AN OPEN SCIENCE FRAMEWORK FOR LONG-TAIL RESEARCH DATA

Application example:

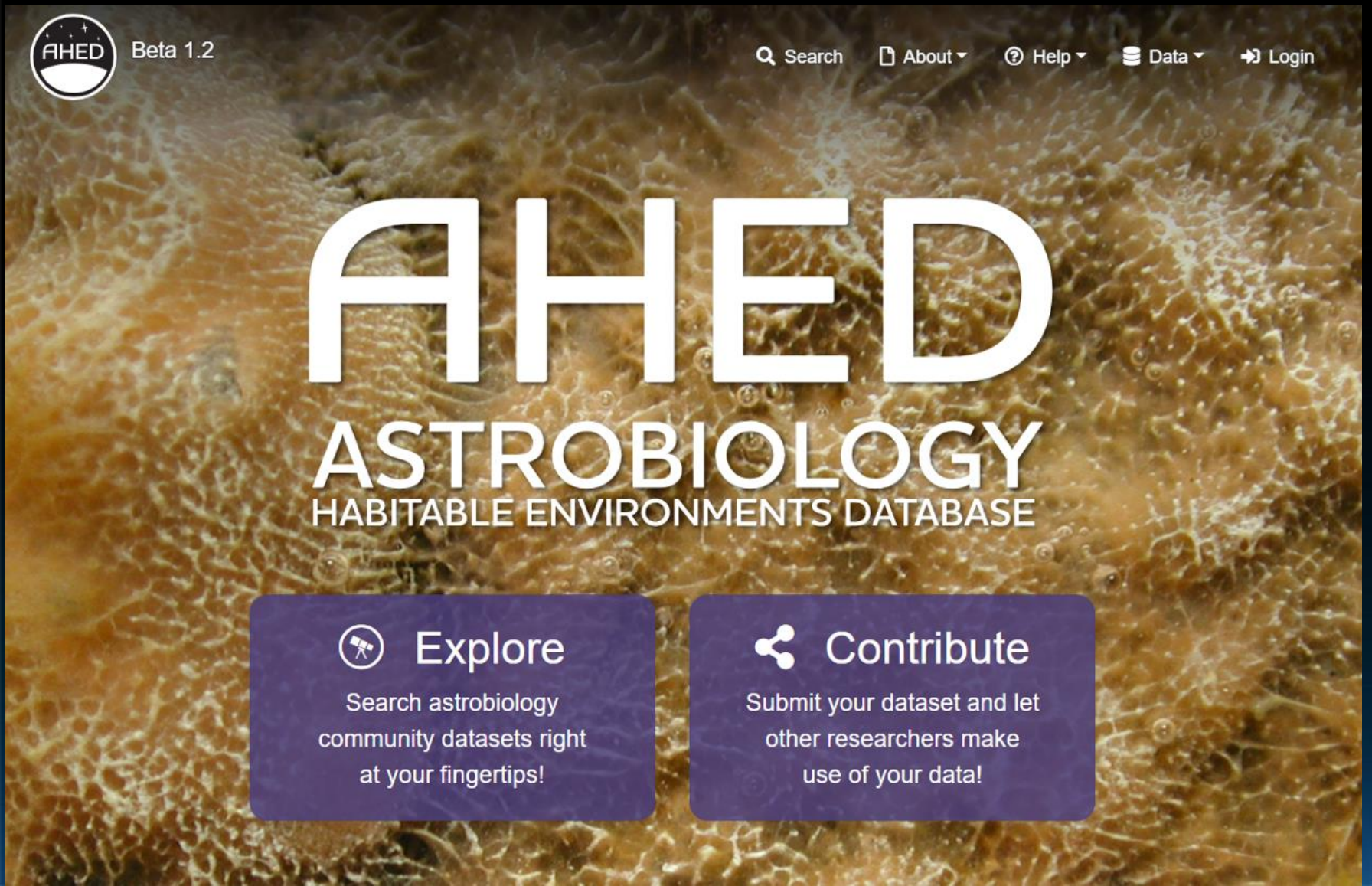
AHED
ASTROBIOLOGY
HABITABLE ENVIRONMENTS DATABASE

NASA

SETI INSTITUTE

Open Data REPOSITORY

What is AHED?

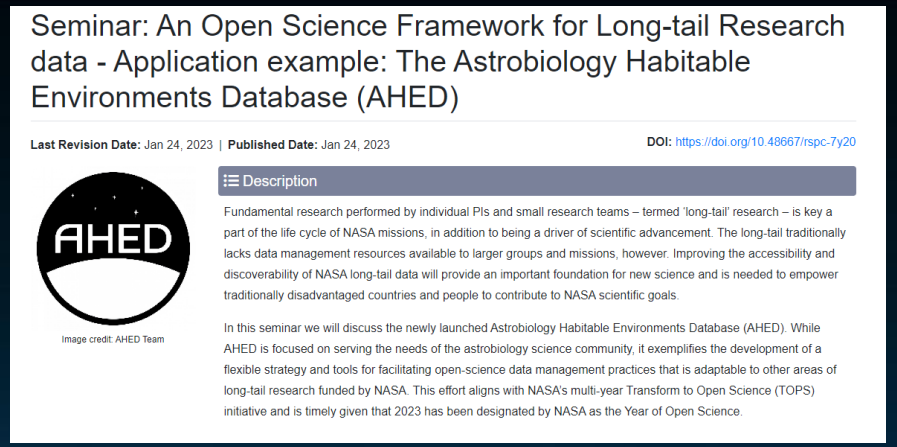


AHED is designed as an open-access, community-driven repository and productivity platform for the storage, discovery, and analysis of data, focusing on the field of astrobiology.

NASA SILICON VALLEY
AMES RESEARCH CENTER

2023 is the Year of Open Science

- **An open access AHED dataset containing seminar materials is available at:**



Seminar: An Open Science Framework for Long-tail Research data - Application example: The Astrobiology Habitable Environments Database (AHED)

Last Revision Date: Jan 24, 2023 | **Published Date:** Jan 24, 2023

DOI: https://doi.org/10.48667/rspc-7y20

AHED

Image credit: AHED Team

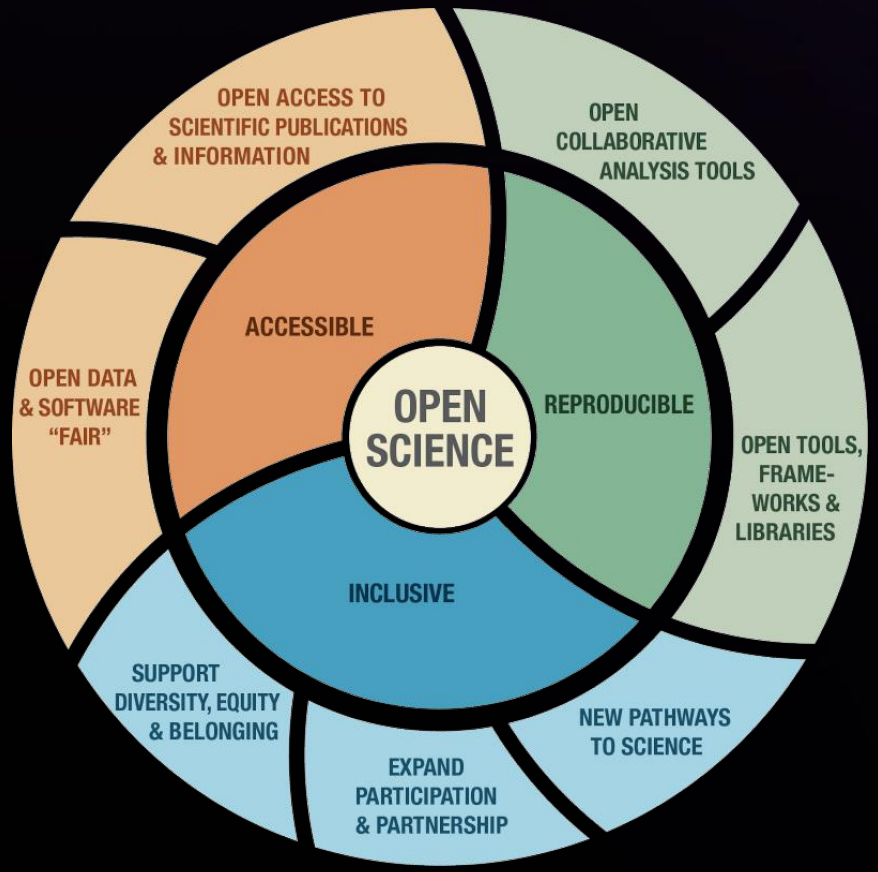
Description

Fundamental research performed by individual PIs and small research teams – termed 'long-tail' research – is key a part of the life cycle of NASA missions, in addition to being a driver of scientific advancement. The long-tail traditionally lacks data management resources available to larger groups and missions, however. Improving the accessibility and discoverability of NASA long-tail data will provide an important foundation for new science and is needed to empower traditionally disadvantaged countries and people to contribute to NASA scientific goals.

In this seminar we will discuss the newly launched Astrobiology Habitable Environments Database (AHED). While AHED is focused on serving the needs of the astrobiology science community, it exemplifies the development of a flexible strategy and tools for facilitating open-science data management practices that is adaptable to other areas of long-tail research funded by NASA. This effort aligns with NASA's multi-year Transform to Open Science (TOPS) initiative and is timely given that 2023 has been designated by NASA as the Year of Open Science.

https://doi.org/10.48667/rspc-7y20

SPD-41a is built on the Open-Source Science Principles of Accessibility, Reproducibility, and Inclusion.

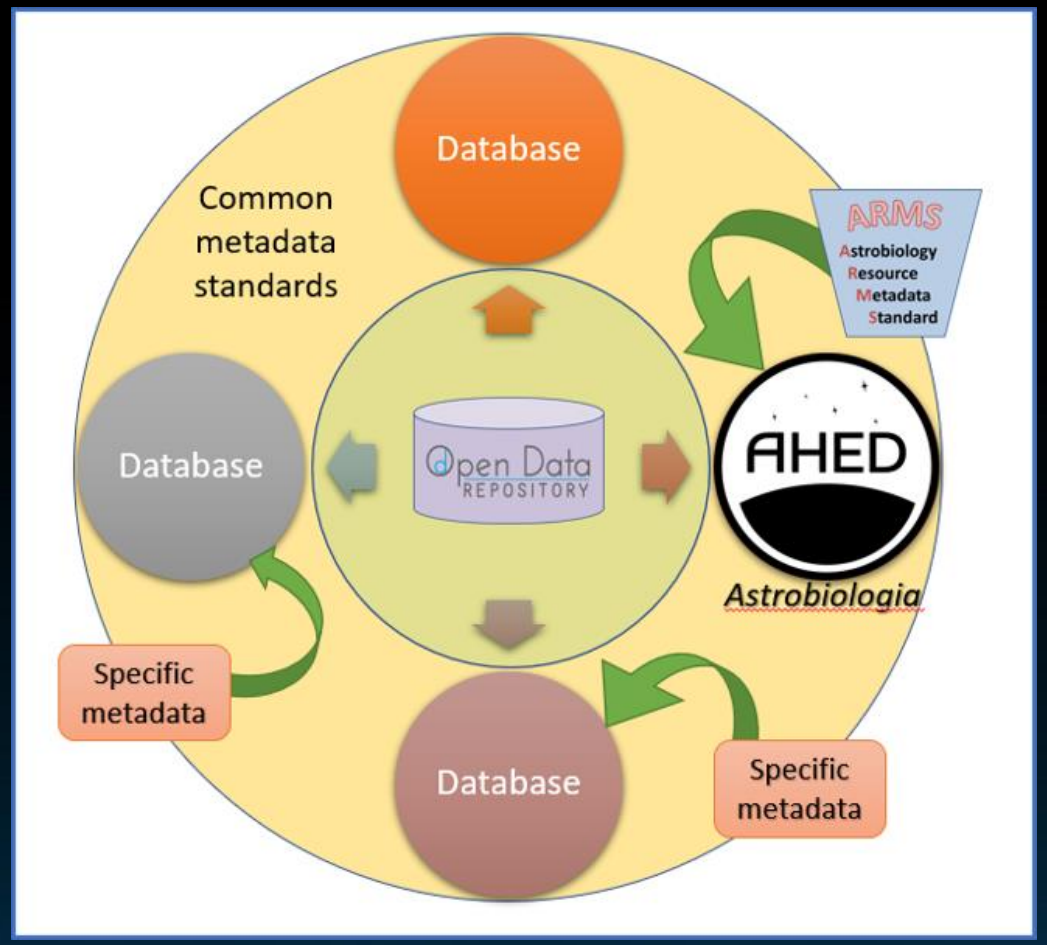


SPD-41a aims to make SMD science as open as possible, as restricted as required, and always secure.

SPD-41a looks to maximize openness while minimizing the burden on researchers.

(Slide from SPD41a Town Hall)

Core Services for Open Science



Ecosystem of databases with ODR as publishing platform. Common and specific metadata standards facilitate discovery and standardization.

What is long-tail research?

'Long-tail' research is performed by individual PIs and small research teams. It is a fundamental part of realizing NASA's scientific goals and a key input feeding into mission life cycles.

AHED Project Goal:

Provide an example of a data management strategy for other long-tail (small teams and individual PIs) research efforts in the Planetary Science.

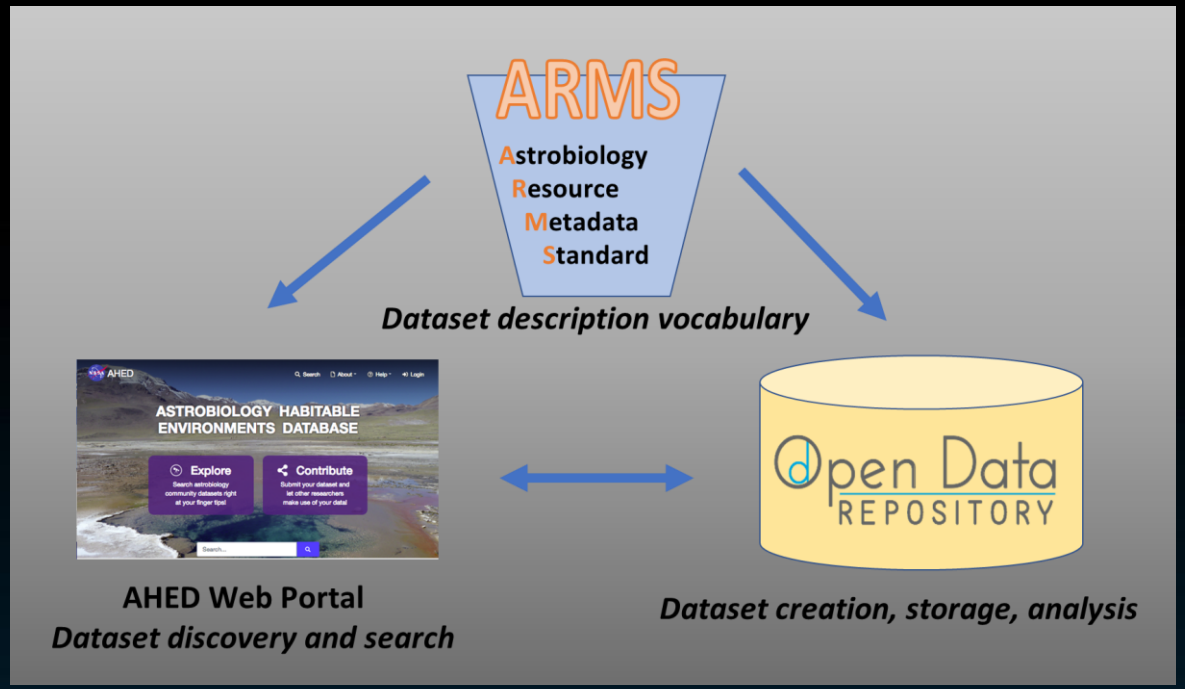
Directorate Strategic Plan Working Group on

Community Data Services

Lead/POC: Thomas Bristow (STX)
Contributors: Chris Dateo (STX), Christiaan Boersma (STA)

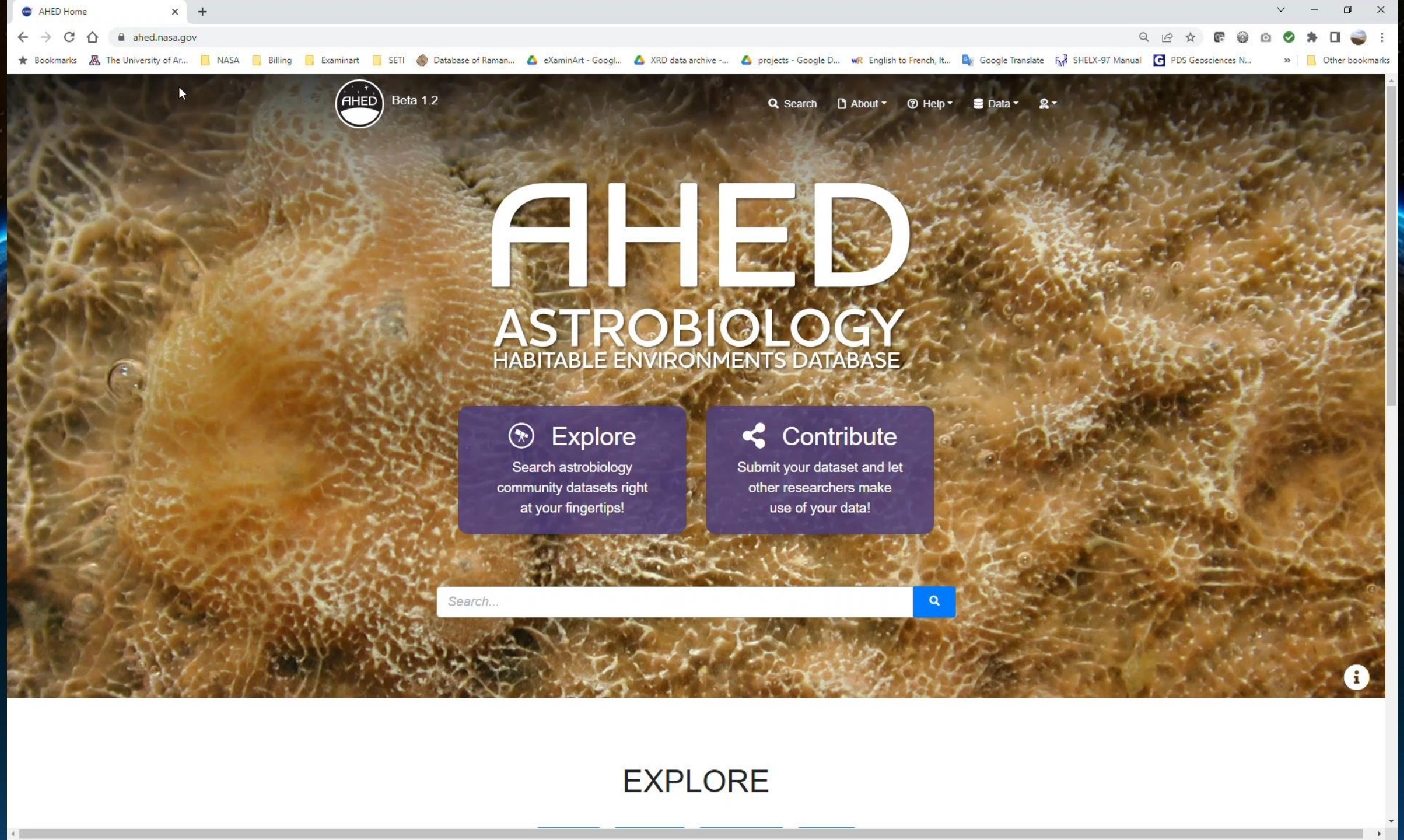
Agenda

- Welcome and Introductions
- Live demo of AHED Portal
- Live demo of dataset creation
- More detail on ARMS
- More detail on ODR
- Wrap up & questions



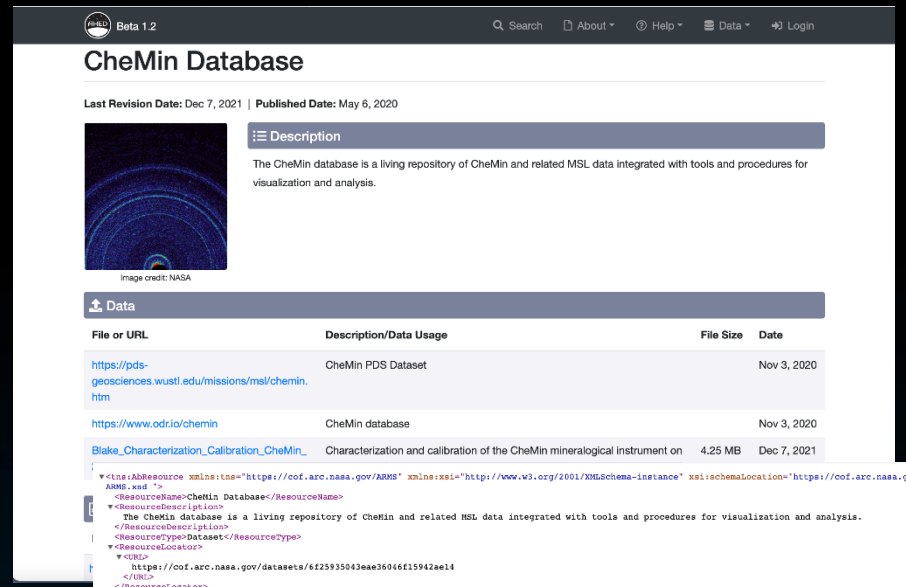
Overview of AHED system components.

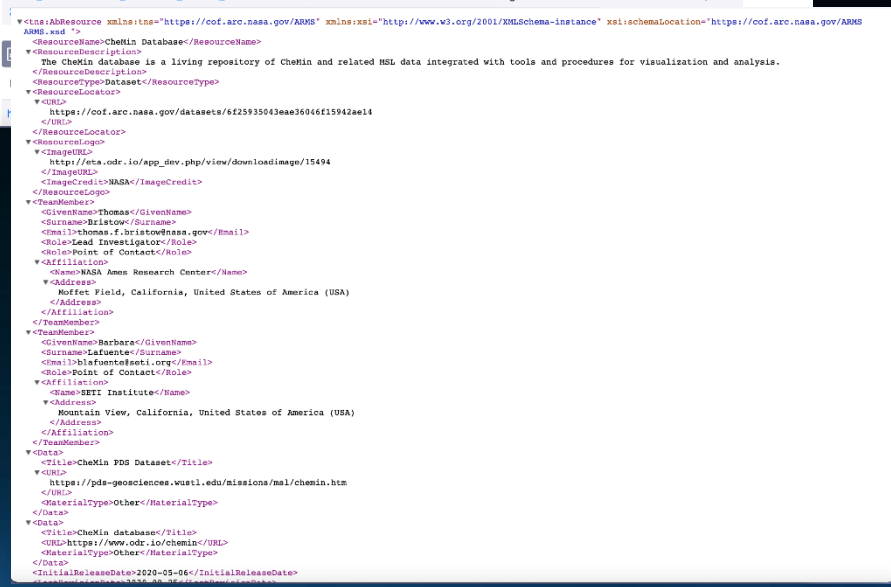
Live demo AHED portal


AHED Home
ahed.nasa.gov
Bookmarks
The University of Ar...
NASA
Billing
Examinart
SETI
Database of Raman...
eXaminArt - Googl...
XRD data archive -...
projects - Google D...
English to French, It...
Google Translate
SHELX-97 Manual
PDS Geosciences N...
Other bookmarks
AHED
Beta 1.2
Search
About
Help
Data
AHED
ASTROBIOLOGY
HABITABLE ENVIRONMENTS DATABASE
Explore
Search astrobiology community datasets right at your fingertips!
Contribute
Submit your dataset and let other researchers make use of your data!
Search...
EXPLORE

ARMS: Astrobiology Resource Metadata Standard

- Dataset description vocabulary to uniformly describe astrobiology ‘resources’.
- Required for all contributed AHED datasets.
- ARMS encourages common structure and metadata for all contributed datasets to facilitate cross-dataset search, integration, discovery and analysis.
- Development approach could be applied to other disciplines to develop corresponding metadata definitions.





ARMS Metadata

- Dataset intrinsic
  - Name
  - Description
  - Location
- Support
  - Team members
  - Funding
  - Project/mission
- Research context
  - Research theme
  - Science discipline
  - Keywords

Astrobiology Keywords

- The ARMS keyword defined vocabulary is a compilation of publication-style keywords extracted from Astrobiology-related journals from the Web of Science for the past 10 years
- After careful curation of the keywords, we developed a 4-tier taxonomy with 757 taggable keywords
- 11 top-level categories
  - astronomical, biological, chemical, geological
  - process
  - environmental, exploration, planetary
  - computational, methods
  - support

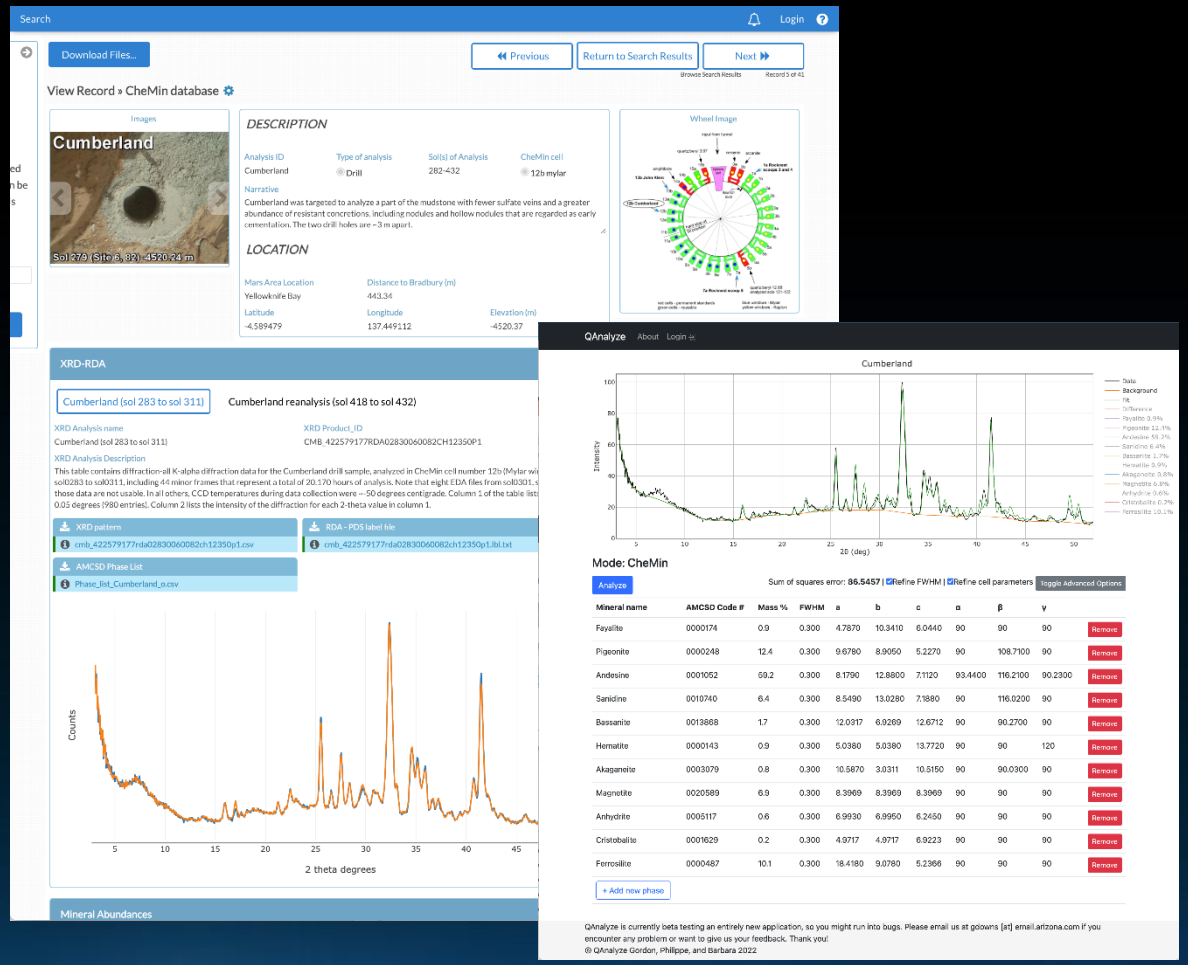
Keyword Validation

- Validated against 848 AbSciCon 2019 abstracts
  - 99% of the abstracts contained 1+ ARMS keywords
  - 67% of the ARMS keywords appeared in the abstracts
  - Cluster analysis revealed 36 potential new keywords, which were added to the lexicon
- Cluster analysis could be re-purposed to discover candidate keywords for new disciplines

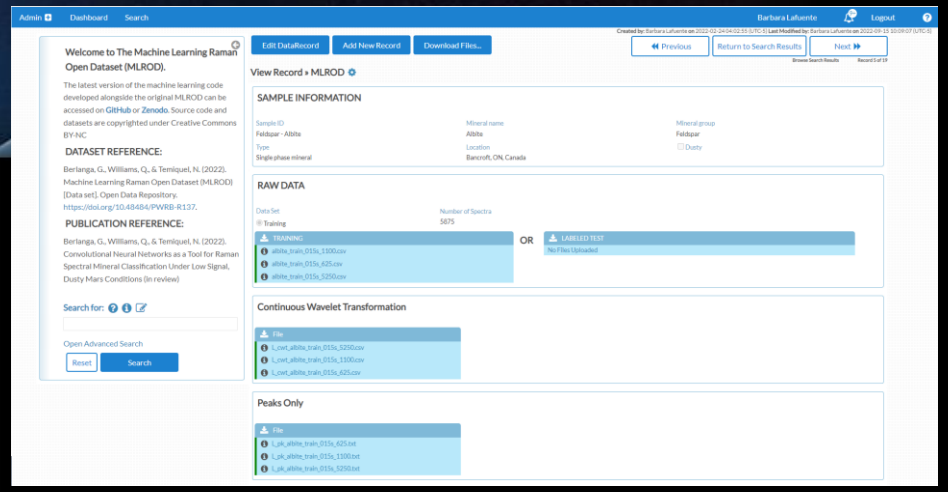
OPEN DATA REPOSITORY

Open Data REPOSITORY

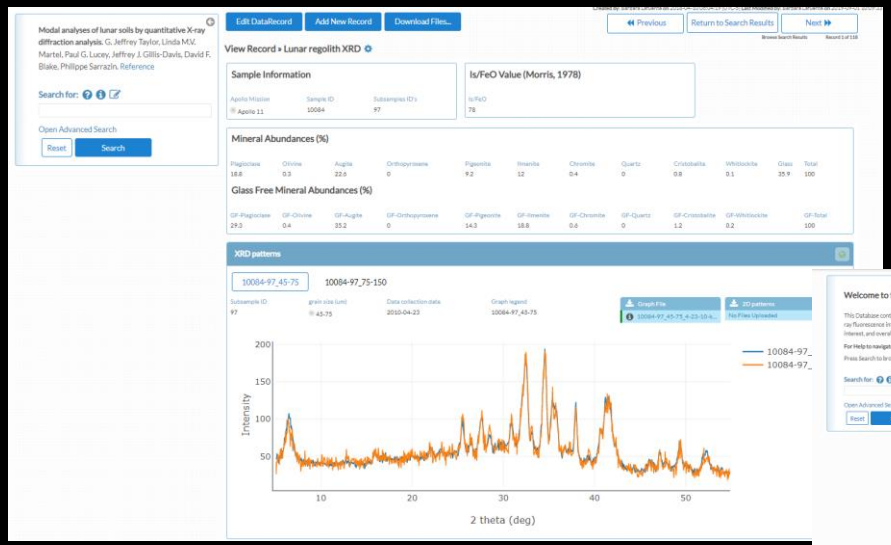
- Dataset Publication Platform
- Provides the backend for AHED
- Design & Create Databases Online
  - Design data entry forms for inputting data
  - Create a layout that displays your data
- Interactive Datasets
  - Graphing capability
  - Plugins for specific functionality
  - Integrate with 3rd party apps
- Manage access through permissions
- Dynamically update data via CSV imports and API
- To learn more about ODR's functionality:
  - https://www.opendatarepository.org



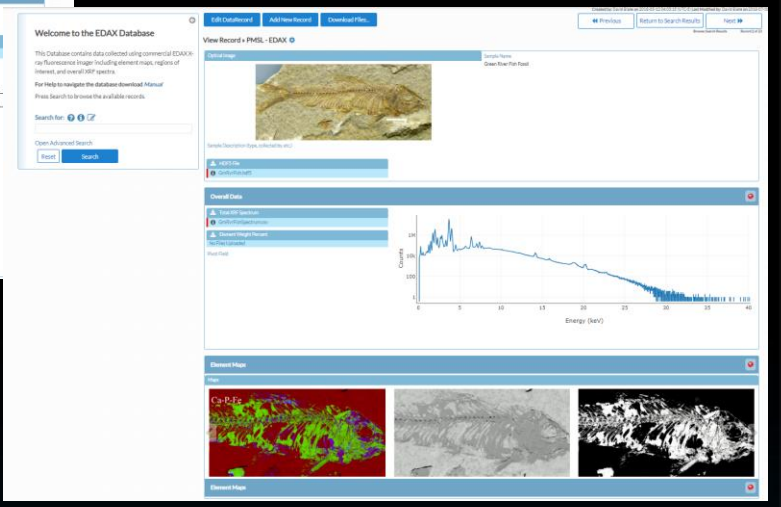
ODR dataset examples



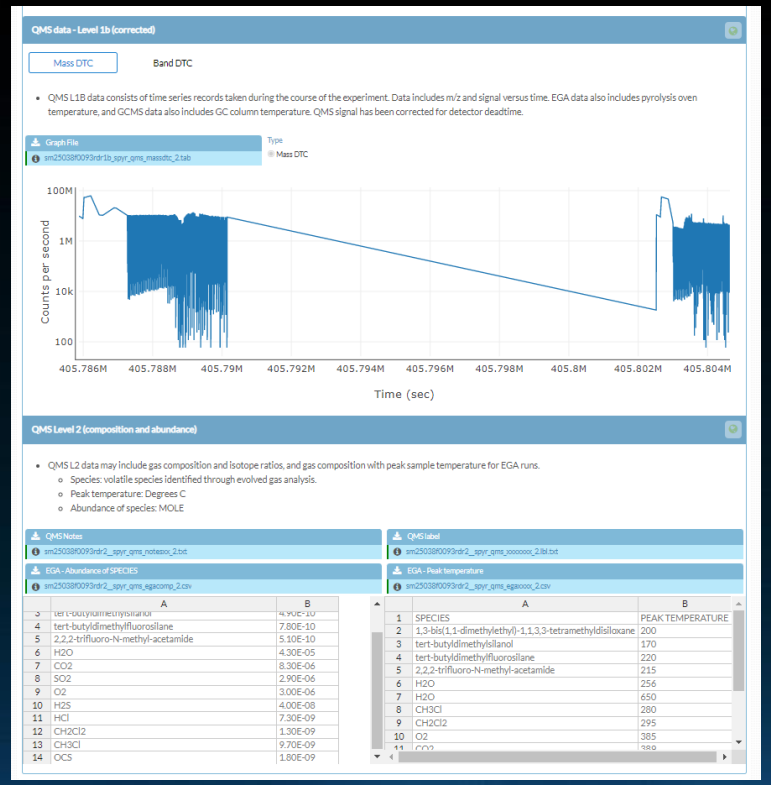
MLROD - https://doi.org/10.48484/PWRB-R137. (Berlanga et al, 2022)



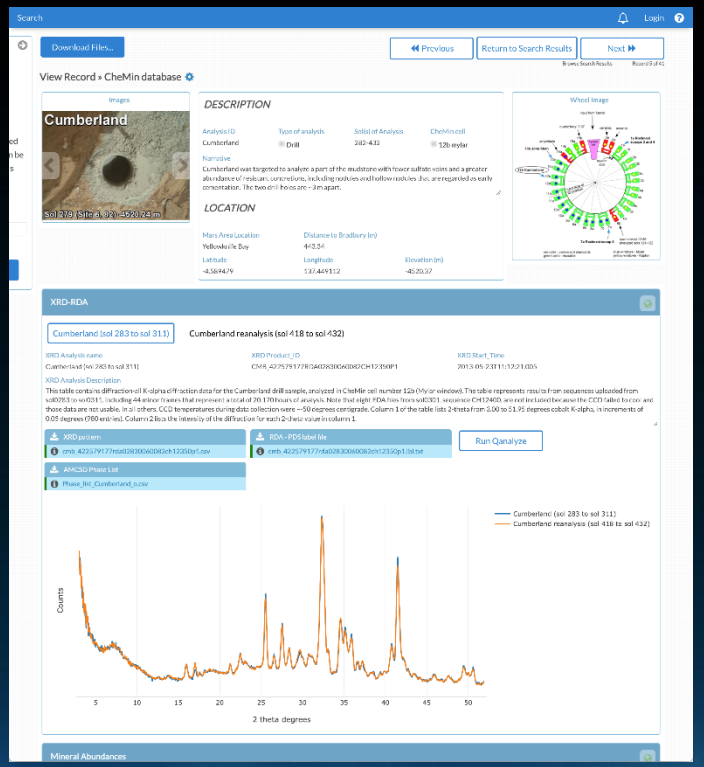
Lunar regolith XRD (Taylor et al, 2019)



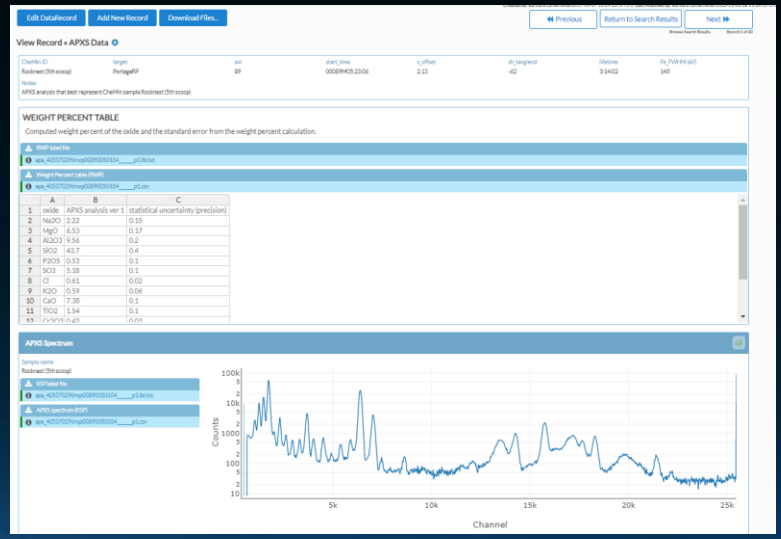
PMSL- EDAX Database (Blake et al, 2019)



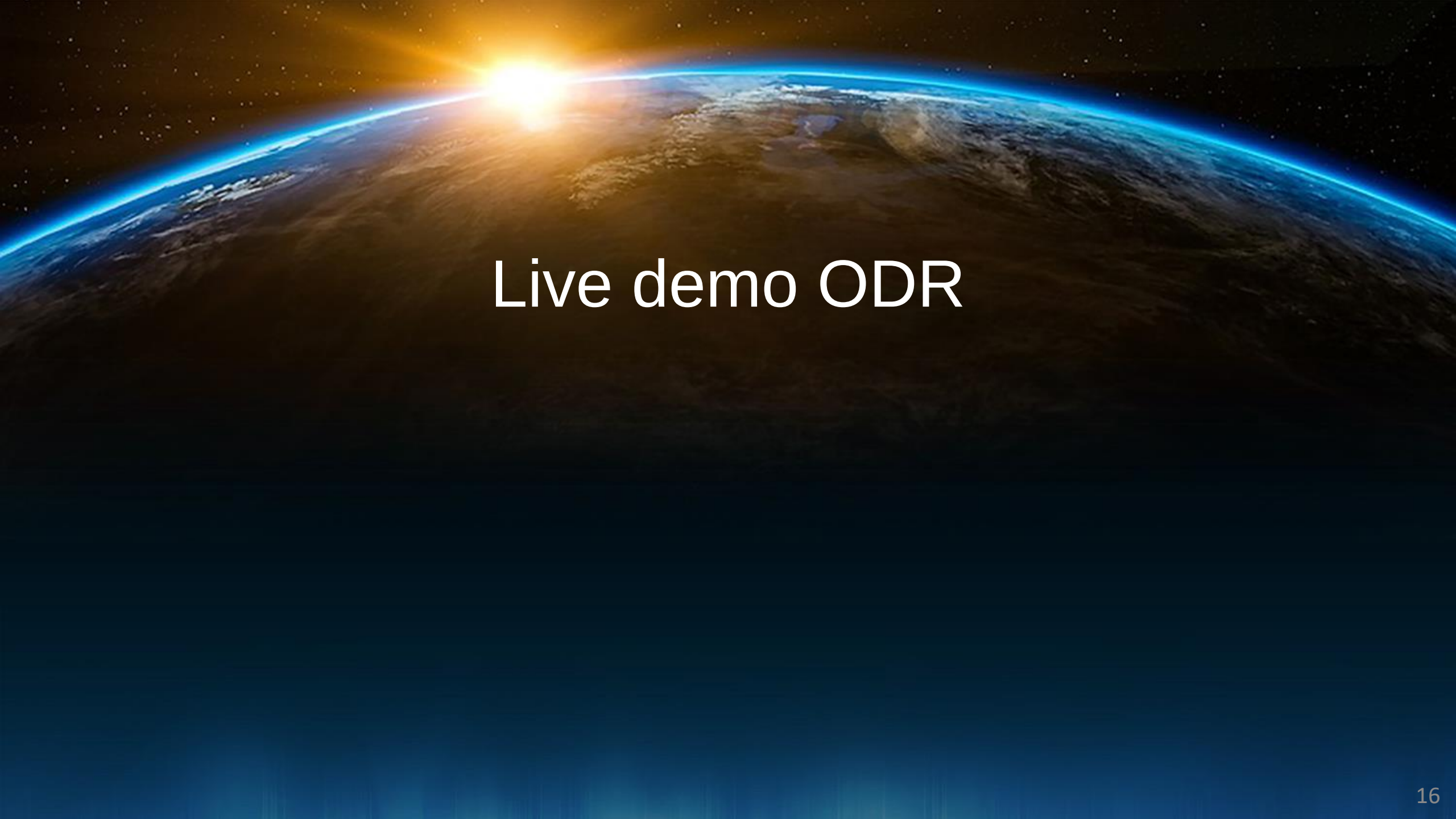
SAM – MSL Database



CheMin – MSL Database



APXS – MSL Database


Live demo ODR

Questions

CONTACT US!

Thomas Bristow (NASA Ames) – thomas.f.bristow@nasa.gov
Chris Dateo (NASA Ames) – christopher.dateo@nasa.gov
Kevin Boydstun (NASA Ames) – kevin.boydstun@nasa.gov
Barbara Lafuente (SETI) – blafuente@seti.org
Nate Stone (ODR) – nate.stone@opendatarepository.org
Mark Vorobets (NASA Ames) – mark.v.vorobets@nasa.gov
Shawn Wolfe (NASA Ames) – shawn.r.wolfe@nasa.gov

AHED

Thank you!

SEMINAR MATERIALS:
https://doi.org/10.48667/rspc-7y20