

THE ASTROMATERIALS DATA SYSTEM: ADVANCING ACCESS AND PRESERVATION OF PAST, PRESENT, AND FUTURE LAB ANALYTICAL DATA OF NASA'S ASTROMATERIALS COLLECTIONS.

K. Lehnert^{1,*}, P. Ji¹, J. Mays¹, J. D. Figueroa¹, A. Johansson¹, L. Profeta¹, L. Song¹, S. Richard², S. Morrison³, A. Ostroverkhova³

¹Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA; ² US Geoscience Information Network (USGIN); ³Carnegie Science Earth & Planets Laboratory, Washington, DC, USA

Introduction: The Astromaterials Data System [1] provides a comprehensive solution for access, sharing, publication, and archiving of laboratory analytical data generated on astromaterials that makes these data Findable, Accessible, Interoperable, and Reusable = FAIR. AstroMat offers services for researchers to share, publish, and archive their data in a data repository that follows community best practices, and maintains a comprehensive synthesis of published geochemical and mineralogical data generated on astromaterials specimens that facilitates advanced data analytics via human- and machine readable interfaces. Astromat is modeled after the successful EarthChem data system [2] with the EarthChem Library and the EarthChem Synthesis. Initial versions of the Astromat Synthesis (AstroDB) and the Astromat Repository (AstroRepo) were released in 2020 and are fully functional.

Astromat Architecture: The Astromaterials Data System (referred to as 'Astromat') is designed as an ecosystem of interconnected applications that provide human- and machine-readable interfaces to the data gathered and managed in AstroMat's databases. The various components of AstroMat are intended for specific purposes and include interfaces for users to search, access, explore, visualize, analyze, and contribute data; software tools for data curators to compile, track, validate, ingest, manage, and annotate data; and machine-actionable interfaces that connect the databases to internal and external software tools.

Astromat Repository: The Astromaterials Data Repository (AstroRepo) accepts user contributions related to geochemical, geochronological, experimental and petrographic studies of extraterrestrial samples. Data is submitted through an online portal and curated by domain experts. The datasets undergo a technical review for metadata completeness, are assigned unique identifiers (DOI), published for public access and use with appropriate licenses, and archived for long-term preservation. Templates are provided for researchers to organize and document different types of analytical measurements. The AstroRepo encourages use of IGSN [3] as a globally unique and persistent sample identifier so data for a given sample can be unambiguously referenced to a given sample and linked across different datasets. AstroRepo is in the process of applying for repository certification by CoreTrustSeal [4].

Astromat Synthesis: Following the successful concept of geochemical data syntheses for terrestrial samples such as GEOROC [5] and EarthChem [6], AstroDB is a database that integrates all past, present, and future laboratory analytical data for NASA's astromaterials collections including the Lunar, Antarctic Meteorite, Cosmic Dust, Genesis, Hayabusa, Stardust, and Microparticle Impact collections. AstroDB data curators compile these data from the literature and other data sources together with a wide range of information pertaining to the provenance of the analytical measurements (e.g., analytical instruments, laboratory, sample preparation procedures, instrument calibration, data reduction) and to the quality of the data. The data and metadata are stored in a database structured according to the Observation Data Model ODM2 [7] that in turn follows the conceptual Observations & Measurements data model [8]. So far, AstroDB includes data from more than 3,000 studies of the Lunar and the Antarctic Meteorite collections. Compilation of data from all other JSC collections will be completed over the next 3 years. New astromaterials data submitted to the AstroRepo will be routinely included in the synthesis. AstroDB data and metadata have been mapped to the Planetary Data System's standard PSD4, so they can be archived at the PDS Geoscience node. AstroDB features graphical interfaces for users to search, access, explore, visualize, analyze, and contribute data (AstroSearch, AstroRef, AstroPlot) and machine-actionable interfaces that connect the databases to internal and external software tools (AstroAPIs, RepoAPIs).

References: [1] <https://www.astromat.org>; [2] <https://www.earthchem.org>; [3] <https://www.igsn.org>; [4] <https://www.coretrustseal.org>; [5] Sarbas, B (2008) The GEOROC database as part of a growing geoinformatics network *Geoinformatics 2008—Data to Knowledge*, pp. 42-43. USGS, 2008. [6] Hsu, L (2017) IEDA EarthChem: Supporting the sample-based geochemistry community with data resources to accelerate scientific discovery *Authorea Preprints*; [7] Horsburgh JS et al (2016) Observations Data Model 2: A community information model for spatially discrete Earth observations. *Environmental Modelling & Software* 79: 55-74; [8] Cox, S (2011) Observations and Measurements-XML Implementation. Version 2.0.