

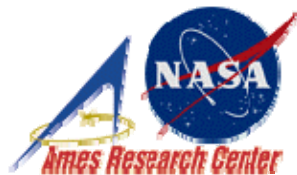


AHED

The Astrobiology Habitable Environments Database

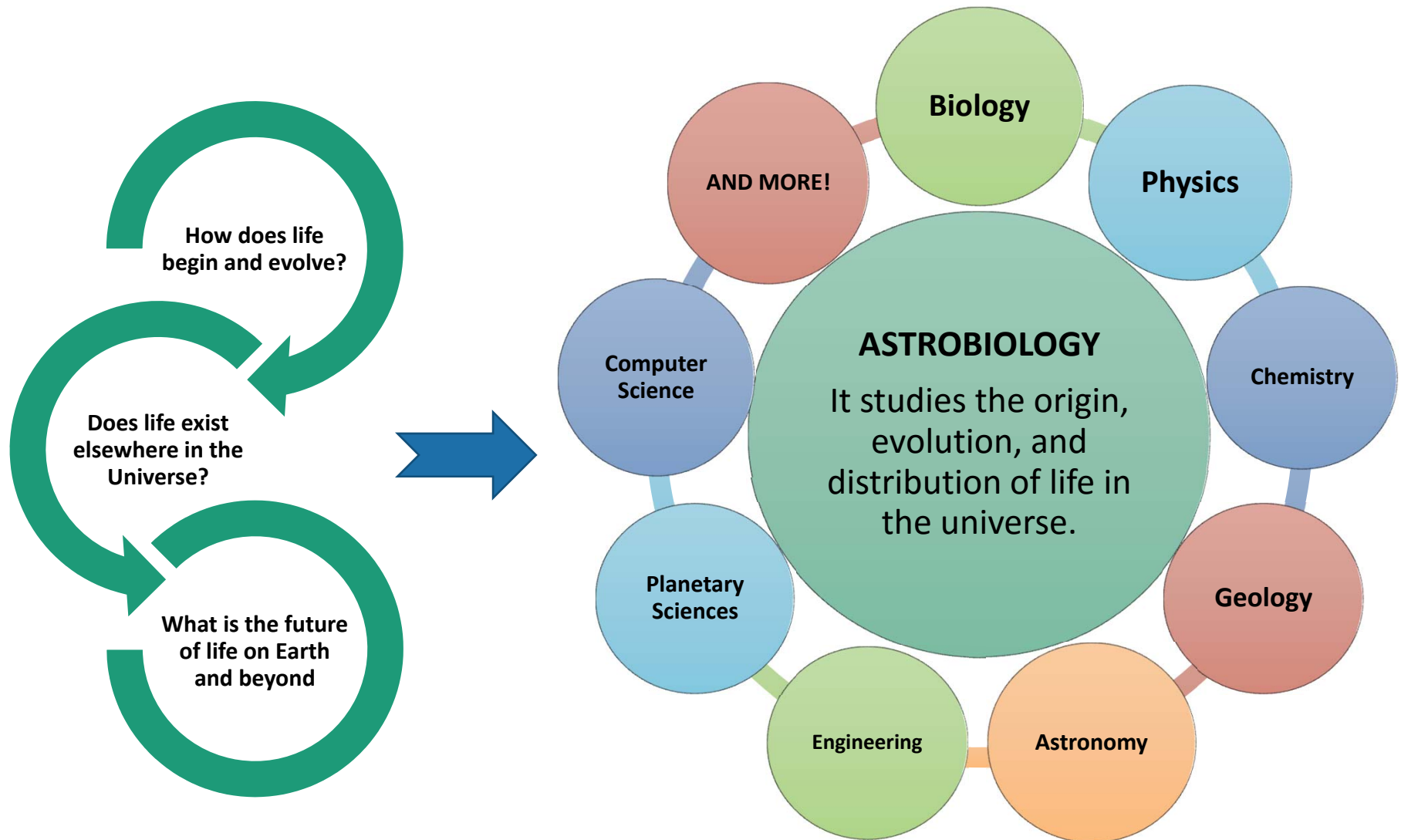
B. Lafuente, N. Stone, R. T. Downs, D. Blake, T. Bristow, M. Fonda and A. Pires

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



AGU Fall Meeting, Thursday 17th December 2015

ASTROBIOLOGY



"Astrobiology is **MULTIDISCIPLINARY** in content and **INTERDISCIPLINARY** in its execution"

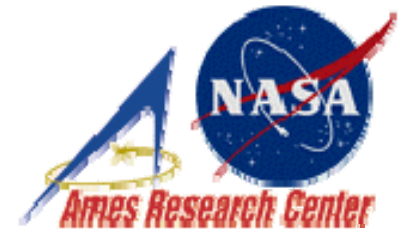
(The NASA Astrobiology Roadmap, Des Marais *et al.*, 2008)

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”

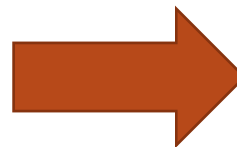
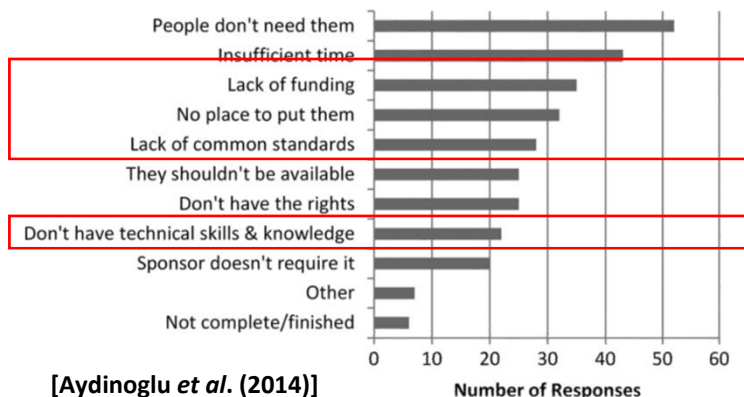
(J. P. Holdren memo from OSTP, Feb 22, 2013)

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

[Aydinoglu et al., 2014]

Barriers on data sharing:

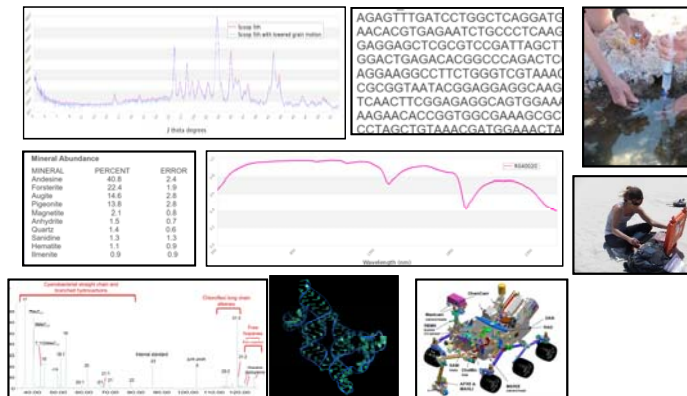


AHED + ODR

[Aydinoglu et al., 2014]

AHED *The Astrobiology Habitable Environments Database*

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



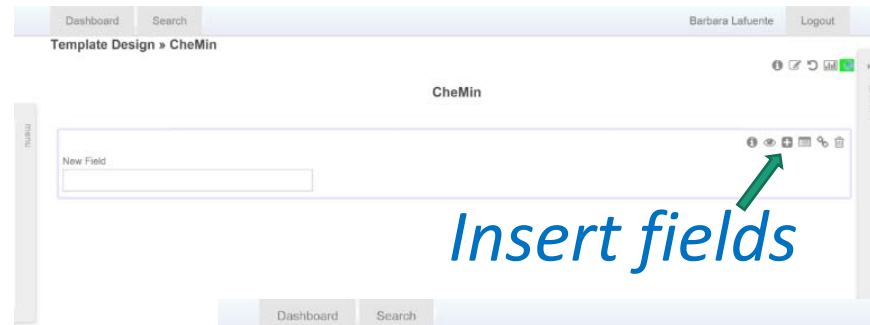
Libraries
Published data
Digital notebook
Planetary missions

Take home message

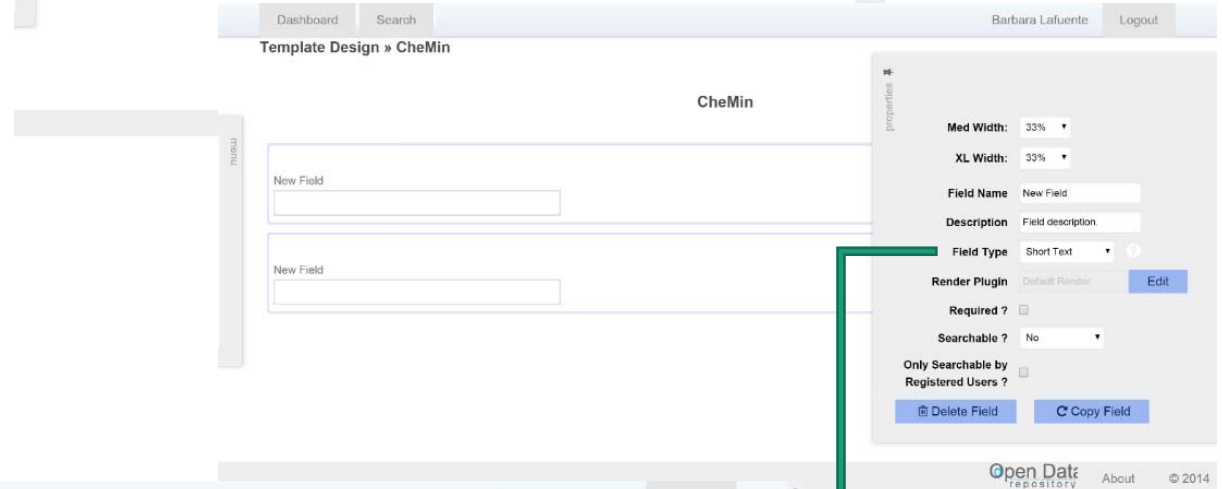
- ✓ **AHED** is a centralized, high quality, user friendly long term repository for archiving and collaborative sharing data relevant to Astrobiology.
- ✓ **ODR** provide a simple tool that allows quick creation of data structures for searching and display.
- ✓ **GOALS:**
 - ❖ Foster **long-term** innovative research by supporting integration and analysis of diverse datasets
 - ❖ Promote **interdisciplinary and collaborative** research amongst widely-distributed investigators in astrobiology.

Project Progress

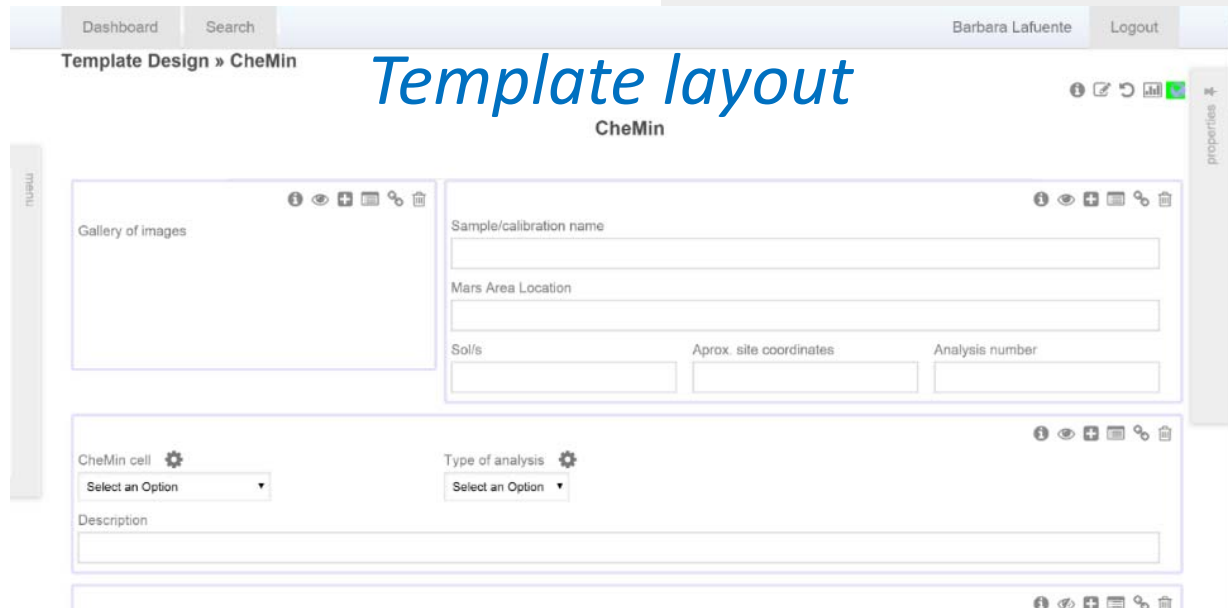
Drag-and-drop



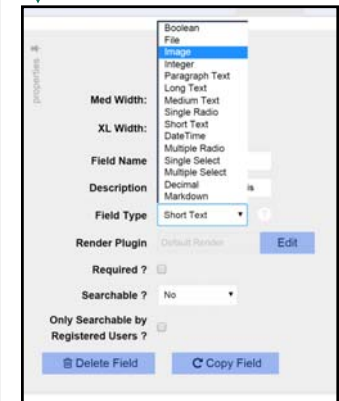
Insert fields



Types of fields



Template layout



Project Progress

Drag-and-drop	😊
CSV Import	😊

[illegible]

Data Type List

Type Name	Number of Records	Search Layout	Field Layout	Datatype Properties	Download Schema	CSV Import	Delete
CROMO-Analysis	106	 Edit	 Edit	 Edit Properties	 Download	 Import	 Delete
CROMO-Wellis	14	 Edit	 Edit	 Edit Properties	 Download	 Import	 Delete

Template

**Populated
database**

[illegible]

Project Progress

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

QV 1,1	2011-11-02	49	0	11.55	0	0.67	0
CSW 1,3	2011-12-05	73					

Showing 1 to 10 of 106 entries

Previous **1** 2 3 4 5 ... 11 Next

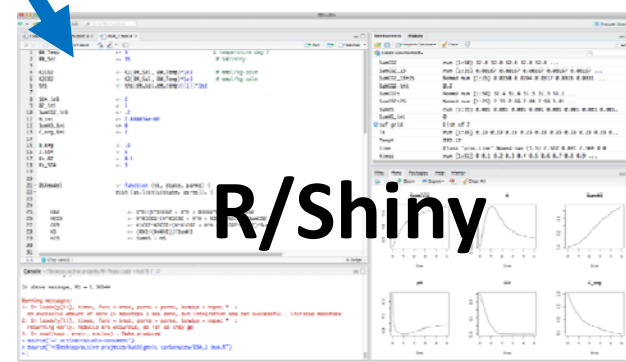
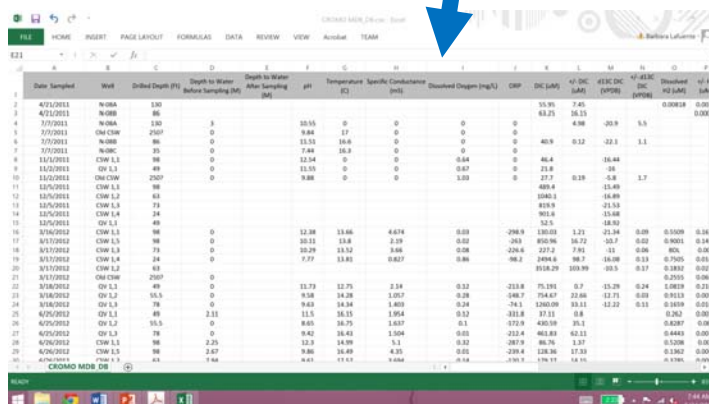
Mass Edit all these datarecords **Mass Export these datarecords into CSV file**

The user can select which fields will be exported

CSV file

Tools for data analysis

R/Shiny



Project Progress

Drag-and-drop	😊
CSV Import	😊
CSV Export	😊
User permissions	😊

Login access

Login

Username:

Password:

☐ Remember me

[Forgot Password?](#)

User roles

User Name	User	Admin	Super Admin
nate@opendatarepository.org	☐	☐	☒
ajpires@email.arizona.edu	☐	☐	☒
test@hotmail.com	☒	☐	☐
downs	☐	☐	☒
barbaralafuente@email.arizona.edu	☐	☐	☒

User permissions

Permissions List > Test User 3 (User)

Type Name	Can View Datatype	Can Edit Records	Can Add Records	Can Delete Records	Can Design Datatype	Can Administrate Datatype
CROMO-Analysis	☐	☐ datafields...	☐	☐	☐	☐
CROMO-Wells	☐	☐ datafields...	☐	☐	☐	☐

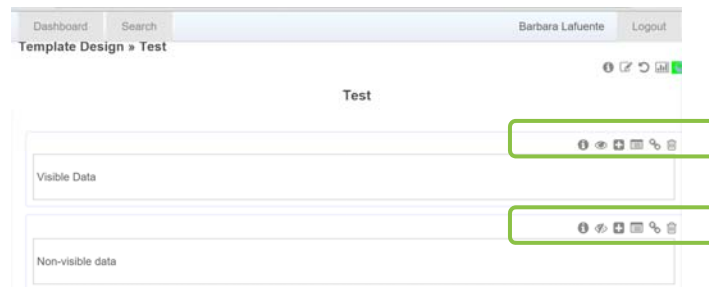
Project Progress

Drag-and-drop	☺
CSV Import	☺
CSV Export	☺
User permissions	☺
Public/non-public	☺

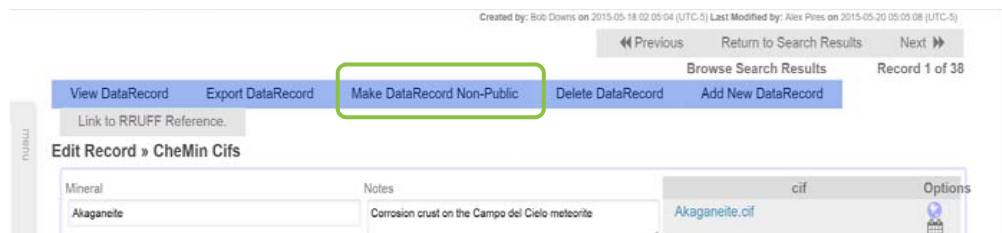
DATABASE visible or non-visible to the public



SELECTED DATA visible or non-visible to the public



INDIVIDUAL RECORDS visible or non-visible to the public

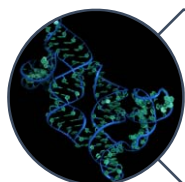


AHED Pilot Databases



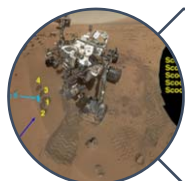
PLRP - Pavillon Lake Research Project DB

- www.odr.io/PLRP ---- 536 records
- Microbialite and water samples



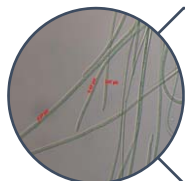
Direct molecular evolution sequence DB

- 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)



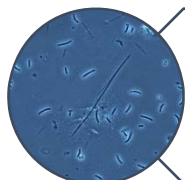
ES Culture Collection DB

- 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

Dashboard Search Barbara Lafante Logout
Created by: Ron Kuter on 2015-07-29 10:07:47 UTC Last Modified by: Ron Kuter on 2015-07-29 10:07:48 UTC

Edit DataRecord 44 Previous Return to Search Results Next 39
Browse Search Results Record 1 of 536

Search Result » PLRP

Sample Identification

#1			
sampleID	name	sampleType	sampleDescription
TP06207	Three Poles	microbialite	

Sample Collection Location

location	samplingArea	samplingSubarea
Pavillon Lake	Pavillon Lake: Central Basin	Central Basin: Three Poles
latitude	longitude	depthInMeters
50.888758	121.739987	35

Sample Collection Timing

collectionDate	collectionYear	collectionMonth
2007-06-22	2007	6
collectionTime	season	

Sample Collection Context

iceConditions	light	station	commandID

Sample Collection Rationale

scienceHypotheses	scienceQuestions
otherRationale	

Sample Processing and Curation

parentSample	replicateSamples	archivalStatus
0	0	
collectionMode	collector	collectionContainerType
collection	storageLocation	sampleAnalyses
initialQuantity	remainingQuantity	quantityUnit
0	0	

Sample Description, Collection, and Handling Comments

sampleDescription
notes
sharing

Microbialite Sample Characteristics

microMacroMorphology	surfaceTexture	predominantColors
colorNotes	evolutionPresent	angularitySizeCategory
wholeStructure	exactStructureSizeCM	structureSizeCategory
	0	

Other Information

photos
contributor
Biddle

Water Sample Characteristics

pH	alkalinity
0	0
hardness	cond
0	
Fe	Na
0	0
Ca	SO4
0	0
NO3	Cl
0	0
DOC	TP
0	0
DOC	TP
0	0
DOC	TP
0	0

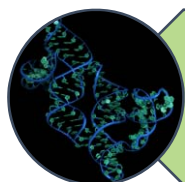
Open Data
© 2014

AHED Pilot Databases



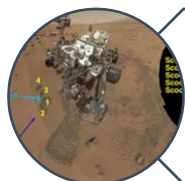
PLRP - Pavillon Lake Research Project DB

- www.odr.io/PLRP ---- 536 records
- Microbialite and water samples



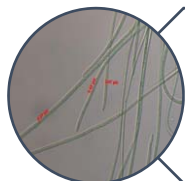
Direct molecular evolution sequence DB

- 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)



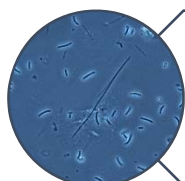
ES Culture Collection DB

- 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

Dashboard

Search

Barbara Lafuente

Logout

Created by: Minna Popovic on 2015-08-14 08:08:03 (UTC-5) Last Modified by: Minna Popovic on 2015-10-07 04:10:19 (UTC-5)

Edit DataRecord

<< Previous

Return to Search Results

Next >>

Browse Search Results

Record 42 of 43

Search Result » inVitro

SUMMARY

Sample ID

MP_M7b_7MK_MAD

Molecule Type

RNA

Function type

Catalyze a chemical reaction

Function

Nuclease

ASSOCIATED PUBLICATION

Title

In vitro evolution of distinct self-cleaving ribozymes in diverse environments.

Author(s)

Popovic, M, Fias PS, Otter MA.

Journal, year, page, etc

In vitro evolution of distinct self-cleaving ribozymes in diverse environments. Popovic M, Fias PS.

CONSTRUCT

Type of randomization

Full

Length of molecules (ntaa)

203

Length of randomized region

90

Comments on construct

SELECTION CONDITIONS

Number of rounds

7

Were the same conditions used for all rounds?

pH

7

Temperature

22

[Mg] (mM)

5

[Ca] (mM)

[Na] (mM)

[Na] (mM)

[K] (mM)

[Li] (mM)

Other ions

Small molecules

Polymers

Other

Comments

Solutions were sparged with Argon and kept under Argon during selection

SEQUENCING TECHNOLOGY

Type

Illumina

Number of reads/clones

11524043

Sequence data

ILLUMINA STATISTICS

Version

1.8

Index No

18

Index

GTCCGC

Reads raw

11524043

Yield (Mbases)

752

% PF

66.4

Reads PF

7916999

% of raw clusters per lane

5.65

% Perfect Index Reads

95.35

% One Mismatch Reads (Index)

4.65

% of >= Q30 Bases (PF)

84.91

Mean Quality Score (PF)

33.85

QF sequence reads

4423202

%Spaced trim

0.370948176

Optimized >

14

Open Data

Results

About

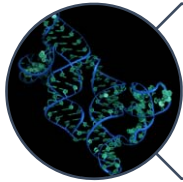
© 2015

AHED Pilot Databases



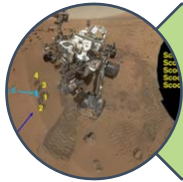
PLRP - Pavillon Lake Research Project DB

- www.odr.io/PLRP ---- 536 records
- Microbialite and water samples



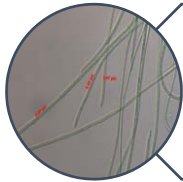
Direct molecular evolution sequence DB

- 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)



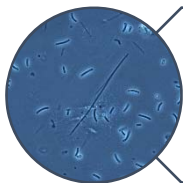
ES Culture Collection DB

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



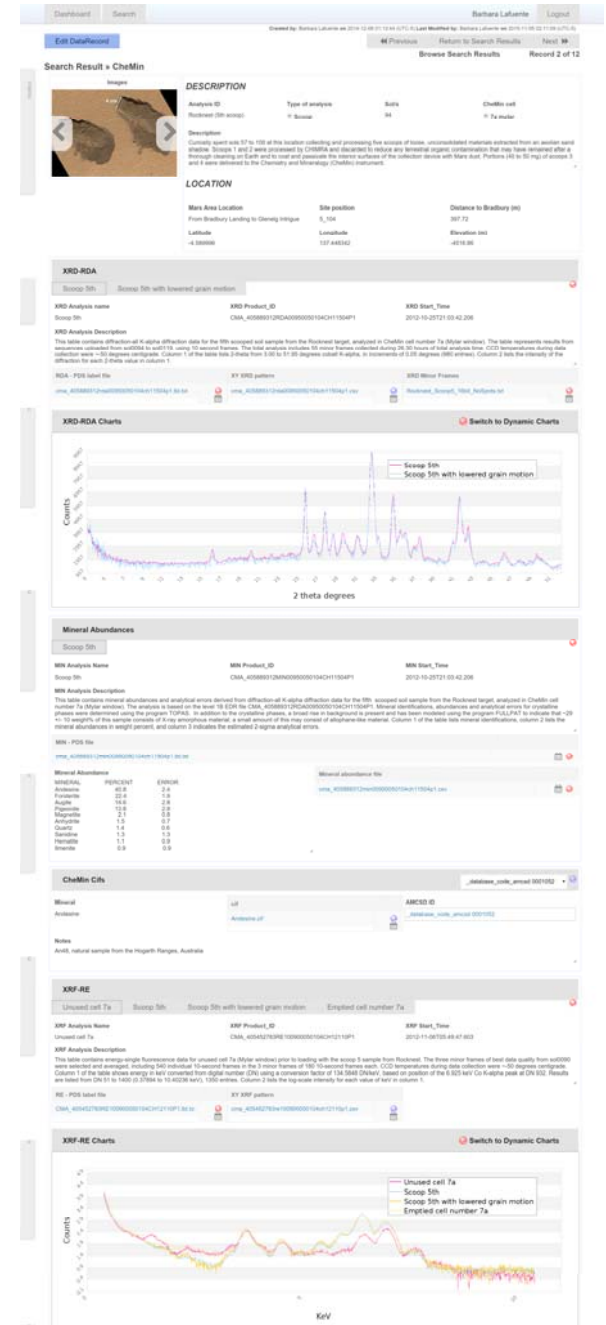
CROMO: Serpentinizing System DB

- 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

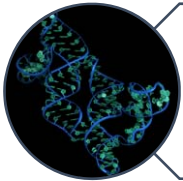


AHED Pilot Databases



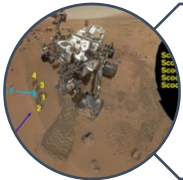
PLRP - Pavillon Lake Research Project DB

- www.odr.io/PLRP ---- 536 records
- Microbialite and water samples



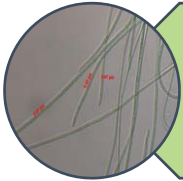
Direct molecular evolution sequence DB

- 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)



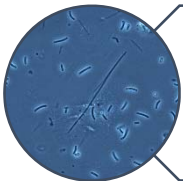
ES Culture Collection DB

- 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

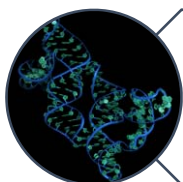
The screenshot displays the ODR Data Portal interface. At the top, there's a navigation bar with 'Data Portal' and 'About' links. Below this, a search bar and filters are visible. The main content area shows search results for the 'ES Culture Collection' database. It includes a table with columns for 'Accession ID', 'Collection ID', and 'File ID'. Below the table, there's a section for 'Microbiology / Phylogeny' featuring a phylogenetic tree. Further down, there's a 'Growth / Isolation Conditions' section with various parameters like 'Media', 'Incubation', and 'Growth conditions'. At the bottom, there's a 'Collection Information' section with details about the collection and a large image of a field of serpentinizing systems.

AHED Pilot Databases



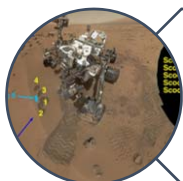
PLRP - Pavillon Lake Research Project DB

- www.odr.io/PLRP ---- 536 records
- Microbialite and water samples



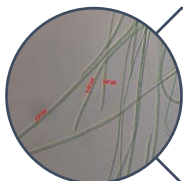
Direct molecular evolution sequence DB

- 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)



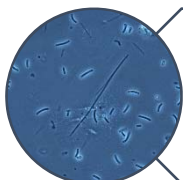
ES Culture Collection DB

- 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

Dashboard Search Created by: Barbara Laflamme on 2015-10-04 10:29 (UTC) / Last Modified by: Barbara Laflamme on 2015-10-11 09:10:23 (UTC-4)

Barbara Laflamme Logout

44 Previous Return to Search Results Next 106

Browse Search Results Record 4 of 106

Search Result > CROMO-Analysis

Sample ID
N-008_7A02011_00.00

CROMO-Wells

Well ID	Bottom of Hole (BOH) (m)	No Images Found
N-008	0	

In Situ Physical data

Date Sampled	Time sampled	Drilled Depth (Ft)	Drilled Depth (Ft)-Notes
2011-07-07	0:00	86	

Depth to Water Before Sampling (M)	Depth to Water After Sampling (M) - Notes	Depth to Water After Sampling (M)	Depth to Water After Sampling (M)-Notes
0			

Temperature (C)	pH	Specific Conductance (mS)
16.6	11.51	0

ORP	Dissolved Oxygen (mg/L)
	0

Dissolved Gases and Dissolved Inorganic Carbon

Dissolved H2 (uM)	Dissolved H2 (uM)-Notes	+/- H2 (uM)	+/- H2 (uM)-Notes

Dissolved CH4 (mM)	Dissolved CH4 (mM)-Notes	+/- CH4 (mM)	+/- CH4 (mM)-Notes
0.002		0.0004	

Dissolved CO (uM)	Dissolved CO (uM)-Notes	+/- CO (uM)	+/- CO (uM)-Notes

DIC (uM)	DIC (uM)-Notes	+/- DIC (uM)	+/- DIC (uM)-Notes
40.9		0.12	

d13C DIC (VPDB)	d13C DIC (VPDB)-Notes	+/- d13C DIC (VPDB)	+/- d13C DIC (VPDB)-Notes
-22.1		1.1	

d13C CH4 (VPDB)	d13C CH4 (VPDB)-Notes	+/- d13C CH4 (VPDB)	+/- d13C CH4 (VPDB)-Notes
-23.96			Value = Detection Limit

Major Inorganic Anion Concentrations

Fluoride (ppm)	Fluoride (ppm)-Notes	+/- Fluoride (ppm)	+/- Fluoride (ppm)-Notes

Chloride (ppm)	Chloride (ppm)-Notes	+/- Chloride (ppm)	+/- Chloride (ppm)-Notes

Nitrite (ppm)	Nitrite (ppm)-Notes	+/- Nitrite (ppm)	+/- Nitrite (ppm)-Notes

Sulfate (ppm)	Sulfate (ppm)-Notes	+/- Sulfate (ppm)	+/- Sulfate (ppm)-Notes

Bromide (ppm)	Bromide (ppm)-Notes	+/- Bromide (ppm)	+/- Bromide (ppm)-Notes

Nitrate (ppm)	Nitrate (ppm)-Notes	+/- Nitrate (ppm)	+/- Nitrate (ppm)-Notes

Phosphate (ppm)	Phosphate (ppm)-Notes	+/- Phosphate (ppm)	+/- Phosphate (ppm)-Notes

Organic Acid Concentrations

Lactate (uM)	Lactate (uM)-Notes	+/- Lactate (uM)	+/- Lactate (uM)-Notes
		0 = BDL	0.08

Acetate (uM)	Acetate (uM)-Notes	+/- Acetate (uM)	+/- Acetate (uM)-Notes
		0 = BDL	0.75

Formate (uM)	Formate (uM)-Notes	+/- Formate (uM)	+/- Formate (uM)-Notes
		0 = BDL	2.36

Propionate (uM)	Propionate (uM)-Notes	+/- Propionate (uM)	+/- Propionate (uM)-Notes
		0.05	

Butyrate (uM)	Butyrate (uM)-Notes	+/- Butyrate (uM)	+/- Butyrate (uM)-Notes
		0.6	

Valerate (uM)	Valerate (uM)-Notes	+/- Valerate (uM)	+/- Valerate (uM)-Notes
		0.55	0.19

Cell Counts? ☐ TRFLP? ☐
16S Sequence? ☐ qPCR? ☐
Metagenomics? ☐

Date	Other Comments	Author
2015-10-05	DIC abundance calculated by IRMS method. Not exactly comparable to FID method values.	Barbara Laflamme

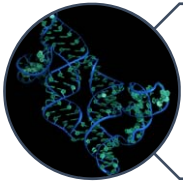
Open Data
© 2014

AHED Pilot Databases



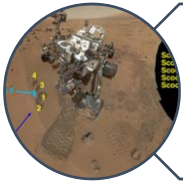
PLRP - Pavillon Lake Research Project DB

- www.odr.io/PLRP ---- 536 records
- Microbialite and water samples



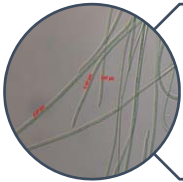
Direct molecular evolution sequence DB

- 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)



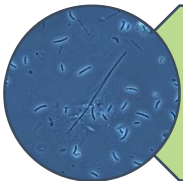
ES Culture Collection DB

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



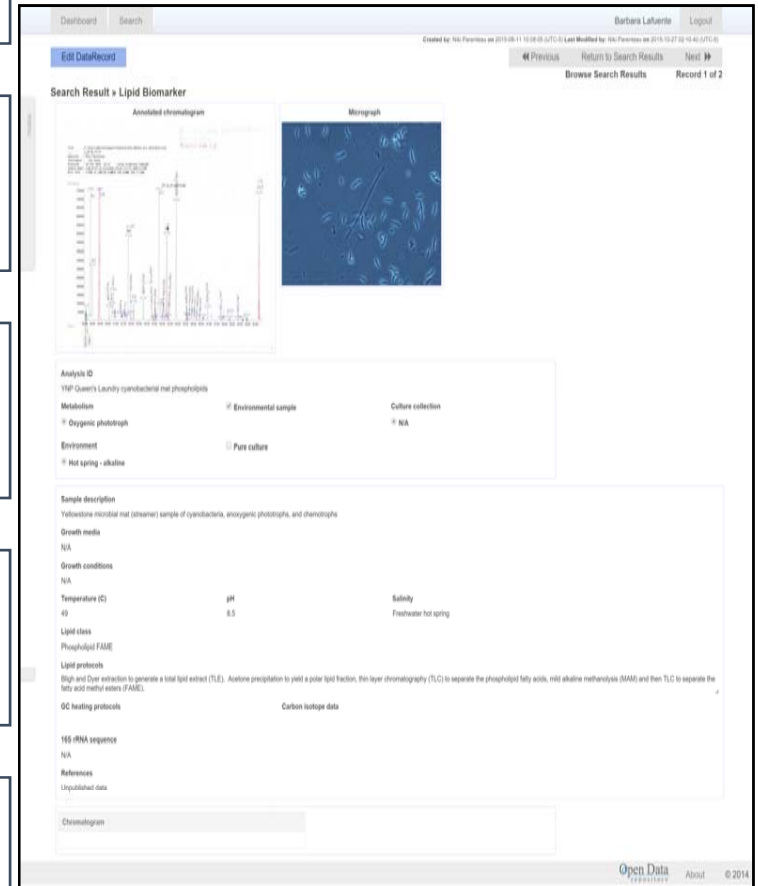
CROMO: Serpentinizing System DB

- www.odr.io/CROMO_analysis ---- 106 records
- Drill samples from serpentinizing systems



Lipid Biomarker DB

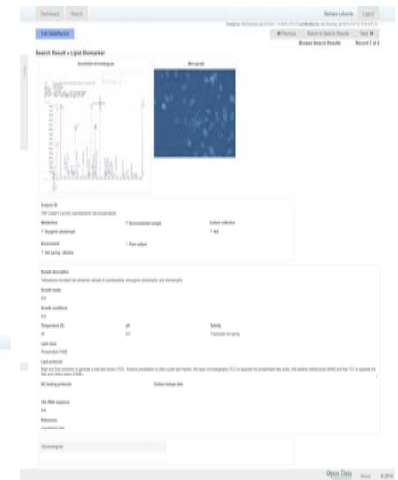
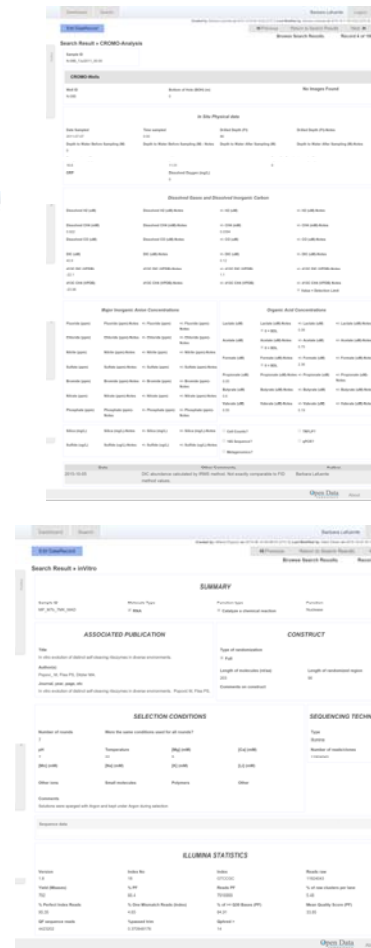
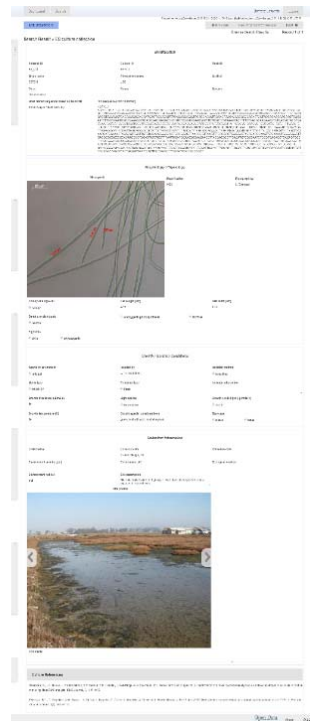
- <http://odr.io/biomarker> ---- 2 records
- Lipids characterize in pure cultures of microbes



AHED Pilot Databases

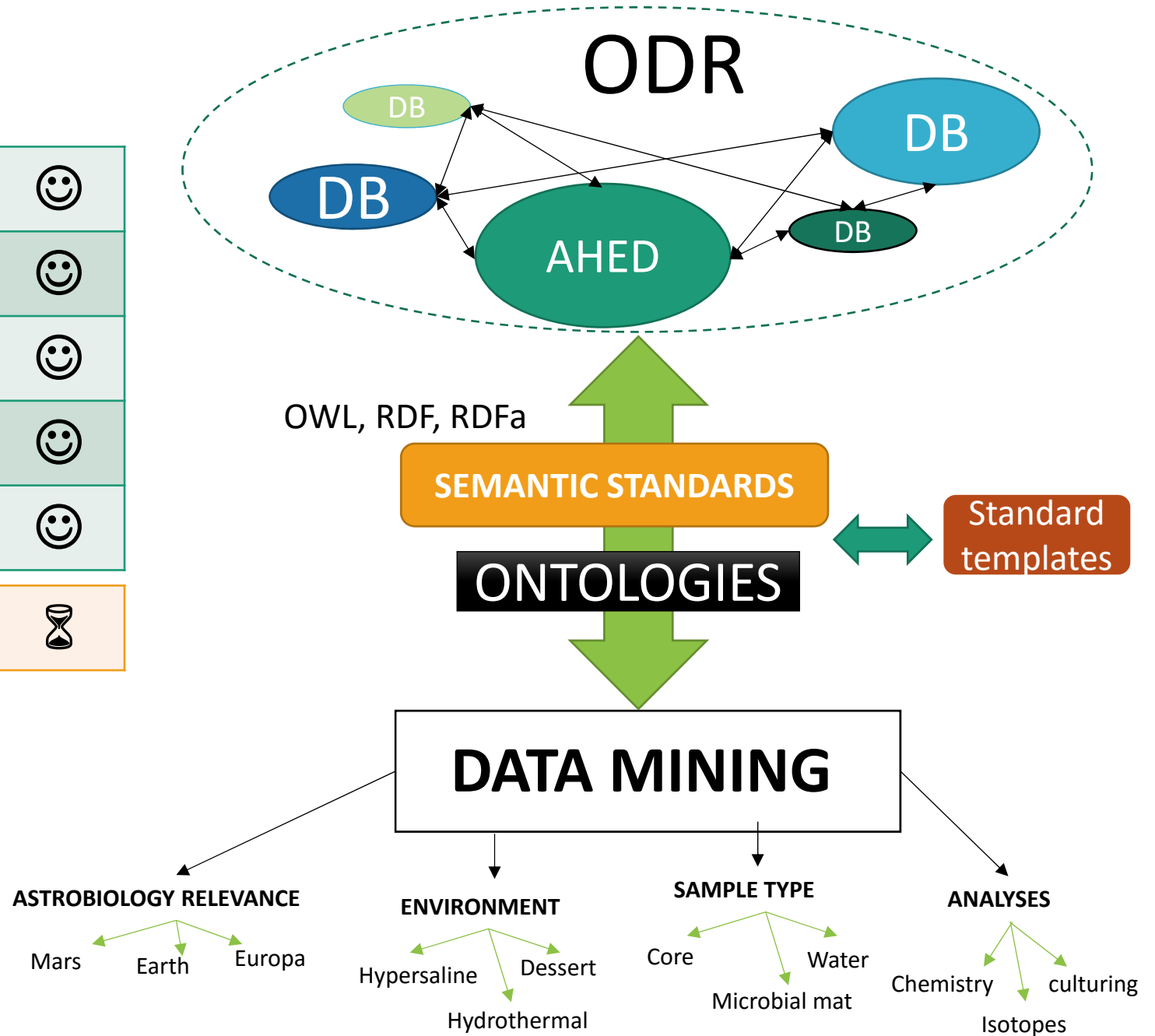
Take home message

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



Road map

Drag-and-drop	😊
CSV Import	😊
CSV Export	😊
Public/non-public	😊
User permissions	😊
Semantic Web	⌚

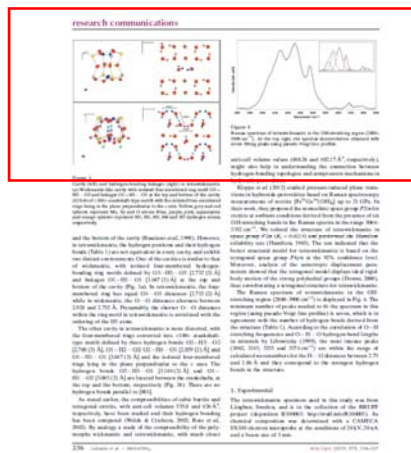


Road map

Drag-and-drop	😊
CSV Import	😊
CSV Export	😊
Public/non-public	😊
User permissions	😊

Semantic Web	⌚
Citation	⌚
User feedback	⌚
NASA's server	⌚

Citation



User feedback



NASA's server

Stability
Security
Service



Minimal downtimes
Ilimited capacity
Continued growth

AHED The Astrobiology Habitable Environments Database

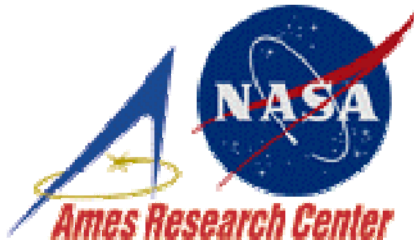
Take home message

- ✓ **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