

As a general comment, there is a lack of reference to open science in the Plan. Achieving greater access to publicly funded research data is a worthy goal, but access that does not include, for example, correct and full metadata, license, and other re-use information, etc. is not as useful to data consumers. These open science requirements are now NASA policy for SMD-funded data per SPD-41a, but there is no equivalent policy for other, non-SMD-funded research data at NASA. Is a plan for increasing conformance to open science NASA-wide forthcoming?

Specific comments on the Plan

1. Regarding “Applicability” section text:

All researchers receiving federal funding will be required to submit Data Management Plans (DMPs), however in some cases it is expected that some data will not be made public. This includes but is not limited to the following categories:

...

- Work which results in Personally Identifiable Information (PII) (i.e. human subject data)

The way this is currently written, it is unclear if exemptions for investigators producing human subject data are 1) not required to submit data management plans or 2) are not expected that their data will be made public (or 1 and 2). Most likely, the intention of this section is that investigators who produce human subject data will not necessarily be expected to make their data public. However, these same investigators should also have to submit Data Management Plans. Data Management Plans describe important information regarding assurance of data integrity, reliability, etc., are a pillar of open science, and composing them does not require that the investigator necessarily grant public access to produced data. Suggest rewriting this section to clarify.

Also, investigators using human subjects may read the exemption text as applying to their whole investigation, when only one portion yields PII. The exemption should only apply to that portion. Rephrasing as “any part of a scientific investigation that results in PII...” would clarify this.

2. This sentence from Part A, Section 3.0 could be further clarified as follows:

“Scientific data **supporting** peer-reviewed scholarly publications resulting from federally funded research shall be made freely available and publicly accessible by default at the time the **supported publication is made public**, and no embargo by a publisher or others should be imposed nor will such be recognized by NASA.”

The logic for this change is that making scientific data publicly accessible is itself a publication, and the word “publication” can now refer to publishing of data (a **data publication**). Historically, scientists have reserved the word “publication” to mean only the kind of predominantly textual/prosaic documents embodied by journal articles, monographs, etc. But, especially with the recent emphasis on the value of publishing datasets, the term **publication** can also now refer to making data publicly available, as well as mixtures of prose and data such as commonly found in conference presentations, webinars, and so on. The slightly reword language will clarify which of the two kinds of publications is being required at what time and in what context.

It should be noted that the title of Part B, “Peer-reviewed publications” is also a bit misleading, as it is intended to refer only to the kinds of textual/prosaic documents that traditionally/historically have often been the subject of peer reviews. However, many journal-led reviews are now including peer-review of supporting datasets when these text documents are submitted for review, and some journals offer both publication and peer-review of data sets. Thus, the title could be interpreted to include scientific data as they also can be “peer-reviewed publications,” but that is not the scope of Part B.

3. “Unique digital object identifiers (DOIs) must be assigned to all datasets supporting peer-reviewed publications”. Since DOIs can resolve to any kind of web resource, it would be useful to include in this statement that the DOIs assigned should resolve to web resources that reveal all data and contextual metadata that is necessary to interpret the data. Consumers of the data who leverage DOIs to access data but are not given a direct and obvious path to all relevant metadata might otherwise misinterpret shared data accessed through a DOI.
4. Section 4.5, Public Access to Scientific Data: This section lists certain scientific data repositories (with hyperlinks) maintained by NASA, but not others. This does not seem sustainable, and could do a disservice to readers of the plan, who may interpret omitted repositories as not being important and relevant points of public access to NASA managed scientific data. Attempting to make the listing complete risks omitting some, and also will not necessarily reflect the NASA scientific data repository landscape going forward. I suggest not including individual repository links in this section and instead only including the catalogs science.data.nasa.gov and data.nasa.gov.

Dan Berrios MD PhD MPH
NASA Ames Research Center
(415) 902 5099 (cell)