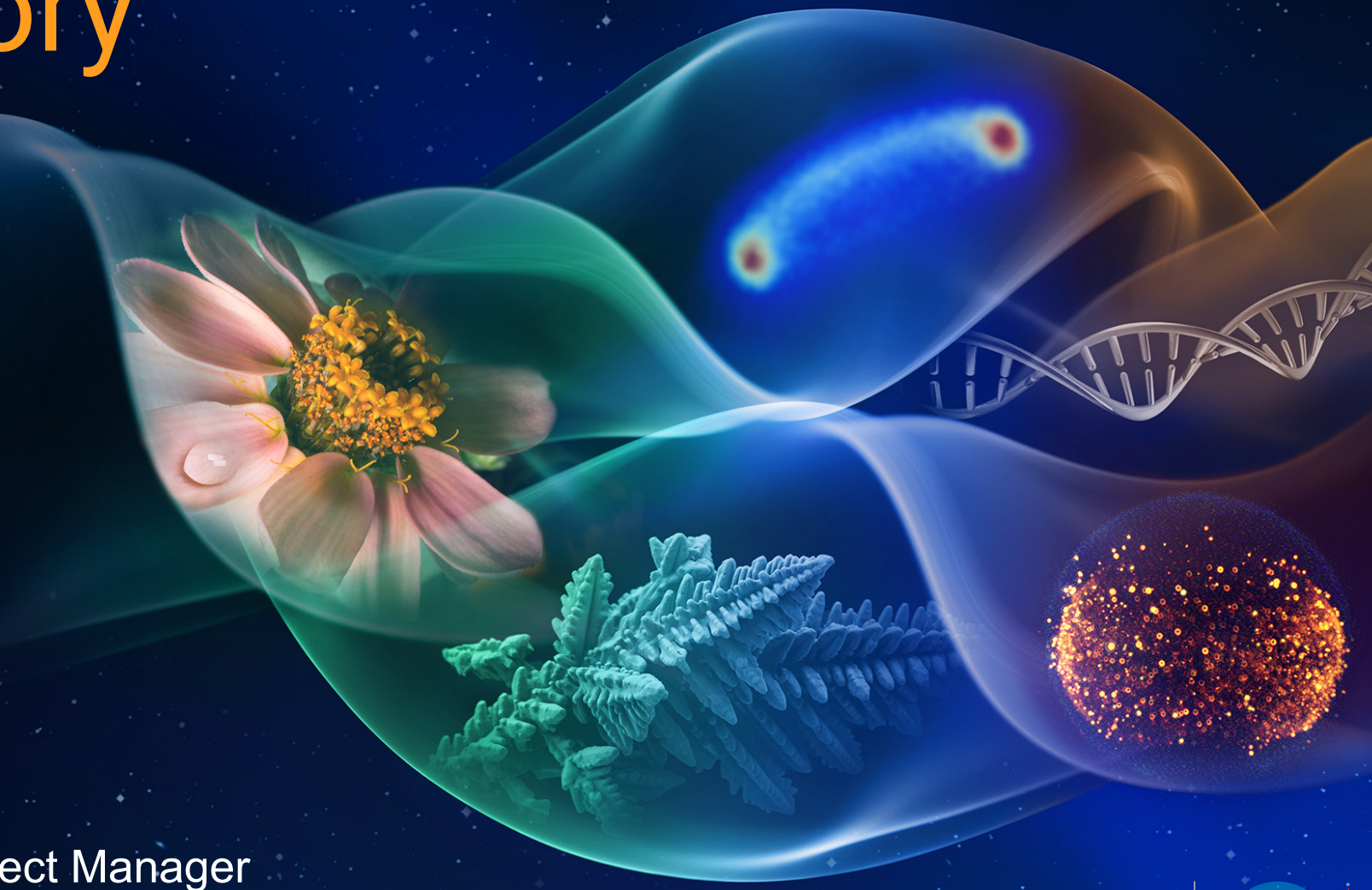
NASA's Open Science Data Repository



Danielle Lopez
ALSDA Deputy Project Manager
NASA Ames Research Center

National Aeronautics and Space Administration

NASA Space Biology Open Science

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)

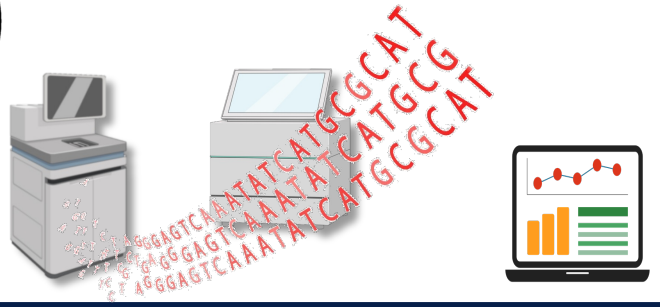
Collection of non-human specimens and space microbial culture

Ames Life Sciences Data Archive (ALSDA)

Collection and curation of mission, project, and imaging data

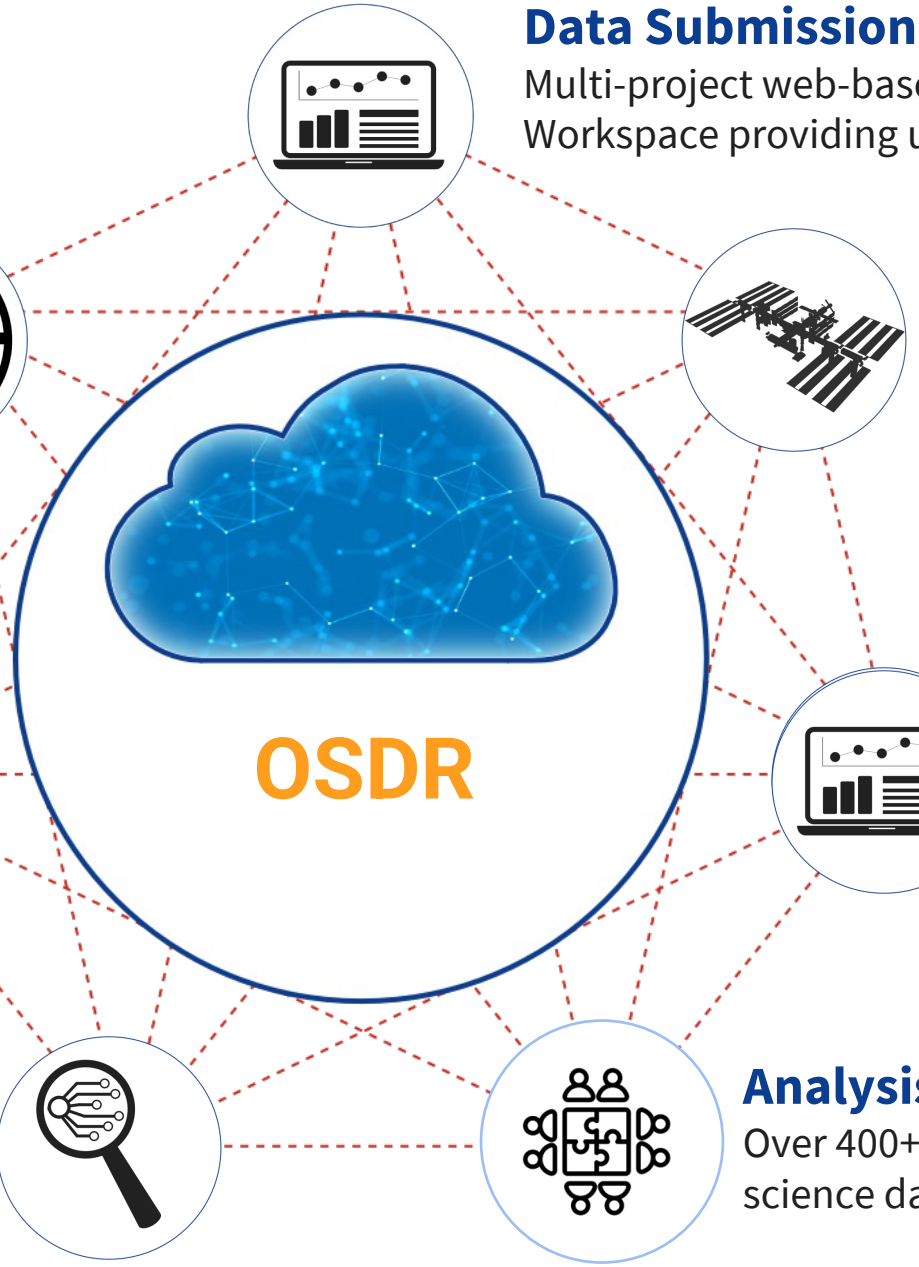
NASA GeneLab (GL)

Collection and curation of omics data



NASA Internal Program

Open-Source Science Programs



Data Submission and Workspace

Multi-project web-based submission portal
Workspace providing users file storage

Commercial Entities

Mission science
Environmental

Data Repository & Visualization

Open access and curated data.
Interactive visualization portal to enable knowledge discovery

AI/ML

Training Models and feedback on ML ready data

Training

Programs developed for high school and university students

New Discoveries

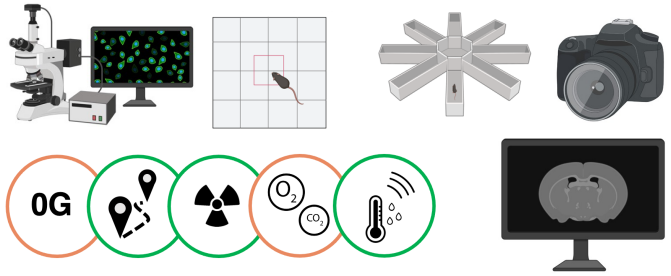
Enabling new publications and research through data reuse

Analysis Working Groups

Over 400+ members analyzing open science data

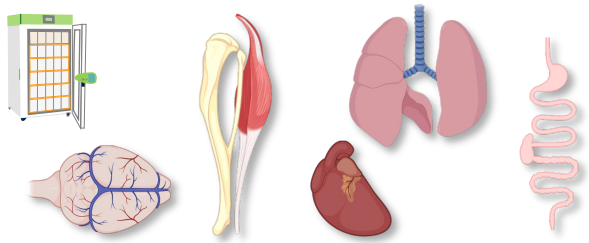
NASA's Open Science Data Repository

Physiological/Phenotypic/Imaging/ Environmental Telemetry Data



Molecular/Omics Data

Biospecimens



ALSDA

GeneLab

NBISC

NASA Open Science Data Repository (OSDR)

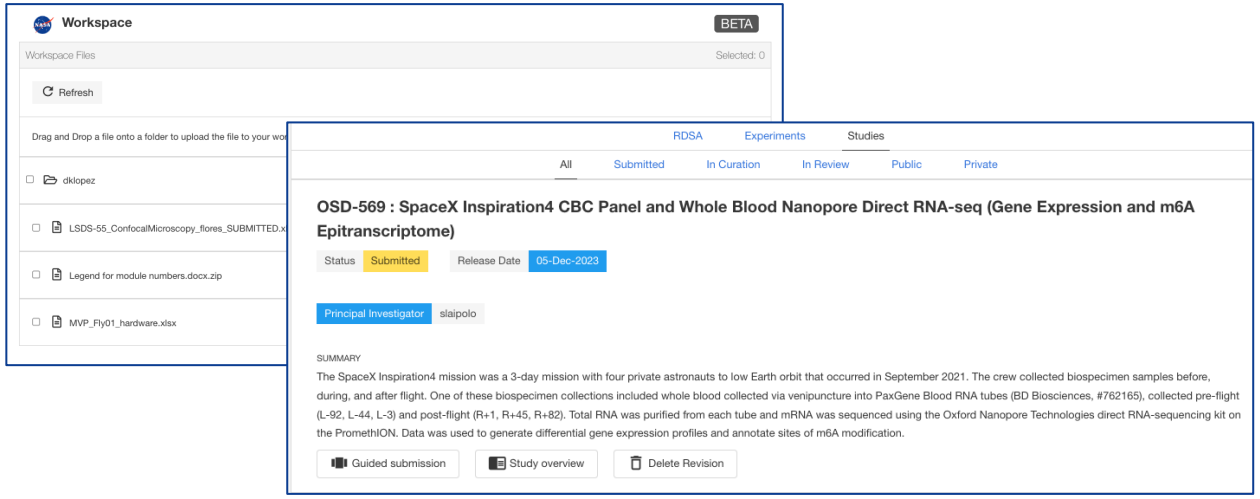
osdr.nasa.gov/bio

- Single Submission Portal (BDME)
- User Interface/Website Tool for RDSAs (Research Data Submission Agreements)
- Workspace
- Maximally Open Access with Necessary Controls for Sensitive Data
- Data Maximally FAIR

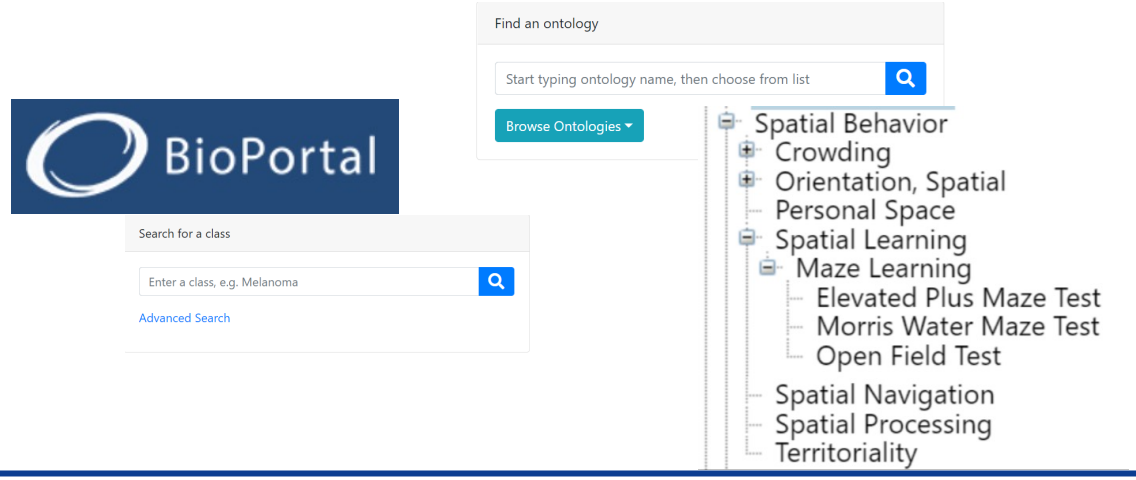
Findable Accessible Interoperable Reusable

Single Submission Portal

Contribute and Work



Standardization: Ontology and Linkages



Sample-Level Curation

- Strain, age, sex, animal husbandry
- Radiation info, sample collection

	Parameter	Parameter	Parameter	Parameter
S_1 data				
S_2 data				
S_3 data				
S_4 data				

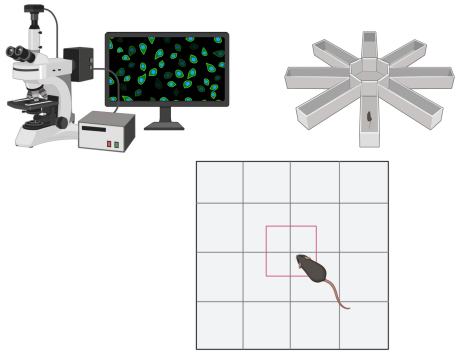
Source Name	Sample Name	Characteristics: Organism	Characteristics: Strain	Comment: Animal Source	Factor Value: Spaceflt
B1	B1_RAT1_Femur	Rattus norvegicus	Fischer 344	Taconic Biosciences (Farms)	Basal Control
B2	B2_RAT2_Femur	Rattus norvegicus	Fischer 344	Taconic Biosciences (Farms)	Basal Control
B3	B3_RAT3_Femur	Rattus norvegicus	Fischer 344	Taconic Biosciences (Farms)	Basal Control
B4	B4_RAT4_Femur	Rattus norvegicus	Fischer 344	Taconic Biosciences (Farms)	Basal Control
B5	B5_RAT5_Femur	Rattus norvegicus	Fischer 344	Taconic Biosciences (Farms)	Basal Control
B6	B6_RAT6_Femur	Rattus norvegicus	Fischer 344	Taconic Biosciences (Farms)	Basal Control
GC1	GC1_RAT13A_Femur	Rattus norvegicus	Fischer 344	Taconic Biosciences (Farms)	Ground Control
GC2	GC2_RAT13B_Femur	Rattus norvegicus	Fischer 344	Taconic Biosciences (Farms)	Ground Control
GC3	GC3_RAT14A_Femur	Rattus norvegicus	Fischer 344	Taconic Biosciences	Ground Control

Assay-Level Curation

- How assay was conducted
- Specs of maze, tracking software details, microscope settings

Scanco mCT40 scanner
Technology Type:
Micro-Computed Tomography
Select Export Columns

Sample Name	Protocol REF	Parameter Value: Scanner	Parameter Value: Volume Of Interest Location	Parameter Value: Volume Of Interest Size	Parameter Value: Volu of Interest Location 2
B1_RAT1_Femur	Image Scan Acquisition	Scanco mCT40	Cortical: Mid-Diaphysis	984 length micrometer	Cancellous: Distal Metaphysis
B2_RAT2_Femur	Image Scan Acquisition	Scanco mCT40	Cortical: Mid-Diaphysis	984 length micrometer	Cancellous: Distal Metaphysis
B3_RAT3_Femur	Image Scan Acquisition	Scanco mCT40	Cortical: Mid-Diaphysis	984 length micrometer	Cancellous: Distal Metaphysis



Open Science Data Repository

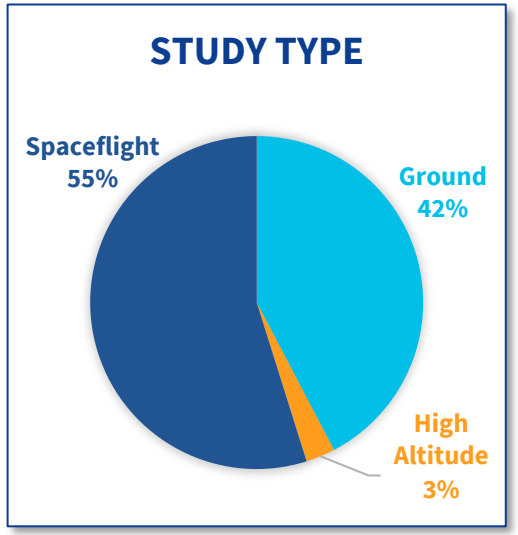
453 Studies

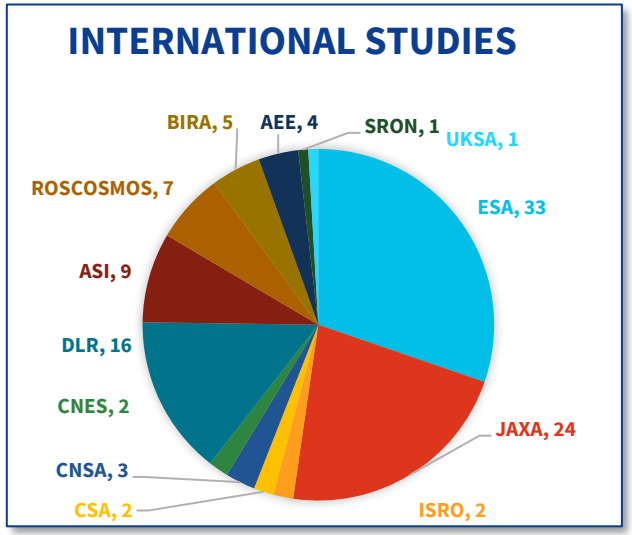
880 Datasets

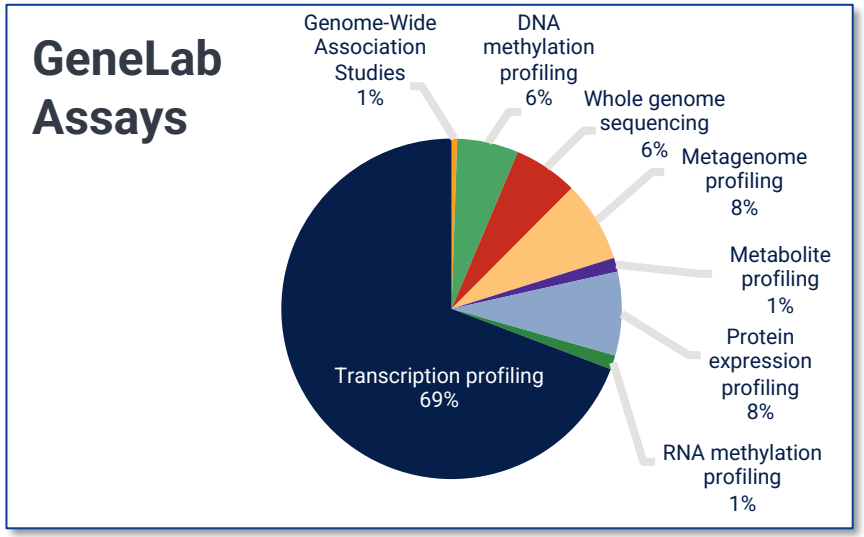
45 Species

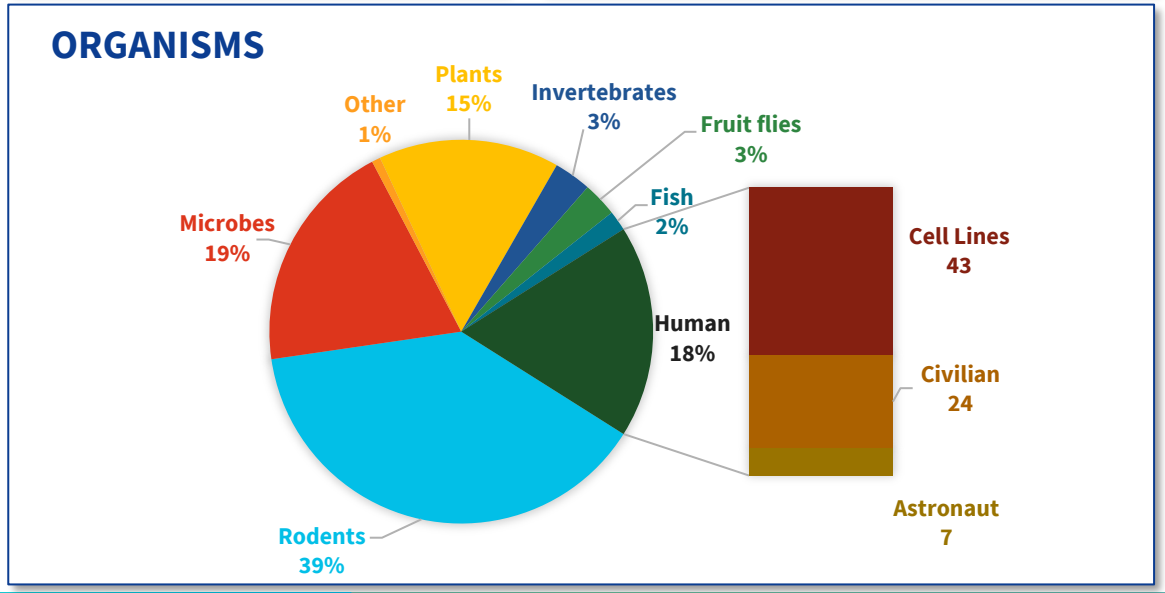
>55 Assays

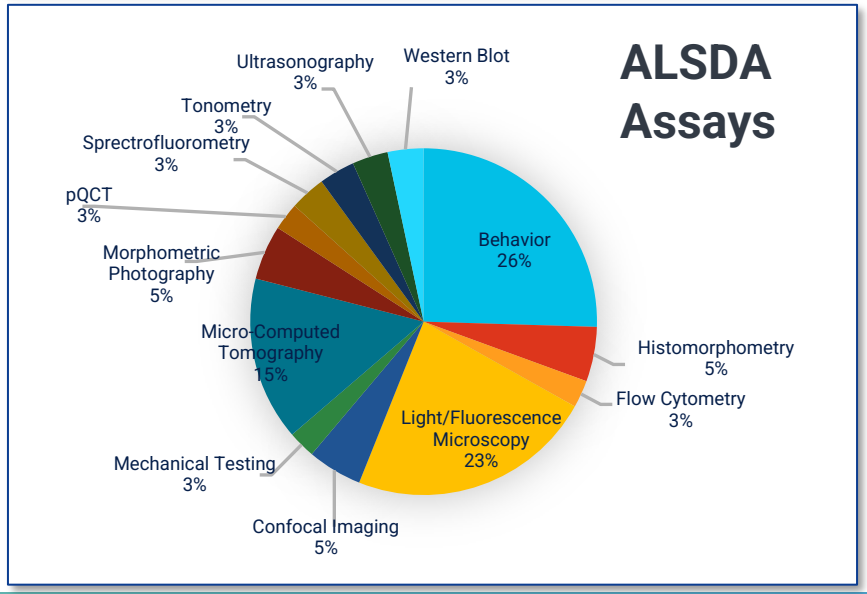
>150TB Data





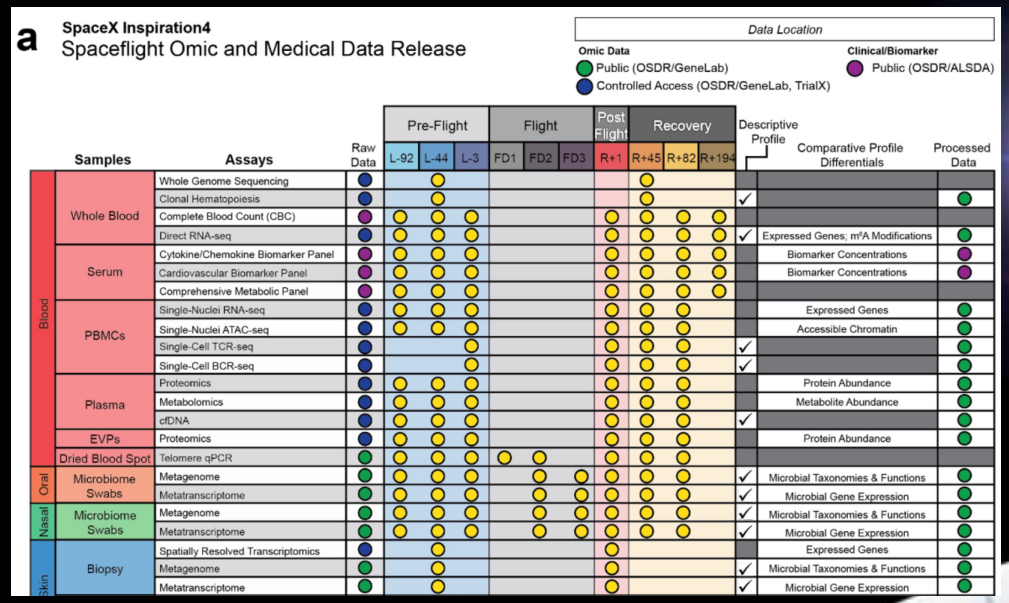






Engaging with commercial entities: I4 Human Data into OSDR

INSPIRATION4



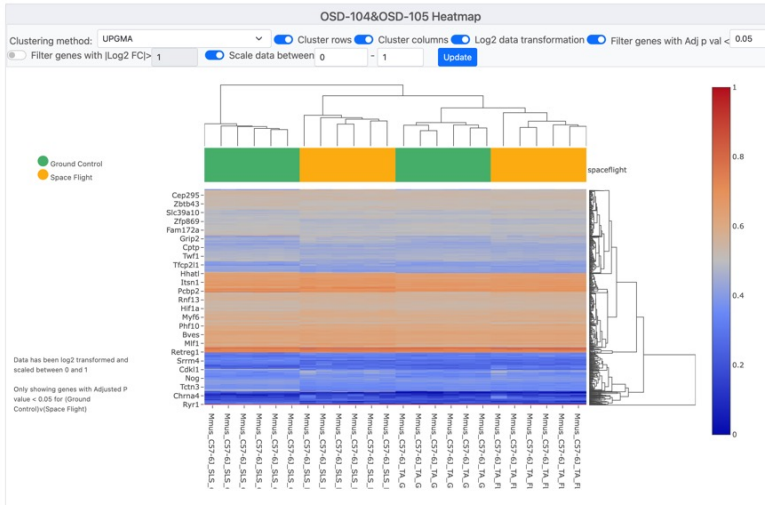
- Credit: Chris Mason's Lab; Overbey et al., 2023
- In Peer Review at *Nature*

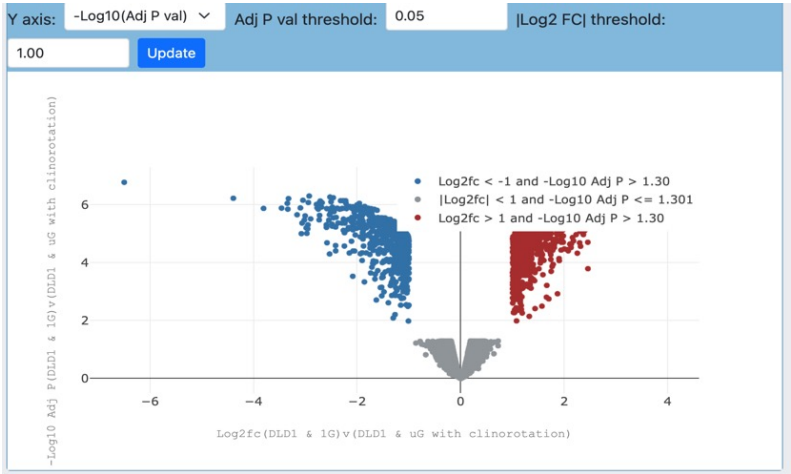
Request Process inspired by NIH's Database of Genotypes and Phenotypes (dbGaP)

1. Request button In OSDR
2. Data Request Package w/ Data Use Agreement
3. PI Verification
4. Review Board by Stakeholders
5. Approved/Denied
6. Secure Transfer
7. Access removed at end of IRB

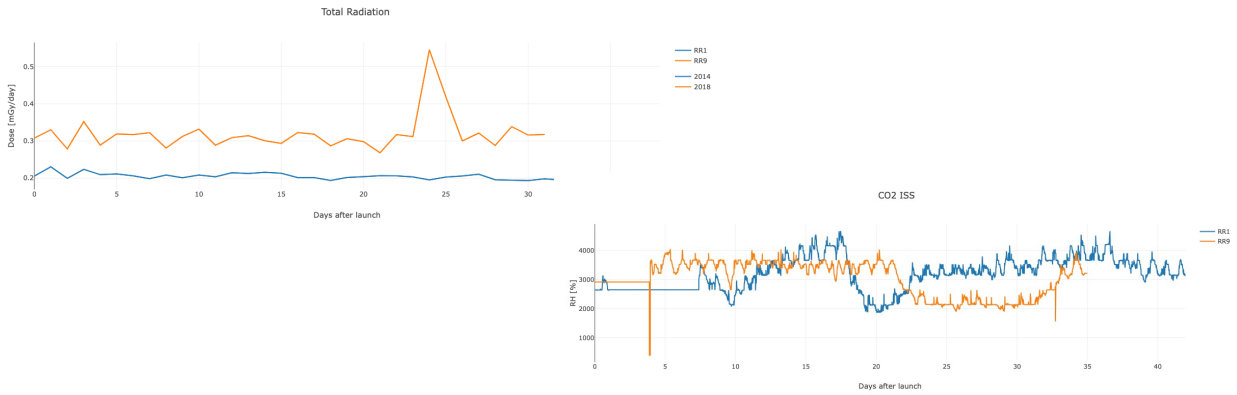
Visualization Applications

MULTI-STUDY VIZ

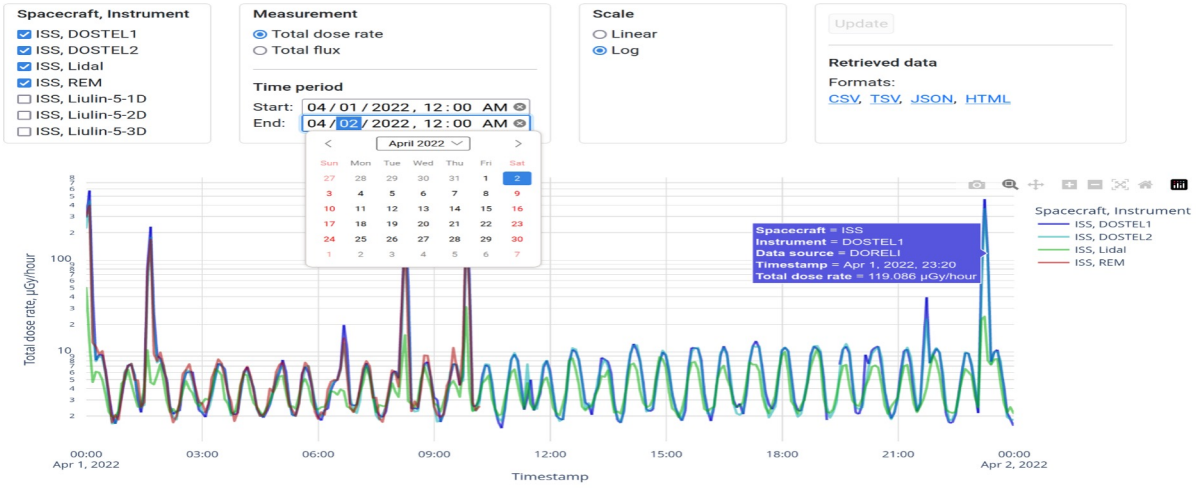




ENVIRONMENTAL DATA APP



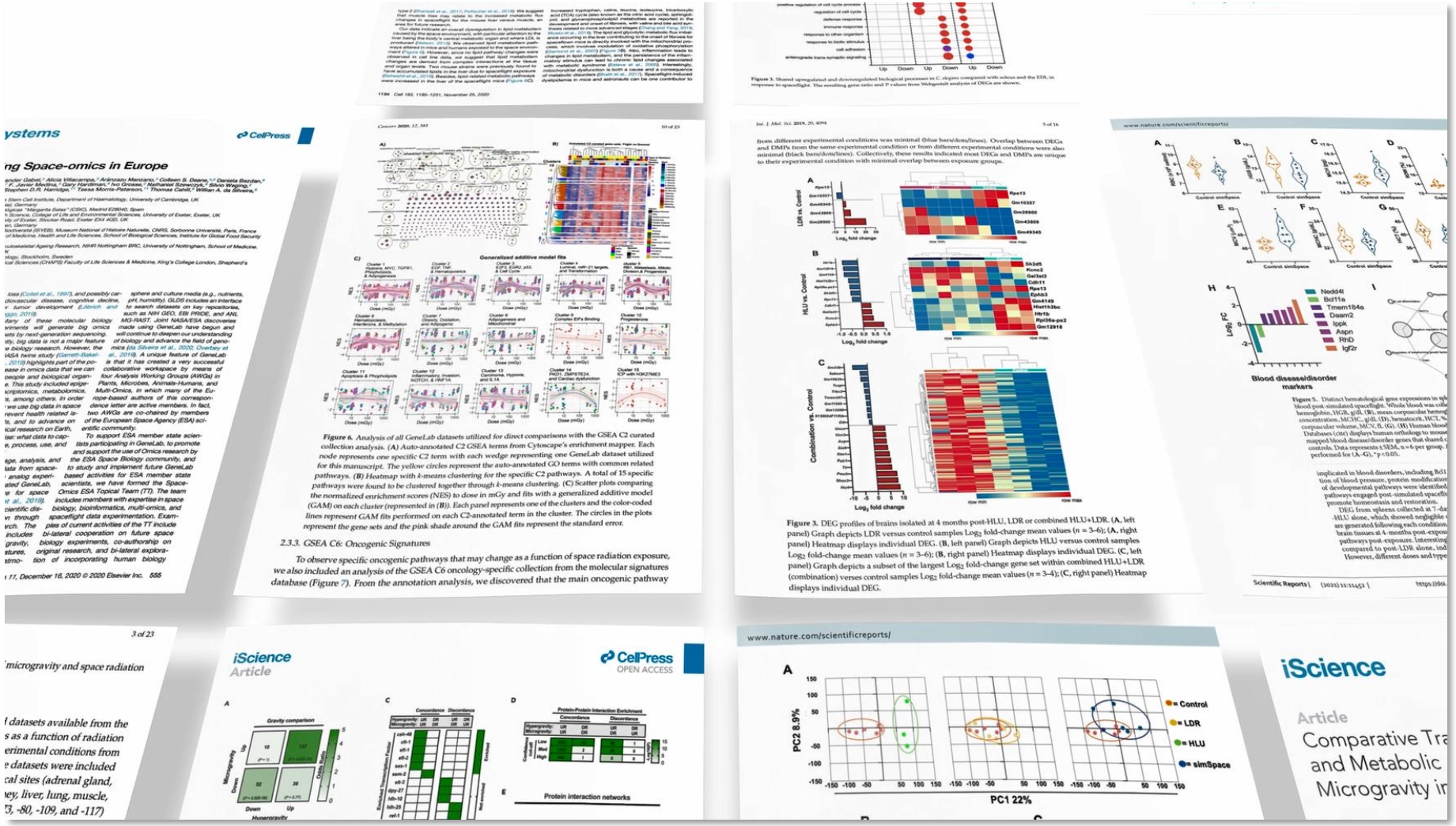
RADLAB



Re-use of Data and Enabling New Discoveries

60+ publications (12 publications produced by the AWGs) using data available in GeneLab.

A coordinated package of **29 scientific papers** published in **five Cell Press journals** featuring **9 papers** utilizing data or resources in GeneLab. *High impact research highlights.*



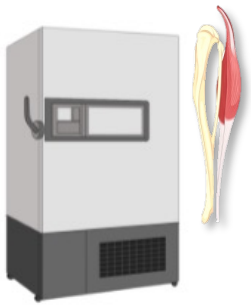
Maximizing Scientific Return

Primary Investigation

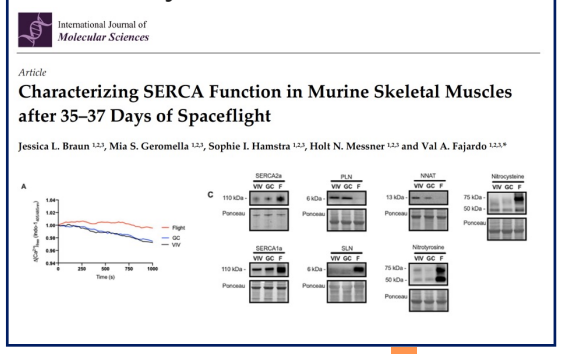
Dissections

Tissue Storage

Tissue Request



Analysis and Publication


International Journal of Molecular Sciences

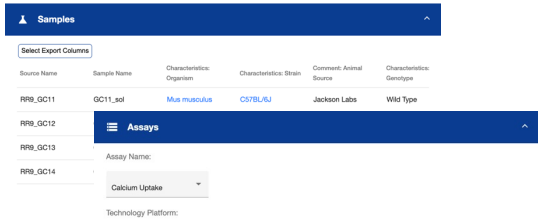
Article

Characterizing SERCA Function in Murine Skeletal Muscles after 35–37 Days of Spaceflight

Jessica L. Braun [1,2,3], Mia S. Geromella [1,2,3], Sophie I. Hamstra [1,2,3], Holt N. Messner [1,2,3] and Val A. Fajardo [1,2,3,*]

PI Results

Submitted & Curated Datasets



433.52 KB

OSD-488 Version 2

Characterizing SERCA Function in Murine Skeletal Muscles after 35-37 Days of Spaceflight from RR-1 and RR-9 (calcium uptake and Western blot in soleus and tibialis anterior)

Data Reused

Soleus (SOL) Muscle Data:

- Proteomics (quantitative mass spectrometry)
- Transcriptomics (bulk RNA-seq)
- Epigenomics (DNA methylation)
- Calcium Reuptake

Tibialis Anterior (TA) Muscle Data:

- Transcriptomics (bulk RNA-seq)
- Epigenomics (DNA methylation)
- Calcium Reuptake

- Combined omics & phenotypic data: RR-9, RR-1

GeneLab Datasets: GLDS-104, GLDS-105

ALSDA Datasets: GLDS-488

- Qlattice xAI (explainable AI) study showing which omics drive the phenotypic
- New results & publication forthcoming!

Kevin Li and Riya Desai, AI4LS Working Group

Open Science Analysis Working Groups

*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.*

ANIMAL

104 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

307 members

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

MICROBES

98 members

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

PLANTS

88 members

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

AI/ML

96 members

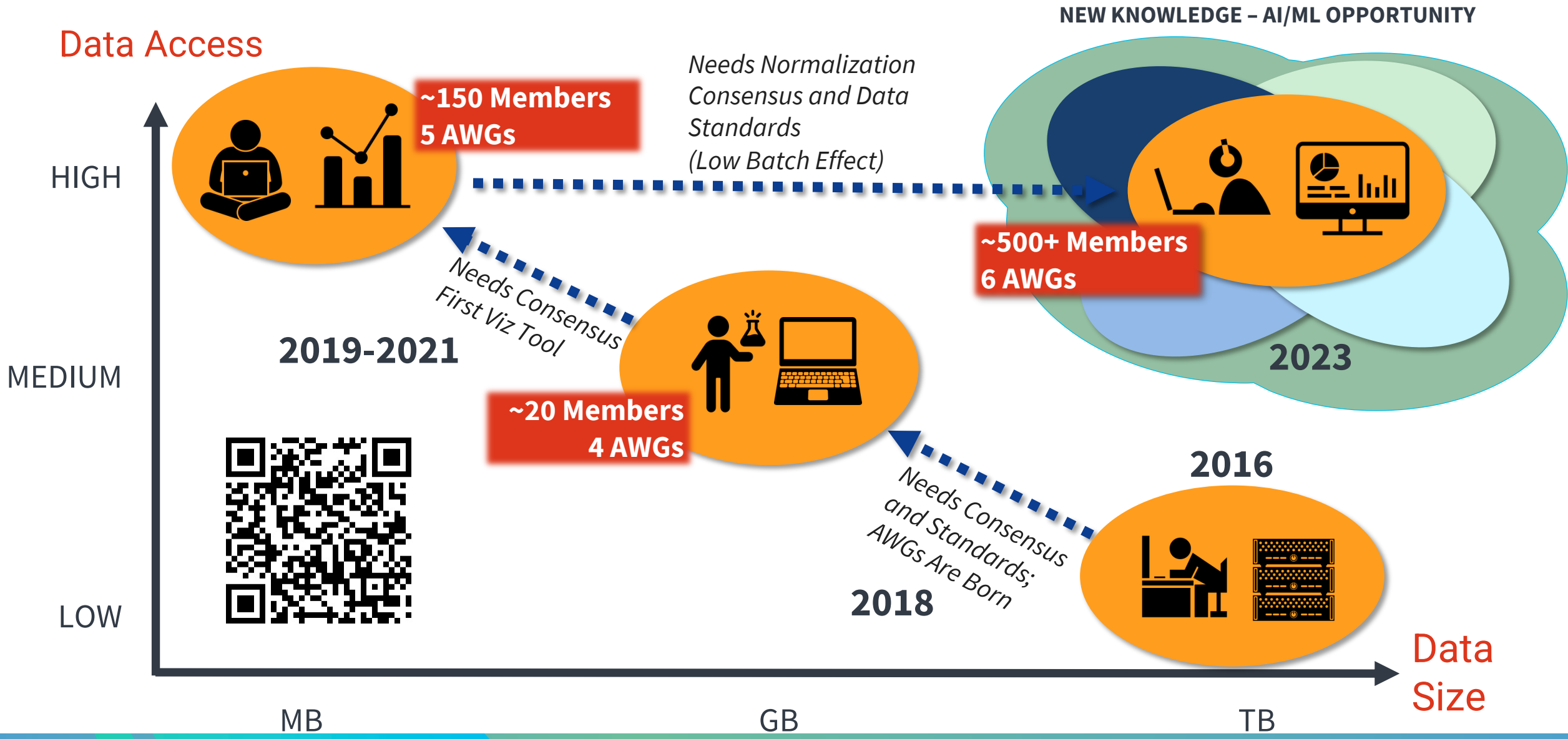
Focuses on developing data AI-readiness guidelines, algorithm and automation development, and developing ethical guidelines to increase trust and explainability surrounding AI in space biology.

ALSDA

138 members

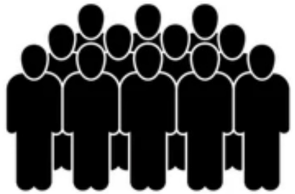
Feedback on science data and metadata standards for physiological, phenotypic, and behavioral datasets to be reusable. Datasets span from raw to processed-results data, and across tabular, bioimaging, and video formats.

A Brief History of the Open Science AWGs



Enabling our scientific community

GL4U Direct Approach
MSI/HBCU Students



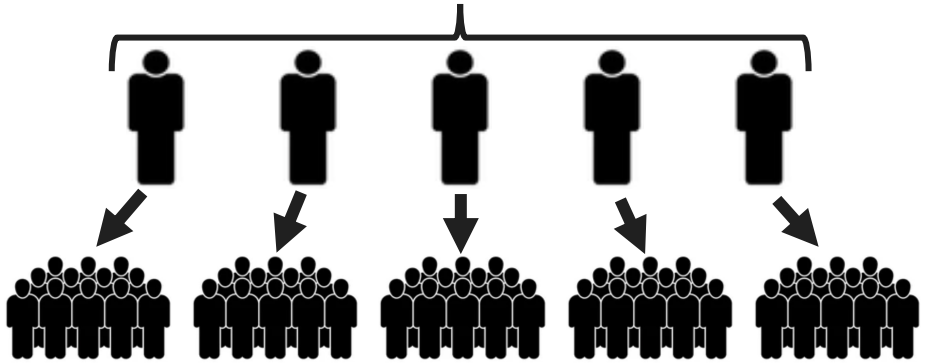
2021 – Space biology & RNAseq

- Virtual 1-week long bootcamp with 14 **SJSU** students
- Collaboration with USRA/SJSU; SJSU compute system
- Space biology- and RNAseq-specific lectures and hands-on instruction using Jupyter Notebooks (JNs)

2023 – Space biology & Amplicon Seq

- In-person 4-day bootcamp with 20-40 **CSULA** students
- Collaboration with JPL/CSULA; NSF ACCESS compute system
- Space biology- and Amplicon Seq-specific lectures and JNs

GL4U Indirect Approach
MSI/HBCU Educators

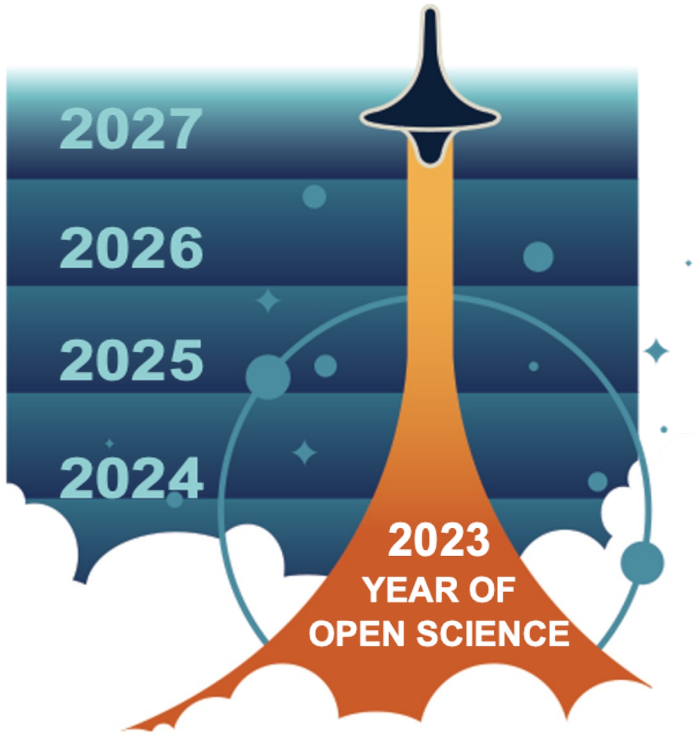


2022 – Space biology & RNAseq

- Virtual ~1.5-week long bootcamp
- Collaboration with JPL; SMCE compute system
- 6 professors and 4 graduate students from **4 HBCUs/MSIs**
- Space biology- and RNAseq-specific lectures and hands-on instruction using JNs
- Resources to teach at home institution

NASA’s 2023 Year of Open Science

NASA's Transform to Open Science (TOPS) is a $40 million 5-year mission to accelerate adoption of open science that kicks of with the NASA’s 2023 Year of Open Science

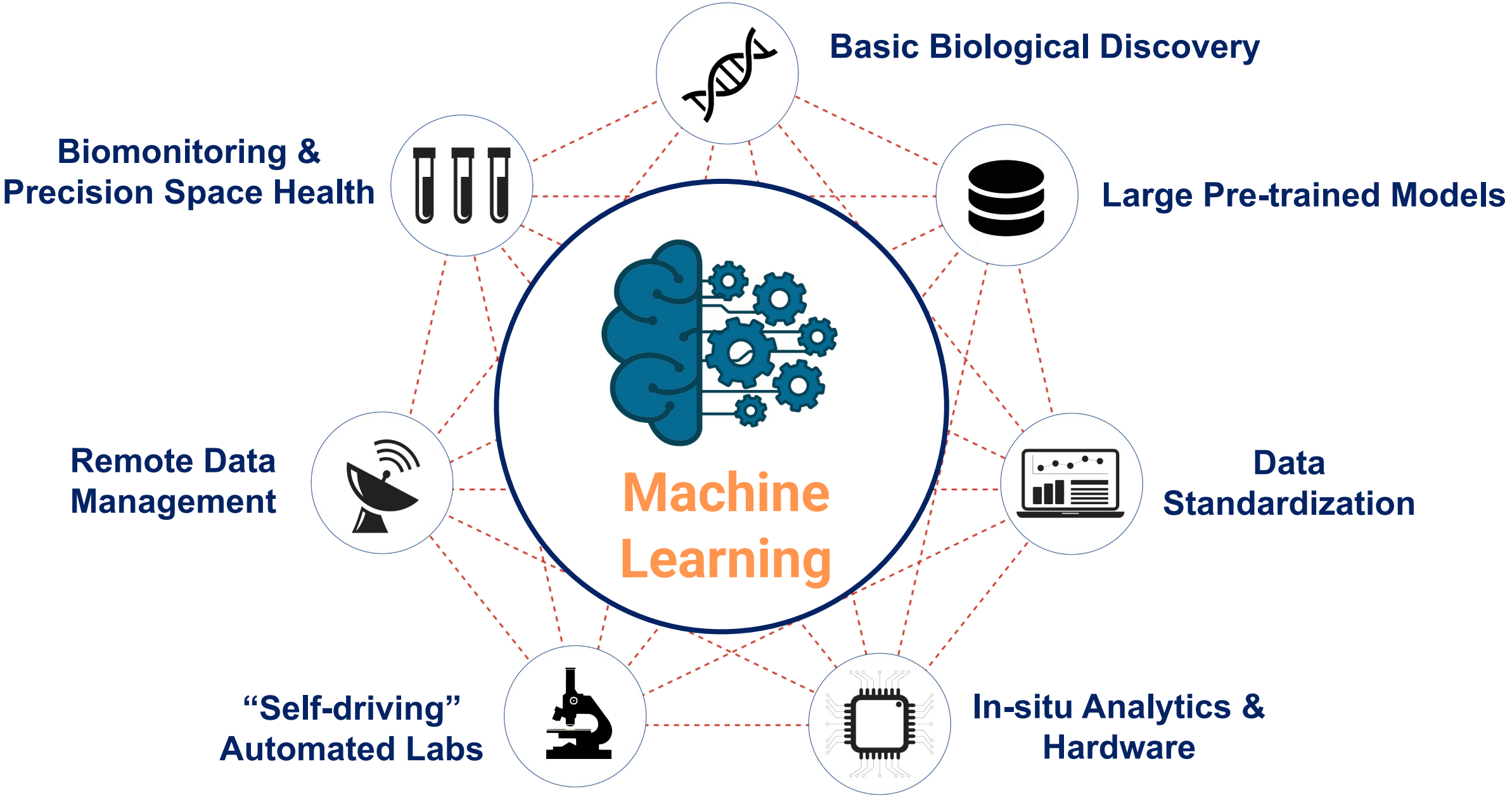


Objectives:

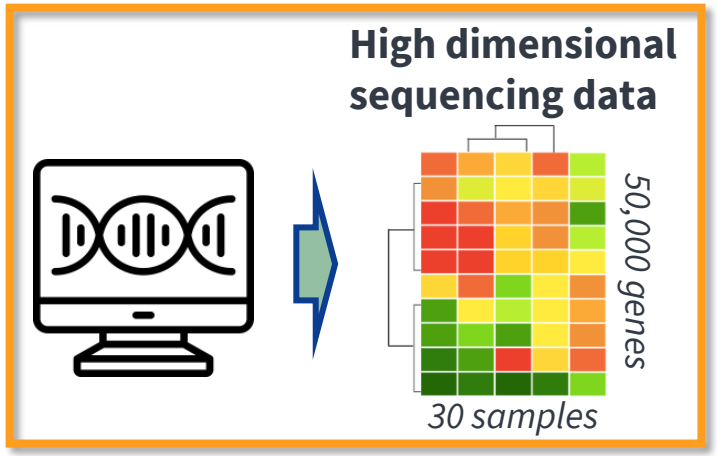
- Increase understanding and adoption of open science
- Accelerate major scientific discoveries
- Broaden participation by historically underrepresented communities

TOPS Email List

TOPS Website


Basic Biological Discovery
Large Pre-trained Models
Data Standardization
In-situ Analytics & Hardware
"Self-driving" Automated Labs
Remote Data Management
Biomonitoring & Precision Space Health
Machine Learning

Space biological data analysis challenges



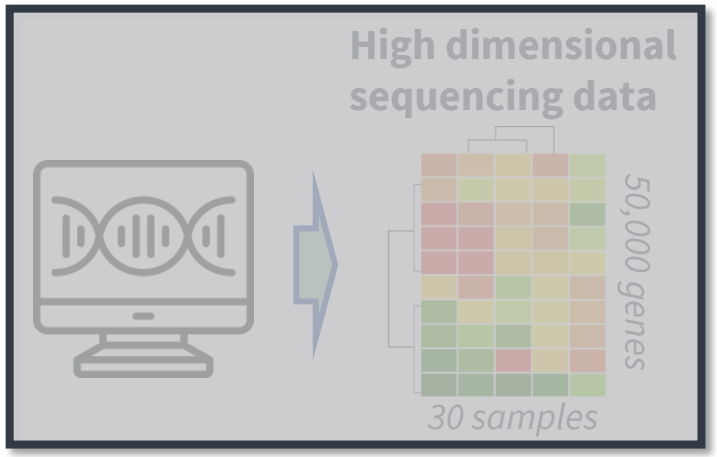
Space Biological Data Challenges

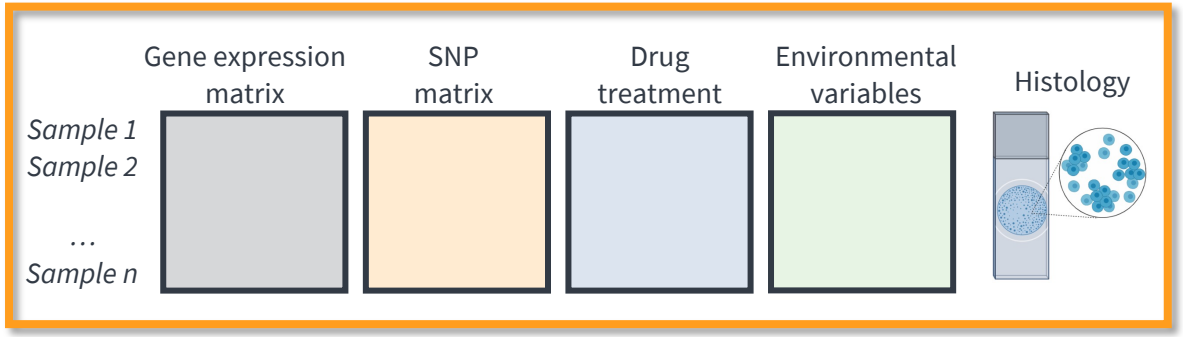
- **Small sample *n***
- **High feature count**
- Heterogeneous data
- Sparse data
- Transfer from model to human

Space biological data analysis challenges

Space Biological Data Challenges

- Small sample *n*
- High feature count
- **Heterogeneous data**
- **Sparse data**
- Transfer from model to human

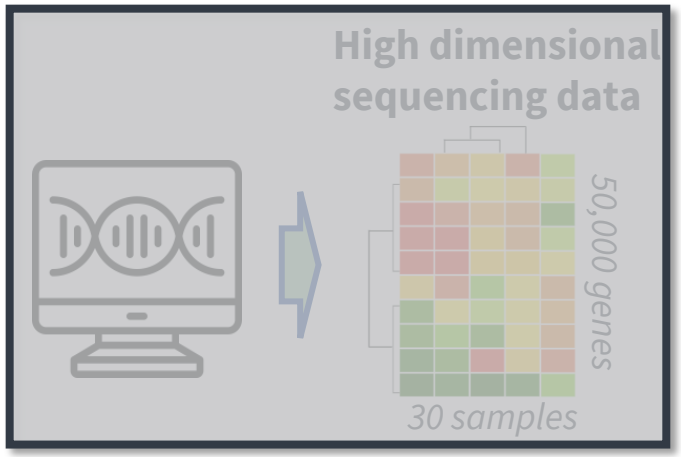


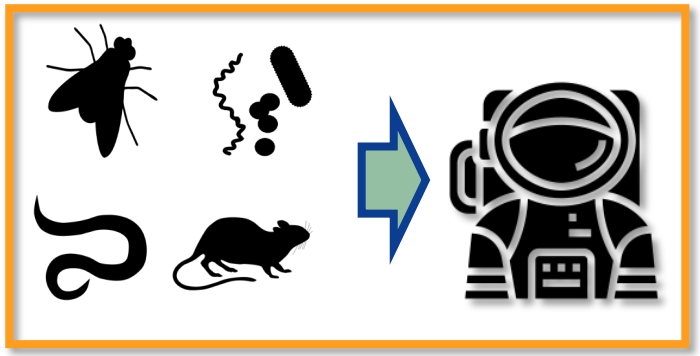


Space biological data analysis challenges

Space Biological Data Challenges

- Small sample *n*
- High feature count
- Heterogeneous data
- Sparse data
- **Transfer from model to human**

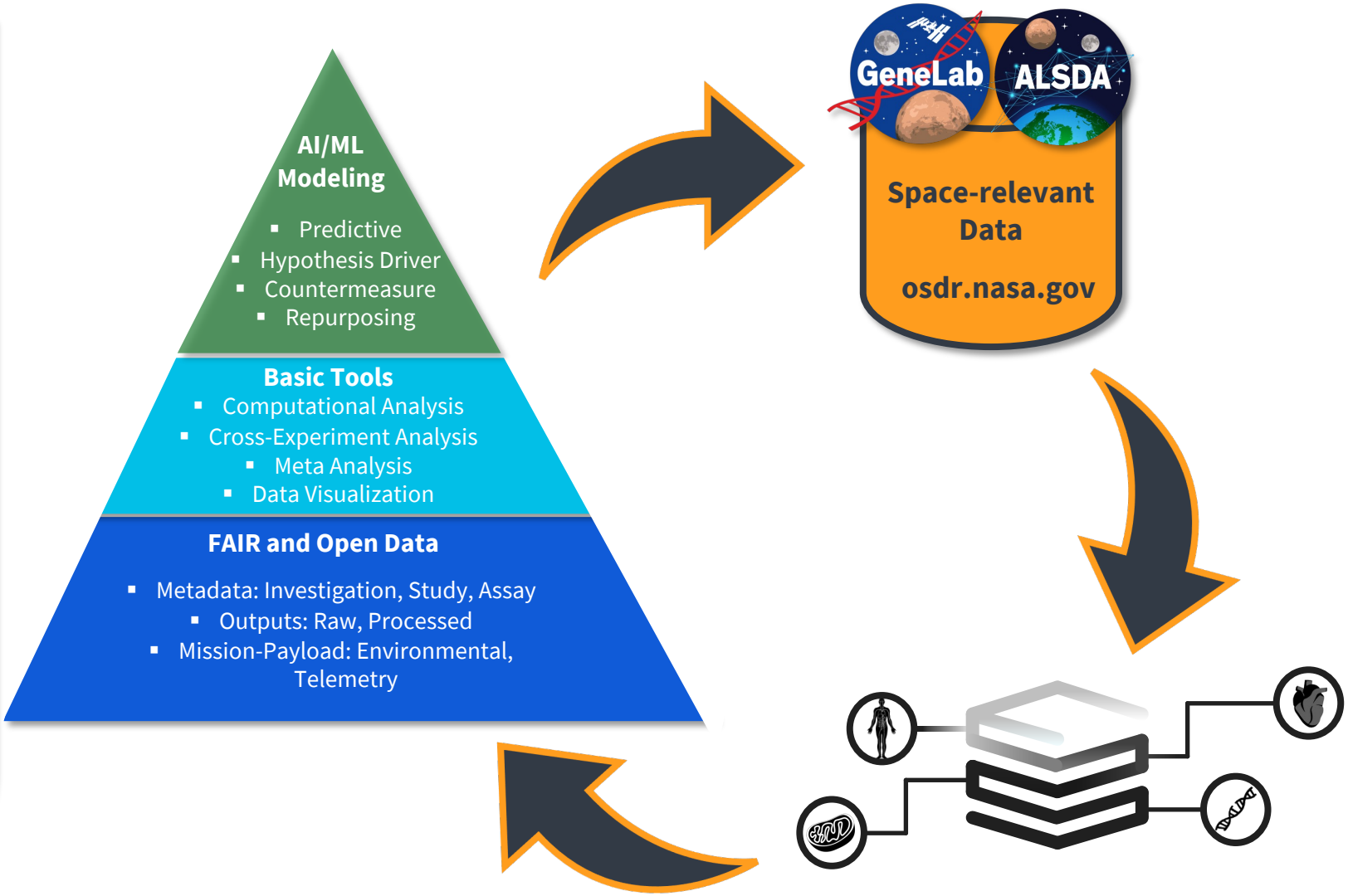




AI for Life in Space working group: AI4LS

Leveraging ML and AI methods to model space biology data from the **NASA Open Science Data Repository** to better understand the complex effects of spaceflight on living systems across hierarchical biological levels.

AI4LS



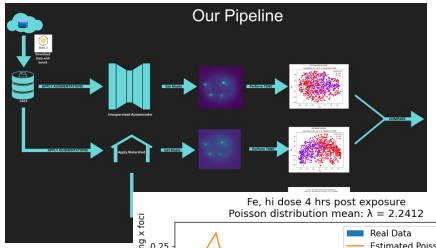
Biological benchmarking datasets

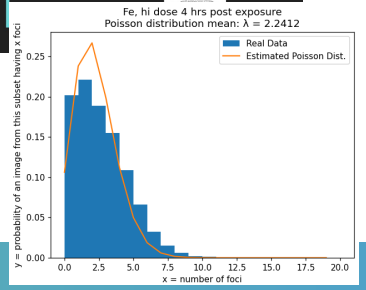
- NASA Science Mission Directorate effort
- Each Division provides ML-ready datasets to benchmark new methods
- Common data types in each domain
- Biological and Physical Sciences provided 2 datasets:
  - RNA sequencing
  - Fluorescence microscopy
- Publicly available on Amazon Registry of Open Data

RNAseq Dataset

Microscopy Dataset

- BPS Microscopy benchmark dataset formed the basis for **UC Irvine CS175** senior ML projects
- Real-world scientific data inspired the students
- ML-ready dataset allowed students to spend time on ML rather than preprocessing





AI/ML Architecture to Support Deep Space Mission Goals

nature machine intelligence

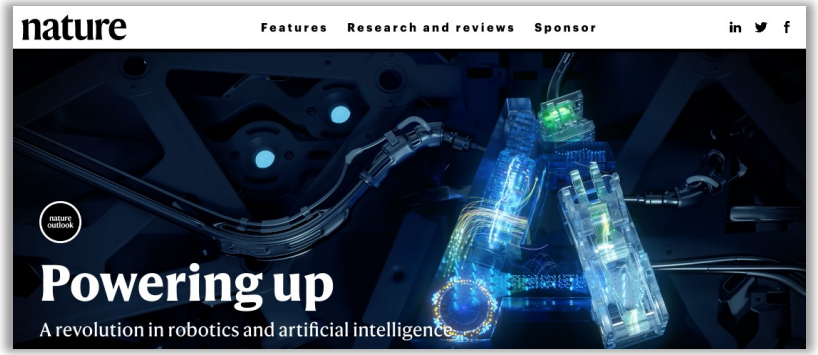
Review article

https://doi.org/10.1038/s42256-023-00617-5

Biomonitoring and precision health in deep space supported by artificial intelligence

Received: 23 December 2021

Ryan T. Scott[1,52], Lauren M. Sanders[2,52], Erik L. Antonser



Scott et al.

Sanders et al.

nature machine intelligence

Review article

https://doi.org/10.1038/s42256-023-00618-4

Biological research and self-driving labs in deep space supported by artificial intelligence

Received: 23 December 2021

Lauren M. Sanders[1,52], Ryan T. Scott[2,52], Jason H. Yang[3], Amina Ann Qutub[4],

Stanford University · SJSU San José State University · Anschutz · The University of Texas UTSA at San Antonio · UCSF · Berkeley Lab Lawrence Berkeley National Laboratory · Mayo Clinic · Baylor College of Medicine · Icahn School of Medicine at Mount Sinai · Salk Institute for Biological Studies · Medical School University of Minnesota Driven to Discover · Johns Hopkins University · Google · Rutgers New Jersey Medical School · Harvard University · KBR · Michigan State University · NASA · Caltech · UNF University of North Florida · The University of Alabama at Birmingham · Weill Cornell Medicine

Scott et al., *Nature Machine Intelligence* 2023
Sanders et al., *Nature Machine Intelligence* 2023

THANK YOU!

osdr.nasa.gov/bio

GeneLab is funded by the NASA Space Biology Program within the NASA Science Mission Directorate's (SMD) Biological and Physical Sciences (BPS) Division.

ALSDA is funded by the NASA Space Biology Program within the NASA Science Mission Directorate's (SMD) Biological and Physical Sciences (BPS) Division and NASA Human Research Program (HRP)

Acknowledgements

Sylvain Costes
Sam Gebre
Lauren Sanders
Ryan Scott
GeneLab Team
ALSDA Team

Questions: Danielle.k.lopez@nasa.gov
Contact for AI4LS: lauren.m.sanders@nasa.gov

genelab.nasa.gov

@NASAGeneLab

facebook.com/NASA.OSDR