OPEN AND REUSABLE DATA FOR ASTROMATERIALS SAMPLES

K. Lehner1, D. Hezel2, P. Ji3, J. Mays1, L. Profeta1, L. Song1, S. Morrison1, J.D. Figueroa1, and A. Johansson1,
1Lamont-Doherty Earth Observatory, Columbia University, USA, lehner@ldeo.columbia.edu, 2University of
Frankfurt, Germany, 3Carnegie Science Earth & Planets Laboratory, USA

Introduction: Laboratory studies of meteorites and astromaterial samples returned by space missions generate
highly valuable and expensive analytical data. These data are part of a much larger ecosystem of research data that
today is regulated by the principles of Open Science, and need to be Findable, Accessible, Interoperable, and
Reusable (=FAIR) [1]. We will provide an overview and update of the Astromaterials Data System that offers a
comprehensive suite of systems and services for the ingestion, curation, management, publication, attribution,
dissemination (human- and machine-readable interfaces), and archiving of laboratory analytical data of astromaterials
samples.

Demands for Open Astromaterials Sample Data. Open Science is a new research paradigm that is envisioned
to accelerate scientific innovation, ensure trust in science through transparency and reproducibility of research
outcomes, and allow future scientific endeavors to build on past research outcomes. Open Science commonly refers
to policies that make results of publicly-funded research – publications, data, samples, software - freely available and
accessible and is anticipated to transform the research process to a more findable, accessible, interoperable, and
reusable (=FAIR) one (Wilkinson et al. 2016). Most funders and publishers today require researchers and authors to
deposit the data generated as part of a research grant and that underpin the findings reported in a scientific publication
in trusted domain repositories. Trusted repositories are those that can demonstrate compliance with international best
practices for data curation through certification with recognized authorities such as CoreTrustSeal.

Astromat Repository. The NASA-funded Astromaterials Data System (https://www.astromat.org) released the
Astromat Repository in 2020, which serves as an open-access repository for datasets (analytical, experimental,
synthesis databases) and other digital content submitted by researchers. The Astromat Repository helps investigators
comply with funders’ and publishers’ policies for FAIR data as a trusted repository. Services include: interactive
online submission with guidance for properly documenting, licensing, releasing, and citing content; review of
submitted content; DOI registration with DataCite; long-term archiving; and user support and training. In order to
achieve the highest level of Interoperability and Reusability of data (the I&R in FAIR), Astromat data curators
carefully review submitted content and interact directly with users to augment and enrich, if necessary, documentation
of data provenance (e.g., methods of data and sample acquisition and processing) and data quality, and to include
identifiers in the metadata (e.g., DOI for cited publications, IGSN for cited samples, funding award numbers, and
mission identifiers), so that datasets can be linked to other digital resources.

Astromat Synthesis. In order to further enhance reusability and impact of Astromaterials sample data, Astromat
gathers, harmonizes, and synthesizes laboratory analytical data from past, present, and future studies into the AstroDB
datastore that can be searched, accessed, and explored via human- and machine-readable interfaces (AstroSearch and
AstroAPI). The Astromat Synthesis enables new ways of mining and analyzing laboratory analytical data for
Astromaterials samples, leading to the creation of new knowledge and maximizing the scientific return of NASA’s
investment into sample return missions, sample curation, and data acquisition on these samples. Astromat works
closely with MetBase (https://metbase.org), the the world’s largest compilation of meteorite data, to ensure that the
two databases are complementary and compatible. Both data systems participate in the international initiative
‘OneGeochemistry’ [2] that aims to develop and promote best practices and data standards for laboratory analytical
data and in initiatives to establish best practices for sample citation and identification [3].

References: