AHED

The Astrobiology Habitable Environments Database

B. Lafuente, N. Stone, R. T. Downs, D. Blake, T. Bristow, M. Fonda and A. Pires

Data Visualization 1: Data Visualization for Information Exploration and Knowledge Discovery I

AGU Fall Meeting, Thursday 17th December 2015
ASTROBIOLOGY

It studies the origin, evolution, and distribution of life in the universe.

“Astrobiology is MULTIDISCIPLINARY in content and INTERDISCIPLINARY in its execution”

(The NASA Astrobiology Roadmap, Des Marais et al., 2008)
About the AHED project

- Mineral spectroscopy database
- Rigid structure
- Any change in structure requires extensive re-programing

- Diverse datatypes
- Flexible structure
- Ability to change design and include new data at any time

PROJECT SUPPORTED BY:

Science-enabling research activity (SERA)
NASA NNX11AP82A, Mars Science Laboratory Investigations
Efforts on data sharing in Astrobiology

“results of federally funded scientific research must be available and useful for the public, industry, and science community”
(J. P. Holdren memo from OSTP, Feb 22, 2013)

Benefits of sharing data:
- Reanalysis of data to verify results
- Reinterpretation of data with a different approaches
- Data integrity and preservation
- Eliminates data redundancy
- Training tool for future researches

Barriers on data sharing:

- [Aydinoglu et al., 2014]

AHED + ODR

[Aydinoglu et al., 2014]
AHED  The Astrobiology Habitable Environments Database

- Morphological data
- Textural and contextual images
- Chemistry
- Biochemistry
- Isotopic data
- Sequencing data
- Mineralogy and crystallography
- etc

Take home message

✓ AHED is a centralized, high quality, user friendly long term repository for archiving and collaborative sharing data relevant to Astrobiology.

✓ ODR provide a simple tool that allows quick creation of data structures for searching and display.

✓ GOALS:
  - Foster long-term innovative research by supporting integration and analysis of diverse datasets
  - Promote interdisciplinary and collaborative research amongst widely-distributed investigators in astrobiology.
Project Progress

Drag-and-drop 😊

Insert fields

Template layout

Types of fields
## Project Progress

- **Drag-and-drop**: Happy face
- **CSV Import**: Happy face

---

### Data Type List

<table>
<thead>
<tr>
<th>Type Name</th>
<th>Number of Records</th>
<th>Search Layout</th>
<th>Field Layout</th>
<th>Datatype Properties</th>
<th>Download Schema</th>
<th>CSV Import</th>
<th>CSV Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORDAO-Analysis</td>
<td>136</td>
<td>Edit</td>
<td>Edit</td>
<td>Edit Properties</td>
<td>Download Schema</td>
<td>Import</td>
<td>Delete</td>
</tr>
<tr>
<td>CORDAO-Writs</td>
<td>14</td>
<td>Edit</td>
<td>Edit</td>
<td>Edit Properties</td>
<td>Download Schema</td>
<td>Import</td>
<td>Delete</td>
</tr>
</tbody>
</table>

---

### Template

- Populated database

---

#### Project Progress Table

<table>
<thead>
<tr>
<th>Date Sampled</th>
<th>Well</th>
<th>Drilled Depth (m)</th>
<th>Depth to Water Before Sampling (m)</th>
<th>Temperature (°C)</th>
<th>Specific Conductivity (μS/cm)</th>
<th>Dissolved Oxygen (mg/L)</th>
<th>pH</th>
<th>ORP (mV)</th>
<th>Zn (μg/L)</th>
<th>Fe (μg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/3/2011</td>
<td>N-08A</td>
<td>130</td>
<td>3</td>
<td>10.55</td>
<td>115.92</td>
<td>3.44</td>
<td>4.48</td>
<td>-22.1</td>
<td>2.7</td>
<td>-30.9</td>
</tr>
<tr>
<td>4/3/2011</td>
<td>N-08B</td>
<td>86</td>
<td>0</td>
<td>52</td>
<td>46.0</td>
<td>0.11</td>
<td>-30.5</td>
<td>-30.9</td>
<td>2.7</td>
<td>-30.9</td>
</tr>
<tr>
<td>7/3/2011</td>
<td>N-09A</td>
<td>75</td>
<td>0</td>
<td>17.46</td>
<td>31.51</td>
<td>1.68</td>
<td>0.86</td>
<td>-30.9</td>
<td>2.7</td>
<td>-30.9</td>
</tr>
<tr>
<td>7/3/2011</td>
<td>N-09B</td>
<td>35</td>
<td>0</td>
<td>12.56</td>
<td>0.18</td>
<td>0.68</td>
<td>0.68</td>
<td>-30.9</td>
<td>2.7</td>
<td>-30.9</td>
</tr>
<tr>
<td>7/3/2011</td>
<td>N-10A</td>
<td>58</td>
<td>0</td>
<td>17.58</td>
<td>0.87</td>
<td>0.68</td>
<td>0.68</td>
<td>-30.9</td>
<td>2.7</td>
<td>-30.9</td>
</tr>
<tr>
<td>7/3/2011</td>
<td>N-10B</td>
<td>81</td>
<td>0</td>
<td>11.53</td>
<td>0.87</td>
<td>0.68</td>
<td>0.68</td>
<td>-30.9</td>
<td>2.7</td>
<td>-30.9</td>
</tr>
<tr>
<td>7/3/2011</td>
<td>N-11A</td>
<td>25.5</td>
<td>0</td>
<td>9.85</td>
<td>0.25</td>
<td>0.68</td>
<td>0.68</td>
<td>-30.9</td>
<td>2.7</td>
<td>-30.9</td>
</tr>
<tr>
<td>7/3/2011</td>
<td>N-11B</td>
<td>25.5</td>
<td>0</td>
<td>9.85</td>
<td>0.25</td>
<td>0.68</td>
<td>0.68</td>
<td>-30.9</td>
<td>2.7</td>
<td>-30.9</td>
</tr>
</tbody>
</table>
**Project Progress**

- Drag-and-drop
- CSV Import
- CSV Export

CSV file

The user can select which fields will be exported

R/Shiny

Tools for data analysis
Project Progress

**Login access**

Username: barbaralafuente@email.ar
Password: ************

Forgot Password?

**User roles**

<table>
<thead>
<tr>
<th>User Name</th>
<th>User</th>
<th>Admin</th>
<th>Super Admin</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:nate@opendaterepository.org">nate@opendaterepository.org</a></td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
</tr>
<tr>
<td><a href="mailto:gpiros@email.arizona.edu">gpiros@email.arizona.edu</a></td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td><a href="mailto:test@hotmail.com">test@hotmail.com</a></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>downs</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td><a href="mailto:barbaralafuente@email.arizona.edu">barbaralafuente@email.arizona.edu</a></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

**User permissions**

Permissions List > Test User 3 (User)

<table>
<thead>
<tr>
<th>Type Name</th>
<th>Can View Datatype</th>
<th>Can Edit Records</th>
<th>Can Add Records</th>
<th>Can Delete Records</th>
<th>Can Design Datatype</th>
<th>Can Administer Datatype</th>
</tr>
</thead>
<tbody>
<tr>
<td>CROMO-Analysis</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>datafields...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CROMO-Wells</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>datafields...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Project Progress

Drag-and-drop

CSV Import

CSV Export

User permissions

Public/non-public

**DATABASE visible or non-visible to the public**

**SELECTED DATA visible or non-visible to the public**

**INDIVIDUAL RECORDS visible or non-visible to the public**
AHED Pilot Databases

PLRP - Pavillon Lake Research Project DB
- [www.odr.io/PLRP](http://www.odr.io/PLRP) ---- 536 records
- Microbialite and water samples

Direct molecular evolution sequence DB
- [www.odr.io/invitro](http://www.odr.io/invitro) ---- 43 records
- Sequence data from in-Vitro evolution experiments

CheMin DB
- [www.odr.io/chemin](http://www.odr.io/chemin) ---- 12 records
- Data from CheMin instrument (MSL-Curiosity)

ES Culture Collection DB
- [http://odr.io/ES_cultures](http://odr.io/ES_cultures) ---- 9 records
- Data from isolated cyanobacteria and heterotrophs

CROMO: Serpentinizing System DB
- [www.odr.io/CROMO_analysis](http://www.odr.io/CROMO_analysis) ---- 106 records
- Drill samples from serpentinizing systems

Lipid Biomarker DB
- [http://odr.io/biomarker](http://odr.io/biomarker) ---- 2 records
- Lipids characterize in pure cultures of microbes
AHED Pilot Databases

PLRP - Pavillon Lake Research Project DB
- www.odr.io/PLRP ---- 536 records
- Microbialite and water samples

Direct molecular evolution sequence DB
- www.odr.io/invitro ---- 43 records
- Sequence data from in-Vitro evolution experiments

CheMin DB
- www.odr.io/chemin ---- 12 records
- Data from CheMin instrument (MSL-Curiosity)

ES Culture Collection DB
- http://odr.io/ES_cultures ---- 9 records
- Data from isolated cyanobacteria and heterotrophs

CROMO: Serpentinizing System DB
- www.odr.io/CROMO_analysis ---- 106 records
- Drill samples from serpentinizing systems

Lipid Biomarker DB
- http://odr.io/biomarker ---- 2 records
- Lipids characterize in pure cultures of microbes
# AHED Pilot Databases

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Website</th>
<th>Record Count</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLRP - Pavillon Lake Research Project DB</td>
<td><a href="http://www.odr.io/PLRP">www.odr.io/PLRP</a></td>
<td>536 records</td>
<td>Microbialite and water samples</td>
</tr>
<tr>
<td>Direct molecular evolution sequence DB</td>
<td><a href="http://www.odr.io/invitro">www.odr.io/invitro</a></td>
<td>43 records</td>
<td>Sequence data from in-Vitro evolution experiments</td>
</tr>
<tr>
<td>CheMin DB</td>
<td><a href="http://www.odr.io/chemin">www.odr.io/chemin</a></td>
<td>12 records</td>
<td>Data from CheMin instrument (MSL-Curiosity)</td>
</tr>
<tr>
<td>ES Culture Collection DB</td>
<td><a href="http://odr.io/ES_cultures">http://odr.io/ES_cultures</a></td>
<td>9 records</td>
<td>Data from isolated cyanobacteria and heterotrophs</td>
</tr>
<tr>
<td>CROMO: Serpentining System DB</td>
<td><a href="http://www.odr.io/CROMO_analysis">www.odr.io/CROMO_analysis</a></td>
<td>106 records</td>
<td>Drill samples from serpentinizing systems</td>
</tr>
<tr>
<td>Lipid Biomarker DB</td>
<td><a href="http://odr.io/biomarker">http://odr.io/biomarker</a></td>
<td>2 records</td>
<td>Lipids characterize in pure cultures of microbes</td>
</tr>
</tbody>
</table>
AHED Pilot Databases

PLRP - Pavillon Lake Research Project DB
- www.odr.io/PLRP ---- 536 records
- Microbialite and water samples

Direct molecular evolution sequence DB
- www.odr.io/invitro ---- 43 records
- Sequence data from in-Vitro evolution experiments

CheMin DB
- www.odr.io/chemin ---- 12 records
- Data from CheMin instrument (MSL-Curiosity)

ES Culture Collection DB
- http://odr.io/ES_cultures ---- 9 records
- Data from isolated cyanobacteria and heterotrophs

CROMO: Serpentining System DB
- www.odr.io/CROMO_analysis ---- 106 records
- Drill samples from serpentinating systems

Lipid Biomarker DB
- http://odr.io/biomarker ---- 2 records
- Lipids characterize in pure cultures of microbes
AHED Pilot Databases

PLRP - Pavillon Lake Research Project DB
- [www.odr.io/PLRP](http://www.odr.io/PLRP) --- 536 records
- Microbialite and water samples

Direct molecular evolution sequence DB
- [www.odr.io/invitro](http://www.odr.io/invitro) --- 43 records
- Sequence data from in-Vitro evolution experiments

CheMin DB
- [www.odr.io/chemin](http://www.odr.io/chemin) --- 12 records
- Data from CheMin instrument (MSL-Curiosity)

ES Culture Collection DB
- [http://odr.io/ES_cultures](http://odr.io/ES_cultures) --- 9 records
- Data from isolated cyanobacteria and heterotrophs

CROMO: Serpentinizing System DB
- [www.odr.io/CROMO_analysis](http://www.odr.io/CROMO_analysis) --- 106 records
- Drill samples from serpentinizing systems

Lipid Biomarker DB
- [http://odr.io/biomarker](http://odr.io/biomarker) --- 2 records
- Lipids characterize in pure cultures of microbes
AHED Pilot Databases

PLRP - Pavillon Lake Research Project DB
- www.odr.io/PLRP ---- 536 records
- Microbialite and water samples

Direct molecular evolution sequence DB
- www.odr.io/invitro ---- 43 records
- Sequence data from in-Vitro evolution experiments

CheMin DB
- www.odr.io/chemin ---- 12 records
- Data from CheMin instrument (MSL-Curiosity)

ES Culture Collection DB
- http://odr.io/ES_cultures ---- 9 records
- Data from isolated cyanobacteria and heterotrophs

CROMO: Serpentinizing System DB
- www.odr.io/CROMO_analysis ---- 106 records
- Drill samples from serpentinizing systems

Lipid Biomarker DB
- http://odr.io/biomarker ---- 2 records
- Lipids characterize in pure cultures of microbes
AHED Pilot Databases

Take home message

✓ Unique URL
✓ Disparate datasets
✓ Different template designs
✓ Some data in common
Road map

<table>
<thead>
<tr>
<th>Feature</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drag-and-drop</td>
<td>😊</td>
</tr>
<tr>
<td>CSV Import</td>
<td>😊</td>
</tr>
<tr>
<td>CSV Export</td>
<td>😊</td>
</tr>
<tr>
<td>Public/non-public</td>
<td>😊</td>
</tr>
<tr>
<td>User permissions</td>
<td>😊</td>
</tr>
<tr>
<td>Semantic Web</td>
<td>🕒</td>
</tr>
</tbody>
</table>

OWL, RDF, RDFa

ONTOLOGIES

DATA MINING

ASTROBIOLOGY RELEVANCE
- Mars
- Earth
- Europa

ENVIRONMENT
- Hypersaline
- Hydrothermal
- Dessert

SAMPLE TYPE
- Core
- Water
- Microbial mat

ANALYSES
- Chemistry
- Isotopes
- culturing
Road map

<table>
<thead>
<tr>
<th>Feature</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drag-and-drop</td>
<td>😊😊😊</td>
</tr>
<tr>
<td>CSV Import</td>
<td>😊😊😊</td>
</tr>
<tr>
<td>CSV Export</td>
<td>😊😊😊</td>
</tr>
<tr>
<td>Public/non-public</td>
<td>😊😊😊</td>
</tr>
<tr>
<td>User permissions</td>
<td>😊😊😊</td>
</tr>
<tr>
<td>Semantic Web</td>
<td>😢😢😢</td>
</tr>
<tr>
<td>Citation</td>
<td>😢😢😢</td>
</tr>
<tr>
<td>User feedback</td>
<td>😢😢😢</td>
</tr>
<tr>
<td>NASA’s server</td>
<td>😢😢😢</td>
</tr>
</tbody>
</table>

Citation

DOI  AHED

User feedback

NASA’s server

Stability

Security Service

Minimal downtimes
Illimited capacity
Continued growth

AGU Fall Meeting, Thursday 17th December 2015
AHED The Astrobiology Habitable Environments Database

Take home message

✓ First goal achieved: system allows flexibility for data storage and visualization.

✓ As the system matures, tools for embracing the semantic web will be added and citation system will allow research data to be used and appropriately cited.

✓ With the use of the NAS, stability, continued growth, online analyses capabilities and the size of the user community will not be limited by the database platform architecture, processor speed, internet connectivity, or data storage.

✓ We expect that as AHED growth, it will be use beyond the Astrobiology community and related scientific fields will be able to connect their research with AHED to make possible a broader framework of databases for hypothesis generation and knowledge discovery.
To contact us:

- Robert Downs: rdowns@u.arizona.edu
- Barbara Lafuente: barbaralafuente@email.arizona.edu
- David Blake: david.blake@nasa.gov
- Tom Bristow: thomas.f.bristow@nasa.gov
- Nate Stone: nate.stone@opendatarepository.org