in the evaluation of future proposals.

If you have any questions that are not answered by the C.1 or this FAQ please send those to meagan.thompson@nasa.gov and cc SARA@nasa.gov