Bringing cosmochemistry data into the Big-Data Era: the AstroMaterials Data System

Lucia Profeta[1] Kerstin Lehnert[1], Peng Ji[1], Juan David Figueroa[1], Annika Johansson[1], Jennifer Mays[1], Shaunna M Morrison[2] and Alexandra Ostroverkhova[2]

1 - Columbia University, Lamont-Doherty Earth Observatory, Palisades, NY, United States; 2 - Carnegie Institution for Science, Earth and Planets Laboratory, Washington, DC, United States

ASTROMAT
Astromaterials Data System

Goals of the Astromaterials Data System

Ensure access and reusability of past, present, & future laboratory data for NASA's astromaterials collections.

Support community efforts for astromaterials data rescue.

Support researchers to share their data in compliance with funders' and publishers' data policies.

LUNAR · ANTARCTIC METEORITES · HAYABUSA · MICROPARTICLE IMPACT · COSMIC DUST · STAR DUST · GENESIS

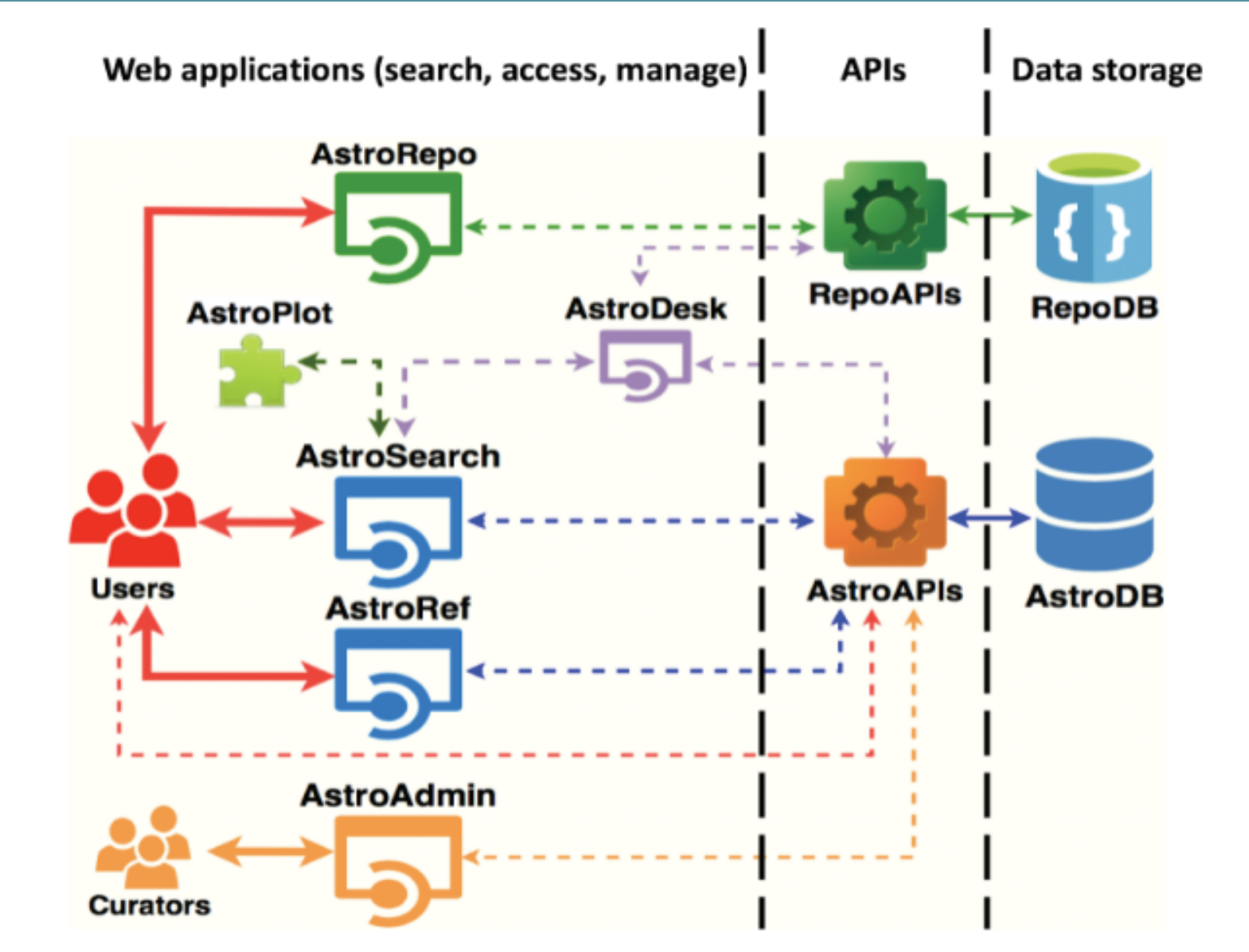https://curator.jsc.nasa.gov

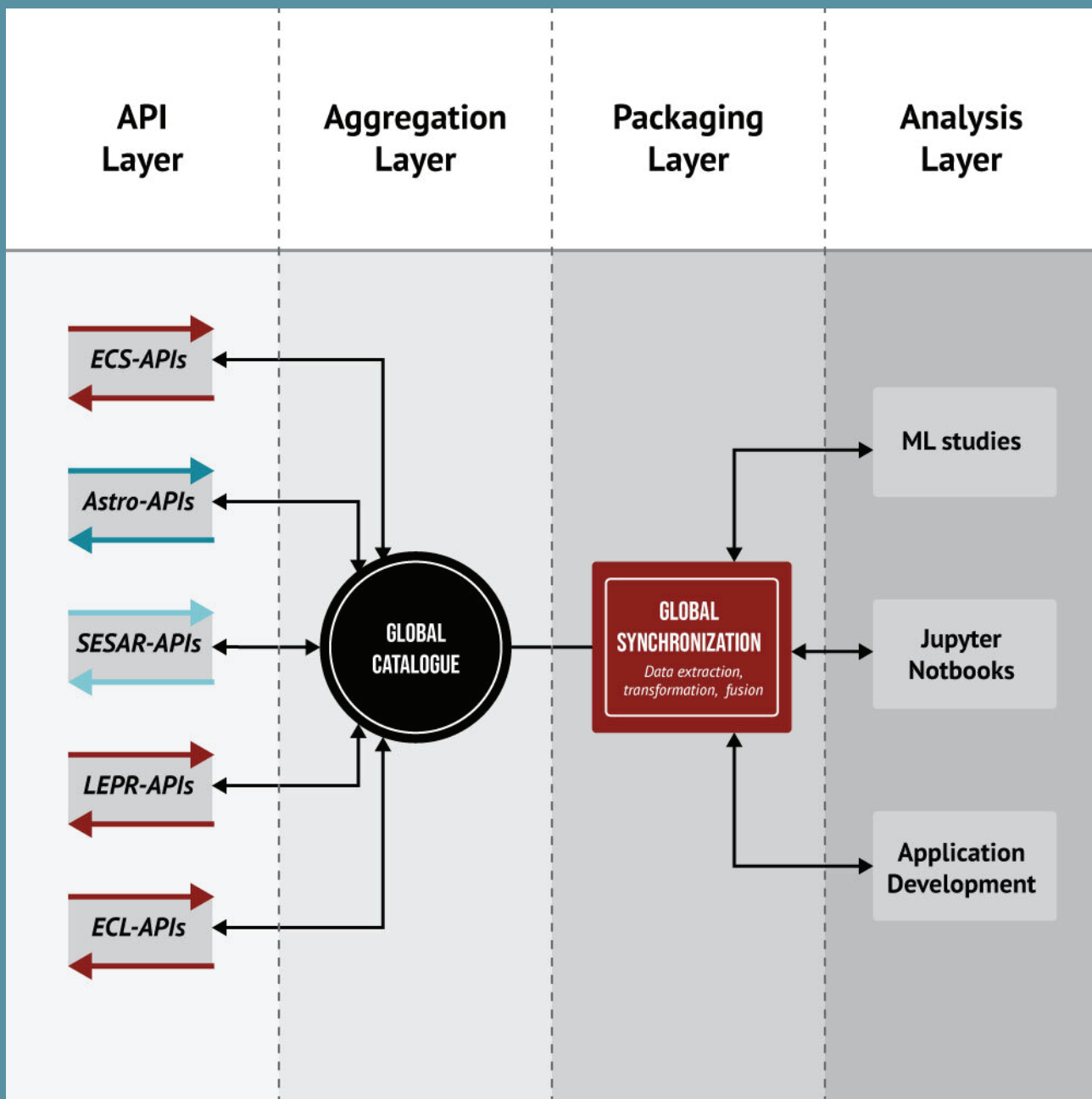## Systems Architecture Built to Support Big Data Studies

AstroMat is a comprehensive data system for laboratory analytical data designed as an ecosystem of interconnected applications that provide human-readable and machine-readable interfaces to the data gathered and managed in AstroMat's databases.

The index store is meant to integrate data from AstroDB tables to provide flexible and fast querying using Elasticsearch. It also communicates with AstroAdmin tools used by curators to input data.
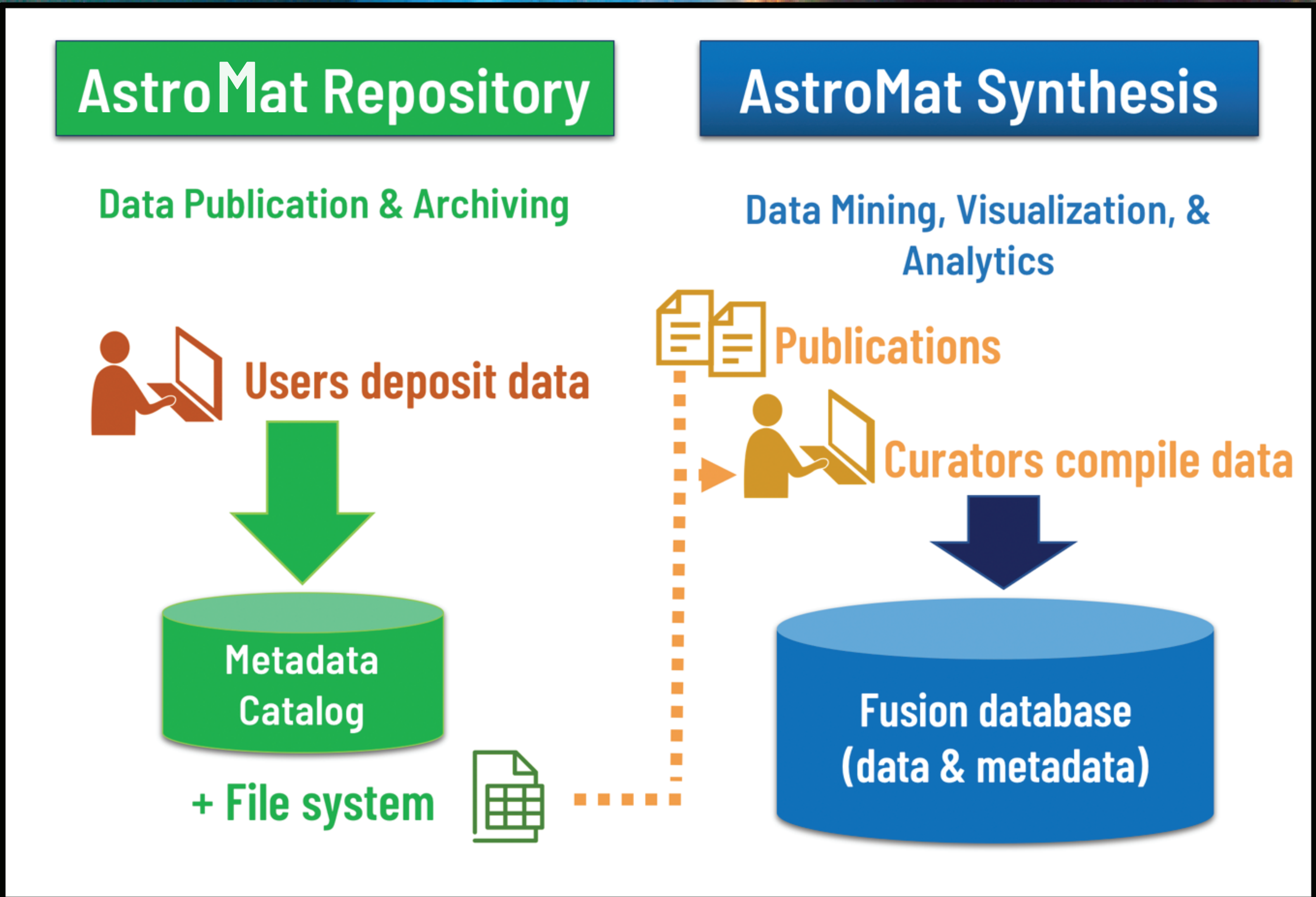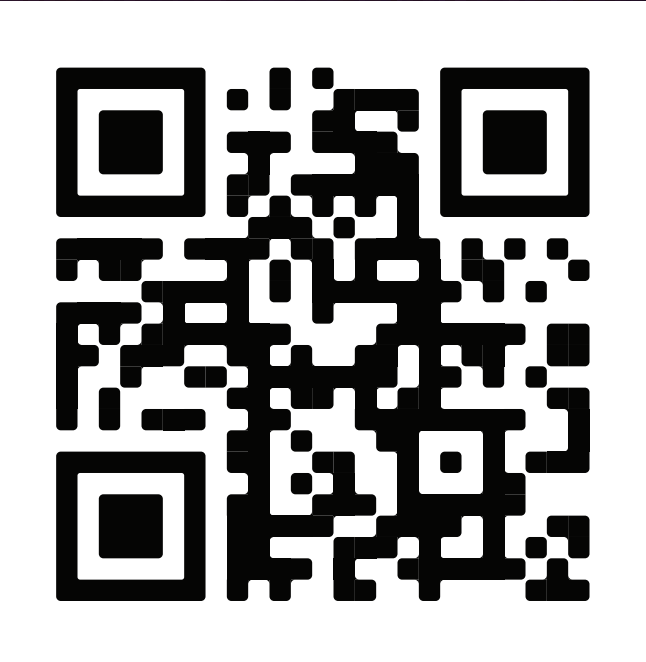

Custom APIs are built on top of Elasticsearch to communicate with user interfaces.

The longterm goal is to create a central access point that can integrate data from multiple sources and package it in user-specified formats to be digestible for machine learning and artificial intelligence applications.

AstroMat offers
Data Publishing & Archiving
Data Access & Mining
for geochemical and petrological data of Astromaterials samples.

AstroMat Repository

Data Publication & Archiving

Users deposit data

Metadata Catalog

+ File system

AstroMat Synthesis

Data Mining, Visualization, & Analytics

Publications

Curators compile data

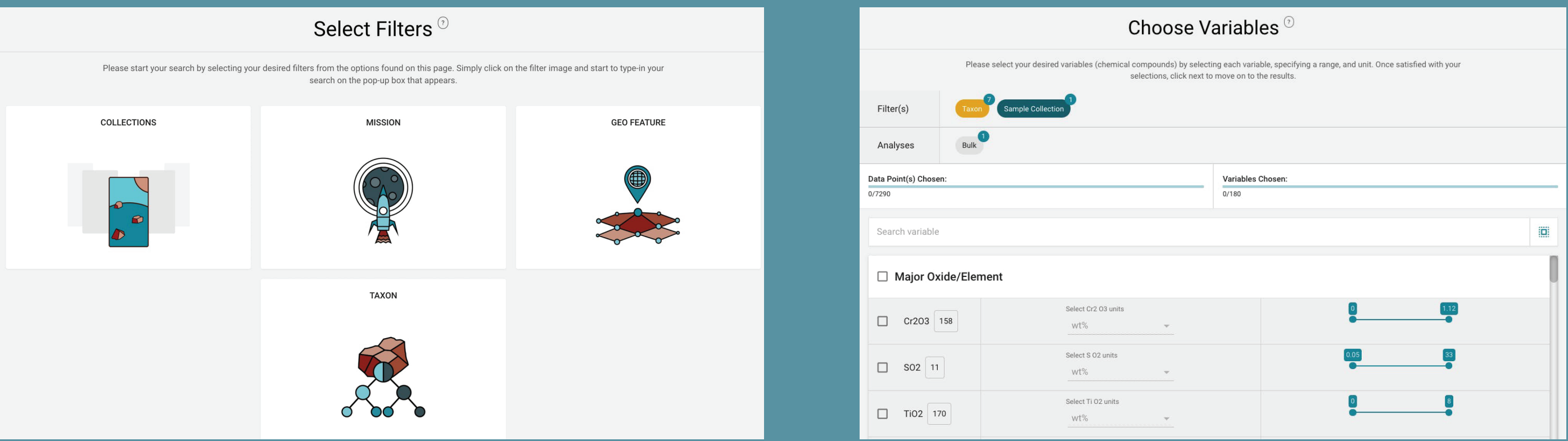Fusion database (data & metadata)

AstroSearch - Integrated Search and Discovery

The AstroMat Search is currently incorporating feedback from usability testing. You can go to our website and sign up to our mailing list to receive more information.

The initial developments for the AstroMat Search interface were inspired by PetDB (the EarthChem Synthesis).

Data can be refined by multiple filters such as collection, mission or geofeatures associated with the samples, as well as taxonomy keywords for rocks and meteorites.

After initial filtering, data can be aggregated based on the analyzed materials and available chemical values.
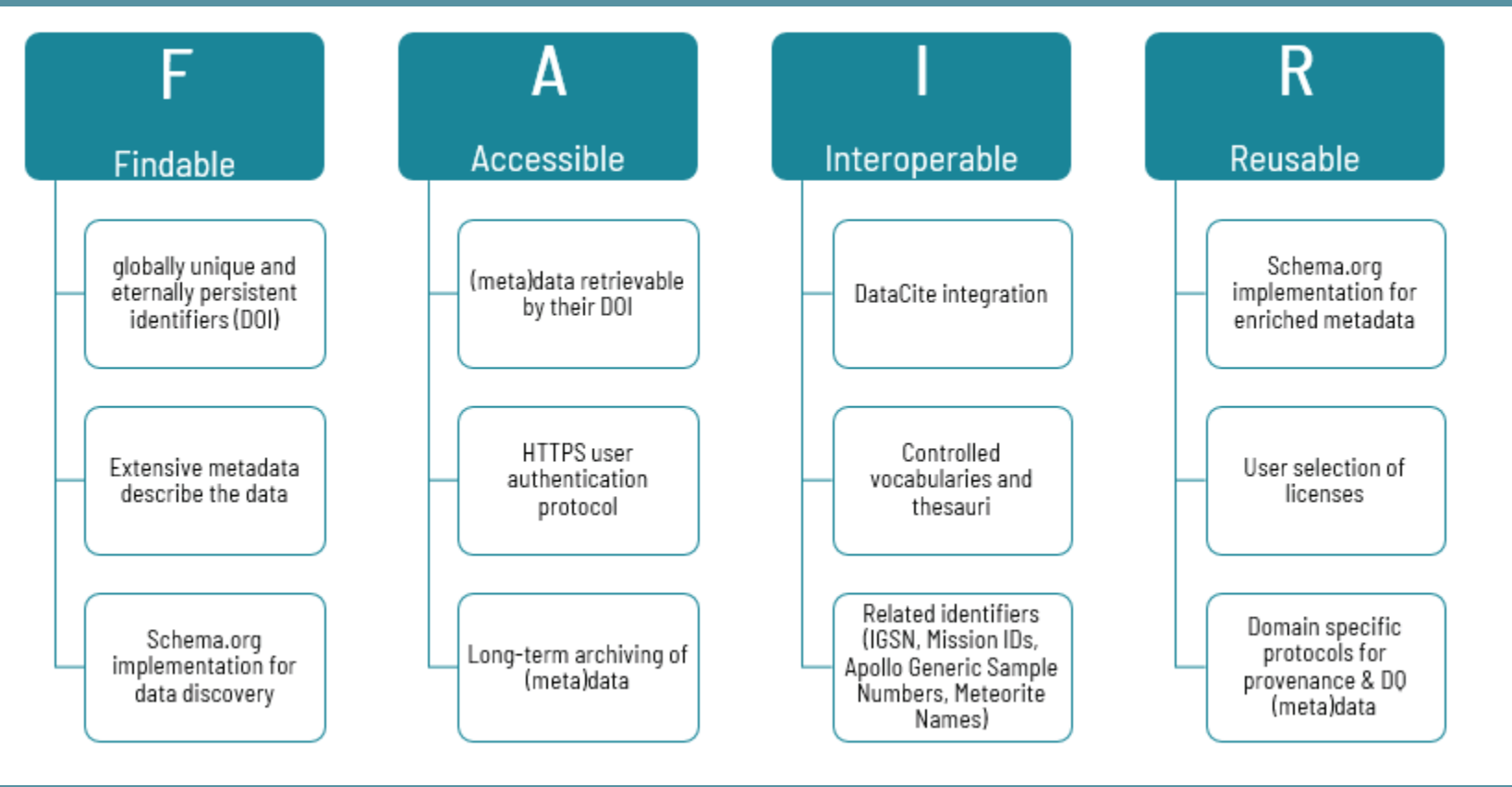

Making Your Data FAIR - the Building Blocks of Big Data

AstroMat strives to improve cosmochemical data sharing for the astromaterials community.

The AstroMat mission mirrors the Core Values of the Planetary Data Ecosystem Independent Review Board (Besse et al., 2021):

• **First, do no harm**: Avoid the law of unintended consequences.

• **FAIR**: Facilitate participation in the PDE by adhering to FAIR data principles of Findability, Accessibility, Interoperability, and Reusability.

• **Open**: Advocate open science practices, including open access, open data, open code, open software/tools, and others.

• **Collaborative**: Encourage international collaboration. Welcome new participants from both inside and outside the professional space exploration community.

• **Effective**: Provide timely, useful support to user communities, especially data producers.

• **Practical**: Pursuit of ideal solutions may sometimes leave the Ecosystem with no solution at all rather than a solution that is sufficient.

AstroMat already addresses many of the IRB recommendations, including:

Recommendation 33: NASA should establish a requirement for the preservation of mission-supported laboratory analyses of returned sample material that makes the information accessible to the planetary science community. – Astromat provides the infrastructure and services that support the preservation and access of laboratory analyses of returned sample material.

Recommendation 47 NASA should support and encourage expanded use of DOI-like identifiers for data, thereby connecting data at various levels of processing to assist users in locating the best version of a data set for their needs. – The AstroMat Repository registers data with DataCite to assign DOI. Astromat also is working to assign persistent, globally unique identifiers (IGSN) to the samples.

Recommendation 49 NASA should fund the development of more analysis-ready data (ARD) products derived from the lower-level products created by NASA missions. – Astromat's Synthesis Database delivers astromaterials data analysis-ready via machine-actionable interfaces.