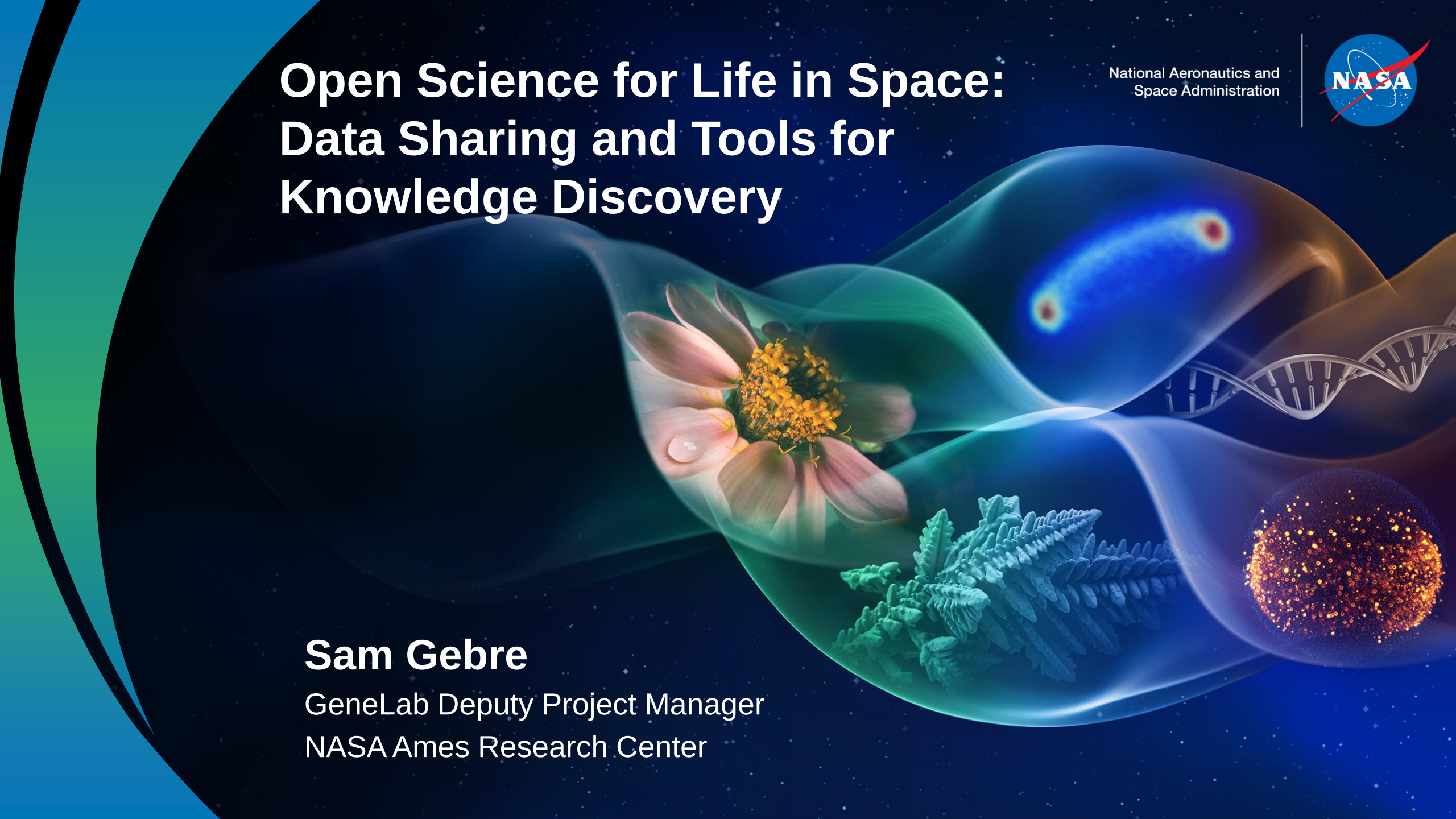
Open Science for Life in Space: Data Sharing and Tools for Knowledge Discovery

National Aeronautics and Space Administration



Sam Gebre

GeneLab Deputy Project Manager

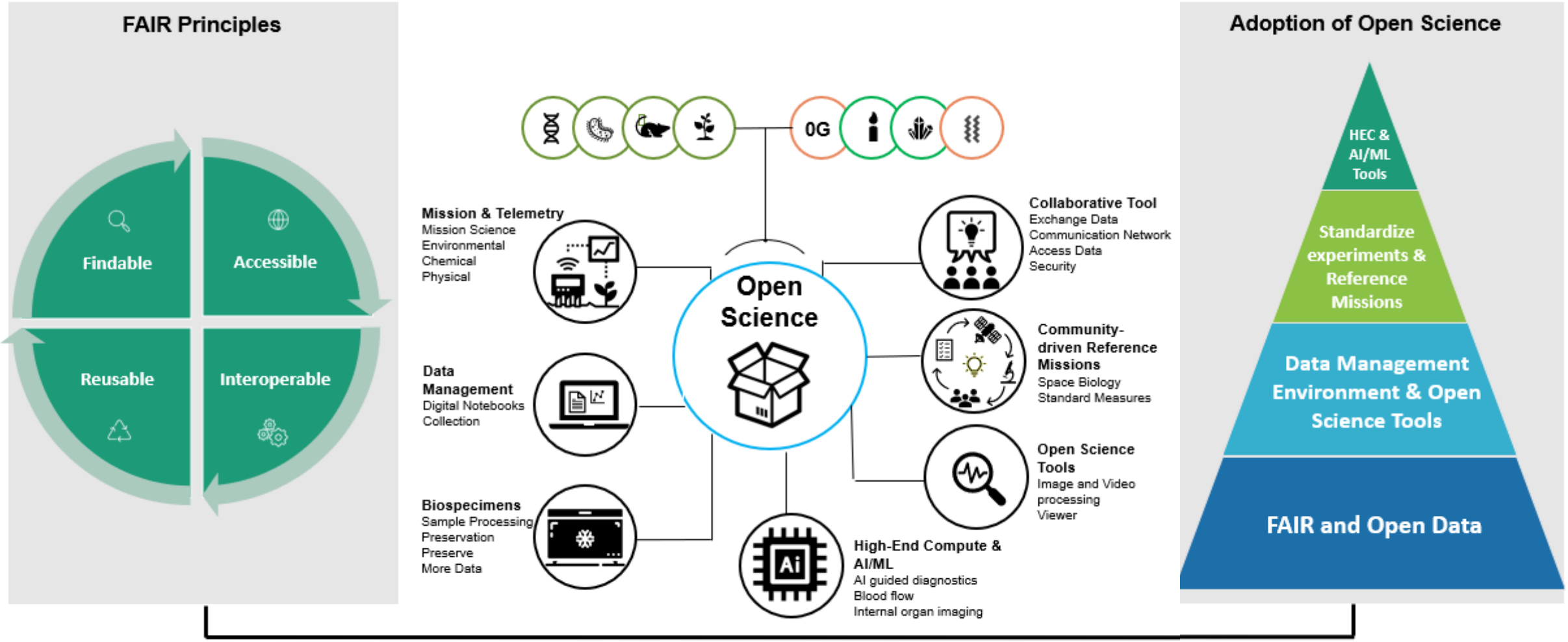
NASA Ames Research Center

Open Science

"We define open science as a collaborative culture enabled by technology that empowers the **open sharing of data, information, and knowledge** within the scientific community and the wider public to accelerate scientific research and understanding."

Ramachandran, R., Bugbee, K., & Murphy, K. J. Moving from Open Data to Open Science. Earth and Space Science, Wiley Publication https://doi.org/10.1029/2020EA001562

How will we get there?



MAXIMIZING THE SCIENTIFIC RETURN OF SPACEFLIGHT EXPERIMENTS

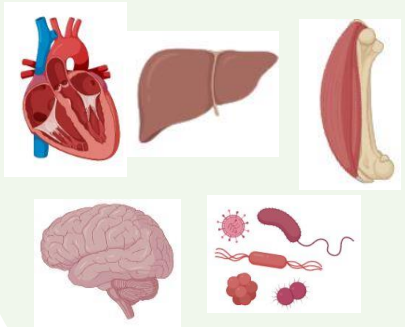
NASA Biological Open Science Resources

Biospecimen Sharing Program (BSP)

- *Dissects and preserves* rodent tissues from Flight and Ground investigations
- *Coordinates* internal tissue sharing

NASA Biological Institutional Scientific Collection (NBISC)

- *Collection* of non-human specimens and space microbial culture



Ames Life Sciences Data Archive (ALSDA)

- *Collects and curates* mission, project and imaging data

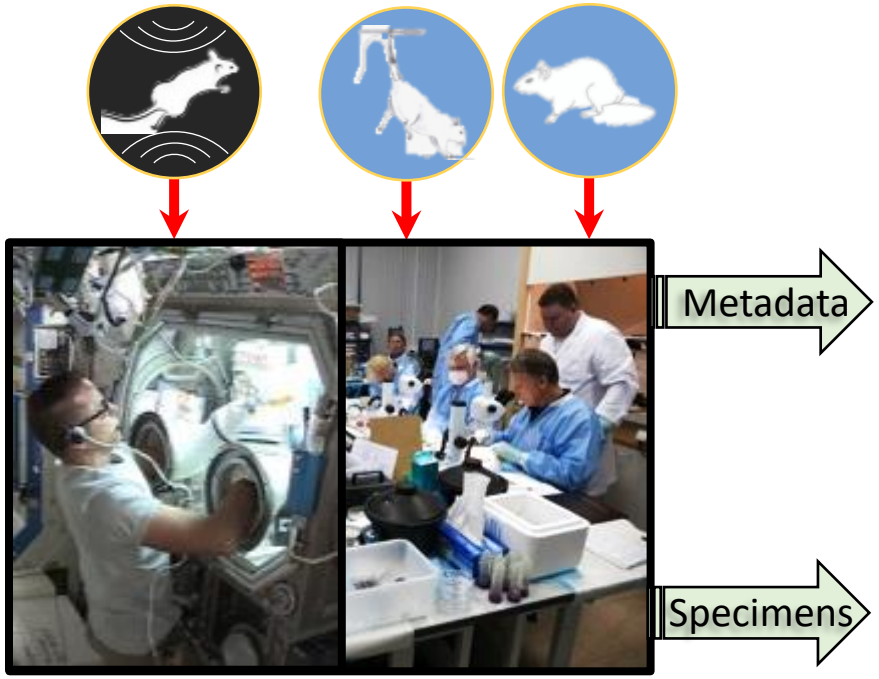
GeneLab (GL)

- *Collects and curates* omics data

NASA Biological Institutional Scientific Collection

Spaceflight

Ground Analogs



Metadata

Specimens

NBISC

Sharing

~35,000 Non-Human Biospecimens*

- Spaceflight and Ground Analog
- Mouse, Rat, Quail, Microbial isolates (expected)
- Searchable catalog (LSDA Website)
- Requests via LSDA Biospecimen Request form

~50,000 Space Radiation Mouse Biospecimens

- Space Radiation Ground Analog
- Newly submitted, will be available soon

Resulted in:

- 33 publications since 2011
- 53 tissue requests since 2016
- 41 GeneLab data sets

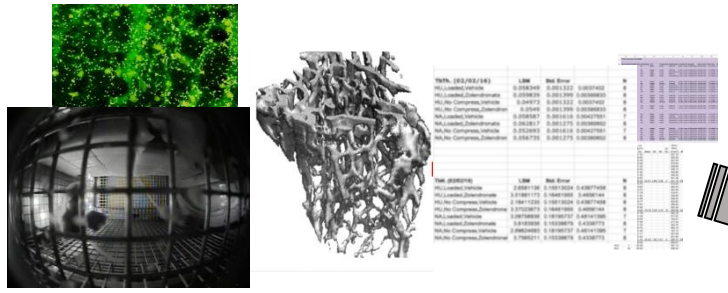
- Fills spaceflight knowledge gaps
- Increases scientific return
- Publications for early career PIs
- Broadens our scientific community
- International collaboration

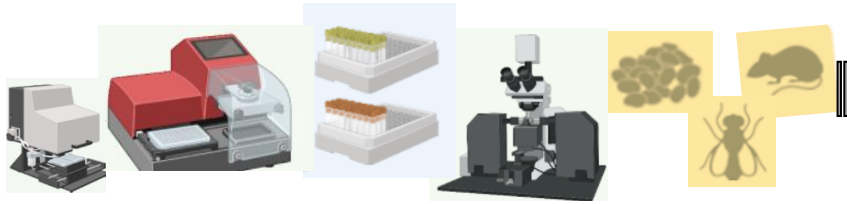
*Specimens mostly from Shuttle and ISS missions. Also stored are specimens from ground analog studies including centrifuge, hind-limb unloaded and partial weight bearing.

Ames Life Science Data Archive

- **Past, Present, Future**
  - International Space Station, Space Shuttle, Bion, Hindlimb Unloading, Radiation-Exposed, Partial Weight Bearing, Rotating Wall Vessel, Clinostat, other
- **Basic & Applied**
  - µGravity, Radiation
  - Behavioral
  - Countermeasures
- **Mouse, Rat**, Drosophila, Yeast, Bacteria, more soon!

Data from PIs, Missions, Experiments





Raw, Processed Data

Assay, Sample Metadata

Mission, Hardware Metadata

Environmental, Telemetry Data

ALSDA

REUSE

Open Science Data Repositories

- Phenotypic-Physiological-Behavioral-Imaging-Level Data
- **Single Submission Portal**
- **Data federated with GeneLab** (Molecular Omics-Level Data)
- Often Same Subjects

(Images from Ronca et al., 2019 https://doi.org/10.1038/s41598-019-40789-y; Cells, Bone, RR Habitat NASA-ARC)

GeneLab

www.genelab.nasa.gov

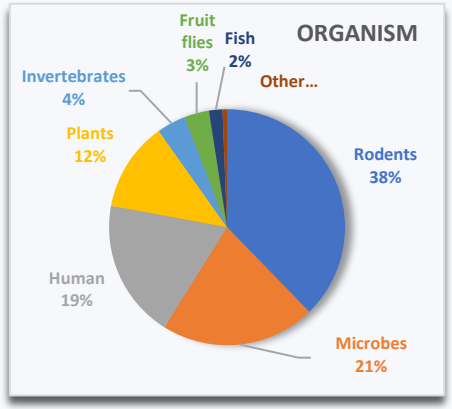
378
Studies

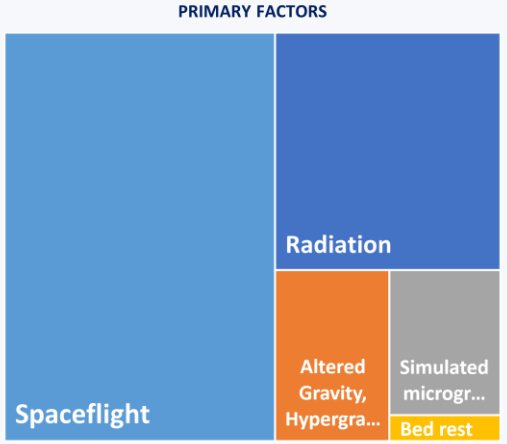
429
Datasets

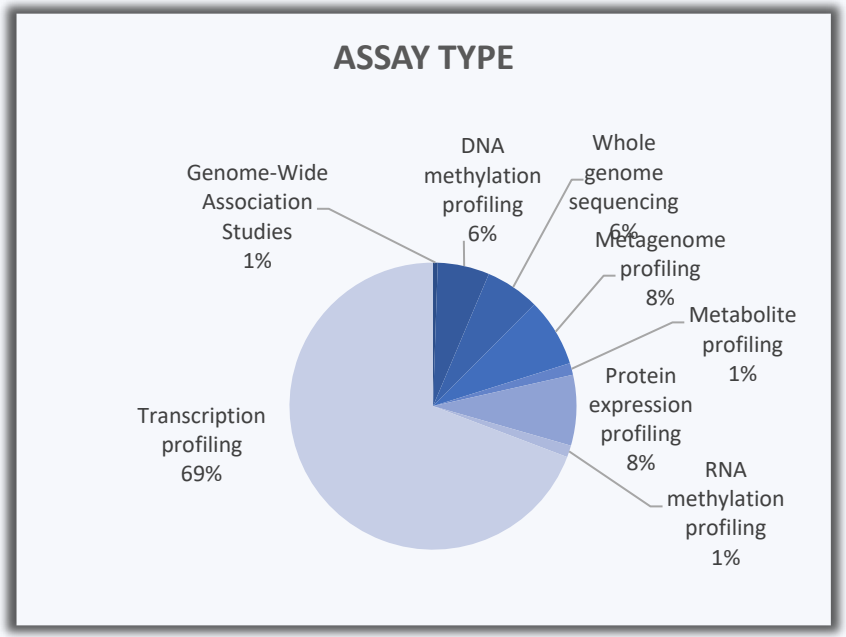
45
Species

>10
Assays

>135 TB
Data







5,847
Biospecimens available

159
Experimental Payloads

1,260
Research Records

<20
Species

164
Omics Studies

Re-use of Data and Enabling New Discoveries

40 enabled publications (**10 publications produced by the AWGs**) using data available in GeneLab.

Mammalian and Invertebrate Models as Complementary Tools for Gaining Mechanistic Insight on Muscle Responses to Spaceflight

by Thomas Cahill [1], Henry Cope [2], Joseph J. Bass [3], Eliah G. Overbey [4], Rachel Gilbert [5,6], Willian Abraham da Silveira [1,7], Amber M. Paul [5,8,9], Tejaswini Mishra [10], Raúl Herranz [11], Sigrid S. Reinsch [5], Sylvain V. Costes [5], Gary Hardiman [1,12], Nathan

An Integrative Network Science and Artificial Intelligence Drug Repurposing Approach for Muscle Atrophy in Spaceflight Microgravity

Vidya Manian*, Jairo Orozco-Sandoval and Victor Diaz-Martinez

Rad-Bio-App: a discovery environment for biologists to explore spaceflight-related radiation exposures

Simon Gilroy

Cell — Supports open access — Submit — Log in — Reg

ARTICLE | VOLUME 183, ISSUE 5, P1185-1201.E20, NOVEMBER 25, 2020

PDF [10 MB]

Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress

Environmental and Molecular Mutagenesis — Environmental Mutagenesis and Genomics Society

BRIEF COMMUNICATION | Open Access

In vitro relationships of galactic cosmic radiation and epigenetic clocks in human bronchial epithelial cells

Jamaji C. Nwanaji-Enwerem, Philippe Boileau, Jonathan M. Galazka, Andres Cardenas

Computational and Structural Biotechnology Journal

Volume 19, 2021, Pages 2223-2235

Reanalysis of the Ma

common gut microb

astronauts induced b

Cell Reports — Open access — Subr

REPORT | VOLUME 33, ISSUE 10, 108406, DECEMBER 08, 2020

PDF [1 MB] Figures

ic Aging and Leukocyte Analysis of Simulated

-500 Mission

Uzoji Nwanaji-Enwerem • Lars Van Der Laan • Jonathan M. Galazka •

iScience — Open access

ARTICLE | VOLUME 23, ISSUE 12, 101734, DECEMBER 18, 2020

PDF [2 MB]

Comparative Transcriptomics Identifies Neuronal and Metabolic Adaptations to Hypergravity and Microgravity in *Caenorhabditis elegans*

Craig R.G. Willis • Nathaniel J. Szewczyk • Sylvain V. Costes • ... Sigrid S. Reinsch • Timothy Etheridge [7] • Catharine A. Conley • Show all authors • Show footnotes

Open Access • Published: November 25, 2020 • DOI: https://doi.org/10.1016/j.isci.2020.101734

Subm

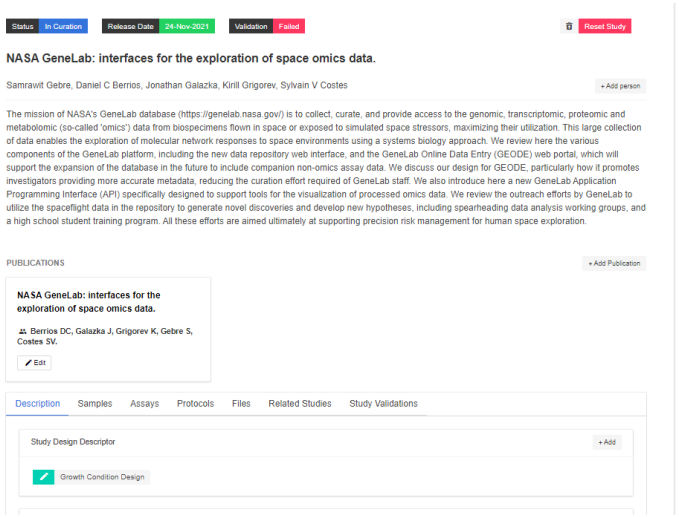
Figures

Open Access | Article

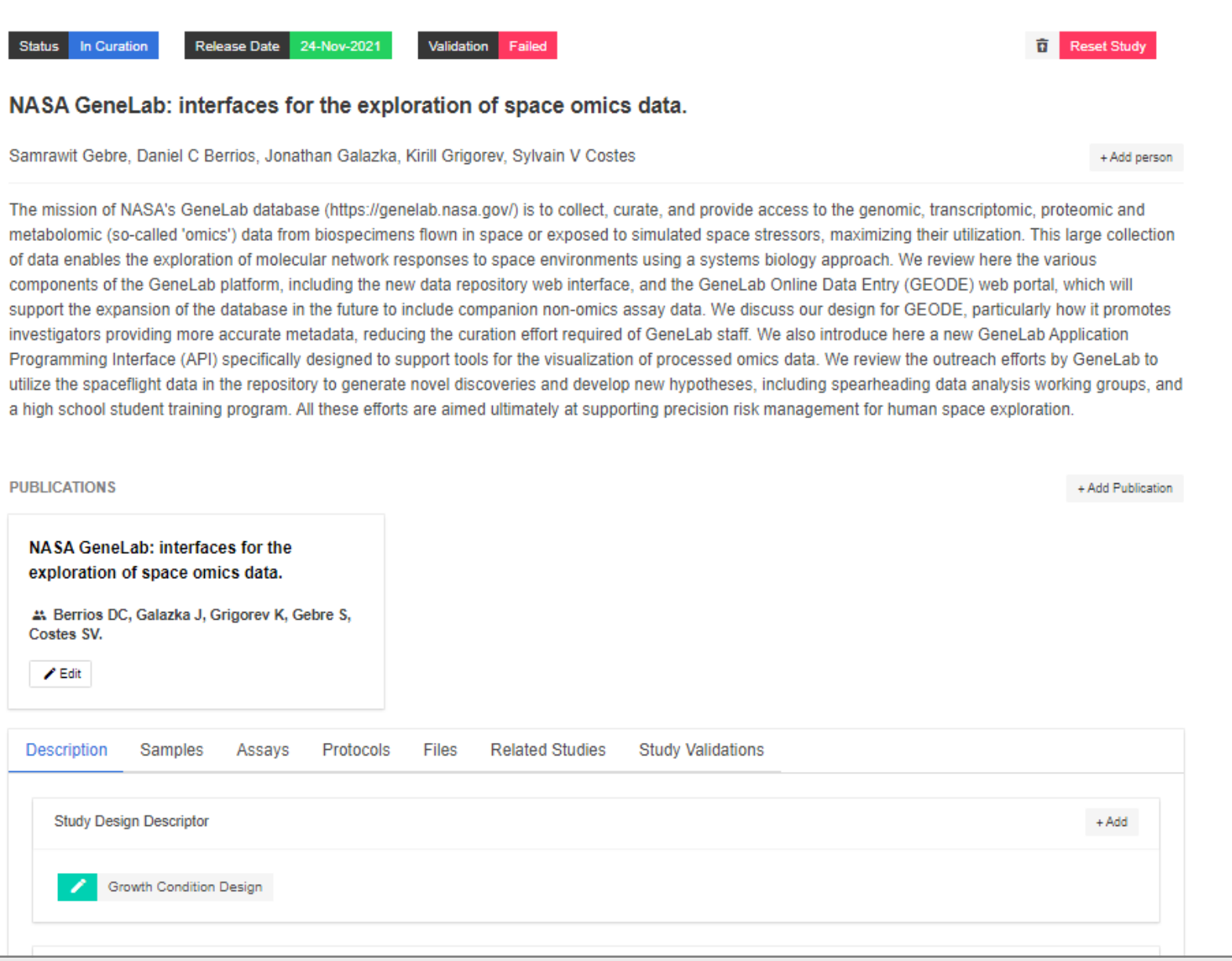
GeneLab Database Analyses Suggest Long-Term Impact of Space Radiation on the Cardiovascular System by the Activation of *FYN* Through Reactive Oxygen Species

by Afshin Beheshti [1,*], J. Tyson McDonald [2], Jack Miller [3], Peter Grabham [4] and Sylvain V. Costes [5,*]

CONTRIBUTE

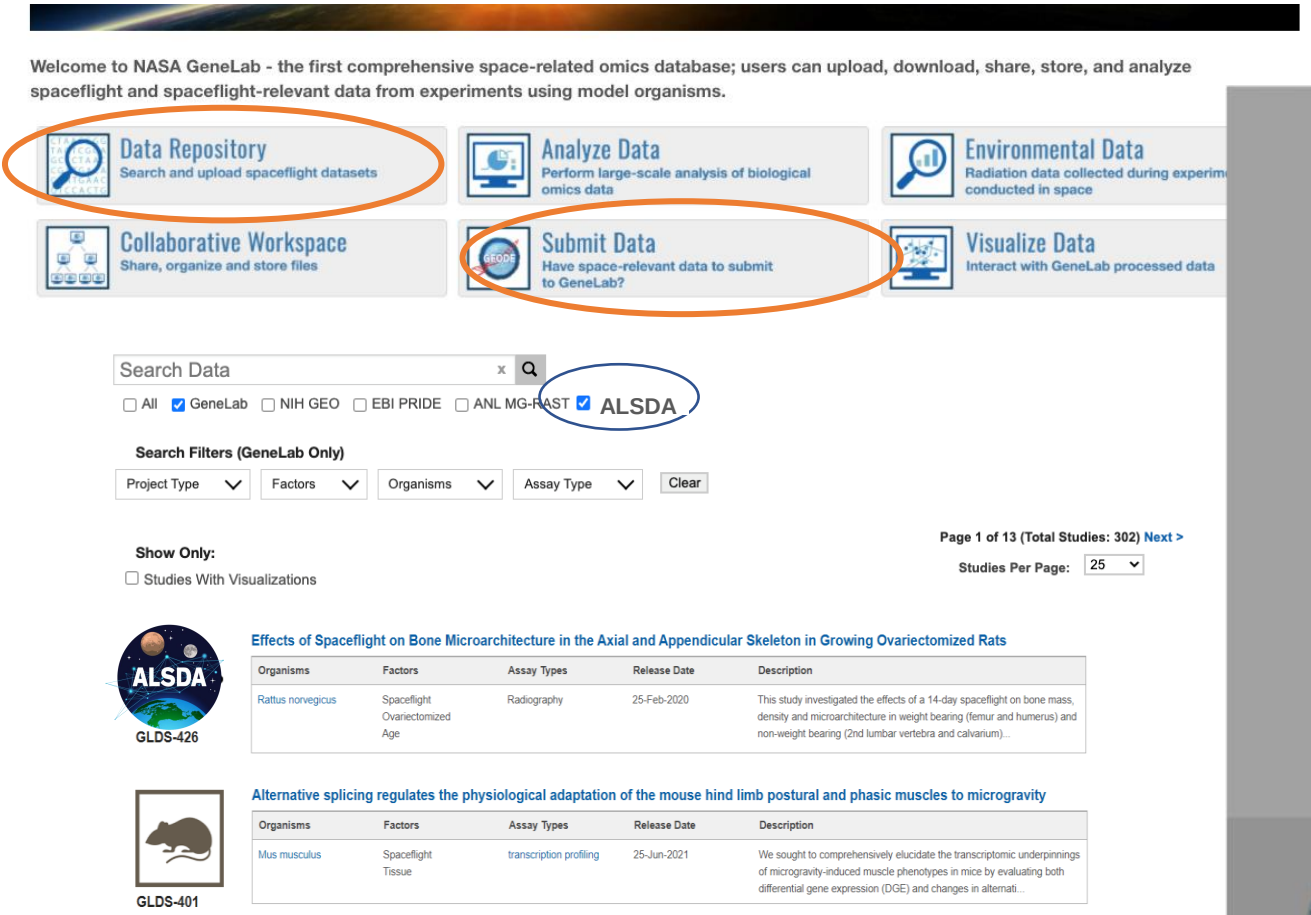

Status In Curation
Release Date 24-Nov-2021
Validation Failed
Reset Study
NASA GeneLab: interfaces for the exploration of space omics data.
Samrawit Gebre, Daniel C Berrios, Jonathan Galazka, Kirill Grigorev, Sylvain V Costes
+ Add person
The mission of NASA's GeneLab database (https://genelab.nasa.gov/) is to collect, curate, and provide access to the genomic, transcriptomic, proteomic and metabolomic (so-called 'omics') data from biospecimens flown in space or exposed to simulated space stressors, maximizing their utilization. This large collection of data enables the exploration of molecular network responses to space environments using a systems biology approach. We review here the various components of the GeneLab platform, including the new data repository web interface, and the GeneLab Online Data Entry (GEODE) web portal, which will support the expansion of the database in the future to include companion non-omics assay data. We discuss our design for GEODE, particularly how it promotes investigators providing more accurate metadata, reducing the curation effort required of GeneLab staff. We also introduce here a new GeneLab Application Programming Interface (API) specifically designed to support tools for the visualization of processed omics data. We review the outreach efforts by GeneLab to utilize the spaceflight data in the repository to generate novel discoveries and develop new hypotheses, including spearheading data analysis working groups, and a high school student training program. All these efforts are aimed ultimately at supporting precision risk management for human space exploration.
PUBLICATIONS
+ Add Publication
NASA GeneLab: interfaces for the exploration of space omics data.
Berrios DC, Galazka J, Grigorev K, Gebre S, Costes SV.
Edit
Description
Samples
Assays
Protocols
Files
Related Studies
Study Validations
Study Design Descriptor
+ Add
Growth Condition Design

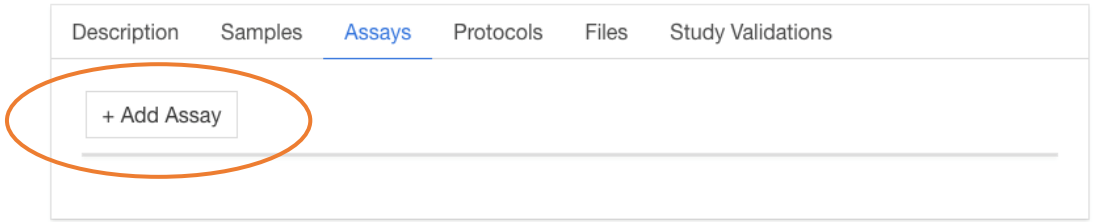
DATA SUBMISSION: ADD, EDIT, SAVE, SUBMIT

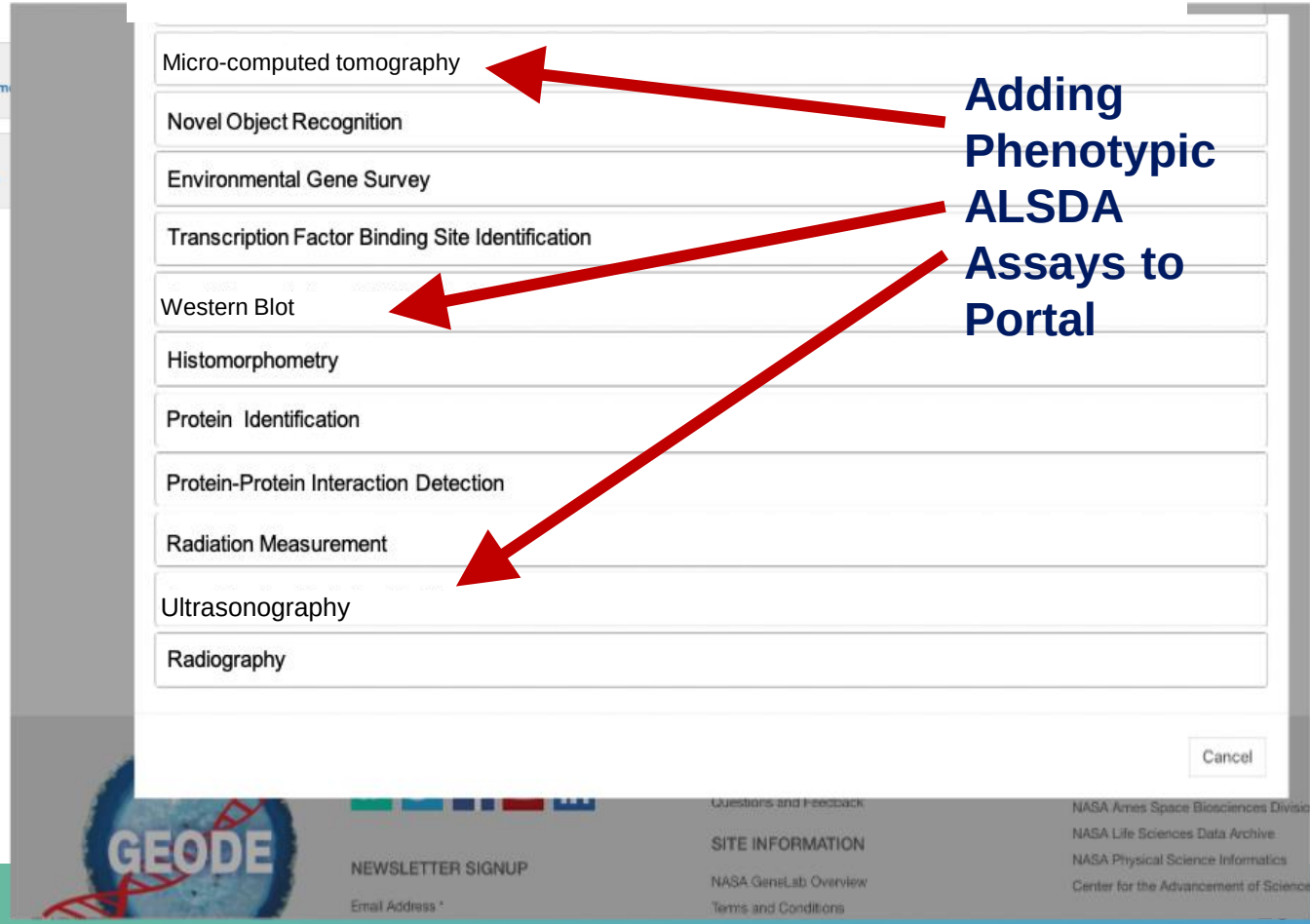


Enhancing Data Submission Portal

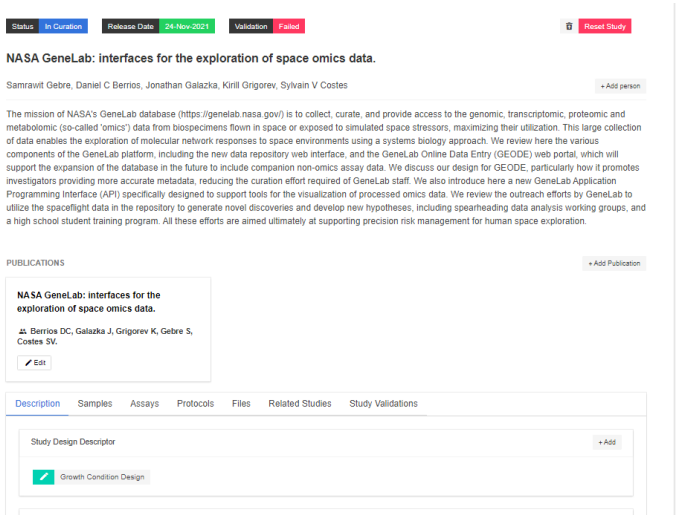
Enhancing submission automation and "self-service"



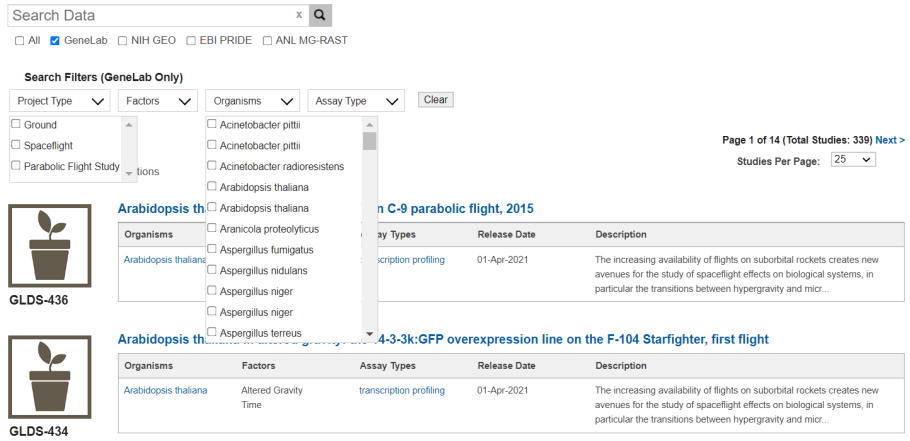




CONTRIBUTE



DISCOVER



DATA REPOSITORY

Search Data

All GeneLab NIH GEO EBI PRIDE ANL MG-RAST

Search Filters (GeneLab Only)

Project Type | Factors | Organisms | Assay Type | Clear

Ground
Spaceflight
Parabolic Flight Study

Acinetobacter pittii
Acinetobacter pittii
Acinetobacter radioresistens
Arabidopsis thaliana
Arabidopsis thaliana
Aranicola proteolyticus
Aspergillus fumigatus
Aspergillus nidulans
Aspergillus niger
Aspergillus niger
Aspergillus terreus

Page 1 of 14 (Total Studies: 339) Next >

Studies Per Page: 25

GLDS-436

Arabidopsis th… …n C-9 parabolic flight, 2015

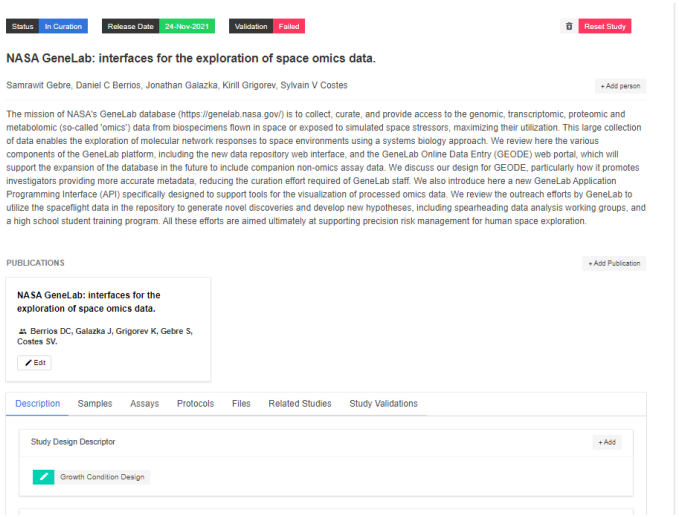
Organisms	…ay Types	Release Date	Description
Arabidopsis thaliana	…scription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and micr...

GLDS-434

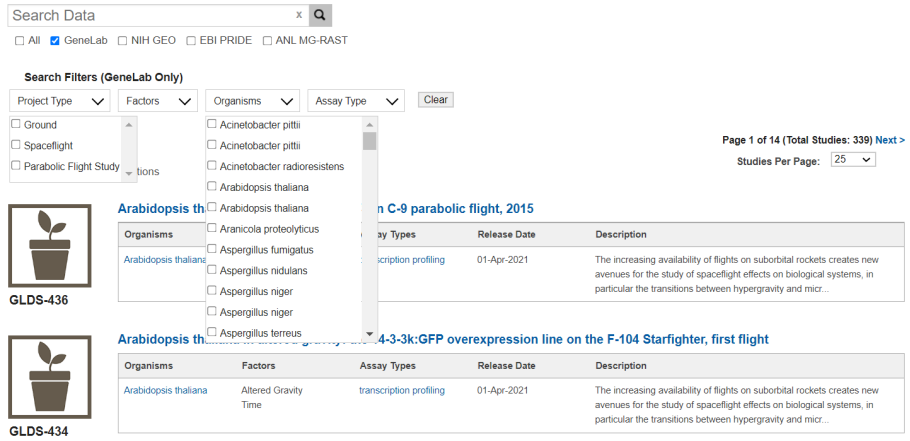
Arabidopsis th… …4-3-3k:GFP overexpression line on the F-104 Starfighter, first flight

Organisms	Factors	Assay Types	Release Date	Description
Arabidopsis thaliana	Altered Gravity Time	transcription profiling	01-Apr-2021	The increasing availability of flights on suborbital rockets creates new avenues for the study of spaceflight effects on biological systems, in particular the transitions between hypergravity and micr...

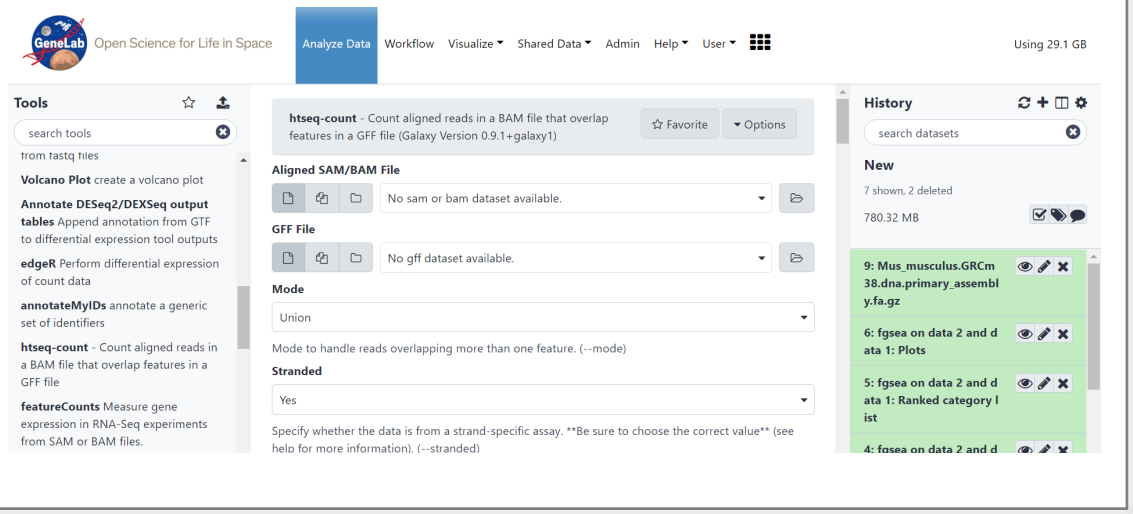
CONTRIBUTE



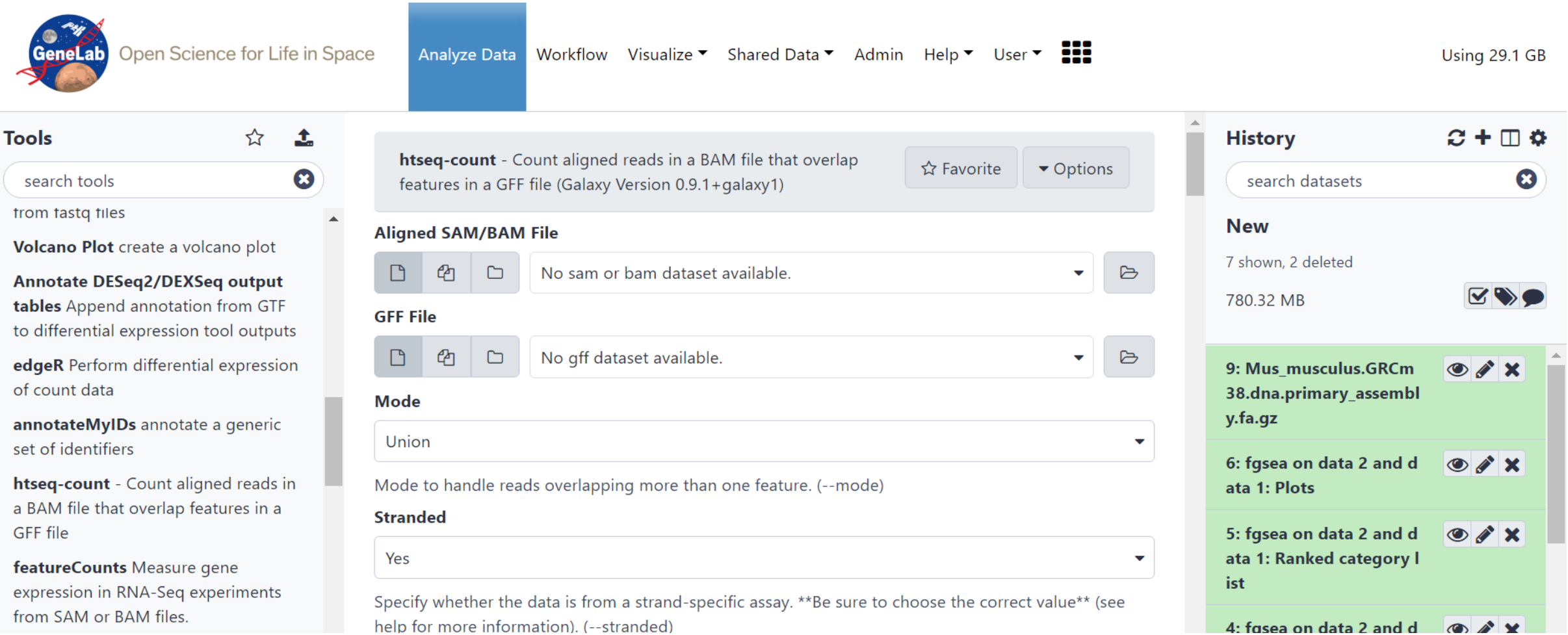
DISCOVER



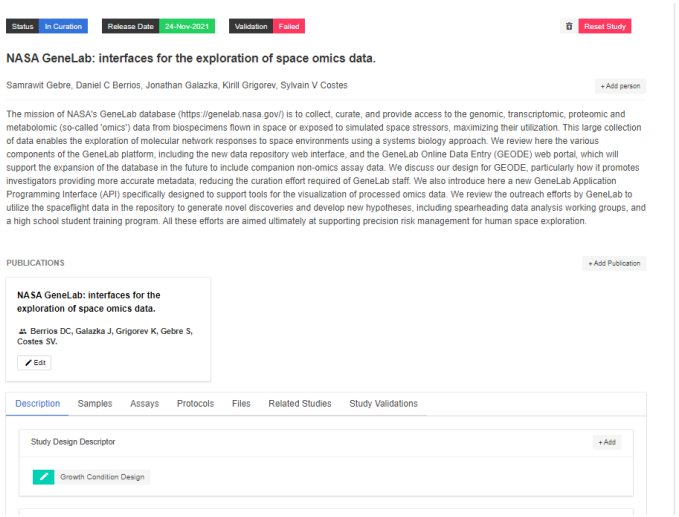
ANALYZE



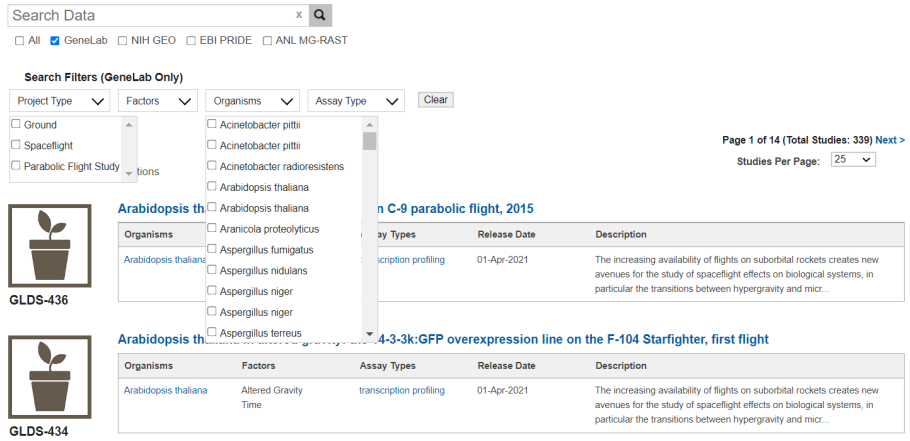
ANALYSIS PLATFORM



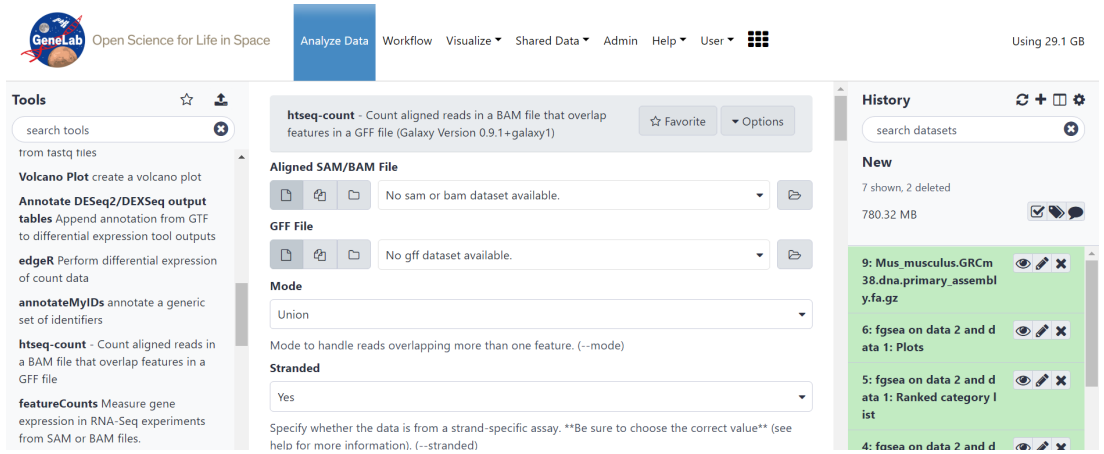
CONTRIBUTE



DISCOVER



ANALYZE



EXPLORE

DATA VISUALIZATION PORTAL

GeneLab Visualization

+ Add Visualization

GLDS-4

Title: Microarray Analysis of Space-flown Murine Thymus Tissue
Organism(s): Mus musculus
Number of Genes: 20217
Number of Samples: 7

Group Selection

Group 1:

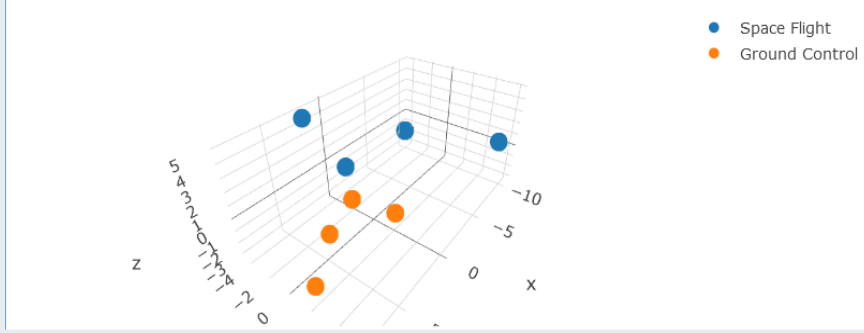
Ground Control

Group 2:

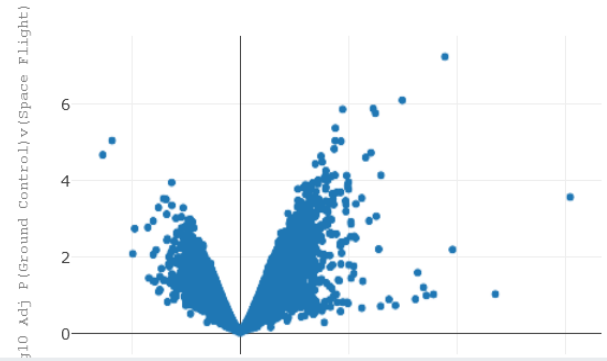
Space Flight

Update

GLDS-4 PCA



GLDS-4 Volcano



GLDS-4 DGE

Maximum p-value:

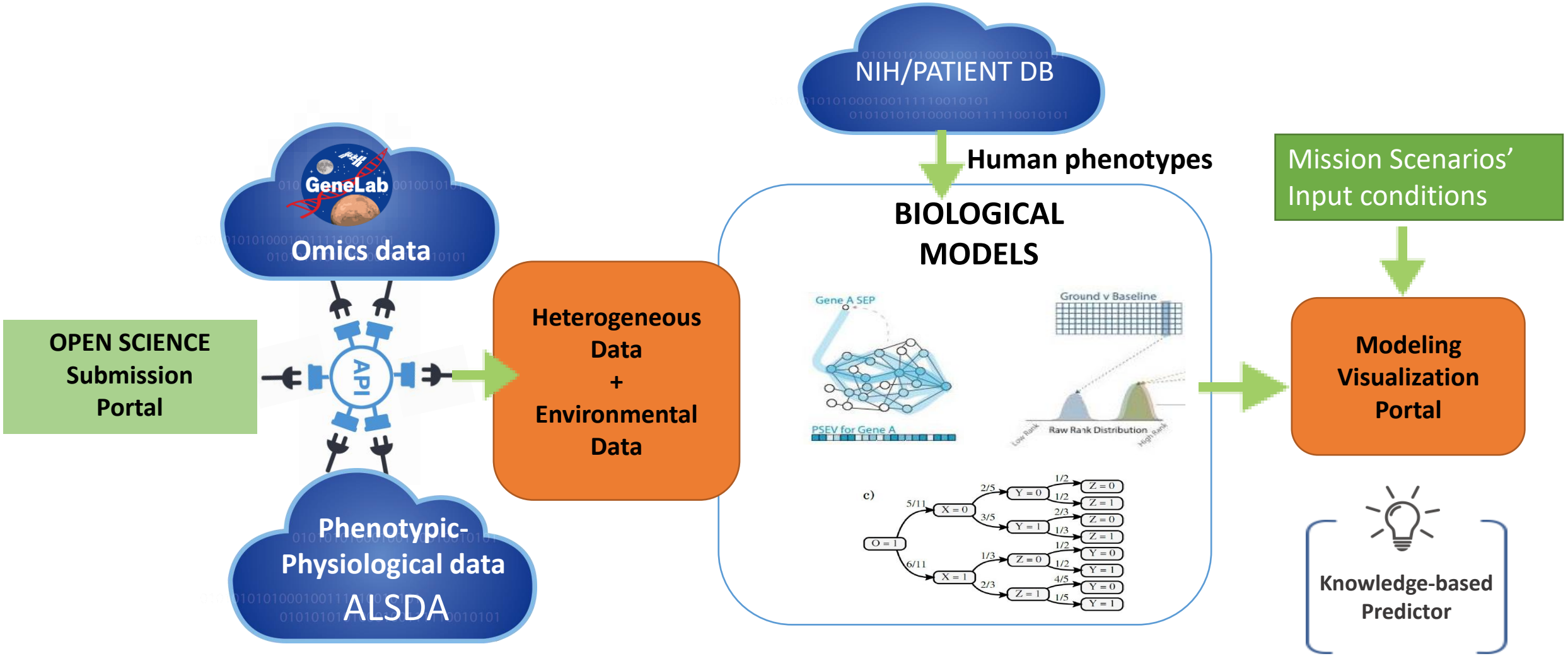
Copy CSV Excel PDF Print

Search:

Symbol	LOG2FC	PVAL	ADJP
Search Symbol	Search LOG2FC	Search PVAL	Search ADJP
0610009B22Rik	0.15407506639823598	0.375179028040943	0.7101000817246831
0610009L18Rik	-0.0314533496186336	0.8000506355583309	0.9247611817918641

Data Reuse Example: Data Modeling

AI ready FAIR data for modeling and risk assessment



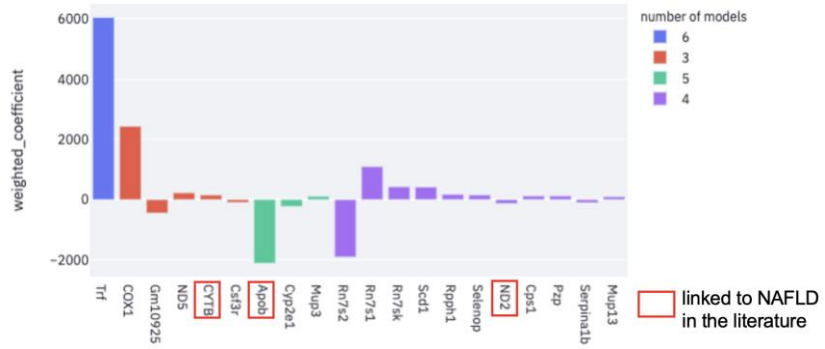
(Image Credits: Nelson et al., 2021, https://doi.org/10.3390/life11010042; Genewein et al., 2020, arXiv:2010.12237v1 [cs.AI])

AI for Life in Space

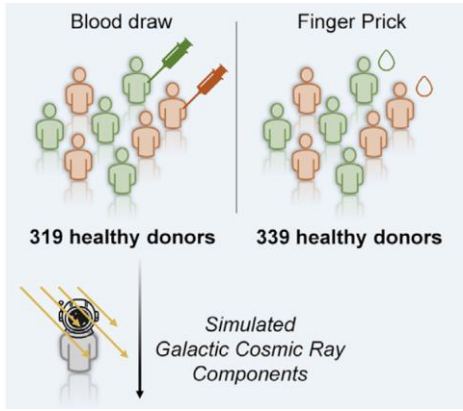
Causal Inference

Genes with a causal relationship to liver disease phenotype in spaceflown rodents

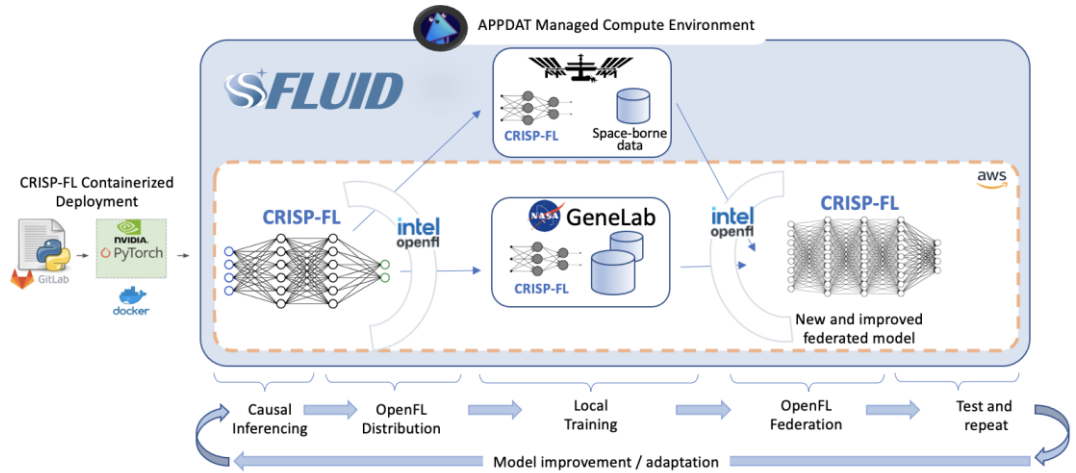
GLDS-47 GLDS-137
GLDS-48 GLDS-168



Genetic variants with a causal relationship to DNA damage and radiation sensitivity in human irradiated immune cells



Federated causal inference between Earth data and spaceborne data



2021 FDL Astronaut Challenge

A novel paradigm for generating human-relevant biomedical insights from observational datasets with limited size by leveraging and augmenting with animal model data

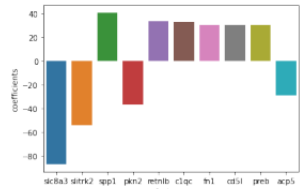


Figure 1: Top-10 IRM features according to feature coefficients with all human and mice gamma-radiation data.

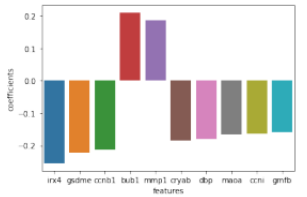


Figure 2: Top-10 IRM features according to feature coefficients with all human and mice heavy-ion radiation data.

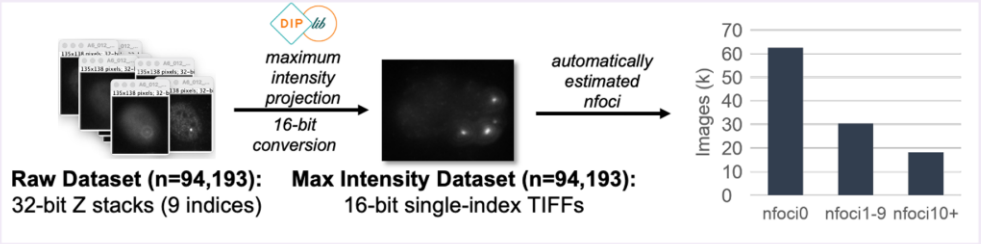
AI for Life in Space

Benchmark Microscopy Dataset

-Huge amounts of microscopy data in space biology, time consuming, manual annotate

Approach:

-Radiation exposure microscopy DNA damage dataset, distinct foci (PI: Costes)



Raw Dataset (n=94,193): 32-bit Z stacks (9 indices)

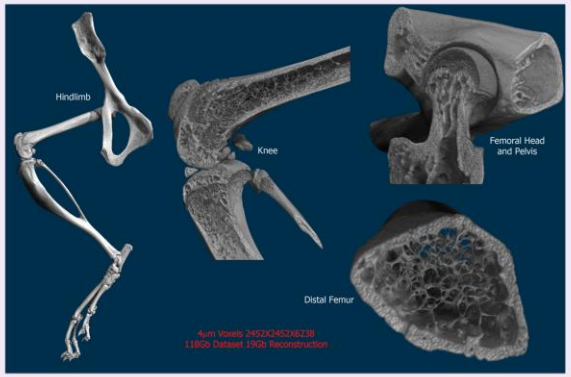
Max Intensity Dataset (n=94,193): 16-bit single-index TIFFs

Benchmark microCT Dataset

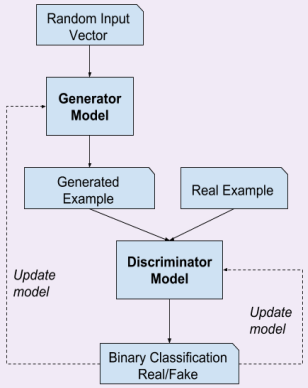
-CT is key to assess space biological bone impacts

Approach:

-Mouse mCT dataset, both spaceflight and group (PI: Almeida)



Benchmark RNA sequencing dataset



-RNA-seq data quantifies all expressed genes in a sample

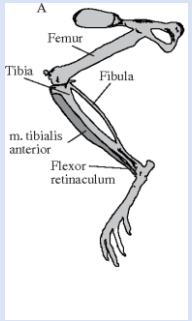
-Classic example of high-D (~50k genes), low-N dataset

Approach:

1. Start with real RNA-seq data from GeneLab
2. Generate synthetic RNA-seq data using generative adversarial network (GAN)
3. Benchmark dataset will be high-D and high-N (synthetic)

ML applications: e.g. causal inference to identify causal genes with relation to a pathogenic target variable

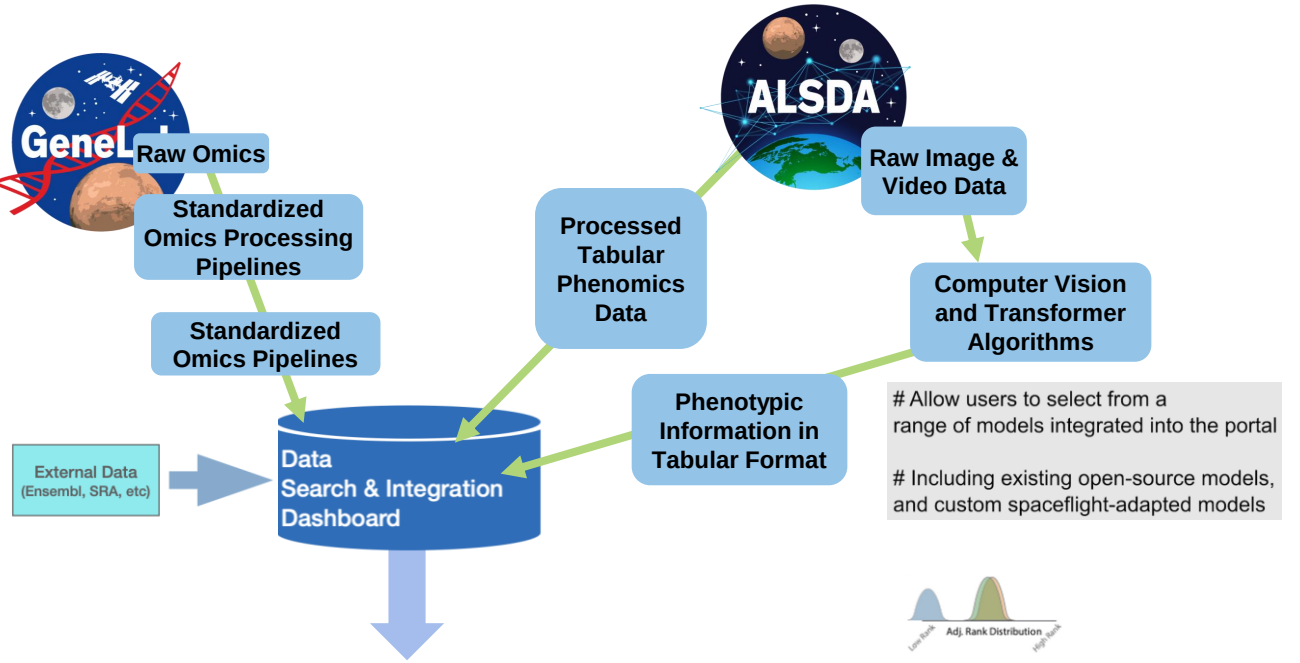
Omics and Phenomic Muscle Data using explainable xAI (Qlattice) to identify drivers of fast vs slow-twitch space muscle impacts



- Transcriptomics (RNA-seq)
- Methylation
- Proteomics
- Calcium uptake

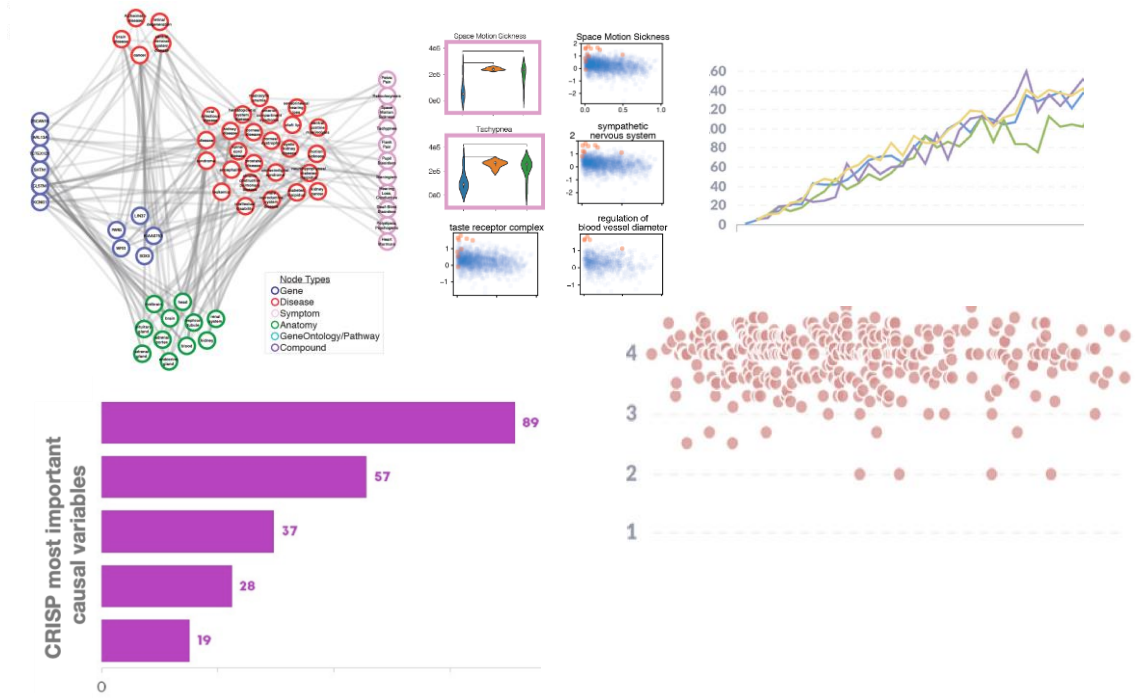
(Credit: BMs: Sylvain Costes & Eduardo Almeida; CRISP: Makenna Myrick, James Casaletto, Sanders et al., 2021; FLUID: Graham Mackintosh, 2022; DAG: Rob Reynolds, et al., 2021; ALSDA Vis, Fu et al., 2021, Svetlana Komarova)

Future: Streamlined Data Integration, Multimodal, Multihierarchical



Analysis Portal

Interactive analysis dashboard of modeling and analysis results



New Modeling Hub

Automated pipelines for a suite of machine learning, knowledge graph, modeling tools

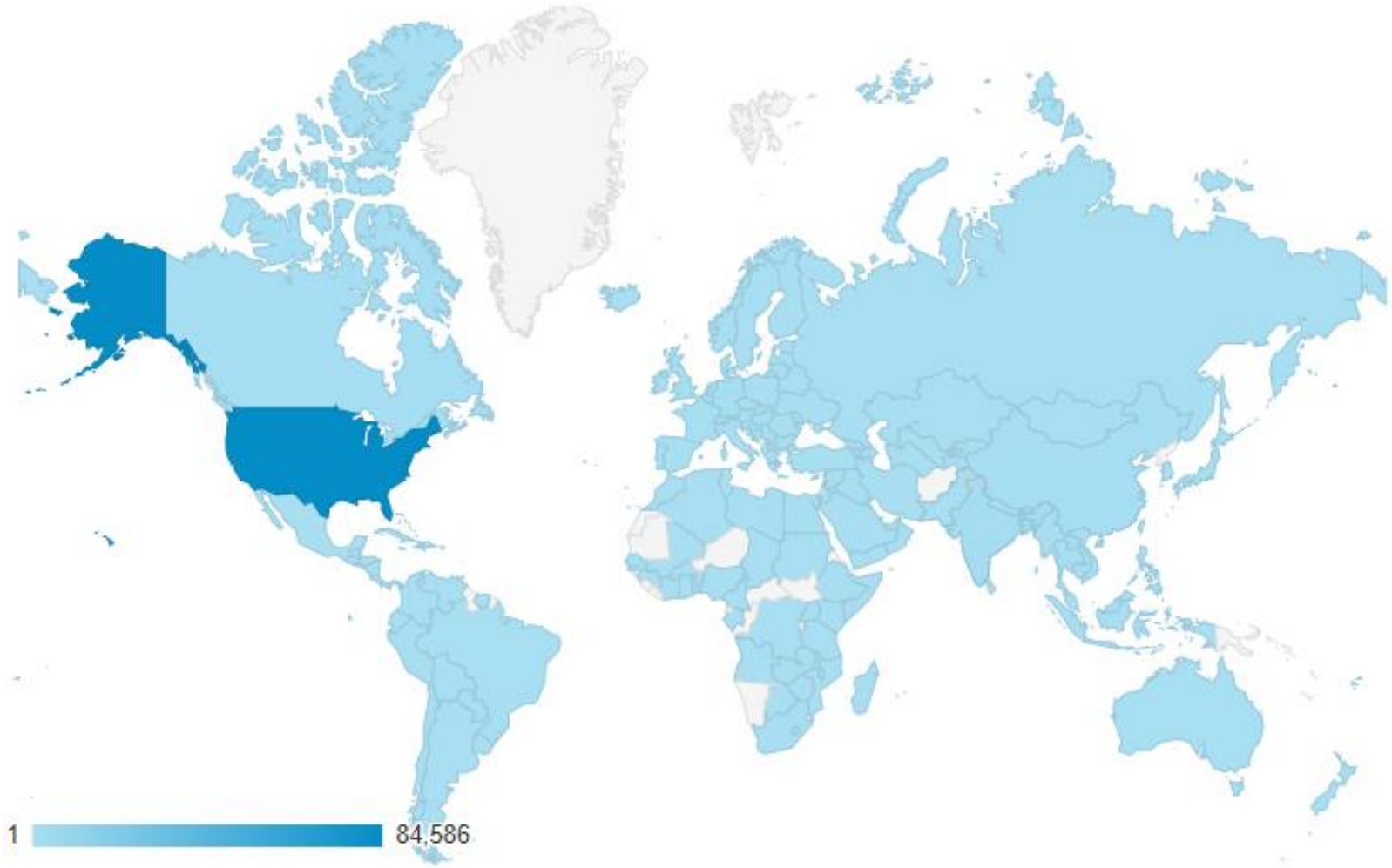
(Credit: Adrienne Hoarfrost, James Casaletto, Lauren Sanders; Images: Nelson et al., 2021, https://doi.org/10.3390/life11010042 Casaletto et al., 2022, CRISP & Liver, in preparation)

SCIENTIFIC COMMUNITY

Making data accessible worldwide

Top 20 Countries utilizing GeneLab:

1. United States
2. India
3. Germany
4. United Kingdom
5. China
6. Canada
7. Singapore
8. Japan
9. Italy
10. Spain
11. France
12. Turkey
13. Australia
14. Brazil
15. South Korea
16. Ghana
17. Netherlands
18. Mexico
19. Russia
20. Philippines



Analysis Working Groups

Analysis Working Groups (AWGs) *consist of* ***400+ scientists*** *from multiple space agencies, international institutions, and industry. Scientists meet monthly with each group to provide feedback, develop standards, and analyze data.*

We invite you to join - https://genelab.nasa.gov/awg/join!

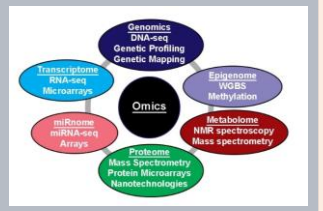
ANIMAL

68 members

Facilitates the use of omics in understanding basic mechanisms by which animals and constituent tissues and cells adapt to the spaceflight environment.

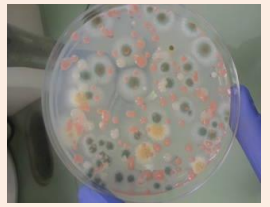
MULTI-OMICS

181 members



Interactions between the different omics to provide complete understanding of the entire system begin studied.

MICROBES



50 members

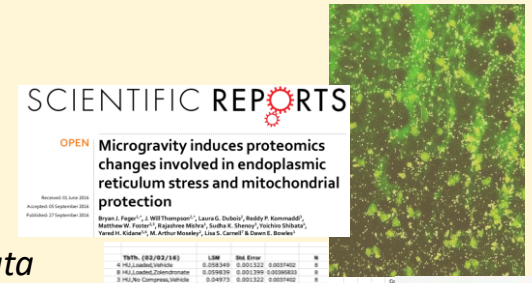
Focuses on analyzing microbial datasets within GeneLab that includes gene-expression, proteomic, metabolomic and environmental metagenomic datasets.

PLANTS

58 members

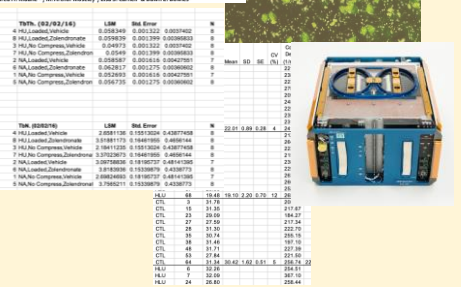
Share and discuss the latest developments in ***Astrobotany*** *– the discipline of botany concerned with interactions between plant biology and space environment.*

ALSDA


SCIENTIFIC REPORTS

OPEN

Microgravity induces proteomics changes involved in endoplasmic reticulum stress and mitochondrial protection



150 members

- *Recruiting members to support data standards and novel methods to analyze cross mission and platform data*
- *Establish standards for various assays' metadata*
- *Provide feedback on database and tools*

Training the Next Generation of Scientists

GeneLab for High Schools (GL4HS): A four-week intensive training summer program for rising high school juniors and seniors to learn bioinformatics and computational biology methods and techniques to analyze space omics data.

Learn more and apply at: https://www.nasa.gov/ames/genelab-for-high-schools

GeneLab for Colleges/Universities (GL4U): For educators and students to learn how analyze omics data using GeneLab standard pipelines and space-relevant data

Access to course materials: https://github.com/nasa/GeneLab-Training/tree/main/GL4U

Space Life Sciences Training Program (SLSTP): Provides undergraduate students entering their junior or senior years, and entering graduate students, with professional experience in space life science disciplines.

Learn more and apply at: https://www.nasa.gov/ames/research/space-life-sciences-training-program

---

Spaceflight Technology, Applications, and Research (STAR): The annual course targets principal investigators (PIs), senior research scientists, and postdoctoral scholars and aims to facilitate their entry to space biology and preparation for conducting spaceflight experiments using NASA and commercial platforms.

Applications open until May 30th: https://science.nasa.gov/biological-physical/programs/star

NASA Postdoctoral Program: If you're an early-career or senior scientist, apply to the NASA Postdoctoral Program to help pursue NASA's mission and experience the world's most diverse technology and expertise.

NEXT ERA OF DATA

Commercial and International Partners

Open access data in GeneLab, **accelerates the pace of scientific discovery** and thus **enables exploration and development of countermeasures** by using systems biology approaches to identify cross-species relevance, mechanisms, biomarkers and potential health risks for LEO and beyond LEO (Lunar and Mars Exploration).

To increase usability and accessibility, space biological research should be available through a common portal. It's important to work closely with our commercial and international partners to ensure data across agencies and platforms are harmonized and comparable. In addition, systems already developed can be utilize not duplicating efforts and costs for data sharing.

For example, **GeneLab already hosts datasets from multiple space agencies (ESA, JAXA, CSA) and from CASIS funded experiments.**

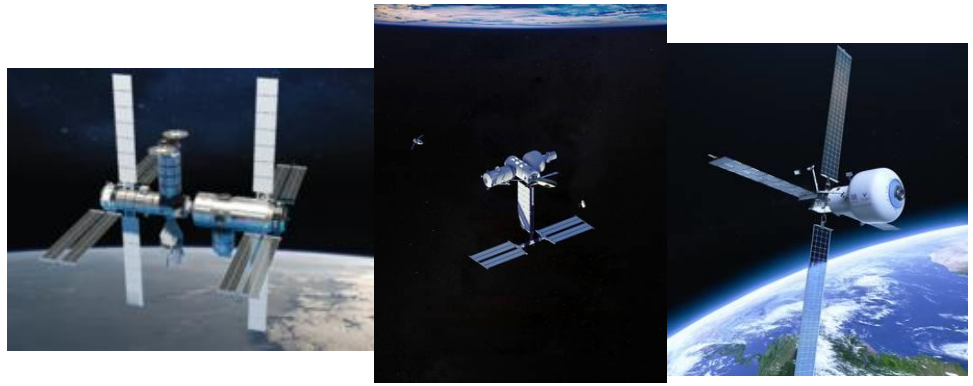
In order to continue maximizing the return of spaceflight experiments, **we encourage commercial and international partners to utilize the GeneLab and ALSDA Data Systems for their research data, both for open and proprietary data.**





Proprietary data

Currently all the data that is published, is openly accessible worldwide. GeneLab/ALSDA has already in place a mechanism for standardize data curation and account-based access. This capability can be expanded in the future for:



Astronaut/Human space-relevant data

- Develop a controlled access data portal
- Create an automatic data access request workflow
- Scientists will be able to translate their findings from model organisms to human outcomes

Commercial LEO Destinations

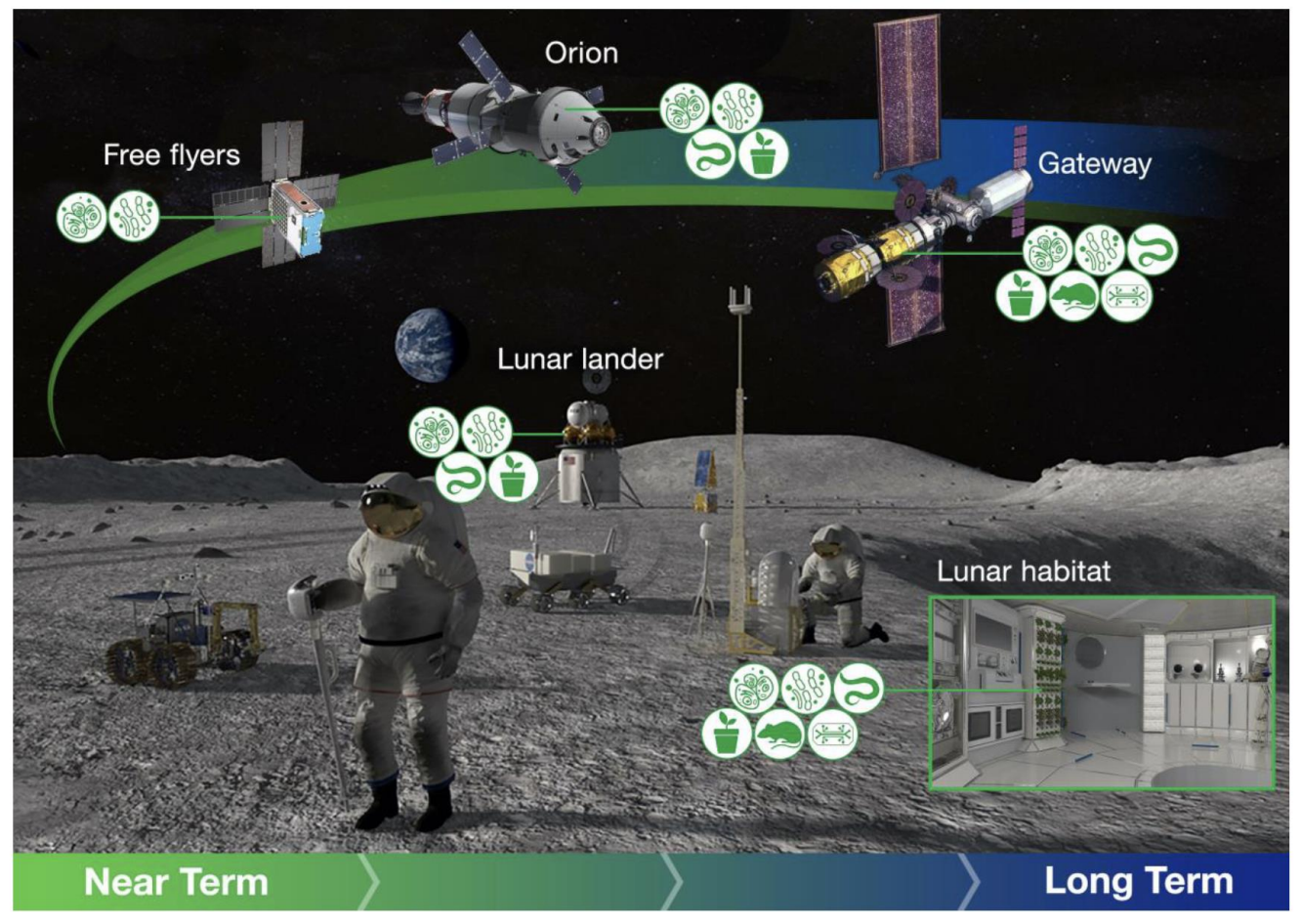
- Develop a controlled access data portal and an automatic data access request workflow for “closed” data
- Enable access to all spaceflight research a centralized data repository

Data should be as open as possible, as closed as necessary.

Metadata can always be open.

-Shelly Stall, TEM April 2020

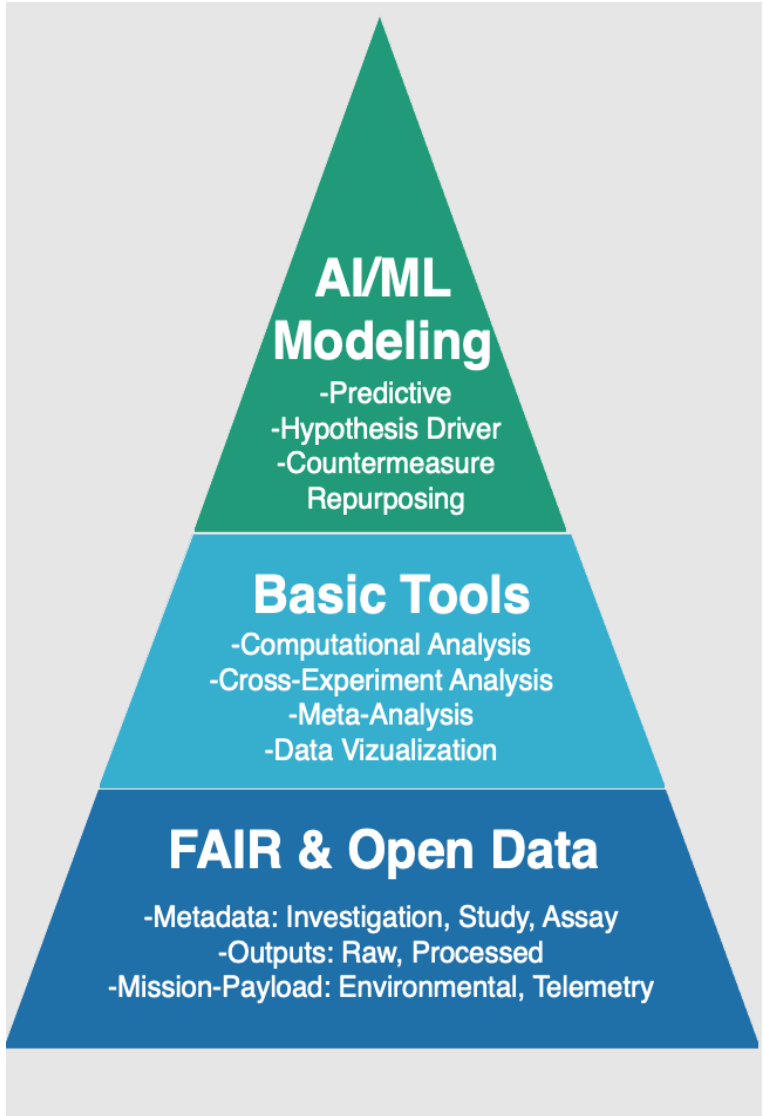
Beyond LEO

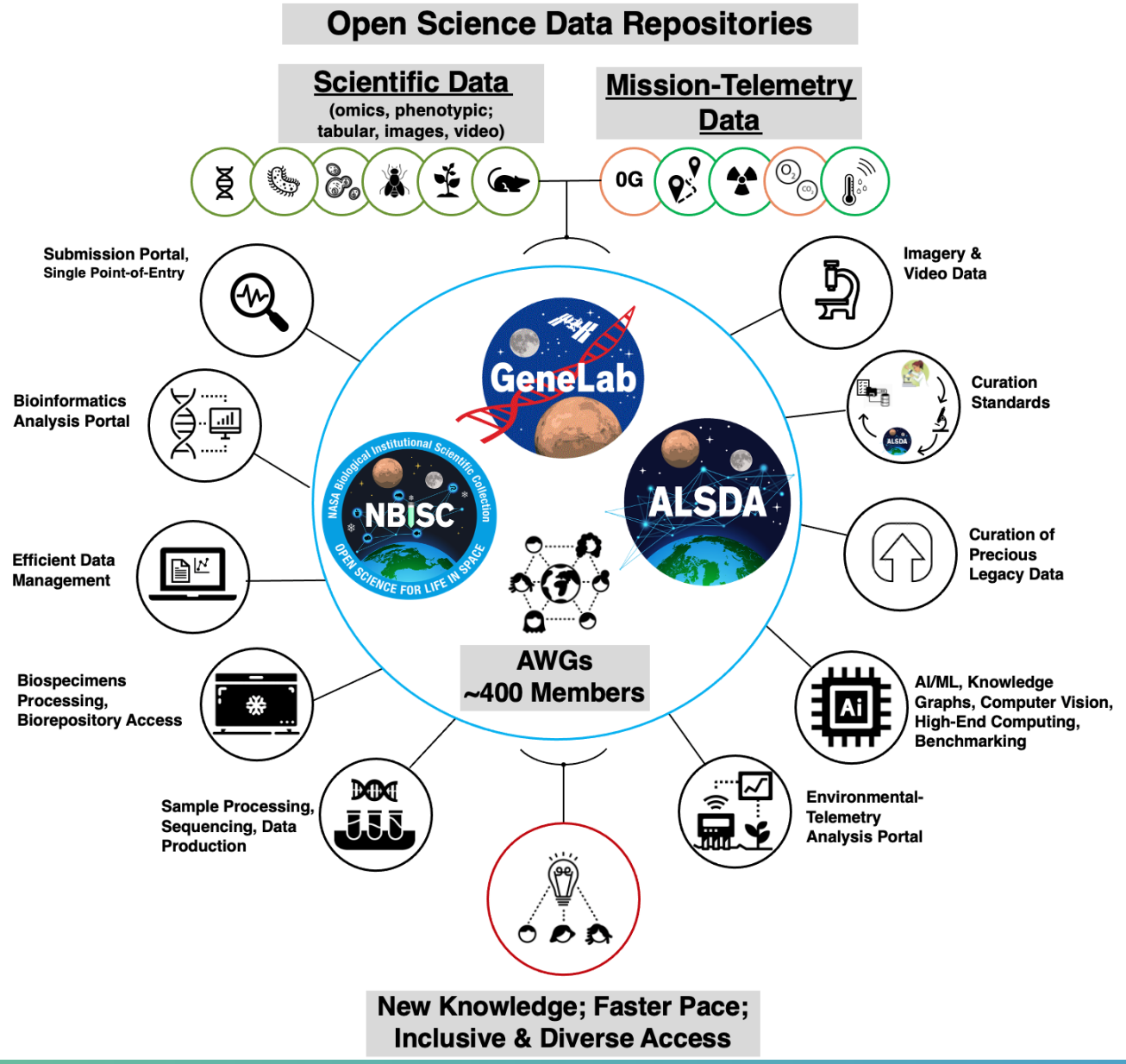


With limited sample return for various platforms, we are preparing new metadata templates and data streams to host and publish data from future lunar experiments and beyond.

(Credit: NASA Biological and Physical Science Division; Beyond LEO Instrumentation & Science Series Science Working Group, 2021 Report, https://ntrs.nasa.gov/api/citations/20210023324/downloads/BLISS%20SWG%202021%20Report%20Final.pdf)

Big Picture for Open Science for Life in Space





THANK YOU!

Open Science for Life in Space projects are funded by the NASA Space Biology program and NASA Human Research Program.

https://www.nasa.gov/ames/research/space-biosciences/alsda

Twitter: @NASAGeneLab
https://genelab.nasa.gov

https://www.nasa.gov/ames/research/space-biosciences/nbisc

https://science.nasa.gov/biological-physical/space-biology-biospecimen-sharing-program

Feel free to email me at samrawit.g.gebre@nasa.gov with any questions or links from this presentation.