THE ASTROMAT SYNTHESIS: A DATA PIPELINE TO GENERATE ANALYSIS READY DATA FOLLOWING FAIR PRINCIPLES

P. Ji, K. Lehnert, J.D. Figueroa, J. Mays, A. Johansson, L. Profeta
[1]Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, United States

Introduction: AstroMat is a comprehensive data system for laboratory analytical data generated by the study of astromaterials samples curated at the NASA Johnson Space Center[1]. It 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[2].

The direction Astromat takes in the era of big data will depend on how we address the following questions frequently asked by scientists: Can I pull all analytical data organized by sample, location and other criteria to deal with particular questions? Can I stream all the analytical data that I need into my Jupyter notebook? How can I integrate the analytical data hosted in Astromat into other tools, eg. Machine learning workflow and computational modeling? The Astromat Synthesis, a data pipeline available to generate analysis ready data following FAIR principles from various data sources, is responding to these questions.

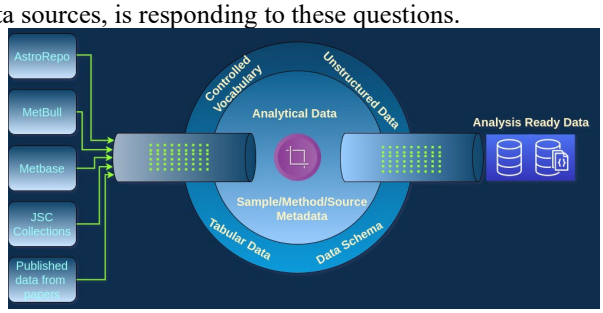

Fig. 1. Astromat Synthesis

Astromat Synthesis: In spite of the vast quantities of analytical data that have been generated on astromaterials, there is no easy way to amalgamate high-quality, analysis-ready data from multiple sources for novel research. Vast amounts of analytical data reside in science articles without connections. Invaluable sample metadata are scattered among official agencies and communities. The Astromaterials Data System (AstroMat) addresses these challenges by: transforming analytical data dispersed in publications and tables with different formats into AstroDB; makes disconnected data connect; consolidates metadata of sample, method and source; generates various indexed JSON document collections, all to meet different scientific and technical needs. **Conclusion:** The Astromat Synthesis is one of the core components of the Astromaterials Data System. Its flexible architecture allows continued integration of data from new sources, and makes data ready for further analysis and synthesis following the FAIR principles.

Acknowledgments: The work described in this abstract was supported by National Aeronautics & Space [80NSSC19K1102].

References:

[1] Lehnert, K., Ji, P., Mays, J., Figueroa, J. D., Johansson, A., Profeta, L., Song, L., and Morrison, S.(2020) The Astromaterials Data System: Advancing Access to Past, Present, and Future Lab Analytical Data of NASA's Astromaterials Collections, Europlanet Science Congress ,EPSC2020-918.

[2] Ji, P., Lehnert, K., Evans, C. A., and Zeigler, R. A.(2018) AstroDB- A Data System for Analytical Data of extraterrestrial samples, American Geophysical Union, Fall Meeting 2018, abstract #IN31A-08.