THE ASTROMATERIALS DATA SYSTEM: ARCHIVING ASTROMATERIALS SAMPLES DATA TO ENABLE OPEN SCIENCE AND NEW DATA-DRIVEN RESEARCH PARADIGMS. K. A. Lehner1, J. Mays1, R. Downs2 P. Ji1, M. Celnick1, S. M. Richard1, A. Johansson1 Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, lehnert@ldeo.columbia.edu, 2Affiliation for second author (full mailing address and e-mail address).

Introduction: Invaluable specimens of rocks, soil, and dust particles have been collected from the Moon, asteroids, and interplanetary space by human and robotic sample collection and return missions such as Apollo, Stardust, Genesis, and Hayabusa to advance our understanding of the Solar System. A new generation of sample-return missions is now underway (OSIRIS-REx) or in preparation (Artemis, Mars) that will involve laboratories across the globe to characterize chemical, physical, and mineralogical properties of the next suite of samples collected in space, generating a never before seen volume of lab analytical data. Rigorous curation, preservation, and open dissemination of these data is vital to leverage the gigantic investments in sample-return missions; support new scientific inquiry and new data-driven research paradigms; help prevent unnecessary use of the precious samples by avoiding duplicate analysis; and provide learning opportunities beyond the immediate astromaterials research community for researchers, students, and instructors from other disciplines and the general public. The Astromaterials Data System (Astromat) is a digital data infrastructure funded by NASA to make astromaterials samples data persistently discoverable and accessible by humans and machines; firmly integrate them into the planetary data ecosystem; and enable researchers to reuse the data with confidence.

Astromat aims to provide reliable, standards-compliant, scalable, and adaptable data services for NASA, the global planetary science community, and the general public in accordance with NASA’s Open Science policies. Astromat’s priority is to assist NASA’s Planetary Science Division in achieving its science goals by operating services that support NASA-funded sample scientists and collection curators with archiving, broadly and openly disseminating, accessing, and using astromaterials samples data.

Status of Astromat Data Services. Since 2018, Astromat has been developing and operating technical and human infrastructure for the preservation, curation, and dissemination of astromaterials samples data, and for the rescue of historic data. Astromat is now offering services that are used by researchers to (1) share and archive data in compliance with data policies of publishers and funders, following best practices for geoanalytical data adopted by the research community, and (2) access and reuse >50 years of published analytical data measured on specimens in NASA’s astromaterials collections. Astromat has designed its systems, services, policies, procedures, and management processes to comply with the FAIR principles for data [1] and with the TRUST principles for Transparent, Responsible, User Focused, Sustainable, and Technologically stable repositories [2]. Astromat consists of two systems as described below.

Astromat Data Archive (ADA). The ADA offers repository services optimized for astromaterials samples data with an online data deposition and metadata entry tool; data templates; review of submitted datasets by curators; DOI registration; licensing; dataset landing pages; and machine-readable catalog metadata. A revision of the ADA is underway to improve and scale the infrastructure to meet the anticipated new demands for sample analysis data of future sample-return missions. Services for archiving sample analysis data of the OSIRIS-REx mission are already in development (see section ‘SAMIS-ADA Pipeline’).

Astromat Data Synthesis is a relational database designed to offer end-users a comprehensive, fully integrated, machine-readable, and analysis-ready data store from which they can extract individual measurements from any number of datasets and publications and create customized data compilations that can be explored, visualized, and analyzed with new computational methods (e.g. Artificial Intelligence, Machine Learning, Neural Network Analysis). The Astromat Data Synthesis today contains nearly 1.2 million data values with rich metadata about the data provenance.

Community Engagement is a critical component of Astromat’s mission. In order to support NASA’s commitment to Open Science, Astromat has been working to establish a cohesive community of data producers, users, and other stakeholders (funders, publishers, professional societies, etc.), who share an interest in practicing open science to facilitate the sharing and use of astromaterials samples data. Among other activities, Astromat is actively engaging with the Extraterrestrial Materials Analysis Group (ExMAG).

Astromaterials Data Archiving Special Study: In March 2022, NASA commissioned the Astromaterials Data Archiving (ADA) Special Study to gather requirements for the archiving of astromaterials samples data from sample-return missions and research projects that can fulfill the demands of NASA’s Open
Science and data archiving policies, international repository standards, and respond to the needs outlined in the report of the Planetary Data Ecosystem Internal Review Board [3] and in the OSIRIS-REx Data Management Plan [4]. Recognizing the particular features of sample data (heterogeneous; hierarchical; additive and distributed data collection), requirements were identified across all stages of the astromaterials data lifecycle that include the technical, organizational, and social aspects of operating trustworthy archiving services. A report on the study, which includes recommendations, was provided to NASA [5]. Astromat’s existing capabilities fulfill many of the requirements pertaining to the collection, acquisition, curation, and publication of astromaterials samples data, but need to be extended and upgraded. Astromat provided an Implementation Plan to NASA in Fall 2022 for enhancing its capabilities to fulfill all requirements of the future astromaterials data archive.

**Implementing the Recommendation of the Special Study.** Astromat will implement required capabilities identified in the Special Study in a phased approach. Phase 1 and 2, which focus on developments needed to fulfill the demands of the OSIRIS-REx Sample Analysis Data Management Plan, started in early 2023 with the development of the new ADA architecture and the SAMIS-ADA pipeline. Phase 3 will start in Fall 2023 with a new funding cycle.

**New ADA Architecture.** The next version of the ADA infrastructure is described by Ji et al. in a separate abstract [6]. It will be scalable to the growing diversity and volume of data, and will provide (semi-) automated data validation, DOI registration, and metadata parsing into the ADA catalog.

**SAMIS-ADA Pipeline.** Since 2021, Astromat has worked with the SAMIS team at the Lunar & Planetary Lab to establish capabilities for routinely transferring OSIRIS-REx sample analysis data from the SAMIS to Astromat. The SAMIS-ADA pipeline includes the transfer of data packages into the archive; parsing of metadata from metadata manifests in the packages into the ADA catalog; external peer-review of data; DOI registration; and generation of dataset landing pages.

**Archiving Standards for Astromaterials Samples Data.** Astromat aims to drive the development and broad adoption of community standards for interoperable and reusable geoanalytical data through close engagement of the geochemistry and scientific data stewardship disciplines. Astromat participates in the global initiative ‘OneGeochemistry’ [7] that has been endorsed by various scientific unions and societies, including the Meteoritical Society, for its efforts to develop data standards for geo/cosmochemistry. OneGeochemistry will develop FAIR Implementation Profiles, and Astromat will engage community members in gathering FAIR Enabling Resources. By meeting the TRUST Principles, Astromat can offer resilient services and continuous improvements to earn and foster trustworthiness among data producers and end-users. Astromat will assess its efforts for compliance with the CARE Principles for Collective benefit, Authority to control, Responsibility and Ethics [8], so that it continually recognizes and respects the needs of indigenous communities and their lands. Astromat will pursue trustworthy digital repository certification [9].

**Peer-Review of Data.** Facilitating the use of astromaterials samples data also includes ensuring that disseminated data products have been reviewed to meet the community’s expectations for data quality and usability. The peer-review process and guidelines for reviewers will be developed. We envision encouraging researchers to publish their analytical protocols and workflows in peer-reviewed journals so they can refer to such method publications in order to streamline the review process. The external review will involve an editorial board composed of community volunteers with expertise in the particular methods utilized to produce the data. A final approval by a Configuration Management Board will be required to verify that the data have met the conditions recommended for public release. Community members will have an opportunity to comment on data they reuse to conduct new studies.

**Community Engagement.** The success of Astromat is dependent upon the community’s use of the data it hosts. Online instructional resources, guidance materials, and personal user support will offer direction and help for producers to deposit data and for data users to search and retrieve data. In-person and online events will be offered to maintain a close connection to the community and to identify opportunities for further enhancements to meet the needs of the community. A Community Advisory Board will further strengthen community input.

**Acknowledgments:** This abstract is based upon work supported by the National Aeronautics and Space Administration under grant no. 80NSSC19K1102.