

Town Hall & Lunch at MetSoc 2023

August 15, 2023

Kerstin A. Lehnert & the Astromat Team

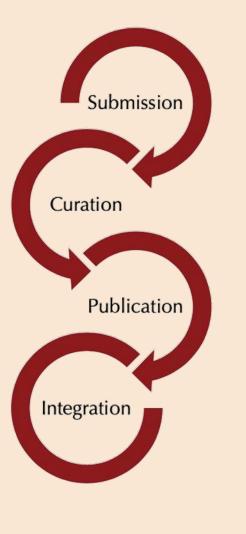
Columbia University, Lamont-Doherty Earth Observatory lehnert@ldeo.columbia.edu

ASTROMAT Astromaterials Data System

www.astromat.org

What is Astromat?





Astromat is a data system that supports the extraterrestrial samples community in making data generated by laboratory analysis of **Astromaterials samples**

Findable, Accessible, Interoperable, & Reusable





Accessible





- Data need to be FAIR in order to fulfill the promises of Open Science.
- FAIR research results maximize the return on NASA's investment in the acquisition and creation of planetary science information (including samples!) as well as the advancement of scientific knowledge.

What makes data FAIR?

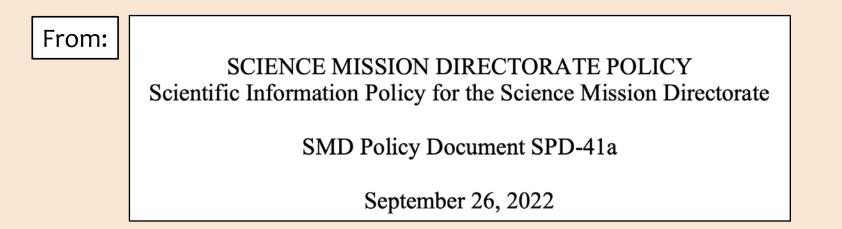


- FAIR data are accessible and understandable to humans & machines.
- FAIR data are deposited and curated in certified trusted repositories to ensure persistent access & preservation.
- FAIR data are citable (registered with Persistent Identifier systems) and should have clear usage licenses.
- FAIR data are documented by rich metadata that support discovery and reuse.

NASA embraces FAIR

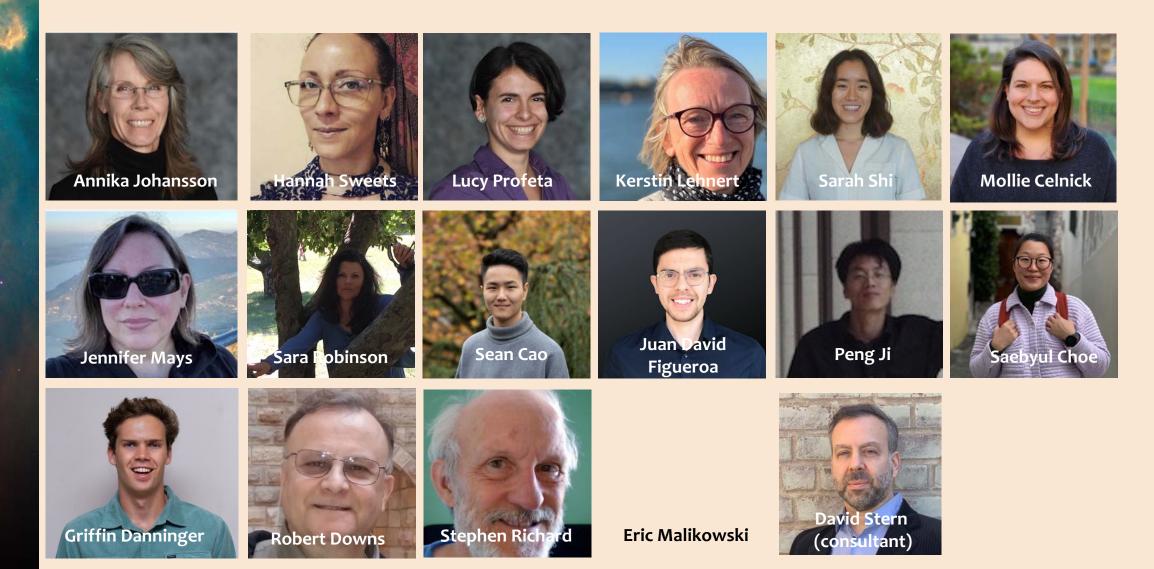


- SMD-funded data shall be made publicly available without fee or restriction of use.
- Data formats shall be machine-readable (i.e., data are reasonably structured to allow automated processing).
- SMD-funded data shall include robust, standards-compliant metadata that clearly and explicitly describe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



The Astromat Team



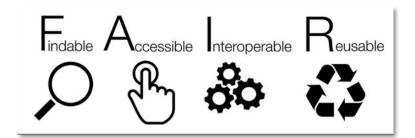


Community-driven & standards-compliant



- Astromat aims to align its services with evolving science needs.
- Astromat engages the community in the development and promotion of community best practices & standards.
- Astromat complies with operational standards for trusted data archives.

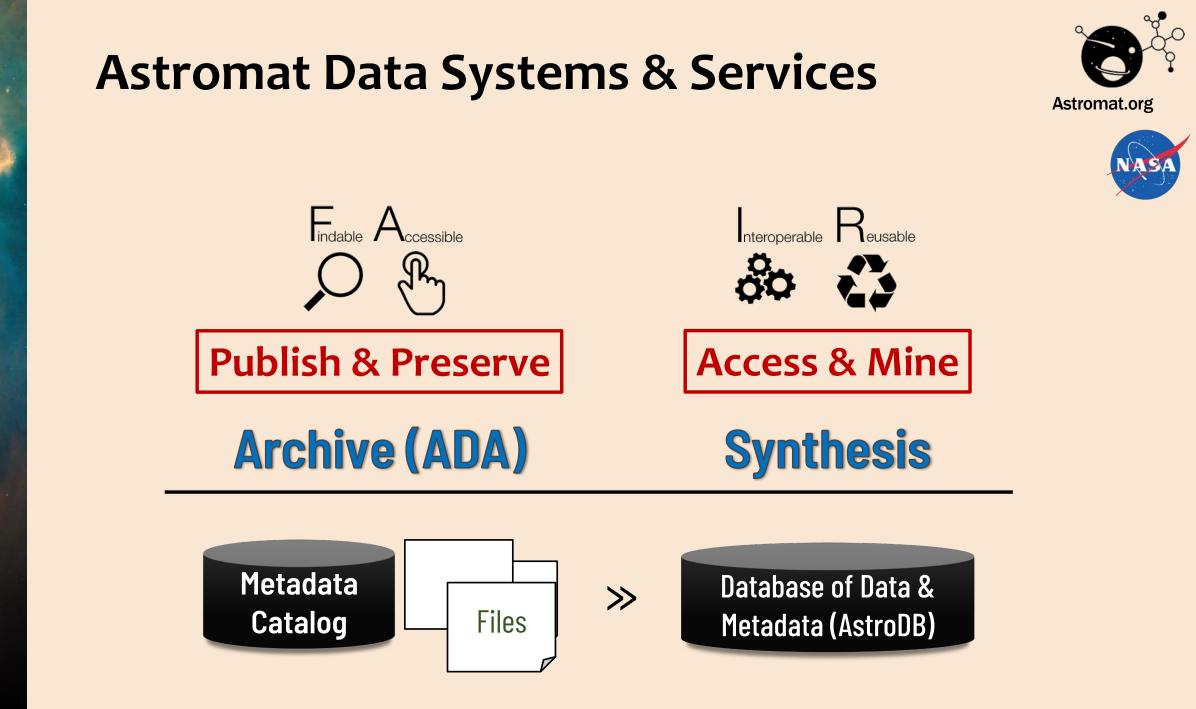
Astromat aims to operate its data systems and services following existing & emerging principles (FAIR, TRUST) and data & repository standards and best practices.



Wilkinson et al. (2016) "The FAIR Guiding Principles for scientific data management and stewardship"



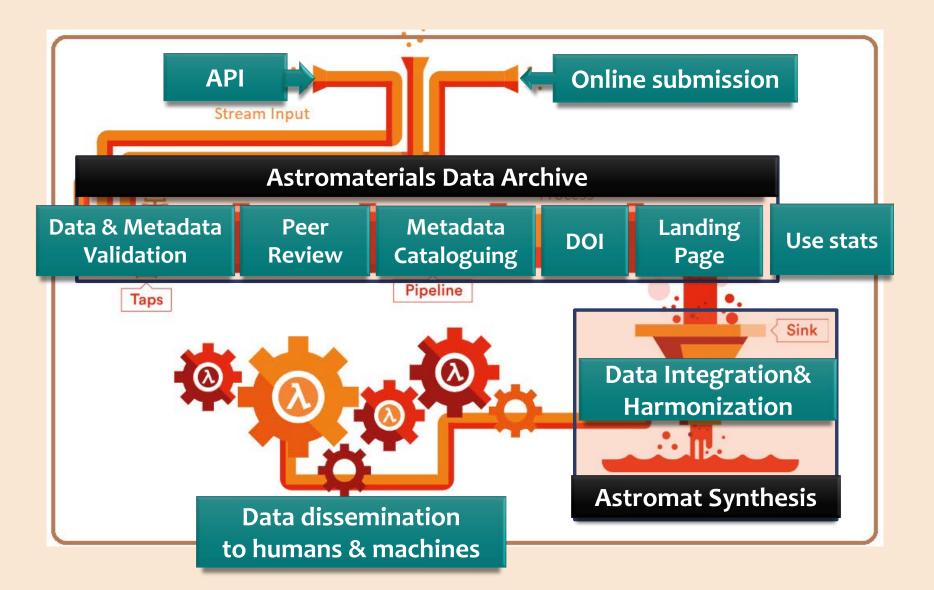
Lin et al. (2020) "The TRUST Principles for digital repositories"



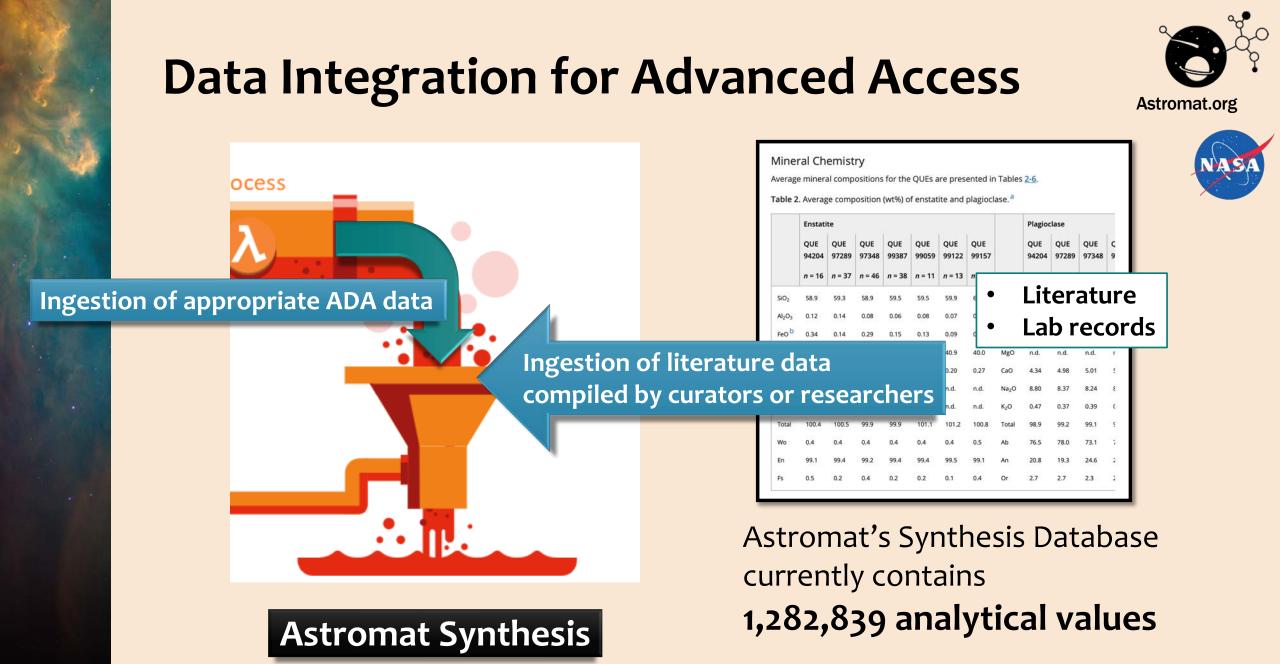
Astromat's Data Pipeline







Credit: https://vibhor-gupta.medium.com/what-are-the-main-building-blocks-of-a-data-pipeline-9cob79a0dcf9



Data Dissemination

Unique identifier

https://doi.org/10.26022/IEDA/111772

SUBMIT DATA . COLLECTIONS

ACCESS DATA .

· ASTROMAT

Creator(s);

Abstract

License

References



Dataset Information DOI: 10.26022/IEDA/111770 60025 Plagioclase Trace Element Data Torcivia, Michael A: ORCID: 0000-0002-9873-9932 Neal, Clive R; ORCID: 0000-0003-0705-3490 Various trace element https://repo.Astromat.org/astro_search.php These data were gathe lunar material. How to cite this datase Torcivia M. A. Neal C. R. 2 https://doi.org/10.26022/IEDA/111770, Accessed 2023-05-21. DOI Creation Date 2022-01-04 Torcivia, M. A., Neal, C., R., Unraveling the components within Apollo 16 Ferroan Anorthosite Suite Cataclastic Anorthosite Sample 60025: Implications for Publication(s) the Lunar Magma Ocean Model. Journal of Geophysical Research: Planets (in prep.) Publication DOI: 2020. (F00679988 Creative Commons Attribution-ShareAlike 4.0 International [CC-RY-SA-4.0] National Aeronautics and Space Administration: 80NSSC17K0467 Funding source(s) National Aeronautics and Space Administration: NNX15AH76G 60025, plagioclase, trace element, LMO User Contributed Keyword(s) Data Available On 2022-01-04 Dataset Resource Type Download File(s) File Name **File Size File Checksun** 1770-1_AMDR_Torcivia_Neal_60025_Plagioclase_Trace_Elements.xlsx 504.89 KB sha1 Download **Related Information** sReferencedBy: DOI: 2020JE006799RF sSupplementedB DOI: 10.26022/IEDA/111771 DOI: 10.26022/IEDA/11177 ALH84007 14C age measuerment JULL, 1998 Rock DOI: 10 https://datasetsearch.research.google.com/ Vocabularies/collection 1 dataset found ALH84007 14C age measuerment JULL,1998 Rock Apollo S Missie Ferroan Anorthosite 60025 Major Element Analysis GET /vocabularies/collectionAgg Retrived all collections appeare in the sample_data index. C Related Article Ferroan Anorthosite 60025 Major https://search.Astromat.org Vocabularies/expedition Element Analysis of Plagioclase repo.astromat.org 2 commons.datacite.org Explore at: Z repo.astromat.org GET /vocabularies/expeditionAgg Retrieve all expeditions appeared in the sample_data index. ALH 84007 https://ref.Astromat.org commons.datacite.org ALH 84007 E xlsx Vocabularies/geoFeature E xlsx Updated Jan 4, 2022 ALH 84008.84 Highly siderophile elem VAN ACKEN.2012 Rock GET /vocabularies/geoFeatureAgg Retrieve all geoFeatures appeared in the sample_data index.

The Last 12 Months:



- Completed NASA Special Study "Requirements for Archiving Astromaterials Samples Data"
 - provided to SMD on August 31, 2022.
 - Downs, R. R., Lehnert, K. A., Mays, J., 2022. <u>https://doi.org/10.26022/IEDA/112690.</u>

• Started implementation of the Astromaterials Data Archive (ADA)

- Initial focus on establishing the technical capabilities & capacity necessary to fulfill the requirements of the OSIRIS-REx Sample Analysis Mission Data Management Plan;
- Proposed plans for Astromat operation & development 2023 2028
 - Collaborative with NASA Mission Cloud Platform
 - Notification of selection received July 2023.

The Next 5 Years



- Development of the full Astromaterials Data Archive;
- Operation of archiving services for Astromaterials samples data;
- Long-term preservation of Astromat data in collaboration with NASA's Mission Cloud Platform team;
- Services for advanced data access;
- Engagement with the community and NASA to ensure that the requirements of SPD-41a are met throughout the development of Astromat.

Future Astromaterials Data Archive (ADA)



- Will have the **capacity** to appraise, curate, and preserve the complete range of laboratory analytical data of Astromaterials samples;
 - Standards compliant data curation framework;
 - (Semi-)automated transfer of Astromaterials samples return-mission data from mission-internal data system (e.g., SAMIS) to the ADA;
 - Interfaces for users to interactively find, access, assess, download, analyze, or integrate Astromaterials samples data;
 - Automated distribution of metadata to publicly accessible catalogs and harvesters.

Standards-compliant Data Curation Framework

- 1. Repository Certification (application to CoreTrustSeal)
- 2. Procedures & policies for external peer review
- 3. External Advisory Committee for community governance
- 4. Development and adoption of **community standards & best practices** for laboratory analytical data

Astromat's Approach to Peer Review



- Peer review has been part of data curation at Astromat, but internal.
 - Validation of metadata completeness & correctness.
 - Guided by FAIR principles and community best practices (note: no comprehensive standards for lab analytical data exist yet);
 - Performed by Astromat data curators with background in geo/cosmochemistry lab analysis;
- Plans for external peer review progressing.
 - Align with procedures & policies of the PDS to the degree possible.
 - Engage the community in defining a realistic process (ExMAG Working Group)
 - Streamline the review process.

Streamlining Peer Review



- Pre-Review: Support for data producers
 - Provide data producers with extensive guidance (templates, training) and support to optimize compliance with standards.
 - Encourage data producers to publish analytical protocols as peer reviewed journal articles.
 - Validate submitted data & metadata automatically before review.
- Review: Provide an online system to manage reviews.
 - Support for reviewers to conduct & submit reviews.
 - Online support for data producers to respond to reviews
- Post-review: Recognize and reward community members who serve as reviewers.

Community Standards & Best Practices



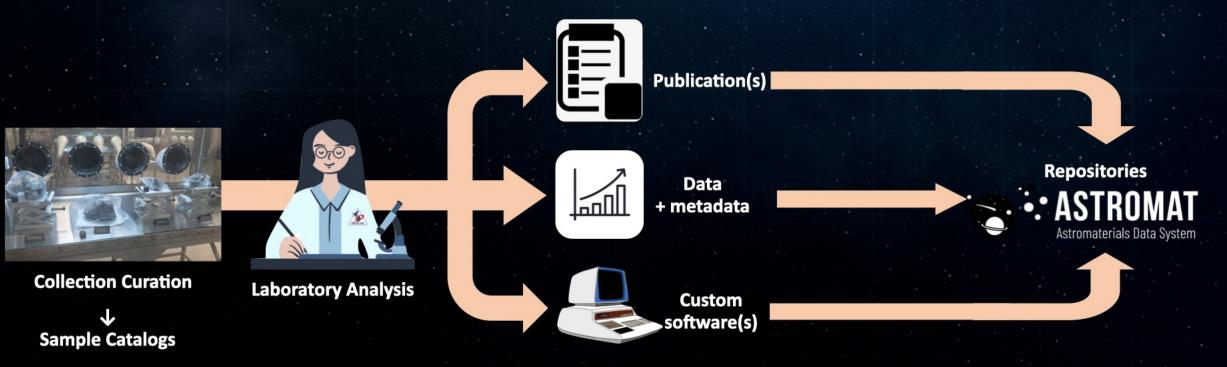
- Facilitate ExMAG WG to make recommendations
 - Evaluate OSIRIS-REx Data Standards
 - Follow-up activities for workshop on "Managing, Publishing, and Archiving Imagery Generated by Geochemical and Cosmochemical Sample Analysis" at Goldschmidt 2023
- Ongoing participation in OneGeochemistry and the WorldFAIR project to advance development of standards for FAIR laboratory analytical data that leverage existing standards
- Continue to coordinate with editors on journal data policies and submission workflows (COPDESS)



	The WorldFAIR Project https:/	/worldfair-project.eu
	Global cooperation o	n n n n n n n n n n n n n n n n n n n
	FAIR data poli and practice	CCY
	and practice	

Update on the OREx-Astromat Special Study and Need for Community Standards in the context of SPD-41a

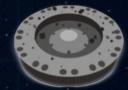
Pierre Haenecour



Extraterrestrial Materials Analysis Group (ExMAG) August Meeting Tuesday 08/01/2023



PSD Policy Supplement to SPD-41A



PLANETARY SCIENCE DIVISION POLICY

Information and Data Management Policy Supplement to SPD-41A

Overview

This PSD Information and Data Management Policy is a supplement to NASA's Science Mission Directorate (SMD) <u>SPD-41A Scientific Information Policy for the Science Mission Directorate</u>. It was developed by NASA's Planetary Science Division (PSD) to assist the planetary science community, NASA Program Officials, and NASA partners in meeting the requirements in SPD-41A for scientific information, which includes data, publications, software, and, for the purposes of PSD, physical samples. Descriptions of how an SMD-funded project will comply with SPD-41A must now be provided in an Open Science and Data Management Plan (OSDMP).

SMD has provided on its <u>Science Information Policy website</u> the <u>Open-Source Science Guidance</u> and an <u>FAQ</u> for general implementation of SPD-41A. This document provides the PSD-specific context for implementing SPD-41A for research, missions, and other activities; this includes information on:

- choosing repositories for data and software;
- developing solicitations and proposals;
- crafting OSDMPs;
- archiving mission data; and
- curating physical samples.

As NASA's approach to the management of scientific information is in transition, this document



ted as needed. PSD acknowledges that additional pository and trainings on best practices in eeded for the community to efficiently meet the olicy. At this time, PSD expects best effort in upport those needs. Additionally, it is possible that ailable, such as NASA-sponsored databases, may not

e available at https://science.nasa.gov/solar-system

ASA will support the evolution of those resources

Scan me to access PDF Archives and repositories should meet the guidelines for an SMD-acceptable Data Repository (SPD-41A, Appendix D)General Policies (work in progress)

SMD-funded data must be made publicly available as rapidly as possible:

- Mission data less than 6 months
- Research data no later than the time of publication (all other data by the end of the performance period)
- PSD Definition of Scientific Information
- All new missions (not reached KDP-B by Sept 26, 2022) and new research activities (after the release of ROSES-2023) are expected to fully comply with this policy

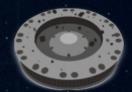


PSD Policy Supplement to SPD-41A



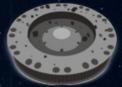
- SMD-acceptable Repositories:
 - Planetary Data System (PDS) = primary archive for NASA's planetary missions
 - NASA Space Science Data Coordinated Archive (NSSDCA) = permanent archive for NASA space science mission data and is the deep archive for PDS data.
 - Astromat = primary NASA-sponsored archive for laboratory analyses of returned samples
 - Additional NASA-supported Discipline Specific Repositories (e.g., NASA Astrobiology Environments Database, NASA Exoplanet Archive)
 - Data can also be archived in non-NASA-supported repositories (e.g., institutional repositories) and in Supplementary Materials in Publications if they comply with SPD-41a and this policy
- All physical materials covered by this policy must be registered with the International Generic Sample Number (IGSN) Organization





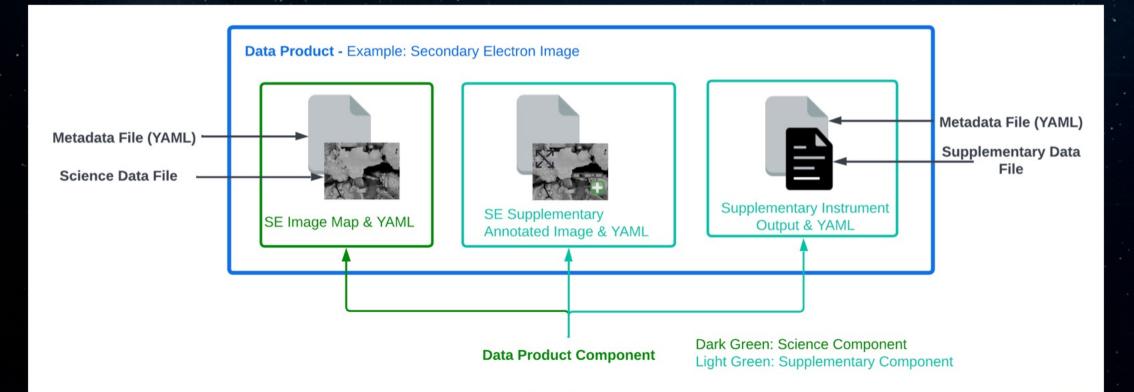
- OSIRIS-REx Sample Analysis Data will be archived with the Astromaterials Data System (Astromat; <u>https://www.astromat.org/</u>)
- Compliance with FAIR principles as far as possible within the timeline given
 - One of the key limitations is the lack of community-agreedupon standards and dictionaries
- What is a Data Standard? Set of rules that details how to effectively
 organize, describe, and format data in a uniform and consistent manner.
- Special Study with OSIRIS-REx and Astromat to develop common data product document structure and standards





What is a Data Product?

Data products are the digital output of scientific analysis or instrument operation.







Software Interface Specification (SIS) (by working group)

> Software Interface Specification Document

Explains the relationship between the **Data Standards Documents** and the **Bundle Delivery Documents** and lists/links the **Bundle Delivery Documents** under this working group. **Bundle Delivery Documents (BDD)** (by analytical technique)



Lists each product under the Analytical Technique, points to the correct **Data Standards Document** and lists any additional formatting or metadata requirements for the data product. Data Standard Documents (DSD) (by data product type)



Defines the acceptable file formats for this data type and lists the required metadata keywords for all products of this type. There will be one DSD for the following types of data formats:

- Document (narrative text, ex: PDFA)
- Tabular Data (CSV)
- Image (or 2D array)
- Data cube (or 3D array)



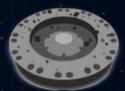


Software Interface Specification (SIS) Documents

organized by working group. These documents define and explain the relationship between the Bundle Delivery Documents (BDD) and the Data Standards Documents (DSD).







Data Standards Documents (DSDs)

define generic and universal standards for each type of data product. This includes data format requirements, required metadata keywords and formats, and labeling and identification requirements.



Data Cube (or 3D array) ↓ NetCDF, CDF, HDF5, mzML Image (or 2D array) ↓ TIFF, PNG, JPEG

Tabular Data ↓ csv, text Data Collection (sets of files that share a single metadata file)





Bundle Delivery Documents (BDD)

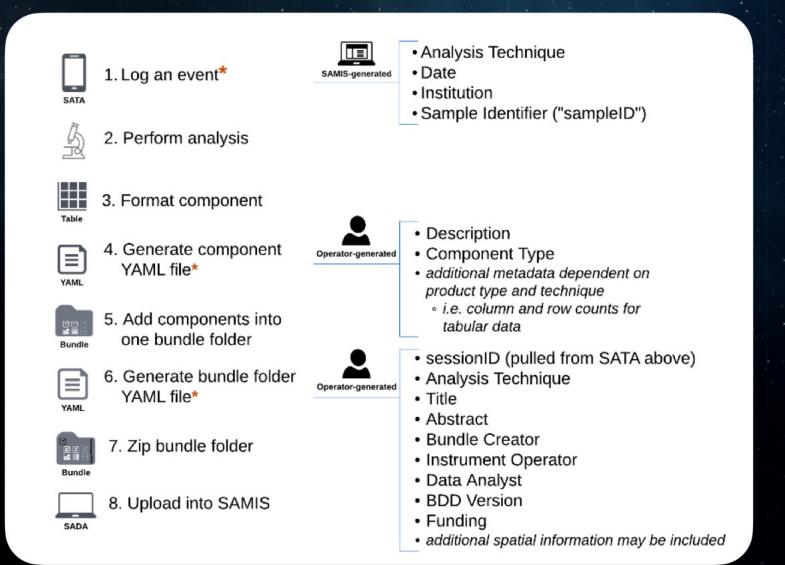
organized by analytical technique, and list data products that will be delivered for data from that technique, including formatting instructions, metadata keywords, or ancillary files or information that is unique to that analytical technique's data product.

EMPA	Raman	ХСТ	VLM	QRIS	GC-MS	LC-MS	VNMIR	NanoSIMS	SLS
μL²MS	FTICR-MS	SS-NMR	GC-C-IRMS	NMR	MC-ICP-MS	EA-IRMS	SIMS	XRD	SEM/FIB-SEM
TEM	EBSD/TKD	XANES	XPS	HR-ICP-MS	SHRIMP	LAF	APT	TIMS	NI-MI
Q-ICP-MS	FINESSE	NG-NS-MS	ICP-OES	GPYC	SCBTCA	DSC	HR-CL	EDS	EELS
NanolR	S-XRF	TGA	LA-ICP-MS	NI-NGMS	RI-TOF- NGMS	DESI- Orbitrap	SThM	PCD-AFM	PSFD
XRF	ARGT	SNMS	AMS	COMPT	SV-RUEC	CAPD	DSSM	LIT	ARM
IC	ToF-SIMS	CE-MS	S-IR						



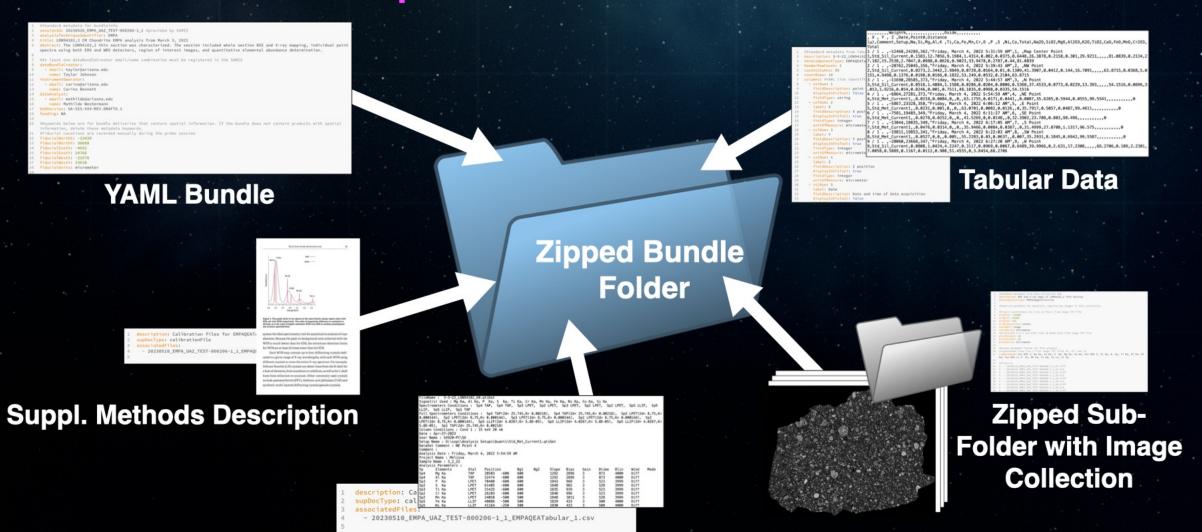
OSIRIS-REx Data Management Plan SAMIS Workflow for Scientists





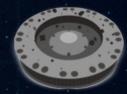


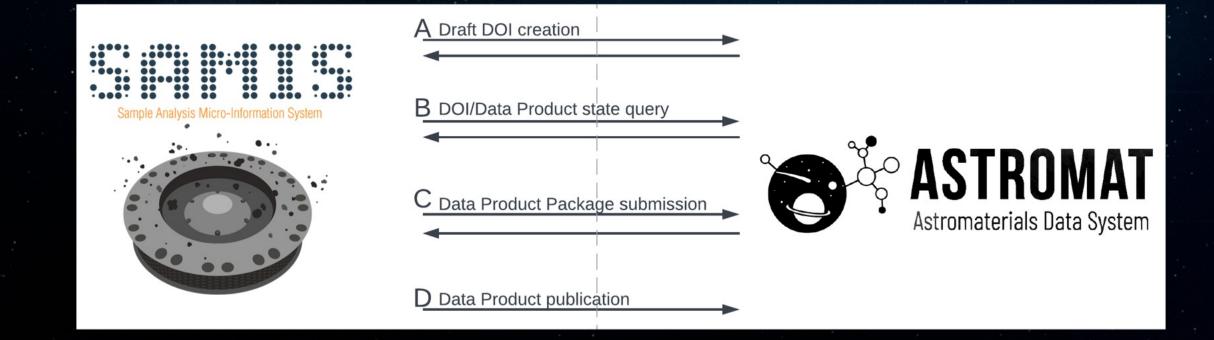
OSIRIS-REx Data Management Plan Example of a data bundle for EMPA



Suppl. Calibration File

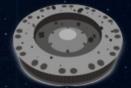
Delivery of Data Bundles to Astromat and DOI Allocations







Need For Community Feedback!



- OSIRIS-REx SAMIS Documents:
 - = best effort to comply with FAIR principles and SPD-41a
 - Developed with input from OREx SAT members (not broader community input)
- ExMAG Working Group on Astromaterials Data Management to:
 - Review the OSIRIS-REx Data Management Documents (data standard documents, and bundle delivery documents) and EarthChem/AstroRepo templates
 - Provide Feedback and advice to Astromat
 - Support Astromat with the development of community-agreedupon standards
- Email me (<u>haenecour@arizona.edu</u>) if you are interested in participating in the working group

FAIR Samples

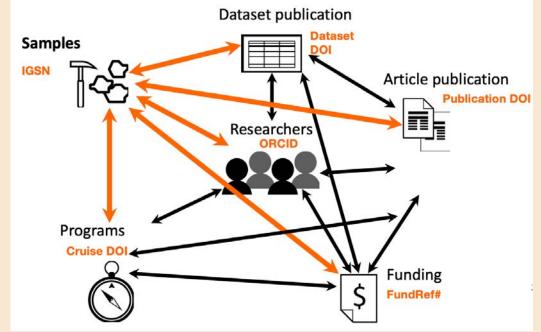


"PSD recognizes that the availability of archives and repositories for physical materials that meet the SPD-41A guidelines for an SMD-acceptable data repository (see SPD-41A, Appendix D) is a work in progress. For now, PSD requires that, at a minimum, all physical materials covered by this policy must be registered with the International Generic Sample Number (IGSN) Organization and the IGSN numbers must be used to cite the physical samples in publications. It is acceptable to register physical samples in the National Science Foundation (NSF) System for Earth Sample Registration (SESAR), which is an IGSN Allocating Agent. Additionally, an archive or repository for physical samples should: have a long-term (at least 25 years) commitment to curation, have a published loan policy and make loans to qualified researchers without discrimination and without fees beyond covering the cost of sample handling, maintain a public catalog with FAIR access, store the physical samples in a safe and secure environment, and publicly provide contact information for the curator responsible for making loans

Applying FAIR Principles to Samples



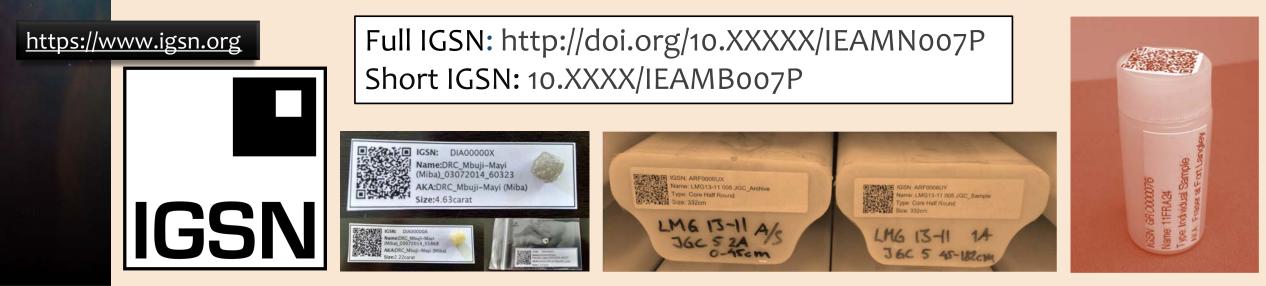
- Samples should be openly accessible, easily discoverable, and documented with sufficient information to make them reusable.
- Samples should be linked to the data derived from them (interoperable) and to the interpretations of these data published in the literature.
 - Use of Persistent Identifiers
 - for samples: IGSN
 - for researchers: **ORCID**
- Rich & standardized metadata following community standards
- Registered in indexed (meta)data repositories
- Clear licenses



IGSN International Generic Sample Number



- A globally unique and persistent identifier for physical samples and specimens
 - federates sample metadata catalogs
 - links publications to online sample metadata profiles
 - allows previously impossible linking and integration of sample-based observations across data systems
 - paves the road toward advanced data mining of sample-based data

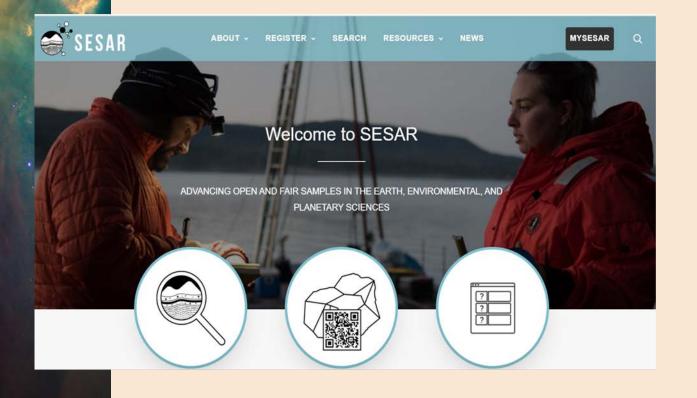




SESAR is funded as part of the IEDA2 Cooperative Agreement



Sesar System for Earth Sample Registrations



SESAR is a community platform that helps make samples more discoverable, accessible, and reusable, and connects samples with the knowledge ecosystem derived from them.

www.geosamples.org

- Registration of sample metadata with DataCite and minting of IGSNs
- Authenticated workspace for sample metadata submission & management
- Curatorial review of metadata and user support
- Sample metadata catalog for discovery/access human- and machine readable

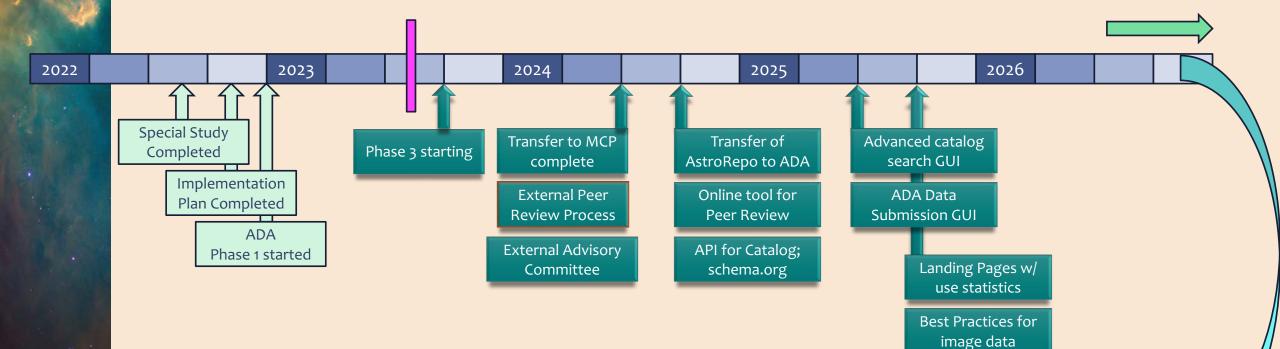
What Data Can I Submit to Astromat?



- Data Types: Laboratory analytical data that describe chemical, physical, petrological, mineralogical, and textural properties of bulk samples, mineral phases, and other components such as chondrules and presolar grains
 - tabular data, documents, spectra, images, and 3D arrays ('data cubes').
- Samples: Data of samples collected by NASA in past, present, and future sample-return missions, and samples that are curated by the JSC <u>Curation Office.</u>
 - Data from Astromaterials samples that do not fall into that definition (e.g., Chang'E-5 samples; meteorites not in the JSC collection) will be accommodated to the degree possible with available resources.

Timeline of Developments





- > Data pipeline from the ADA to the Astromat Synthesis completed (9/2026)
- > Improved user interfaces for data submission workflows deployed (3/2027);
- > Upgraded infrastructure for enhanced access and reusability of imagery done (9/2027);

Pick your topic: ASK NOW!



- Accepted types of data
- Data formats
- Submission process
- Journal policies & Astromat
- Sample registration
- Development timeline & availability of services
- Help with Data Management Planning, Letters of Collaboration



Thank you!

ASTROMAT Astromaterials Data System

www.astromat.org

Contact us with any feedback or questions:

info@astromat.org

Photo credit: NASA, ESA, and the Hubble Heritage Team (STScI/AURA)