



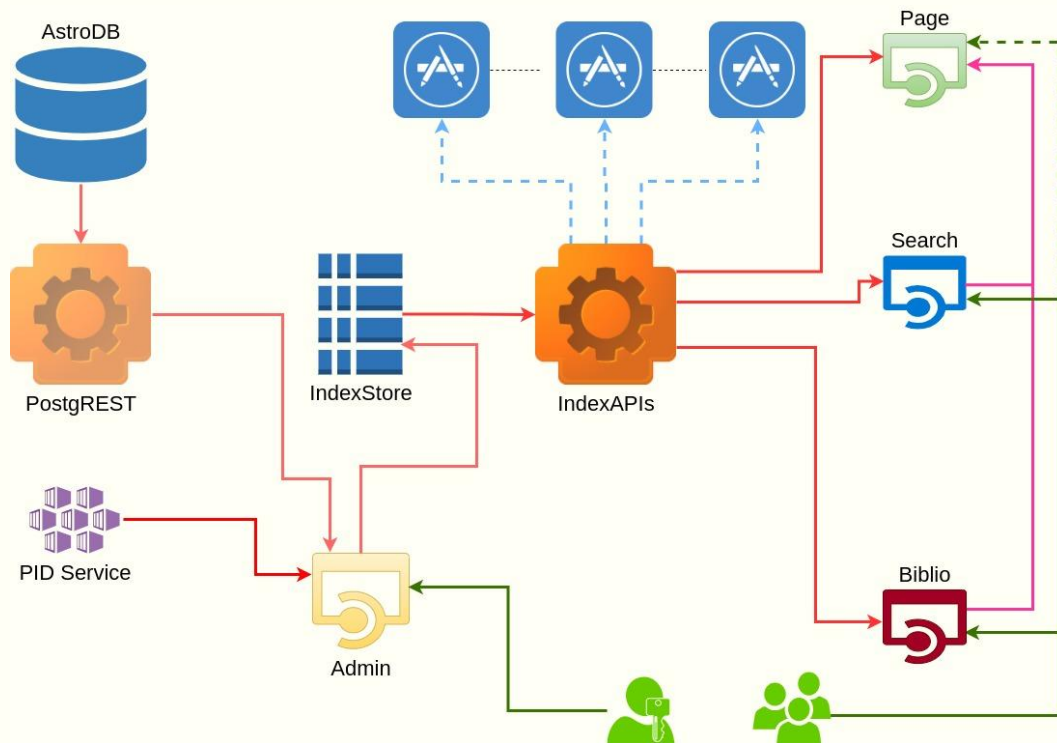
THE ASTROMAT SYNTHESIS:

“A Data Pipeline To Generate Analysis Ready Data Following Fair Principles”

Peng Ji ¹(pengji@ldeo.columbia.edu), Kerstin Lehnert, Juan David Figueroa, Jennifer Mays, Annika Johansson, Lucia Profeta

Lamont-Doherty Earth Observatory, Columbia University, New York, USA

What is Astromat



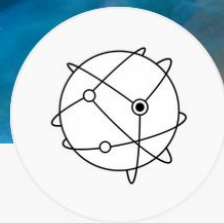
AstroMat is a comprehensive data system for laboratory analytical data generated by the study of astromaterials curated at the NASA Johnson Space Center. It is designed as an ecosystem of interconnected applications that provide human- and machine-readable interfaces to the data gathered and managed in AstroMat's databases.



Astromaterials Data System



ACCESS DATA



SUBMIT DATA



COLLECTIONS

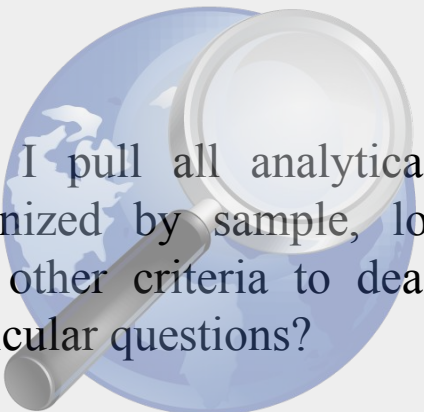
www.astromat.org

info@astromat.org

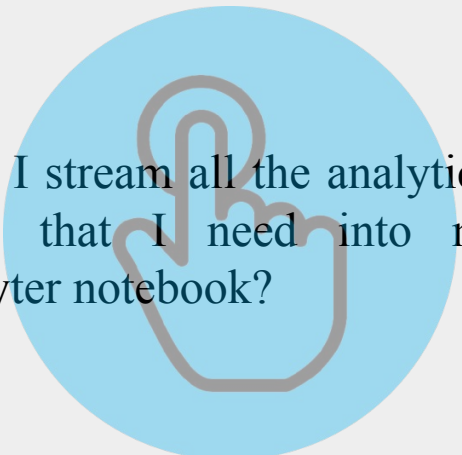
777,896

NUMBER OF ANALYTICAL VALUES IN ASTROMAT'S DATABASE

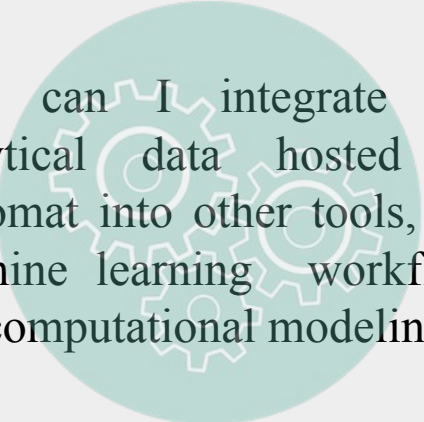
Real Challenges



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?



FAIR Lab Data

Lab Experimental Data




Lab Analytical Data



Findable



Persistent Identifiers (PIDs)




Rich metadata



Indexed data repositories



PIDs in metadata



Accessible



Standard communications protocol



Open, free protocol



Authentication, where necessary




Metadata is always available



Interoperable



Vocabularies



Vocabularies are FAIR



Linked metadata



Reusable




Metadata have multiple attributes



Usage license



Provenance



Community standards



Source : Australian National Data Service (ANDS)



Astromat Synthesis

- Contains published geochemical and petrological data for Astromaterials from scientific papers and data publications
- Analysis Ready Data (ARD): integrated & harmonized by data curators
- Rich data documentation for full FAIR compliance
- Sample metadata: e.g., description, classification, identifiers, related samples
- Method metadata: e.g., lab, instrument, data quality, data manipulation
- Source metadata: bibliographic metadata, DOI & BibCode
- Data access at the granularity of individual samples & measurements.
- Human & machine readable interfaces for data search & retrieval.



Data Ingestion, Integration, Preparation



Astromat Synthesis

