

# Managing Analytical Data from Pristine Returned Samples in Compliance with NASA’S Data Strategy: The Astromaterials Data System

Kerstin Lehnert<sup>1</sup>, Lucia Profeta<sup>1</sup>, and Jennifer Mays<sup>1</sup>

<sup>1</sup>Lamont-Doherty Earth Observatory of Columbia University

November 24, 2022

## Abstract

Analytical studies of astromaterials samples returned by NASA space missions generate unique and highly valuable data that contribute fundamentally to our knowledge and understanding of the origin and evolution of Earth, our solar system, and the universe. These data need to be openly accessible and curated in a manner that maximizes their reuse in and utility for future science and that ensures their quality and long-term preservation. In several recent strategic documents and reports, NASA recognizes this need [1] and is adjusting its science information policies [2]. In 2020, NASA charged the Planetary Data Ecosystem Independent Review Board (PDE-IRB) to conduct a review of the planetary data landscape and make recommendations for improving access to and use of planetary science data by the science community [3]. This presentation will highlight features and services of the Astromaterials Data System that align the IRB’s recommendations. The Astromaterials Data System (Astromat) is a data infrastructure that has been funded by NASA since 2018 to curate, archive, and publish analytical data that are generated from astromaterials samples collected by NASA missions and curated at the Johnson Space Center in the Astromaterials Research & Exploration Science Division. Astromat’s mission is to: preserve astromaterials data and ensure their long-term access and reusability for new science endeavors; restore legacy data of astromaterials samples acquired in the past; synthesize historic and new data into a comprehensive, analysis-ready data store that allows scientists to use new technologies such as Machine Learning and Artificial Intelligence to explore and mine these data in previously impossible ways. Astromat operates a data repository where researchers can deposit their data for archiving and publications, specifically to comply with new journal policies and guidelines for Open and FAIR data and Data Management Plans required by funders. The repository follows international best practices. Astromat also maintains the Astromat Synthesis, a relational database that integrates legacy and new data into a harmonized data collection that allows users to find and extract data at the granularity of individual analytical measurements and combine these into customized new compilations for advanced data analysis. [1] SMD’s Strategy for Data Management and Computing for Groundbreaking Science 2019-2024. [2] Scientific Information policy for the Science Mission Directorate, SMD Policy Document SPD-41 (August 2021). [3] Besse, S., et al. (2021). LPI Contributions 2549, 7070.

# MANAGING ANALYTICAL DATA FROM PRISTINE RETURNED SAMPLES IN COMPLIANCE WITH NASA'S DATA STRATEGY: THE ASTROMATERIALS DATA SYSTEM



80NSSC19K1102

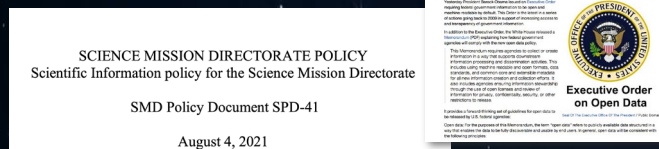
K. Lehnert, L. Profeta, J. Mays

Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

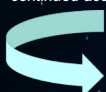
- Analytical studies of astromaterials samples returned by NASA space missions generate unique and highly valuable data that contribute fundamentally to our knowledge and understanding of the origin and evolution of Earth, our solar system, and the universe.
- These data need to be openly accessible and curated in a manner that maximizes their reuse in and utility for future science and that ensures their quality and long-term preservation.

## Open Data

In 2013, the Obama Administration released an Executive Order that required federal government information to be open and machine-readable by default.



In several recent strategic documents and reports, NASA recognizes the benefits of open data and is adjusting its science information policies. In 2021, the new Science Mission Directorate Policy SPD-41 states: "Results of federally funded research and development need to be shared openly in order to maximize the benefit and reach of the information. Data need not only to be archived but also to be curated – that is, the data are assured to have continued accessibility and usability for multiple decades."



- Applies to Publications – Data – Software
- ✓ produced by SMD-funded missions.
  - ✓ produced by investigations funded via research awards.

## FAIR Data

- SMD-funded data shall be made publicly available without fee or restriction of use.
- Data formats shall be machine-readable (i.e., data are reasonably structured to allow automated processing).
- SMD-funded data shall include robust, standards-compliant metadata that clearly and explicitly describe the data.
- SMD-funded data shall be reusable with a clear, open, and accessible data license.

### The FAIR Principles

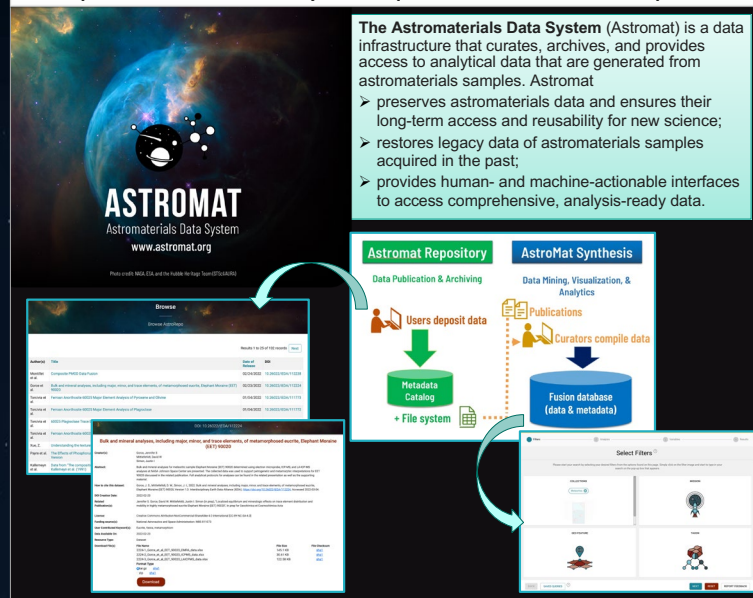


### FAIR Data

- ✓ should be accessible and understandable to humans & machines.
- ✓ should be deposited in certified trusted repositories to ensure persistent access & preservation.
- ✓ should be citable (registered with unique persistent identifier systems) and should have clear usage licenses.
- ✓ should be documented by rich metadata that support discovery and reuse.



AstroMat aims to facilitate discovery, access, mining, publication, and preservation of laboratory data acquired on astromaterials samples.



AstroMat's systems, services, policies, procedures, and management have been designed to comply with the FAIR and TRUST\* principles for repositories. AstroMat also follows best practices for laboratory analytical data established in the geochemistry community and FAIR physical samples that are emerging through various initiatives. (\* TRUST = Transparent, Responsible, User Focused, Sustainable, Technology)



In Nov 2020, the PSD commissioned the IRB to conduct a thorough review of the Planetary Data Ecosystem, which includes the user community and the organizations and facilities involved in planning, obtaining, analyzing, preserving, and sharing data from planetary space missions.

<https://science.nasa.gov/researchers/science-data>

Important IRB recommendations are already fulfilled by the Astromaterials Data System:

Recommendations of the IRB	AstroMat Compliance
R9: NASA should seek CoreTrustSeal certification, and thereby WDS membership, for the PDS data nodes. NASA should encourage CoreTrustSeal certification for other PDE elements that serve as data repositories	AstroMat has already completed its application for CoreTrustSeal Certification.
R33: NASA should establish a requirement for the preservation of mission-supported laboratory analyses of returned sample material that makes the information accessible to the planetary science community	AstroMat provides the infrastructure and services that support the preservation and access of laboratory analyses of returned sample material.
R47: NASA should support and encourage expanded use of DOI-like identifiers for data, thereby connecting data at various levels of processing to assist users in locating the best version of a data set for their needs.	The AstroMat Repository registers data with DataCite to assign DOI. AstroMat also is working to assign persistent, globally unique identifiers (IGSN) to the samples.
R49: NASA should fund the development of more analysis-ready data (ARD) products derived from the lower-level products created by NASA missions.	AstroMat's Synthesis Database delivers astromaterials data analysis-ready via machine-actionable interfaces.