When is "FAIR" F.A.I.R?

ANNE RAUGH

PDS SMALL BODIES NODE, UNIVERSITY OF MARYLAND (*ARAUGH@UMD.EDU*)

"We've been FAIR since before there was a FAIR."

HMM…

Open and F.A.I.R are *not* the same thing.

"OPEN" IS ABOUT DATA RIGHTS AND LICENSING.

"F.A.I.R." IS ABOUT MECHANICS.

"FAIR" is primarily concerned with *programmatic* processing.

FAIR RECOGNIZES THAT DATA IS DIVERSE AND SCATTERED ACROSS CYBERSPACE. SOFTWARE PROCESSING LEVELS THE PLAYING FIELD.

The FAIR Principles at *go-fair.org*

“F.A.I.R” is not a binary state.

FAIRNESS IS A SPECTRUM. THERE ARE VARIOUS WAYS TO INCREASE FAIRNESS AND THEY CAN BE APPLIED INCREMENTALLY.

The FAIR Principles at *go-fair.org*

"We've been FAIR since before there was a FAIR."

Findability

If a user must know which specific archive's service to query to search for a data resource, the data are not *Findable*.

First Steps:

- Assign DOIs to data resources.
- Provide rich metadata in the DOI record.
- Include tagged metadata (schema.org, for example) on the landing page for the data resource.

Accessibility

If the globally unique identifier (the DOI, e.g.) cannot be used to fetch the metadata for the data resource from the repository interface, the (meta)data are not *Accessible*.

First Steps:

- Implement a DOI retrieval option in the local interface.
- If the only interface(s) requires human interaction, prioritize API development.

Interoperability

If a human must read a document to find or understand the metadata, the data are not *Interoperable*.

First Steps:

- Provide metadata in machine-readable formats.
- Use standard vocabularies defined by recognized authorities that are programmatically actionable (UAT keywords, for example) wherever possible.

The FAIR Principles at *go-fair.org*

Reusability

If the metadata for the resource do not indicate the copyright holder and license, the data are not *Reusable*.

First Steps:

- If the data are known to be in the worldwide public domain, indicate that clearly in the metadata.
- Provide *both* copyright holder and license in programmatically accessible metadata otherwise.
- Use standard digital licenses (CC licenses, e.g.) with formal references.

Take-aways

- Open is good. FAIR is good. Open *and* FAIR is the goal.
- Increasing FAIRness can and should be an ongoing process.
- Do the easy things first.
- Seek assistance and support for the hard things.

The FAIR Principles at *go-fair.org*

Questions?

The FAIR Principles at *go-fair.org*

What is FAIR? An evolving NASA view

12 July 2023

Mark A. Parsons
https://orcid.org/0000-0002-7723-0950
mark.parsons@uah.edu
@chutneyboy
University of Alabama in Huntsville
NASA Chief Science Data Office

IMPACT
Interagency Implementation and Advanced Concepts Team

Why do we care about FAIR?

- NASA Science Information Policy (SPD-41a) says data should be FAIR.
- The principles have helped to focus conversation. It's a great foundation.
- It's a clever acronym. Perhaps too clever by half.

What is FAIR?

- 15 principles focussed on what machines (not humans) need at a base level to ensure the broad objectives of FAIR can be achieved.
- Of course we must consider human concerns, but that is not the focus of the principles. They provide a technical baseline that allows us to commonly work at a higher level of abstraction.

The 15 Principles

F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
    A1.1 the protocol is open, free, and universally implementable
    A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
    R1.1. (meta)data are released with a clear and accessible data usage license
    R1.2. (meta)data are associated with detailed provenance
    R1.3. (meta)data meet domain-relevant community standards

What do we know about FAIR?

- A survey of NASA Scientific Data Repositories (see more detailed summary)
  - 16 of 34 repositories responded (47% response rate):
  - 4 Planetary, 1 heliophysics, 6 Astrophysics, 3 Earth Science, 2 Bio
  - More responses still welcome
- All respondents are working on FAIR
- Large variability in what that means and costs
- Most use their own assessment or process
  - GeneLab uses FAIR Evaluator and the GO FAIR Foundation Qualification.
  - LSDA (not an SMD repo) uses FAIR Data Point and FAIR Evaluator.
  - The Astromaterials Data System uses FAIR Implementation Profiles within a custom process.
- Many use existing community standards and assume that does the trick, e.g. IVOA, UMM, PDS, TRUST, Core Trust Seal …

What do we know about FAIR?

- **Persistent Identifiers** an active area but variable adoption: 7 have PIDS on all data, 4 have PIDs on more than half, and 5 have PIDS on less than half to none.
- **Rich metadata:** All but one have defined metadata standards. Those for Earth Science and PDS seem the most comprehensive. Compliance with the standards and general "richness" is uncertain, but much NASA data is very richly described at least for humans.
- **Licenses:** NASA data are generally very openly accessible. All the repositories have at least human readable usage guidance, several have machine-readable licenses, and many allow programmatic access through an API.

What do we know about FAIR?

- **User engagement:** Most of the repositories have user working groups and help desks. They also attend relevant conferences and solicit general feedback. Several repositories conduct user surveys. Of particular note is the ACSI (American Customer Satisfaction Index) survey conducted annually by the Earth Science DAACs.

Challenges

- Resource constraints, especially specialized resources.
- Legacy data (Levels of service?)
- Provider compliance with standards and guidelines
- Tracking data use and citation (Make Data Count?)
- Lack of standards for I and R, especially semantics, especially challenging for specialized data
- FAIR data in the cloud

Thank You

Contact me at mark.parsons@uah.com @chutneyboy

Development of FAIR Guidance in the Earth Sciences Division

Robert R. Downs[1] and Ge Peng[2]

[1]NASA Socioeconomic Data and Applications Center (SEDAC)
ESDIS Standards Coordination Office (ESCO)
Center for International Earth Science Information Network (CIESIN)
Columbia Climate School, Columbia University

[2]Earth System Science Center/NASA MSFC IMPACT
University of Alabama in Huntsville

Prepared for Presentation to
Enabling FAIR Data at NASA — A Planning Webinar
13 July 2023, 12:00 - 1:00 p.m. EDT

OFAIR WG – Synopsis

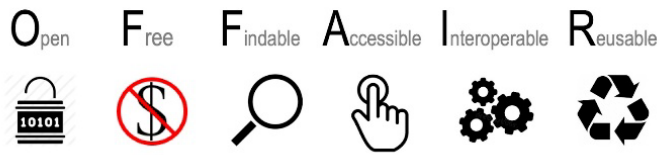
- Recognizing the need to continually improve practices for data management, stewardship, and distribution, the NASA ESDSWG on Making NASA SMD Data Open, Free, Findable, Accessible, Interoperable, and Reusable (O'FAIR) is establishing guidance for data producers and distributors to enable the use of NASA Earth science data as open and FAIR

Adapted from : Making NASA SMD-funded Earth Science Data FAIR Working Group. NASA Earthdata wiki. 2023.

OFAIR WG – Establishment

- ESDSWG 2022 Meeting participants proposed the Earth Science Data System Working Group (ESDSWG) on Making NASA SMD-funded Earth Science Data Open, Free and FAIR (O'FAIR WG)

- Objective: The O'FAIR Working Group aims to synthesize community FAIR practices to provide principle-by-principle guidance on how to apply existing practices to ensure or enable NASA SMD-funded Earth science open and free data/information are also findable, accessible, interoperable, and (re)usable.

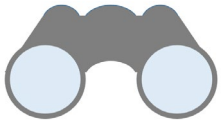


- Technical Chairs: Ge Peng and Robert R. Downs
- ESDS/ESDIS POC: Francis Lindsay

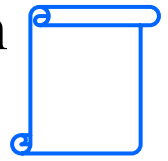
Adapted from : Making NASA SMD-funded Earth Science Data FAIR Working Group. NASA Earthdata wiki. 2023.

OFAIR WG – General Approach

- Define the scope and targeted audience by examining SPD-41a, FAIR Guiding Principles, and related information
- Explore and identify gaps, recommended practices and improvements by reviewing FAIR, data management and stewardship practices, adoption of standards, and FAIR assessments at NASA and across national and international research data communities



- Produce an inception report and draft the guide document by synthesizing international and NASA FAIR practices, obtaining consensus and documenting existing practices and assessments
- Develop and produce guide document and submit to ESDSWG as the WG outcome and to ESCO for review and publication

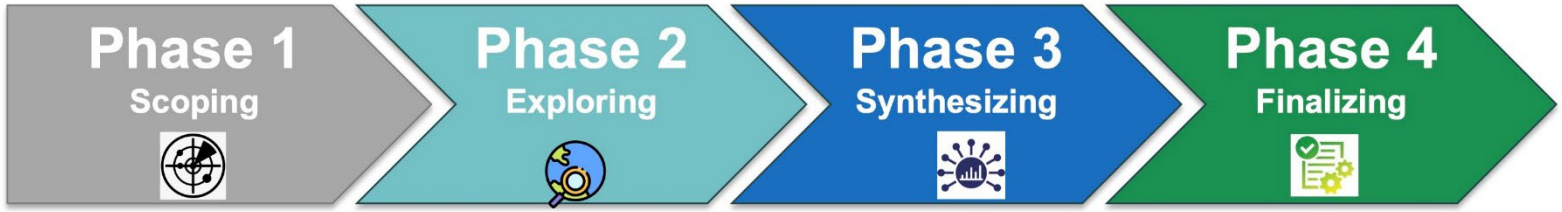


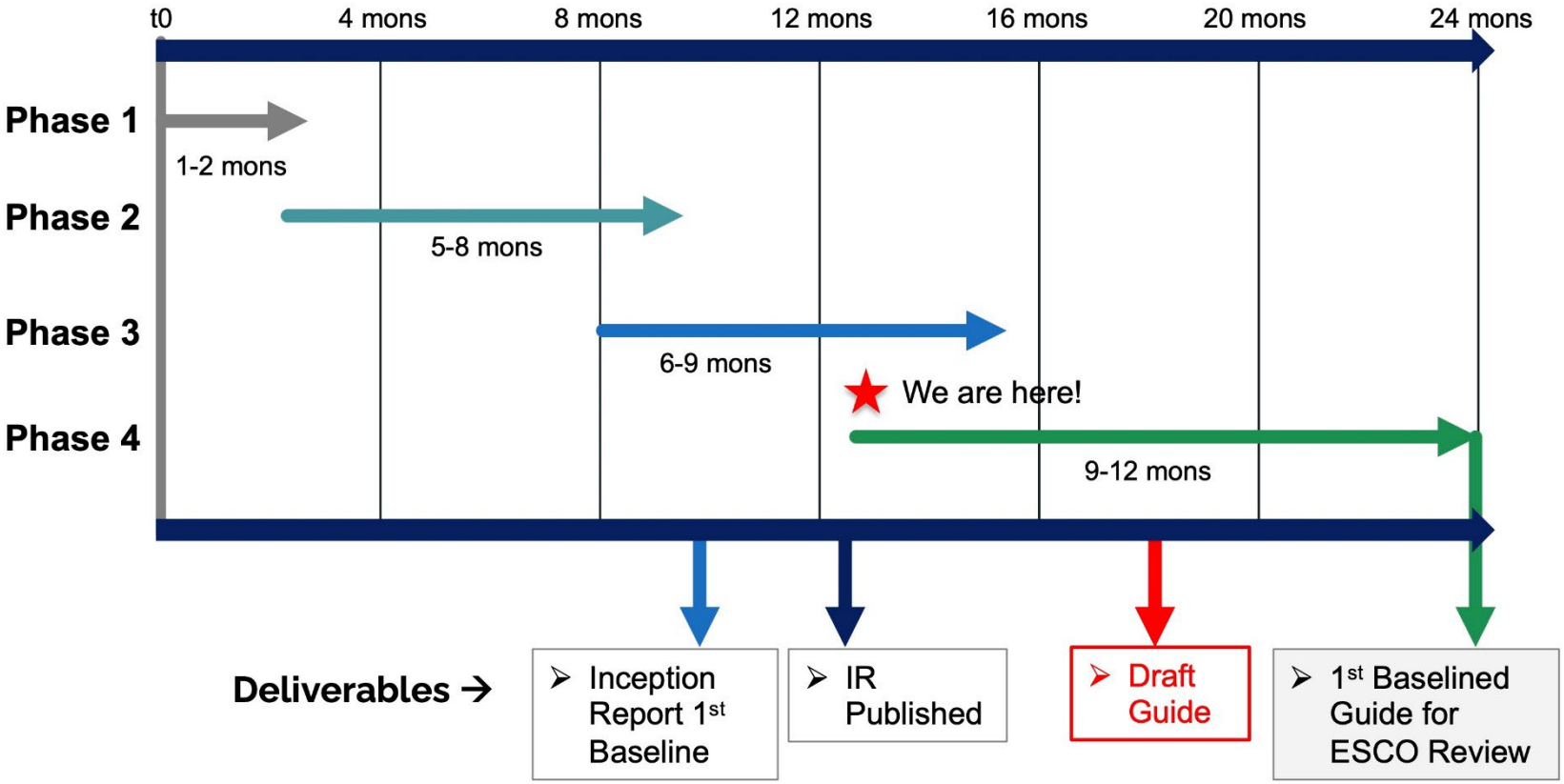
Adapted from : Making NASA SMD-funded Earth Science Data FAIR Working Group. NASA Earthdata wiki. 2023.

OFAIR WG – Key Activities

- Inaugural meeting
  - Identification of contributors and interests
- Monthly telecons - Invited presentations and discussions
  - SPD-41a. Steve Crawford, SMD Science Data Officer
  - FAIR US. Melissa Cragin, SDSC,UCSD; Rice University
  - Enhancing Atmospheric Composition Data FAIRness. Gao Chen, ASDC
- FAIR Practices Collection Spreadsheet
  - Resources on FAIR adoption practices
- In-person break-out meeting at ESDSWG Annual Meeting
  - Discussion on scope and contents of guide document
- First WG deliverable: Inception Report released
  - high-level overview of community FAIR practices

O'FAIR WG – Current Status



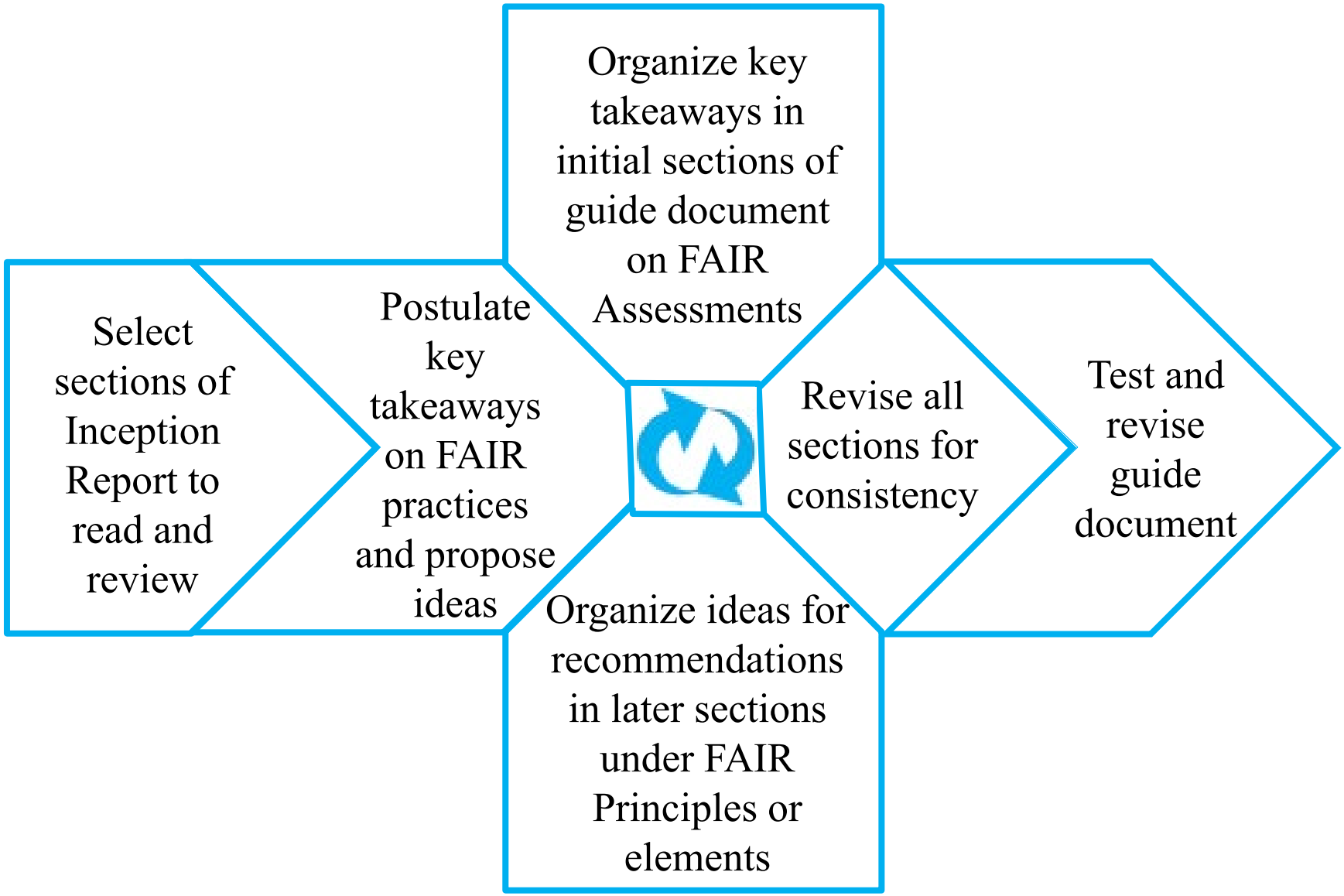


Adapted from : Making NASA SMD-funded Earth Science Data FAIR Working Group. NASA Earthdata wiki. 2023.

O'FAIR WG – Inception Report

- Reviews application of FAIR Principles and provides an overview of FAIR for Earth science data
- Published publicly via the ESDIS Standards Coordination Office (ESCO)
  - Peng, et al. 2023. An Overview of Community FAIR Practices – NASA O'FAIR WG Inception Report. Document ID: NASA-OFAIR-ESDSWG-DOC-0001. Version: v01r00-20230508. https://doi.org/10.5067/DOC/ESCO/ESDSWG-0001V1

OFAIR WG – Guide Development

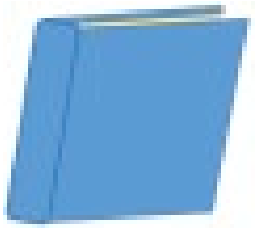


OFAIR WG – Draft Guide Outline

A PRACTICAL GUIDE

1. EXECUTIVE SUMMARY
2. INTRODUCTION
3. BACKGROUND
4. FAIR PRINCIPLES, SPD-41a, AND OTHER PRINCIPLES
5. STATE OF THE ART IN FAIR-NESS ASSESSMENT
6. CURRENT STATE OF FAIR-NESS ASSESSMENT OF NASA'S EARTH SCIENCE DATA
7. MAKING NASA EARTH SCIENCE DATA FAIR
8. DISCUSSION
9. CONCLUSION
10. ACKNOWLEDGEMENT
11. REFERENCES

APPENDIX A. TERMS AND DEFINITIONS



Thank you!

rdowns@ciesin.columbia.edu

ge.peng@uah.edu

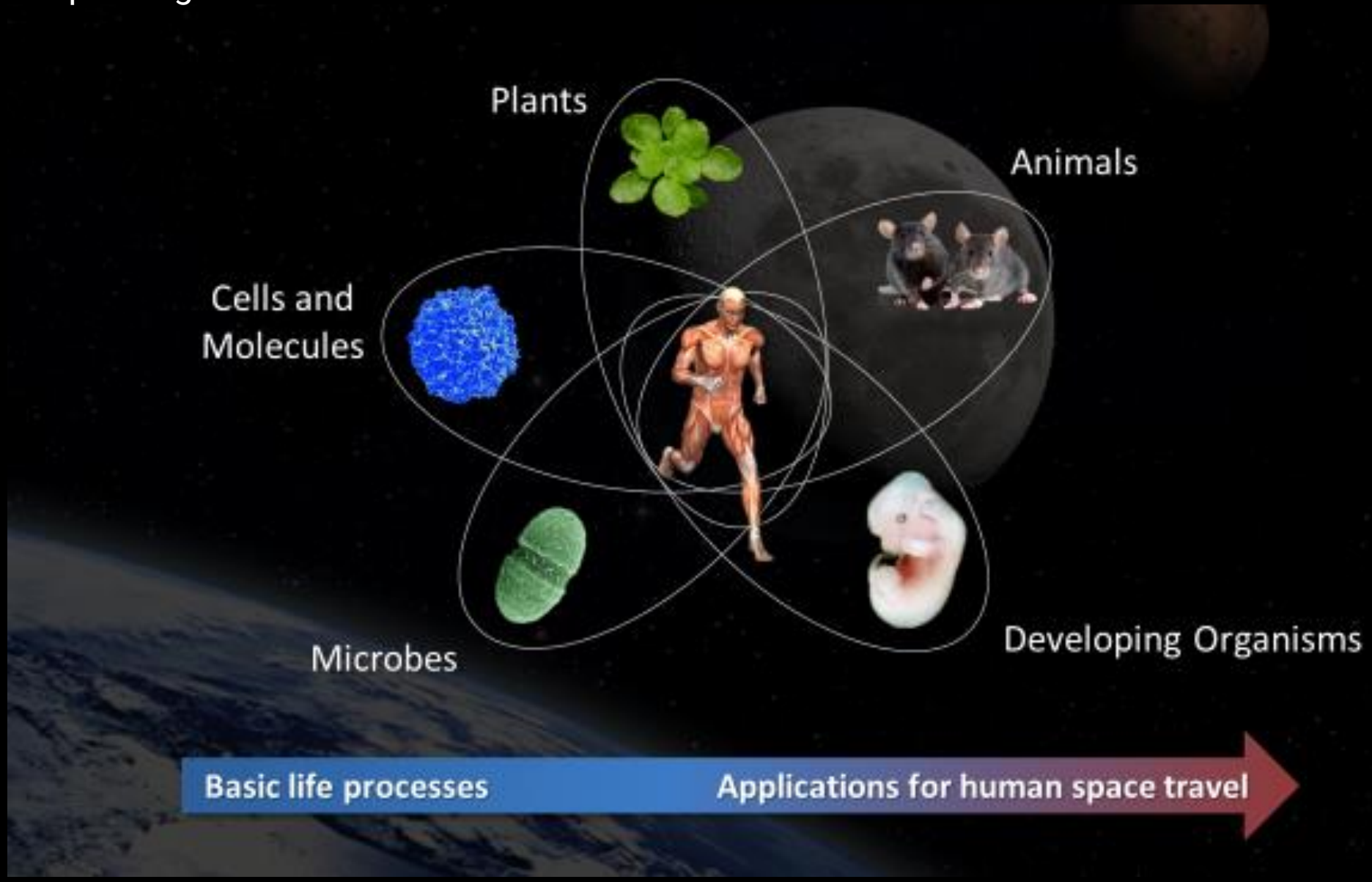
Implementation of FAIR in the life sciences

Samrawit Gebre
GeneLab Deputy Project Manager
NBISC Project Manager
NASA Ames Research Center

National Aeronautics and Space Administration

NASA

Space Biology Program: conducts research to use the space environment to advance our knowledge of how gravity affects the design and function of living organisms, and to understand how biological systems accommodate to spaceflight environments



NASA Biological Open Science Resources

Biospecimen Sharing Program (BSP)

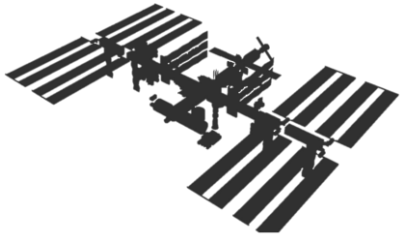
Dissection and preservation of rodent tissues from Flight and Ground investigations. Coordination of internal tissue sharing

NASA Biological Institutional Scientific Collection (NBISC)

Collection of non-human specimens and space microbial culture

Ames Life Sciences Data Archive (ALSDA)

Collection and curation of mission, project, and imaging data



NASA GeneLab (GL)

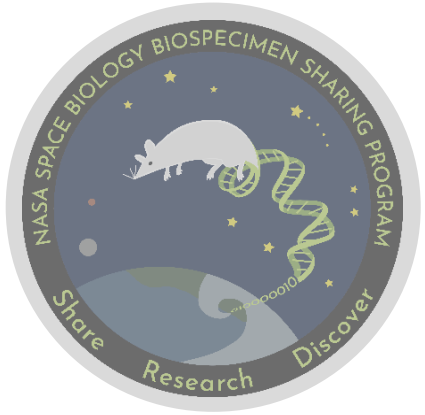
Collection and curation of omics data

NASA Internal Program

Open Source Science Programs — Available Globally

NASA Biological Open Science Resources

Biospecimen Sharing Program (BSP)



Dissection and preservation of rodent tissues from Flight and Ground investigations. Coordination of internal tissue sharing

NASA Internal Program

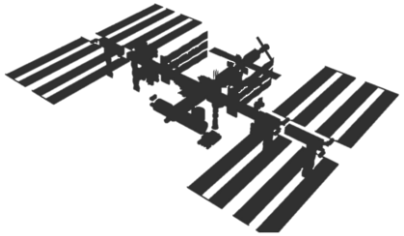
NASA Biological Institutional Scientific Collection (NBISC)

Collection of non-human specimens and space microbial culture

Ames Life Sciences Data Archive (ALSDA)

1994

Collection and curation of mission, project, and imaging data



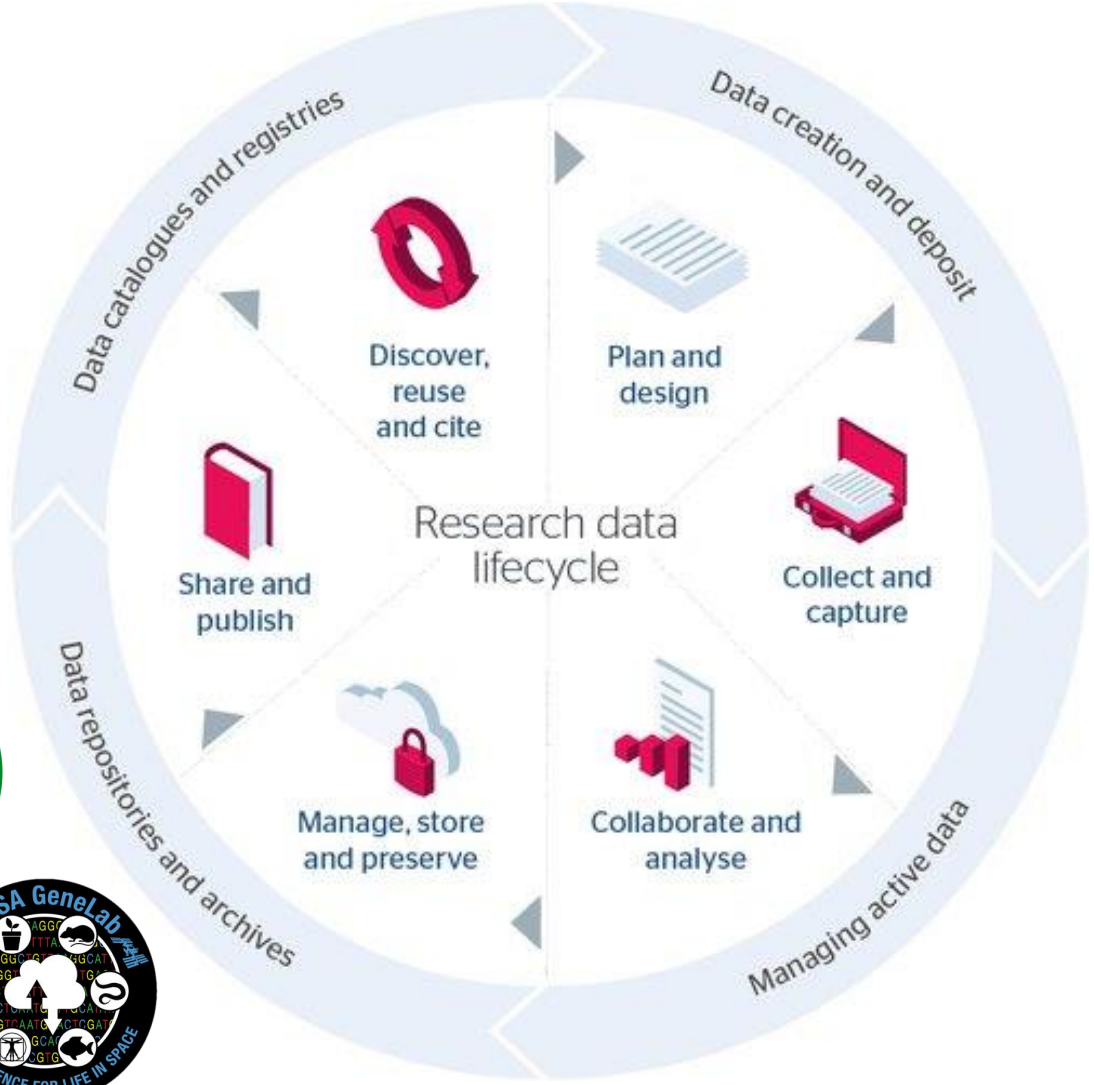
NASA GeneLab (GL)

2014

Collection and curation of omics data

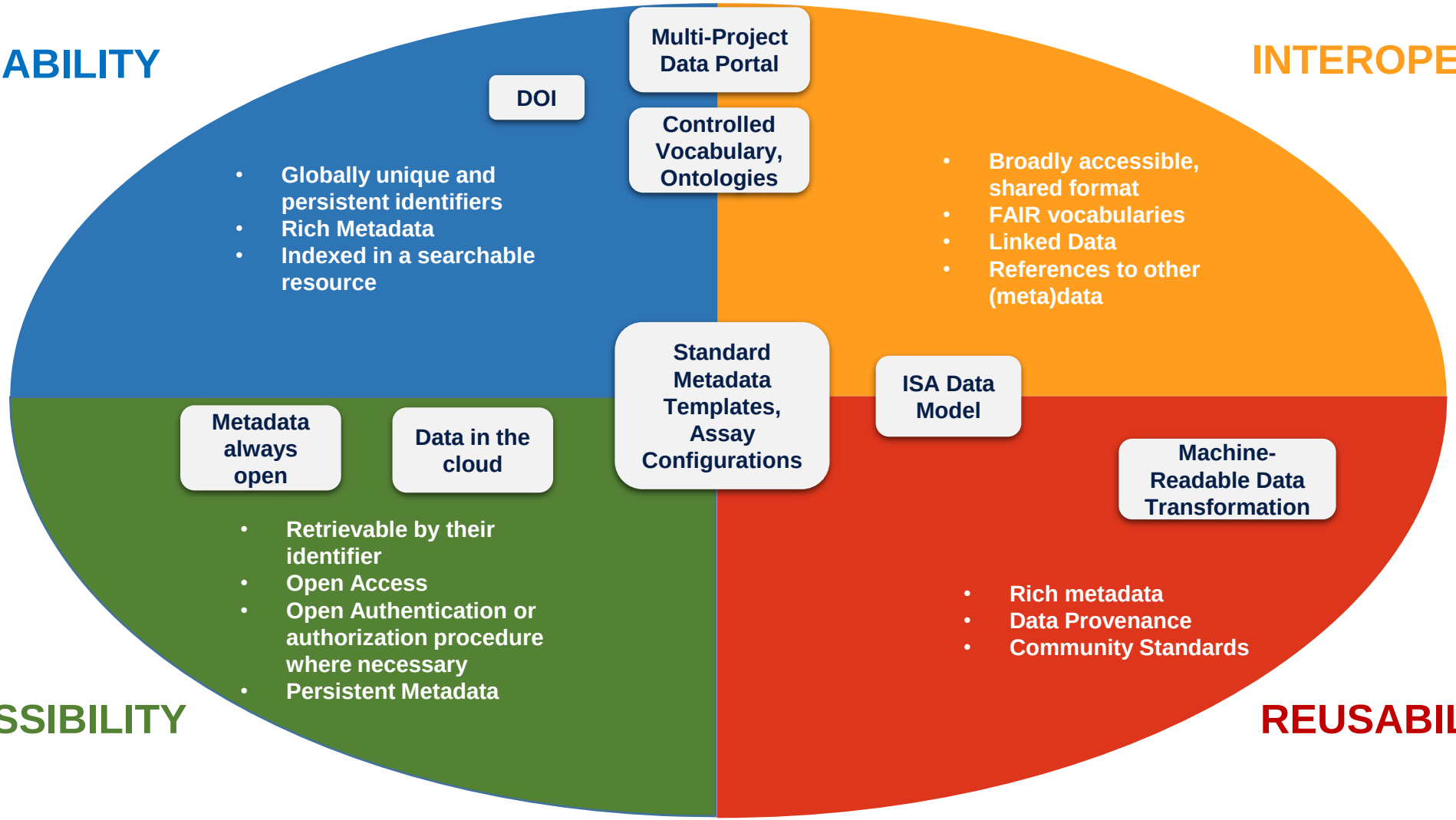
Open Source Science Programs — Available Globally

Research Data Life Cycle



Credit: https://beta.jisc.ac.uk/guides/research-data-management-toolkit

FAIR Progress: GeneLab + ALSDA



(Adapted from Berrios et al., 2018; PMID: 30815061)

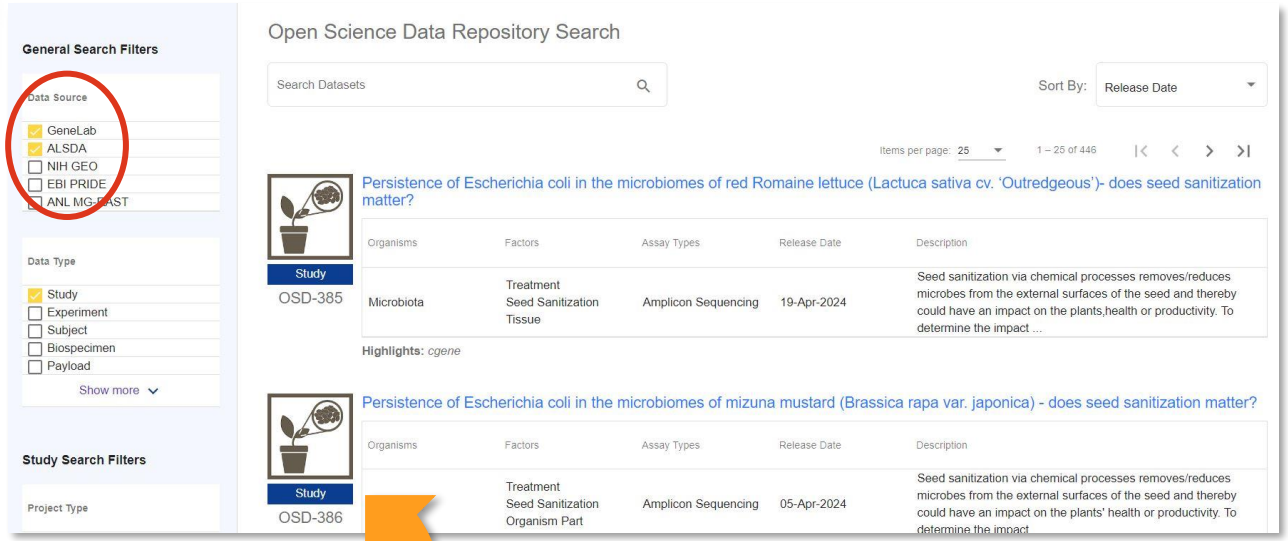
Centralizing information of different types of data

osdr.nasa.gov/bio

ALSDA

GeneLab

Common online search



Mission Telemetry Data

Ascent

Flight

Return

Ground Control

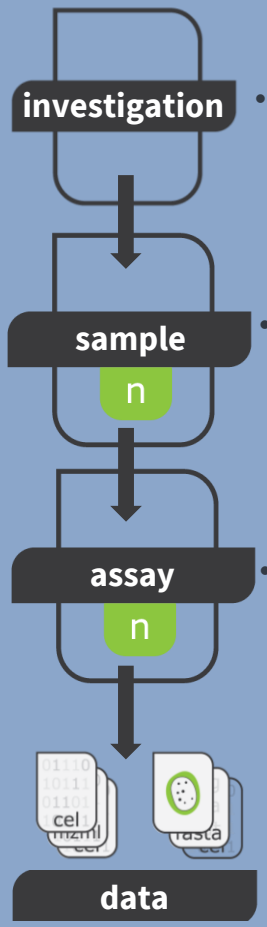
Relative Humidity, Temperature, CO_2, O_2, Airflows, Acceleration, Radiation, Acoustics, etc.

API

Breakdown of a Reusable Dataset: Architecture, Curation, *FAIR*ness

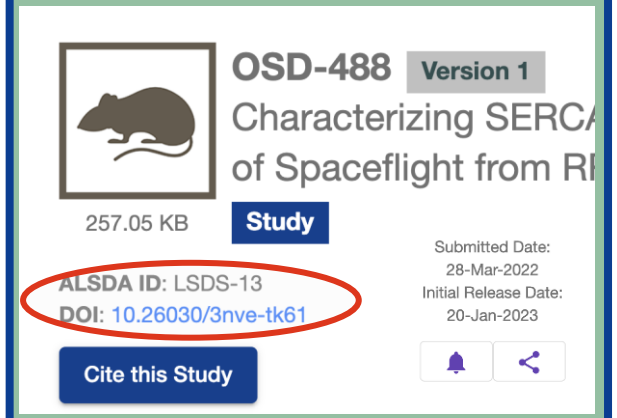
OSDR *Meta*/Data Model

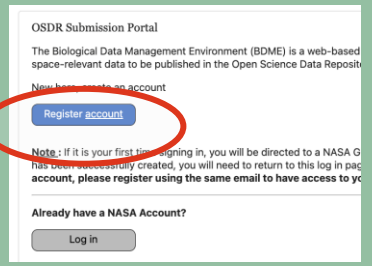
Leverages/inspired by "ISAtab"



- **investigation**
  - Contacts, description, study factors, provenance
- **sample** (n)
  - Table w/individual subject-sample parameters, study design, independent variables, sample collection metadata
- **assay** (n)
  - Table w/individual subject-samples by assay with device name, settings, targets, parameters, reagents, **standards** SME-driven
- **data**
  - Various formats to maximize reuse, machine-readable
  - Tabular, text, images, video
  - Molecular-omics, physiological-phenotypic-behavior-imaging
  - Raw, submitted, processed, transformed for reusability

Open Science Dataset Example





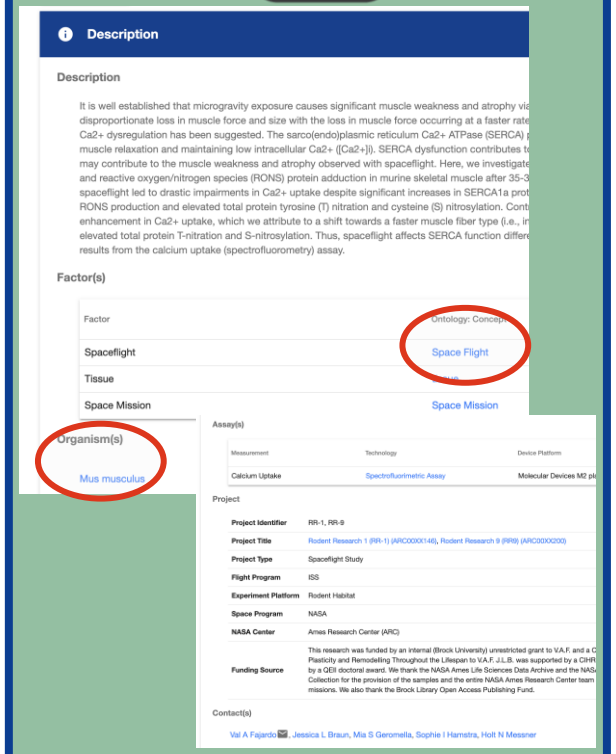
Findable

- Global Unique Identifiers

Accessible

- Secure/Integrated Credentials

investigation



Interoperable

- Metadata Normalization
  - ie., mus musculus
  - ie., Space Flight

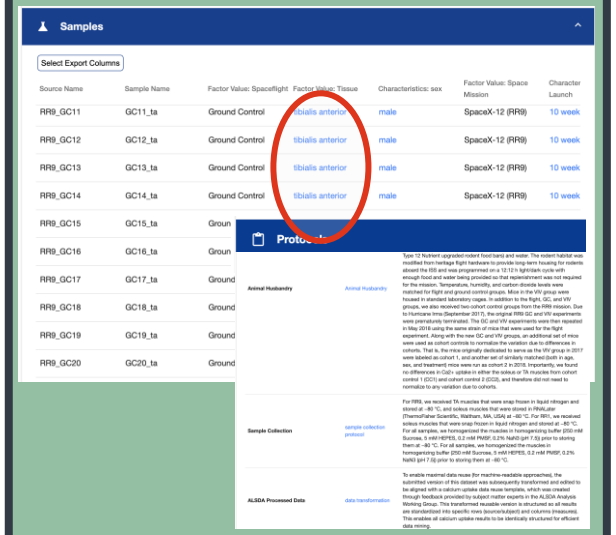
(Berrios et al., 2020, PMID: 33080015; González-Beltran et al., 2014, PMID: 25472428; Sansone et al., 2008, PMID: 18447634; Ray et al., 2019, PMID: 30329036; isa-tools.org)

Breakdown of a Reusable Dataset: Architecture, Curation, *FAIR*ness

sample

n

- Individual Subject/Sample Independent Variable Metadata
- Study Design Metadata



Interoperable

- Metadata Normalization
  - i.e. muscle names / time
- Tibialis anterior

investigation

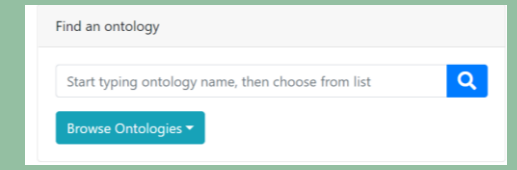
sample

n

BioPortal

OBO Foundry

Ontobee



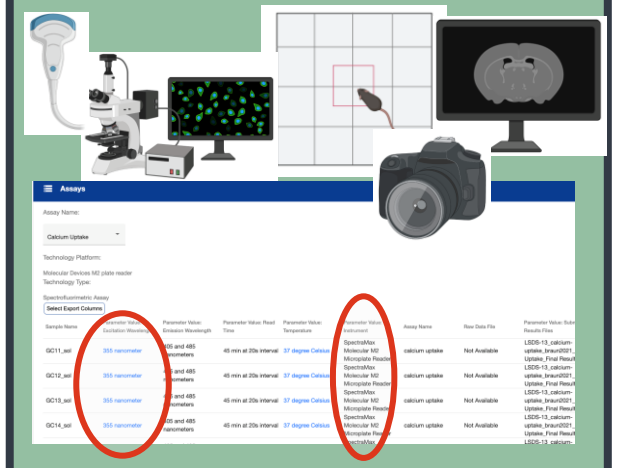
Findable Accessible Interoperable Reusable

- Rich Metadata
- Metadata Schemas
- Community Standards

assay

n

- Dependent Variable Metadata
- How assay was conducted
- Specs of device, settings, software, processing details, etc

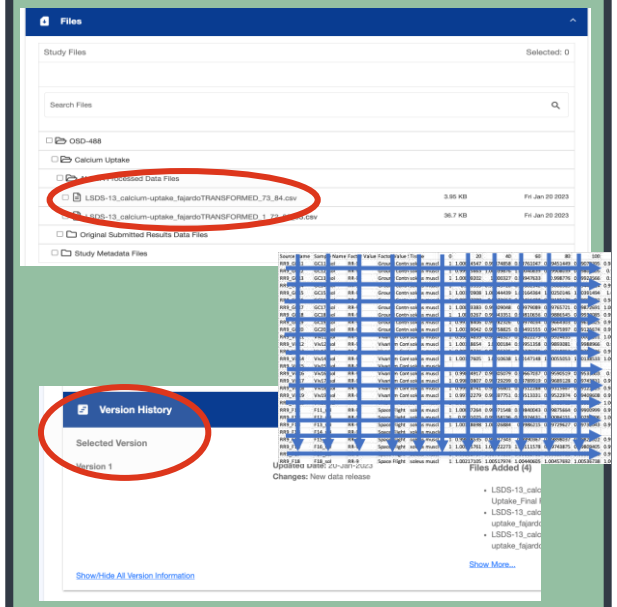


Interoperable

- Metadata Normalization
  - i.e., OCT (optical coh tom)
  - OCT (media/sectioning; optimum cutting temperature)

data

- Raw, Processed; Submitted, Transformed
- Tabular, text, imaging, video
- AWG standardized nomenclature
- Maximally human & machine-readable



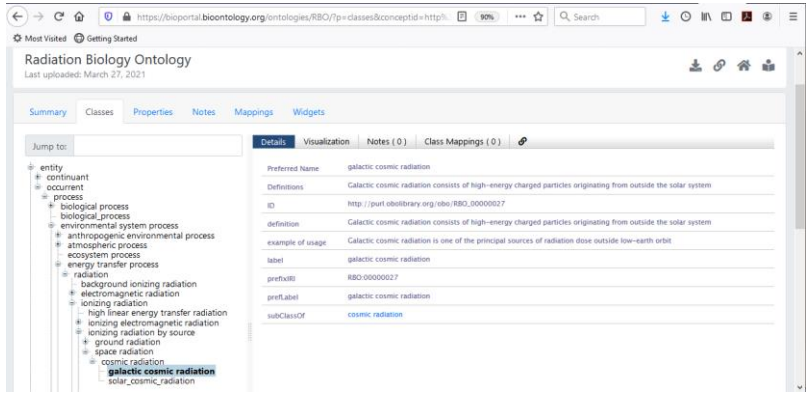
Reusable

- Community Standards
- Data Provenance

(Berrios et al., 2020, PMID: 33080015; González-Beltran et al., 2014, PMID: 25472428; Sansone et al., 2008, PMID: 18447634; Ray et al., 2019, PMID: 30329036; isa-tools.org)

Community Standards - FAIR

Radiation Biology Ontology (RBO) is an ontology for the effects of radiation on biota in terrestrial and space environments.



Launched and maintained by members of OSDR

Paul Schofield, University of Cambridge
Luke Slater, University of Birmingham
Jack Miller, Lawrence Berkeley National Laboratory
Daniel Berrios, NASA Ames Research Center
Sylvain V. Costes, NASA Ames Research Center

Open Science Analysis Working Groups

Animal, Multi-Omics, Microbe, Plant, ALSDA, AI/ML

*Consist of **500+ scientists** from multiple space agencies, international institutions, and industry. Scientists meet monthly with each group to **provide feedback, develop standards, and analyze data**.*

Curation Standards

Use current Minimum Information Standards for Assays

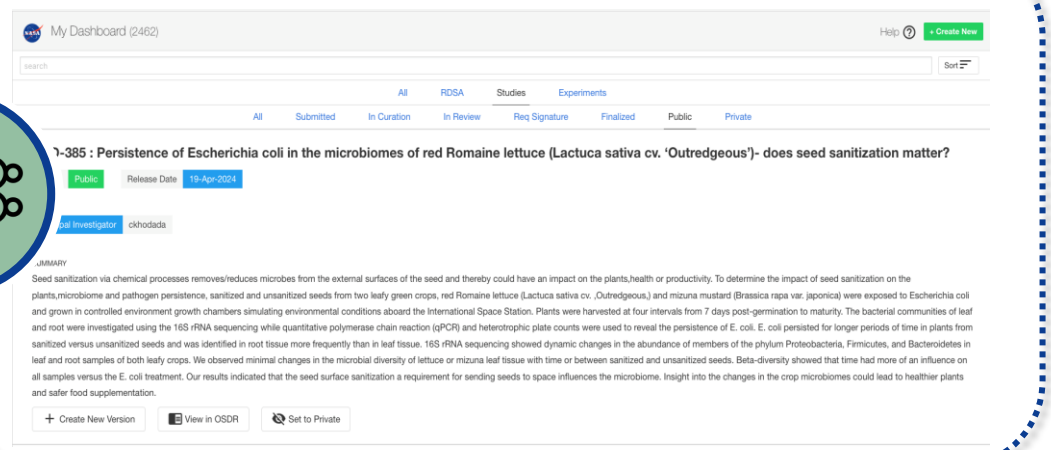
- MIxS
- MIAPE
- MIAME

Develop templates for Assays using our scientific community and references

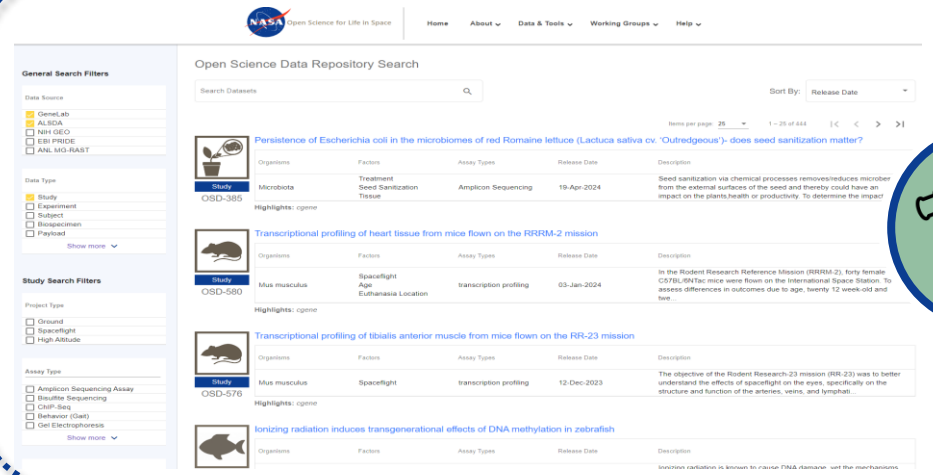
Configuration Name ie., measurement then technology	Status
Behavior (Elevated Plus Maze)	Active
Flow Cytometry (Flow Cytometry)	Active
Behavior (Novel Object Recognition)	Active
Calcium Uptake (Spectrofluorimetry)	Active
Protein Quantification (Western Blot)	Active
Molecular Cellular Imaging (Light/Fluorescence Microscopy)	Active

OSDR: Access & Explore Data

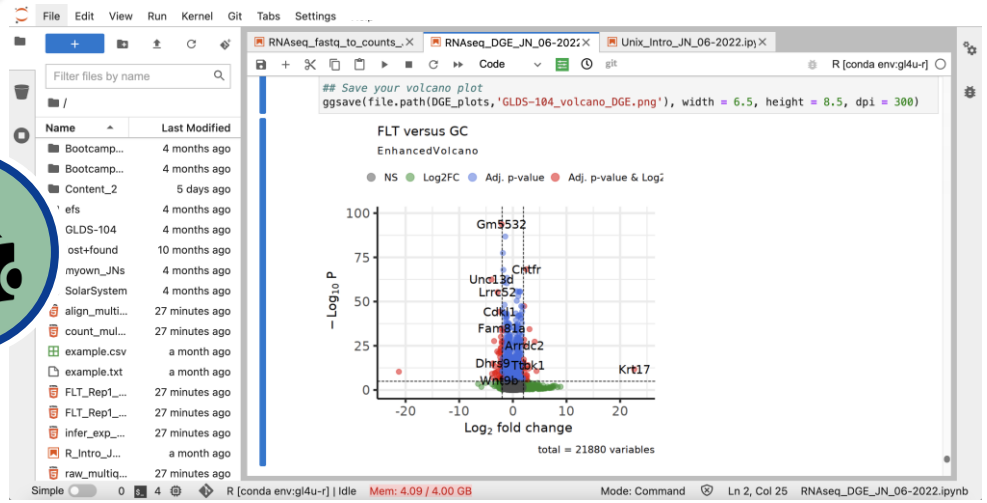
CONTRIBUTE



DISCOVER



LEARN



EXPLORE

Re-use of Data and Enabling New Discoveries

59 (10+ by AWGs) publications using data available in OSDR.

FAIR Assessment

● Pass
◉ Partial pass
○ Fail

FAIR Principle	Metric	GEO[4]	ENA[7]	MG-RAST[6]	Metabolights[8]	GLDS
F1. (meta)data are assigned globally unique and persistent identifier	FM-F1A	●	●	◉	●	●
F1. (meta)data are assigned globally unique and persistent identifier	FM-F1B	●	●	○	●	●
F2. data are described with rich metadata (defined by R1 below)	FM-F2	●	●	●	●	●
F3. metadata clearly/explicitly include identifier of data it describes	FM-F3	●	●	●	●	●
F4. (meta)data are registered or indexed in a searchable resource	FM-F4	●	●	○	◉	●
A1. (meta)data are retrievable by identifier using a standardized communications protocol	N/A					
A1.1 the protocol is open, free, and universally implementable	FM-A1.1	●	●	●	●	●
A1.2 the protocol allows for an authentication and authorization procedure, where necessary	FM-A1.2	●	●	●	●	●
A2. metadata are accessible, even when data are no longer available	FM-A2	●	●	○	○	●
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	FM-I1	○	○	○	◉	◉
I2. (meta)data use vocabularies that follow FAIR principles	FM-I2	○	○	○	◉	◉
I3. (meta)data include qualified references to other (meta)data	FM-I3	○	○	○	○	○
R1. meta(data) are richly described with a plurality of accurate and relevant attributes	N/A					
R1.1. (meta)data released with clear, accessible data usage license	FM-R1.1	●	●	●	●	●
R1.2. (meta)data are associated with detailed provenance	FM-R1.2	●	●	◉	●	●
R1.3. (meta)data meet domain-relevant community standards	FM-R1.3	●	●	○	●	●
Overall FAIRness Score		11	11	6	10.5	12

The FAIR principles, corresponding draft FAIRness metrics, and semi-quantitative FAIRness ratings for select omics data systems. Metrics were those developed by the GO FAIR Metrics[23] group

Berrios DC, Beheshti A, Costes SV. FAIRness and Usability for Open-access Omics Data Systems. AMIA Annu Symp Proc. 2018 Dec 5;2018:232-241.
PMID: 30815061; PMCID: PMC6371294.

Research Data Life Cycle

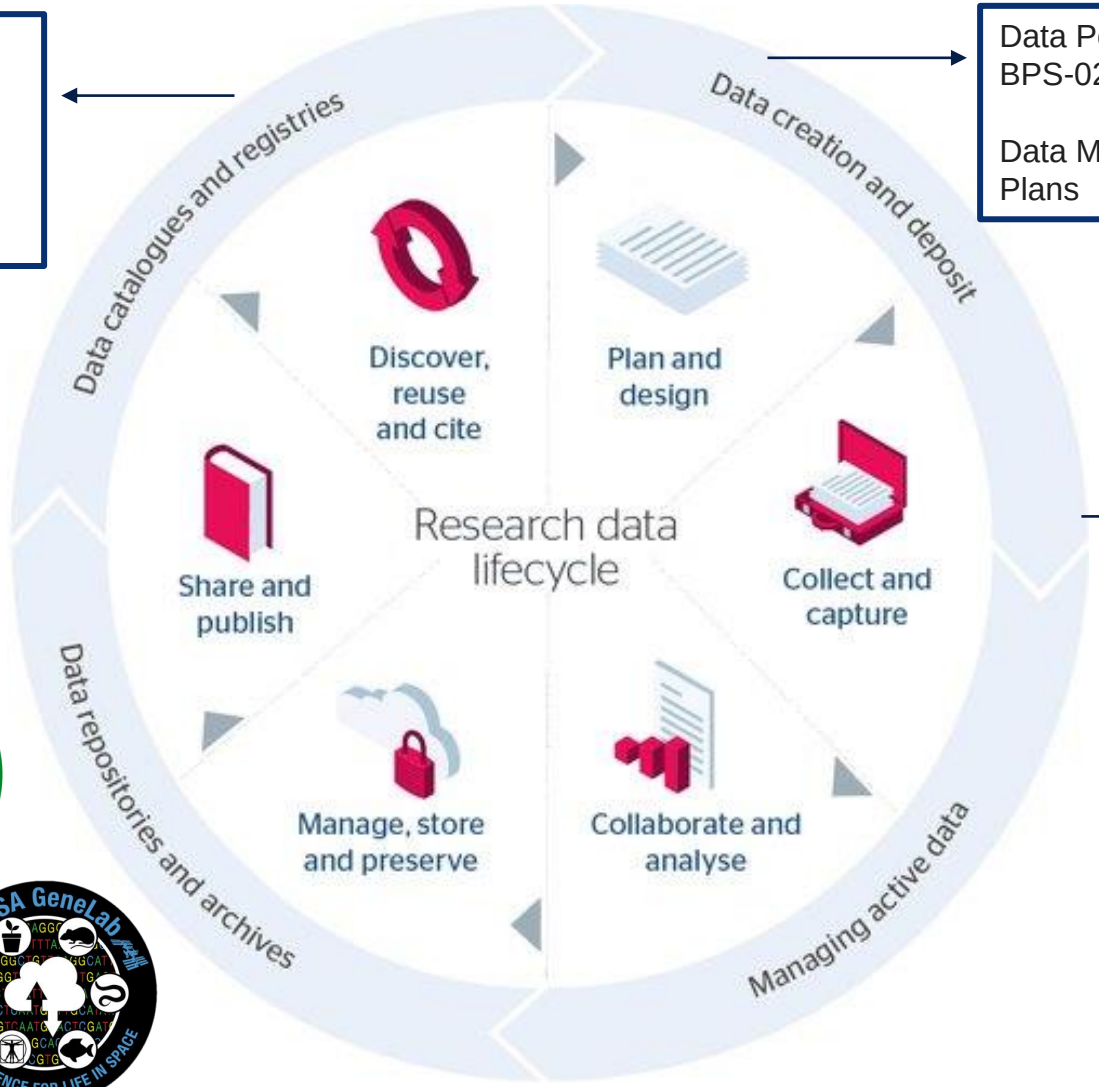
FAIR

Data Policy (SMD-41a, BPS-021)

Data Management Plans

Enable Discovery through Visualization Tools

Analysis Working Groups



FAIR

Ongoing Data Submission & Storage

FAIR

Data Systems

Credit: https://beta.jisc.ac.uk/guides/research-data-management-toolkit

What’s Next

- We will be using the information and feedback from discussions at this webinar and upcoming workshops to select the appropriate tool(s) to evaluate FAIRness.
- Continue to enhance our data systems to ensure (meta)data is FAIR
- Continue working with the scientific community on FAIR
- Should we evaluate other principles such as TRUST? Certification for repositories such as CoreTrustSeal?
- Continue to communicate budget needs to maintain and enhance FAIR

THANK YOU!

TISSUE REPOSITORY

SCIENTIFIC COMMUNITY

OPEN SCIENCE

DATA REPOSITORY

GENELAB OMICS + **ALSDA** PHENOTYPIC

Open Science for Life in Space projects are funded by the Biological and Physical Sciences Division