**DATA SHARING IN ASTROBIOLOGY: THE ASTROBIOLOGY HABITABLE ENVIRONMENTS DATABASE (AHED).** B. Lafuente[1], T. Bristow[1], N. Stone[2], A. Pires[3], R. M. Keller[1], R. T. Downs[3], D. Blake[1] and M. Fonda[1]. [1]NASA Ames Research Center, Mountain View, CA (thomas.f.bristow@nasa.gov), [2]Open Data Repository, Gray, ME [3]University of Arizona, Tucson, AZ.

**Introduction:** Astrobiology is a multidisciplinary area of scientific research focused on studying the origins of life on Earth and the conditions under which life might have emerged elsewhere in the universe. NASA uses the results of Astrobiology research to help define targets for future missions that are searching for life elsewhere in the universe.

The understanding of complex questions in Astrobiology requires integration and analysis of data spanning a range of disciplines including biology, chemistry, geology, astronomy and planetary science. However, the lack of a centralized repository makes it difficult for Astrobiology teams to share data and benefit from resultant synergies. Moreover, in recent years, federal agencies are requiring that results of any federally funded scientific research must be available and useful for the public and the science community.

The Astrobiology Habitable Environments Database (AHED), developed with a consolidated group of astrobiologists from different active research teams at NASA Ames Research Center, is designed to help to address these issues. AHED is a central, high-quality, long-term data repository for mineralogical, textural, morphological, inorganic and organic chemical, isotopic and other information pertinent to the advancement of the field of Astrobiology.

**Objectives**: AHED aims to promote the field of Astrobiology and increase scientific returns from NASA funded research by enabling data sharing, collaboration and exposure of non-NASA scientists to NASA research initiatives and missions.

The main goal of AHED is the creation of a single repository that has the flexibility to deal with the diversity of Astrobiology datasets, while allowing a degree of standardization necessary for more rapid database creation, fulfillment of data archiving mandates, as well as facilitating data discovery and mining through efficient search.

**Characteristics:** AHED is a collection of databases storing information about samples, measurements, analyses and contextual information about field sites where samples were collected, the instruments or equipment used for analysis, and people and institutions involved in their collection.

In the current implementation, metadata for each AHED database is defined separately and independent of other AHED databases. In the version under development, AHED will be structured based on framework of metadata templates. A published AHED metadata standard will sit at the highest level of this scheme, defining metadata requirements of AHED subscribing databases. Curation groups and users will create a library of AHED database templates to allow other scientists and researchers to make compatible, but flexible, database designs tailored to their datasets (Fig. 1). Eventually, the template system will allow these specifications to be published in commonly accepted metadata formats such as the Dublin Core Initiative's metadata standard (http://dublincore.org). All AHED databases will conform to the AHED metadata standard, allowing data mining and search through the AHED web portal.

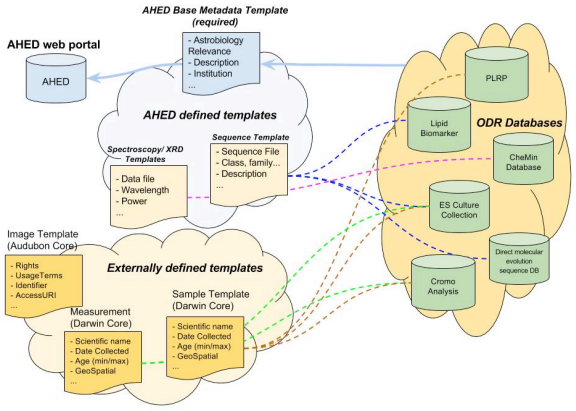


Figure 1. AHED framework of metadata templates.

**Infrastructure:** AHED will provide public and open-access to Astrobiology-related research data through a user-managed web portal implemented using open-source software created by the Open Data Repository (ODR)[1]. At the same time, public definition of the AHED metadata standard will allow other platforms and software to annotate datasets in a way that makes them discoverable and searchable by the AHED web portal. Tracking and publishing changes to the AHED metadata standard allows repository and database software to prompt database curators and owners to keep databases in compliance with the latest version of the AHED metadata standard.

*ODR's Data Publisher.* Astrobiology researchers often work within small communities or operate individually with unique data sets that don't easily fit into

existing database structures. ODR constructed its Data Publisher software to allow researchers to create databases with common metadata structures and subsequently extend them to meet their individual needs and data requirements.

The software accomplishes these tasks through a web based interface that allows collaborative creation and revision of common metadata templates and individual extensions to these templates for custom data sets. This allows researchers to search disparate datasets based on common metadata established through the metadata tools, but still enables project-specific data and analyses to be stored alongside the required common metadata. The software produces web pages that can be made publicly available at the researcher's discretion so that users may search and browse the data in an effort to make interoperability and data discovery a human-friendly task while also publishing semantic data for machine-based discovery. Once relevant data has been identified, researchers can utilize the built-in application programming interface (API) that exposes the data for machine-based consumption and integration with existing data analysis tools (e.g. R, MATLAB, Project Jupyter[2]).

*ODR Functionality.*

Drag-and-drop design: From the master template, administrators of databases can add different field types and modify the layout at any time during the lifetime of the database.

Graphing system: The ODR platform provides a diversity of graph types based on PlotlyJS (https://plot.ly/javascript/) (Fig. 2). This graphing system creates pre-rendered, static versions of each chart and stores them for display. Once the page is loaded, a user can click on any pre-rendered graph and switch to an interactive display that allows zooming, focusing on a single point or line, and many other features.

Large file upload: The system utilizes Flow.js (https://github.com/flowjs/flow.js), which enables browser-based large file uploads Also, users can upload multiple large files simultaneously allowing support for researchers who work with large data sets (e.g. genetic sequence data and high resolution images).

CSV import: ODR provides a CSV import function that will automatically generate a template and populate databases from a spreadsheet, allowing users to import large sets of data in a very short time.

Permission system: A powerful and versatile permission system protects confidentiality and helps preserve data integrity and provenance by ensuring only the users who are authorized can see data and make changes.

Citation: A citation system will allow research data to be used and appropriately referenced by other researchers after the data are made public.

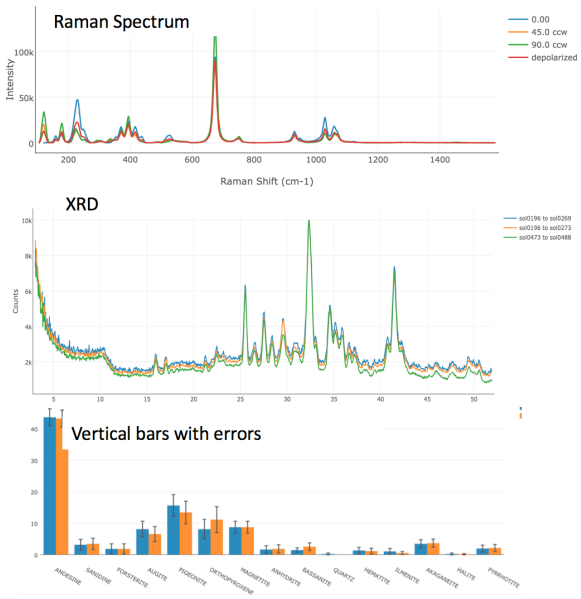


Figure 2. Example of graphs plotted in ODR.

**References:** [1] Stone N. et al. (2015) AGU, abstract IN44A-08. [2] Pérez F. et al. (2007) *Computing in Science & Engineering*, 9.

**Acknowledgment:** We gratefully acknowledge the support for this study by the Science-Enabling Research Activity (SERA) and NASA NNX11AP82A, Mars Science Laboratory Investigations.