



Extraterrestrial Materials Analysis Group Response to RFI: Implementation and Changes to Science Policy Document (SPD)-41

Barbara Cohen¹, ExMAG Chair
Jemma Davidson², ExMAG Vice Chair
Pierre Haenecour³, Co-chair of ExMAG Facilities & Informatics Subcommittee
Rhiannon Mayne⁴, Co-chair of ExMAG Facilities & Informatics Subcommittee

This response was also reviewed by all current members of the Extraterrestrial Materials Analysis Group (ExMAG).

¹NASA Goddard Space Flight Center, Greenbelt MD 20771 (Barbara.A.Cohen@nasa.gov)

²Buseck Center for Meteorite Studies, School of Earth and Space Exploration, Arizona State University, Tempe AZ 85287-6004 (jdavidson@asu.edu)

³Lunar and Planetary Laboratory, The University of Arizona, Tucson, AZ 85721 (haenecour@arizona.edu)

⁴Monnig Meteorite Collection, 2950 West Bowie St, SWR 244, Texas Christian University, Fort Worth, TX, 76109 (r.g.mayne@tcu.edu)

Brief Summary: NASA's Science Mission Directorate SPD-41 Scientific Information Policy is a consolidation of existing U.S. Government policies and laws applicable to SMD scientific information. As part of the SMD Request for Information on proposed additions to SPD-41a, the Extraterrestrial Materials Analysis Group (ExMAG) is providing analysis, findings and recommendations to facilitate the implementation of the policy and reduce the burden put on the sample analysis community as part of the Planetary Science Division (PSD). Our recommendations include the following:

- **SPD-41a should include information and guidance on the plan, timeline, and funding for the policy implementation.**
- **SPD-41a should provide clear direction for investigators on current, NASA-compliant data repositories available to archive planetary sample analysis data.**
- **SPD-41a should specify that spacecraft mission activities involving sample analysis follow the same data archiving policy requirements as research data rather than mission data.**

Background: The information produced as part of NASA’s scientific research activities represents a significant public investment. Results of federally funded research and development need to be shared openly to enable transformational open science through the continuous evolution of science data and computing systems for NASA’s Science Mission Directorate (SMD). The proposed updated SMD information policy, SPD-41a (https://science.nasa.gov/science-red/s3fs-public/atoms/files/draft_SMD-information-policy-v2.x.pdf) is a consolidation of existing U.S. Government policies and laws applicable to SMD scientific information. The core values of the policy are to adhere to existing laws, policies, and regulations to maximize the openness of useful scientific information produced by NASA SMD funding, to preserve the scientific information produced by NASA SMD funding, and to grow the community that can access NASA’s scientific information. SMD has released a Request for Information (RFI) on proposed additions to SPD-41a.

The collection, curation, and analysis of astromaterials are core elements of NASA’s Planetary Science program. NASA-funded efforts include human and robotic sample collection and return missions (*e.g.*, Apollo, Stardust, Genesis, OSIRIS-REx, ANSMET), curation of these samples and other collections (*e.g.*, meteorite, interplanetary dust particles, micrometeorites, space-exposed hardware), and research efforts for sample analysis, including laboratory facilities. The data generated from these diverse efforts result in a broad range of raw data files, as well as calibrated and higher-level data products.

The Extraterrestrial Materials Analysis Group (ExMAG) is a NASA-chartered, community-based, interdisciplinary group providing a forum for discussion and analysis of matters concerning the collection, curation, and analysis of extraterrestrial samples. ExMAG is providing our comments on SPD-41a as it pertains to planetary science sample collection, curation, and analysis efforts. In this effort, we drew upon the expertise of ExMAG members and solicited community input via the Astromaterials Data Management in the Era of Sample-Return Missions Community workshop, held in Tucson (AZ) on November 8-9, 2021.

Analysis & Findings: Below, we provide our findings on potential issues with the implementation of specific policies listed in SPD-41a, followed by a substantiating discussion.

- 1) SPD-41a lacks an implementation plan, timeline, and information on SMD funding to support the policy implementation.
- 2) SPD-41a does not address the state and needs of community agreed-upon archival standards and NASA-compliant archives for data resulting from the analysis of samples to ensure compliance with FAIR Guiding Principles.
- 3) SPD-41a contains specific requirements for mission data that would negatively impact laboratory analysis of samples as part of sample-return missions.

1) SPD-41a lacks an implementation plan, timeline, and information on SMD funding to support the policy implementation.

While SPD-41a is already included in ROSES-22 and newly SMD-funded investigators will be required to comply with the policies, it does not provide a vision or plan on how the policies will be implemented, enforced, or funded. This information is key to proposers to ensure they properly scope their proposal plan and budget to account for the archiving responsibilities.

2) SPD-41a does not address the state and needs of community agreed-upon archival standards and NASA-compliant archives for data resulting from the analysis of samples to ensure compliance with FAIR Guiding Principles.

The General Policies section of SPD-41a states that “SMD-funded data shall include robust, standards-compliant metadata that clearly and explicitly describe the data.” However, the sample science community does not have common agreed-upon archival standards and systems for all the data and metadata derived from the analysis of astromaterials, including mission-returned samples, meteorites, micrometeorites, and interplanetary dust particles. Before the implementation of SPD-41a it is critical to develop community agreed-upon archival standards and metadata dictionaries for all the data and metadata derived from the analysis of astromaterials that follow FAIR (findable, accessible, interoperable, and reusable) guiding principles. The key strength of sample return missions, *i.e.*, the sheer breadth, and ever-expanding types of laboratory measurements that can be applied to the samples, is the key challenge for any data archiving system for the sample science community. Exacerbating the issue is that many SMD-funded analytical instruments produce data in proprietary formats that are not compliant with FAIR Guiding Principles.

SPD-41 further requires that “Information shall be archived in repositories that can maintain the information for a period of at least 25 years.” While the curation and management of NASA-held sample collections of astromaterials at Johnson Space Center has well-defined archival standards and information sharing for NASA collections, there is currently no long-term, NASA-funded, data repository to archive the broad diversity of data types generated by the sample analysis community of researchers (*e.g.*, text, tables, two- and three-dimensional images, spectra, data cubes, etc.). Currently, the majority of sample analysis data are contained in supplemental material deposited with journal publications, in formats of the authors’ choosing, which may not conform with the FAIR principles.

The Astromaterials Data System (AstroMat) is a data infrastructure to store, curate, and provide access to laboratory data acquired on samples curated in the Astromaterials Collections at the Johnson Space Center. However, AstroMat is currently not designed to archive the full range of data types, volumes, and formats generated for laboratory analysis of astromaterials. Furthermore, AstroMat development and hosting is dependent on short-term funding (regular proposal cycle), making it challenging to guarantee that all SMD information submitted to it will be maintained for at least 25 years.

The Planetary Data System (PDS) is a long-term archive of digital data products returned from NASA's planetary missions, and from other kinds of flight and ground-based data acquisitions. While the PDS has implemented the PDS4 archiving standard with a series of data dictionaries, they are not developed (and are not directly applicable) to metadata and data produced from the analysis of astromaterials. The PDS is currently not funded or set up to store all astromaterials analysis data, nor is it clear which node(s) would be charged with managing this broad range of data types.

SPD-41a mentions that repositories are preferred (but not required) to have peer review and public review. However, it is important to note that there is currently no common peer review process for all planetary science data repositories.

The policy also requires the use of persistent identifiers for all investigators supported by SMD funding and all SMD-funded data collections. A key finding from the Astromaterials Data Management workshop is that internationally recognized persistent identifiers, *e.g.*, International Geo Sample Number (IGSN), Digital Object Identifiers (DOIs) and Open Researcher and Contributor ID (ORCID) should be applied to all astromaterials collections and all information resulting from the analysis of the samples (data, metadata, software, and publications). Such identifiers would ensure the findability, cross-referencing and interoperability of information between multiple repositories.

3) SPD-41a contains specific requirements for mission data that would negatively impact laboratory analysis of samples as part of sample-return missions.

As described in the final report of the NASA Planetary Data Ecosystem Independent Review Board (<https://science.nasa.gov/researchers/science-data>), “there is no requirement levied upon sample return missions for the archival of mission-supported laboratory analytical data and metadata.” To be compliant with FAIR Principles, SPD-41a should also include clear guidance for archiving all curatorial and preliminary examination data from returned-sample missions.

However, sample analysis data collected as part of spacecraft missions needs to be treated separately and distinctly from traditional spacecraft mission data. SPD-41a states that “there shall be no period of exclusive access to mission data. A period after the data have been obtained may be allowed for activities such as calibration and validation of the data. This period shall be as short as possible and shall not exceed six months.” On the other hand, research data is only required to be made available “no later than the publication of the peer-reviewed article that describes it.” Laboratory analysis of returned samples during the sample analysis phase of a mission involves the coordination between many analytical techniques, and data sets from multiple instruments are often co-dependent from each other for publication. The release of mission data involving sample analyses (*e.g.*, data from laboratory analyses of returned samples) no later than six months after its acquisition would put unnecessary burden on mission teams and would potentially force them to release data before it has been fully reviewed, characterized, and published by the team. These issues would be alleviated if sample-analysis mission data were to follow the same policy requirements as research data.

Given the policy shortcomings listed above, the Astromaterials community will not be able to fully comply with SPD-41a. We provide below a list of recommendations for modifications to the policy document to address these shortcomings, and suggestions for the implementation of the policies to reduce the burden put on the community to comply with the data information policy.

Recommendations

- SPD-41a should include information and guidance on the plan, timeline, and funding for the policy implementation.
- SPD-41a should provide clear direction for investigators on current, NASA-compliant data repositories available to archive planetary sample analysis data.
- SPD-41a should specify that spacecraft mission activities involving sample analysis follow the same data archiving policy requirements as research data rather than mission data.

Suggestions for Policy Implementation:

- NASA should release to the astromaterials community the NASA catalog of data referenced in SPD-41a.
- As recommended in the final report of the Planetary Data Ecosystem Independent Review Board, create a shared, common taxonomy, controlled vocabulary, high level data dictionary, and/or glossary of terms across the Planetary Data Ecosystem.
- Develop community agreed-upon archival standards, non-proprietary formats, and metadata dictionaries for all the data and metadata derived from the analysis of astromaterials, along with consistent peer-review processes for metadata dictionaries and data submitted to repositories.
- Support the implementation of common persistent identifiers, e.g., International Geo Sample Number (IGSN), Digital Object Identifiers (DOIs) and Open Researcher and Contributor ID (ORCID), for all astromaterials sample collections, SMD data, and SMD-funded investigators.
- Provide funding for all SMD-funded scientific activities to comply with the scientific information (data, software, and publications) policy requirements, including providing long-term support for the astromaterials data repository(ies) to be compliant with maintenance of all the information for a period of at least 25 years and develop community agreed-upon archival standards and metadata dictionaries.
- Provide training (online and at conferences/workshops) for investigators on how to archive their data to meet FAIR principles and use the NASA compliant data repositories. The Coalition for Publishing Data in the Earth and Space Sciences (COPDESS) has identified a lack of clear and detailed guidelines, workflows, and policies that help authors, editors, and data curators to navigate FAIR data sharing and archiving.