Planning and Implementation of Data Archiving for Sample Return Missions: Lessons Learned from OSIRIS-REx

P. Haenecour[1], C. A. Bennett[1], K. A. Lehnert[2], T. L. Johnson[3], J. Mays[2], J. J. Barnes[1], H. C. Connolly, Jr.[1,4,5], and D. S. Lauretta[1].

[1]*Lunar and Planetary Laboratory, The University of Arizona, Tucson, AZ. USA (haenecour@arizona.edu).*

[2]*Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA.*

[3]*Dust Data Management, LLC.*

[4]*Dept. of Geology, Rowan University, Glassboro, NJ, USA.*

[5]*Department of Earth and Planetary Sciences, American Museum of Natural History, New York, NY, USA.*

Introduction. The information produced as part of NASA's scientific research activities represents a significant public investment. Results of federally funded research and development must be shared openly, following FAIR (findable, accessible, interoperable, and reusable) guiding principles to enable transformational open science through the continuous evolution of science data and computing systems for NASA's Science Mission Directorate (SMD). The SMD information policy, SPD-41a, consolidates existing U.S. Government policies and laws applicable to scientific information and, in many instances, is already part of NASA solicitations for funding [1]. The requirements of SPD-41a have been incorporated into ROSES-2023 solicitations, and new missions (pre–Key Decision Point-B) are expected to comply with SPD-41a fully. Grants awarded before ROSES-2023 and existing missions are not required to adopt the new policy requirements, but they are encouraged to do so if feasible with available resources [1].

NASA funds human and robotic sample collection and return missions (e.g., Apollo, Cosmic Dust, Stardust, Genesis, OSIRIS-REx, ANSMET); curation of these samples and other astromaterials (e.g., meteorites, micrometeorites, interplanetary dust particles); 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 and calibrated and higher-level data products that today need to be deposited and published in a publicly accessible archive that follows national and international guidelines for trusted repositories (e.g., US National Science and Technology Council guidelines 'Desirable Characteristics of Data Repositories for Federally Funded Research'; CoreTrustSeal Standards and Certification Board 2022).

NASA OSIRIS-REx Mission. The OSIRIS-REx Sample Analysis Team (SAT) comprises over 240 members, 65 analytical techniques, and 265 individual instruments distributed across dozens of laboratories worldwide. The team developed a comprehensive information system called the Sample Analysis Micro-Information System (SAMIS) to manage all the resultant data and maximize the scientific outputs [2, 3]. SAMIS collects, stores, processes, and distributes all data pertaining to the sample analysis phase of the mission (Figure 1). This encompasses data about the physical state of the samples as well as raw data (measurement signal as the instrument records it in a format that is accessible without the use of proprietary software and with as minimal processing as possible), calibrated data (application of all corrections and calibrations to raw data to produce a science-ready data product), and documentation of all algorithms, software, and/or other relevant analytical data used to produce calibrated and higher-level data products. The SAMIS team developed this system in collaboration with SAT members and the Astromaterials Data System (Astromat; https://www.astromat.org) [4, 5] to establish standardized formats and requirements for all data. These standards ensure that the data can be processed by SAMIS and involve the conversion of proprietary formats into non-proprietary formats and the addition of contextual metadata files to the data bundles. SAMIS also serves as an intermediate platform to aggregate and store all SAT members' data before submitting them to Astromat. Once verified, the data products are assigned permanent DOIs by Astromat, which become immediately available to the team. SAMIS then automatically enhances these products by adding archive-quality metadata, which includes persistent, resolvable identifiers such as institution Research Organization Registry (ROR), Open Researcher and Contributor Identifiers (ORCIDs), and specifics about the instrument(s). Implementation of the IGSN as a persistent identifier for samples is in progress. Finally, SAMIS automatically sends all archive-ready sample analysis data bundles to Astromat, where they are reviewed and stored until SAT members are ready to release them publicly. The data products can be easily cited by their DOIs in peer-reviewed publication(s). Figure 2 shows an example of an OSIRIS-REx data bundle landing page.

The Astromaterials Data System. Astromat is the primary NASA-sponsored archive for laboratory analyses of returned samples [4, 5]. As mentioned above, Astromat has been working closely with the SAT and SAMIS to develop a more comprehensive data management structure and requirements that are currently specific to the OSIRIS-REx mission's needs and restricted timeline but could be updated and expanded to the broader sample analysis community and future sample return

missions (e.g., Artemis, Mars Sample Return). Astromat is also responsible for developing and implementing data peer-review procedures [5]. A working group of the Facility & Informatics Subcommittee of the Extraterrestrial Materials Analysis Group (ExMAG) serves as a community advisory resource for Astromat to provide feedback on development work. Astromat also participates in the OneGeochemistry initiative to coordinate its protocols and policies with the development of broader international data standards.

Initial Lesson Learned from OSIRIS-REx Sample Analysis Data Management. The comprehensive OSIRIS–REx Sample Analysis Plan and the large number and diversity of analytical techniques on the mission [6] portend the complexity and challenges associated with such peer-review procedures. With the SAT actively analyzing Bennu samples and the recent release of the first peer-reviewed publication from this effort [7], we will report the initial lessons learned and challenges associated with managing and archiving sample analysis data.

Acknowledgment. Supported by NASA under Contract NNM10AA11C issued through the New Frontiers Program.

Figure 1. The SAMIS homepage.

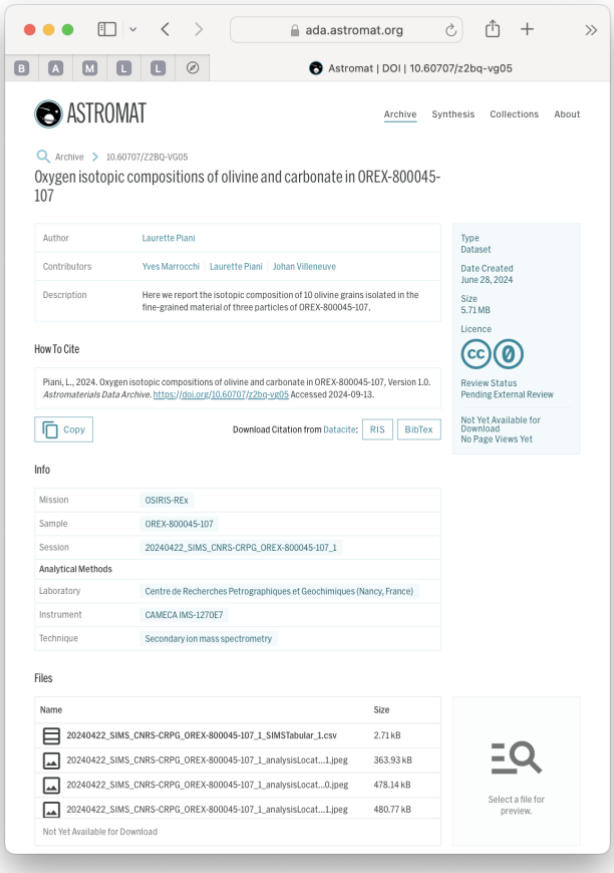

Figure 2. Example of a landing page of OSIRIS-REx sample analysis data on the Astromat website (https://ada.astromat.org/doi/10.60707/z2bq-vg05).

References

[1] https://science.nasa.gov/researchers/open-science/science-information-policy/. [2] Bennett et al. (2022), 85th Annual Meeting of The Meteoritical Society, abstract #6498. [3] Johnson et al. (2023), 6th Planetary Data Workshop, abstract #7077. [4] https://smd-cms.nasa.gov/wp-content/uploads/2023/05/PSDInformationDataManagementPolicy.pdf. [5] Lehnert et al. (2024), 87th Annual Meeting of The Meteoritical Society, abstract #6470. [6] Lauretta et al. (2023), https://arxiv.org/abs/2308.11794. [7] Lauretta & Connolly et al. (2024), Meteoritics & Planetary Science 59, 2453–2486, https://doi.org/10.1111/maps.14227.