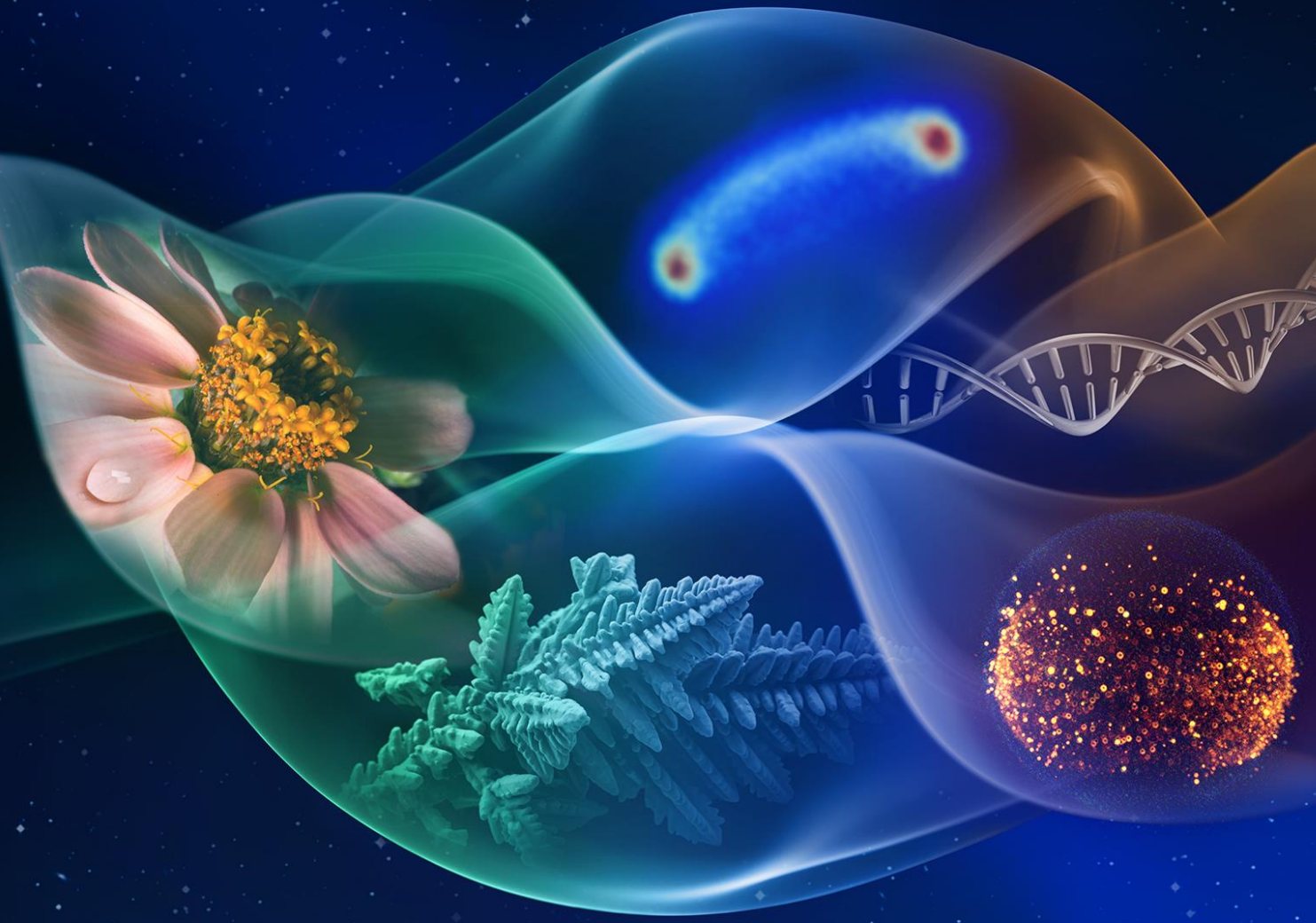
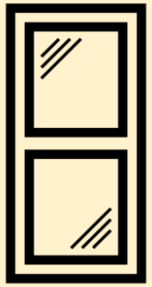
BPAC
Open Science



Sam Gebre
GeneLab Deputy Project Manager
NBISC Project Manager

Open Science

Building **trust in the scientific process** through
transparency, accessibility, inclusivity, and reproducibility

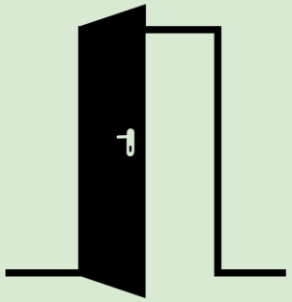


Open (**Transparent**) Science

Both the scientific process and results should be visible, accessible and understandable.

Open (**Accessible**) Science

Data, tools, software, documentation, publications should be accessible to all (FAIR).



Open (**Inclusive**) Science

The process and participants should welcome participation by and collaboration with diverse people and organizations.

Created by Gregor Cresnar
from Noun Project

Open (**Reproducible**) Science

The scientific process and results should be open such that they are reproducible by members of the community.

Open Science at BPS

- BPS recognizes the importance of ***collecting tissues*** **and** ***archiving data, metadata, computational tools***, and ***samples from both spaceflight and ground studies*** to enable Open Science and future experiments.

- Strive to help scientists discover and access datasets to perform primary, secondary, and meta-analyses.

- Open Science Projects strive to implement the "FAIR" principles to ensure all data are: **Findable Accessible Interoperable Reusable.**

- Increasing opportunities for collaboration while promoting **scientific innovation**, **transparency**, and **reproducibility.**

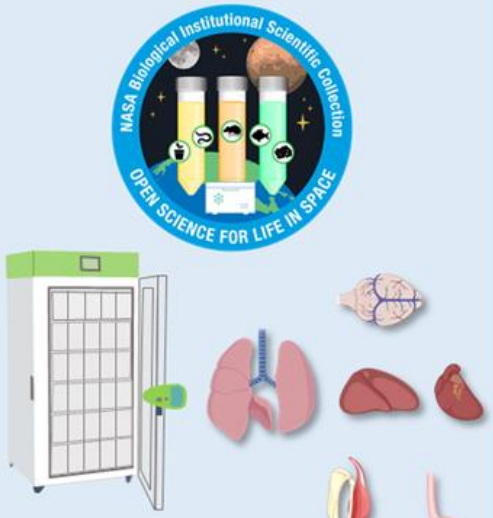
- With the suite of open BPS repositories, scientists can use existing datasets to make new discoveries, propose future investigations, or influence research trends.

Open Science Projects

Physical Sciences Informatics System

Hosts µG Physical Science Investigations

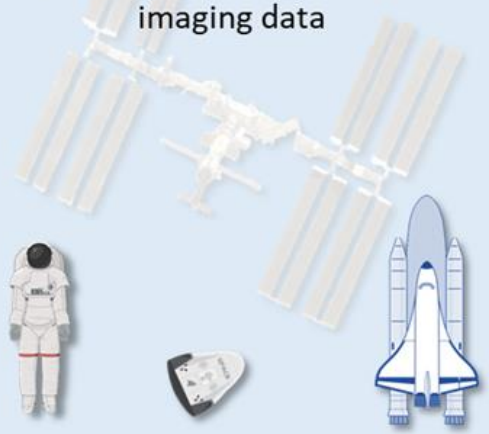
NASA Biological Institutional Scientific Collection (NBISC)



Collection of biospecimens from non-human subjects exposed to spaceflight and associated ground studies

Ames Life Sciences Data Archive (ALSDA)

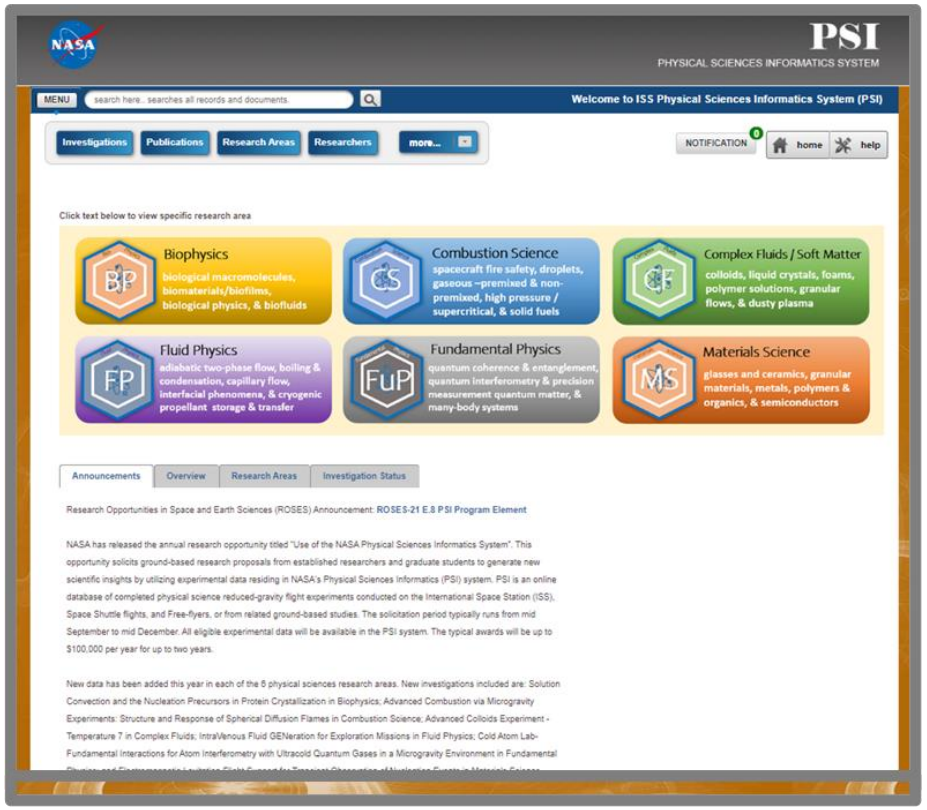
Collects and *curates* physiological, mission, and imaging data



GeneLab (GL)

Collects and curates omics data from space-relevant samples

Physical Sciences Informatics System (PSI)



https://www.nasa.gov/psi

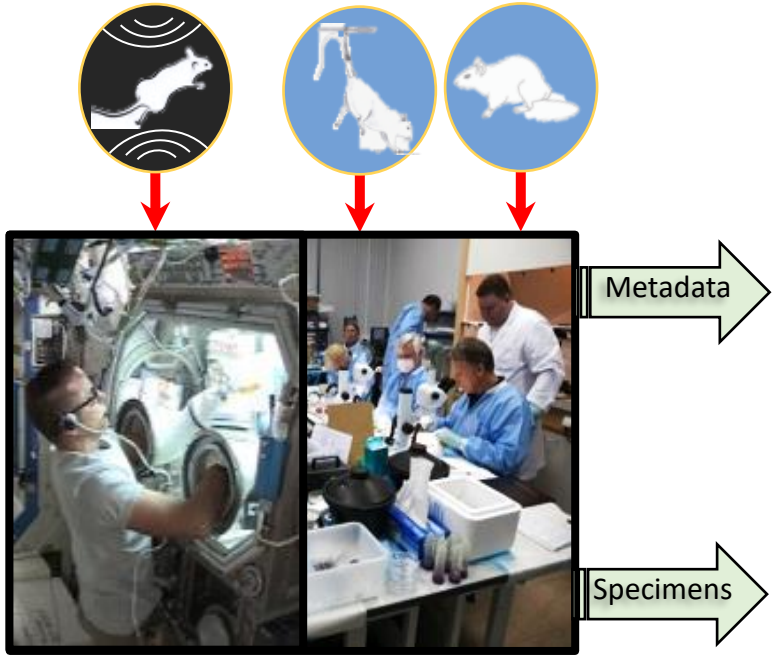
87 Investigations in 6 Research Areas

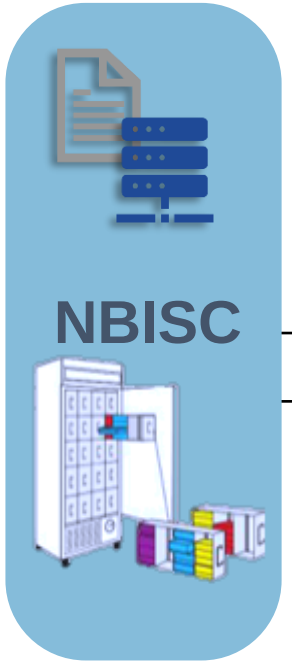
>22 TB data

1,557 users, primarily in academia

NASA Biological Institutional Scientific Collection (NBISC)

Spaceflight Ground Analogs





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
- 62 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.

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

Biological Data Repositories

Collects, curate, and distributes data from NASA funded non-human life sciences investigations

Data Collections:

Mission
Payload
Operations
Vehicle
Hardware
Experiments
Personnel
Subjects
Biospecimens

Established in 1994

859 Experiments

- 1975 to 2022
- Pubications

356 Datasets

- Associated data

Coming soon Studies

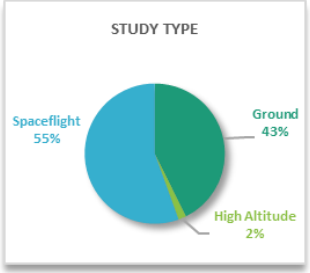
15 TB 300 cf Volume

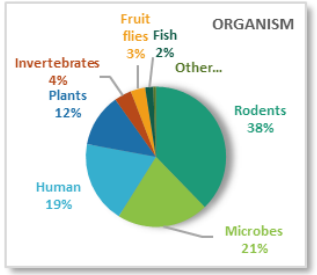
- PI Data: 1.3 TB
- Mission Data: 2.7 TB
- Images/Video: 11 TB

396 Studies

447 Datasets

45 Species





>10 Assays

>135 TB Volume

GeneLab, an open science **multi-omics** repository, covering:

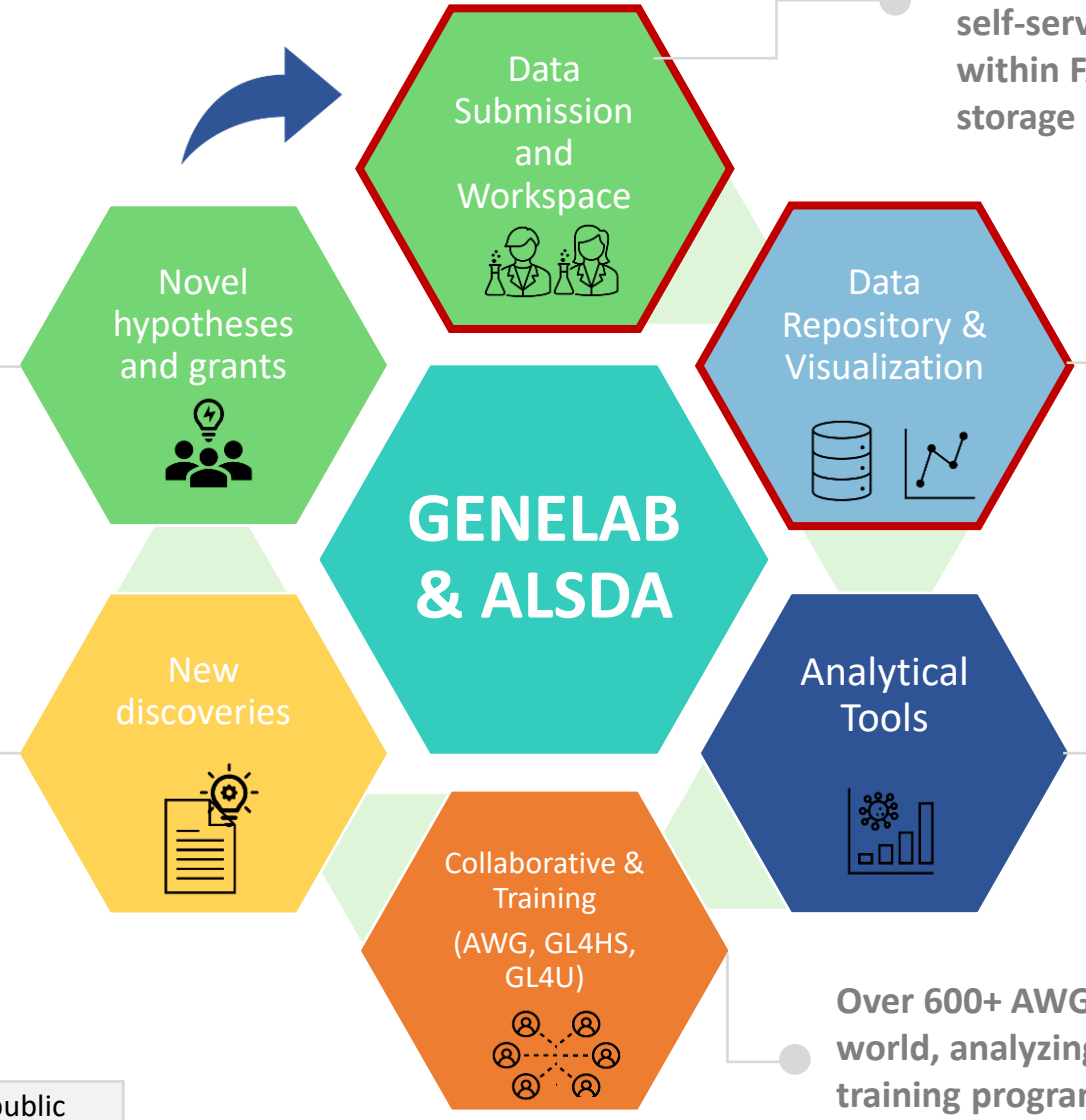
Transcriptomics
Metagenomics Epigenomics
Proteomics
Metabolomics

Studies comprise of data from **model organisms** including:

Microbes
Plants
Fruit flies
Rodents
Nematodes
Humans

Established in 2014

Biological Data Repository Lifecycle



Multi-project web-based submission portal to support self-service metadata curation and data submission within FAIR Guidelines. Workspace providing users file storage in S3 and sharing features.

Open access and well curated data repositories including data from various model organisms from microbes to plants to humans covering the full range of biological assay technologies. Interactive visualization portal to enable knowledge discovery.

Toolshed of bioinformatics tools with user-friendly interface for data analysis at any level (students, citizen scientists, PIs).

Over 600+ AWG members from around the world, analyzing open science data. Successful training programs developed for high school and university students.

Open access data enables discovery of new hypotheses and new ideas for grant proposal. Data from those new research experiments are generated and deposited back into the data repositories.

Enabling new publications and research through *data reuse* and performing meta-analysis.

GeneLab & ALSDA Systems Integration, public release Nov 2022

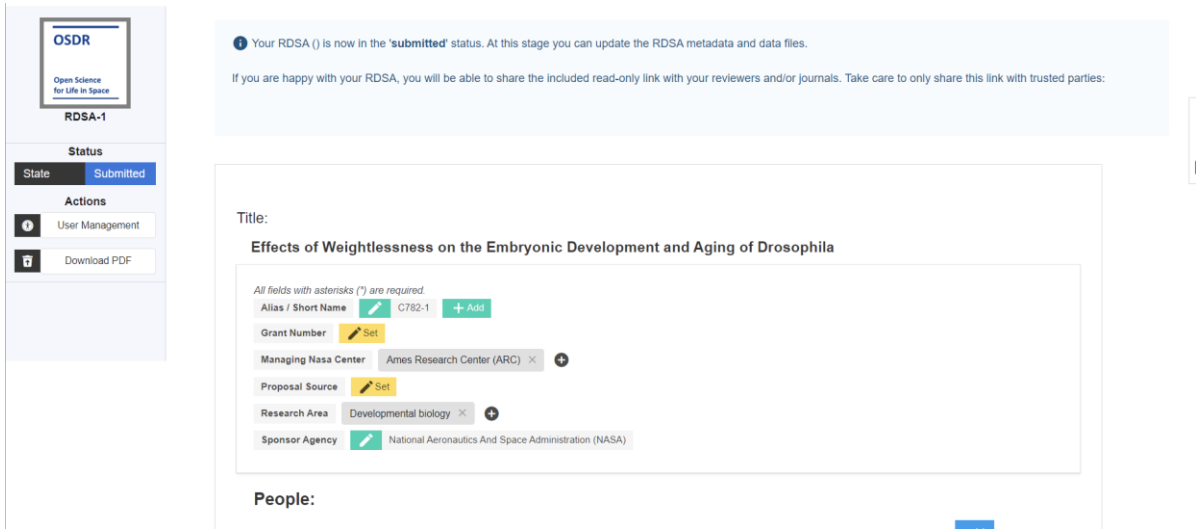
BPS Open-Source Science Initiatives - Access

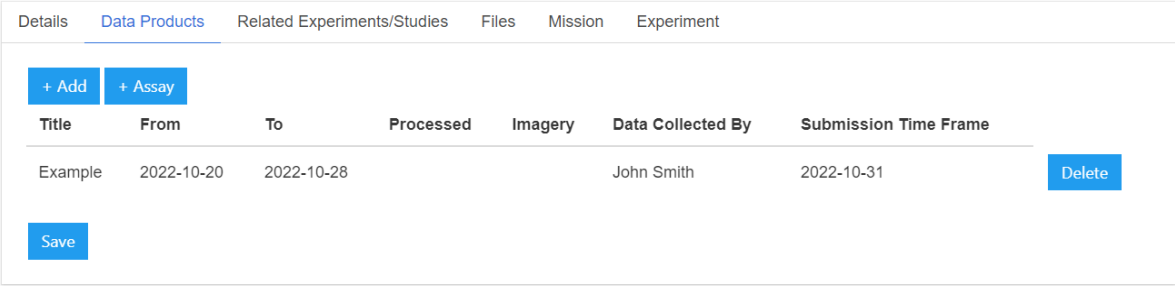
Increasing accessibility of data and knowledge

- Integrating biological data repositories to enable data accessibility, interoperability, and reusability. New centralized data systems coming November 2022.
- Fostering community, collaboration, and exchanging of ideas through Analysis Working Groups and User Communities.
- Incorporating BPS data to Open AWS Registry to increase ease and accessibility to data through open AWS S3 bucket.
- Adhering to the new SMD SPD-41 policy, BPS is developing a data and information policy to increase scientific research throughput and transparency of government funded research.

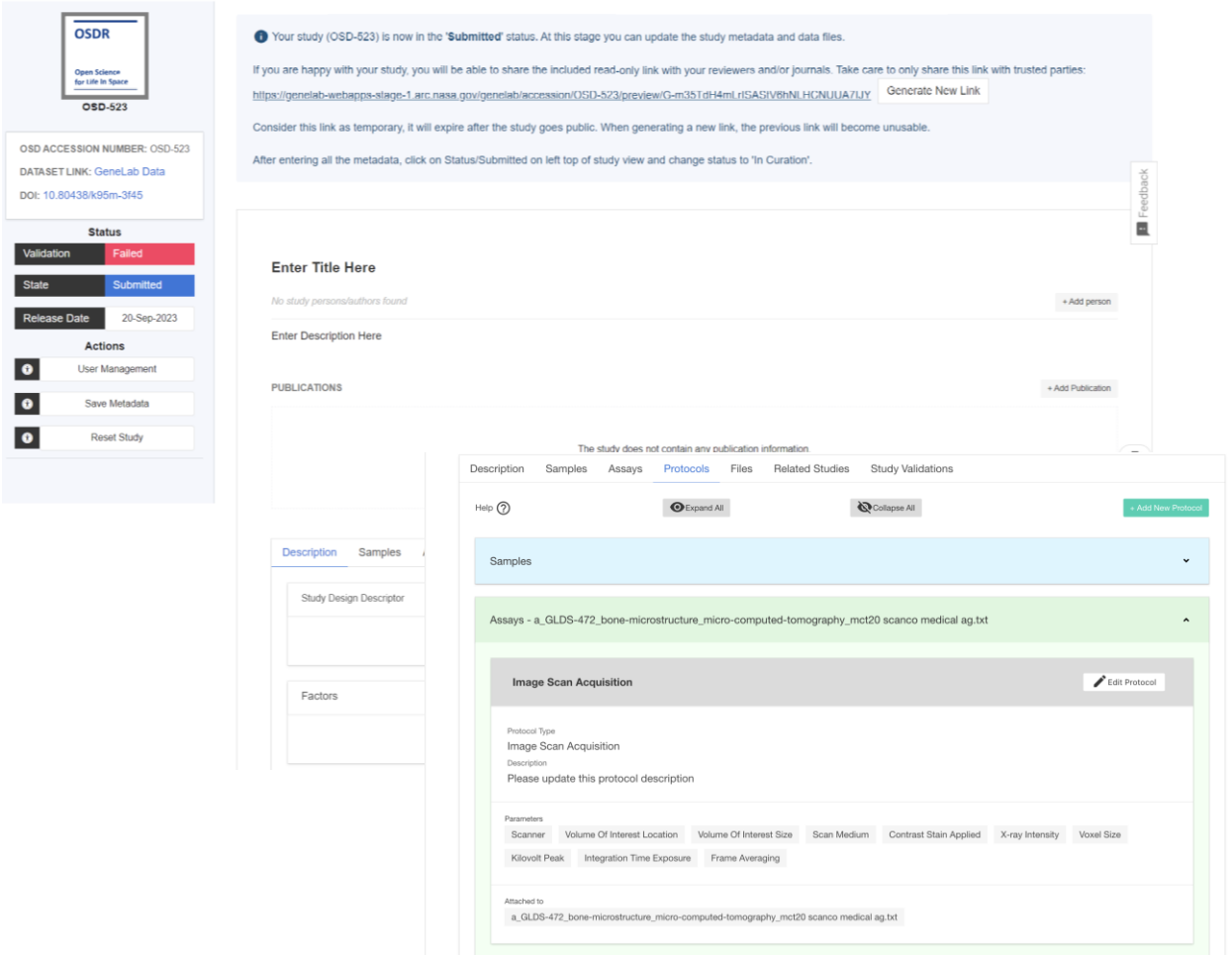
Integrated Data Systems - Submission

Research Data Submission Agreement User Interface - RDSA form is used to define the data, assays and/or tissues that will be transferred and timeframes for submittal.

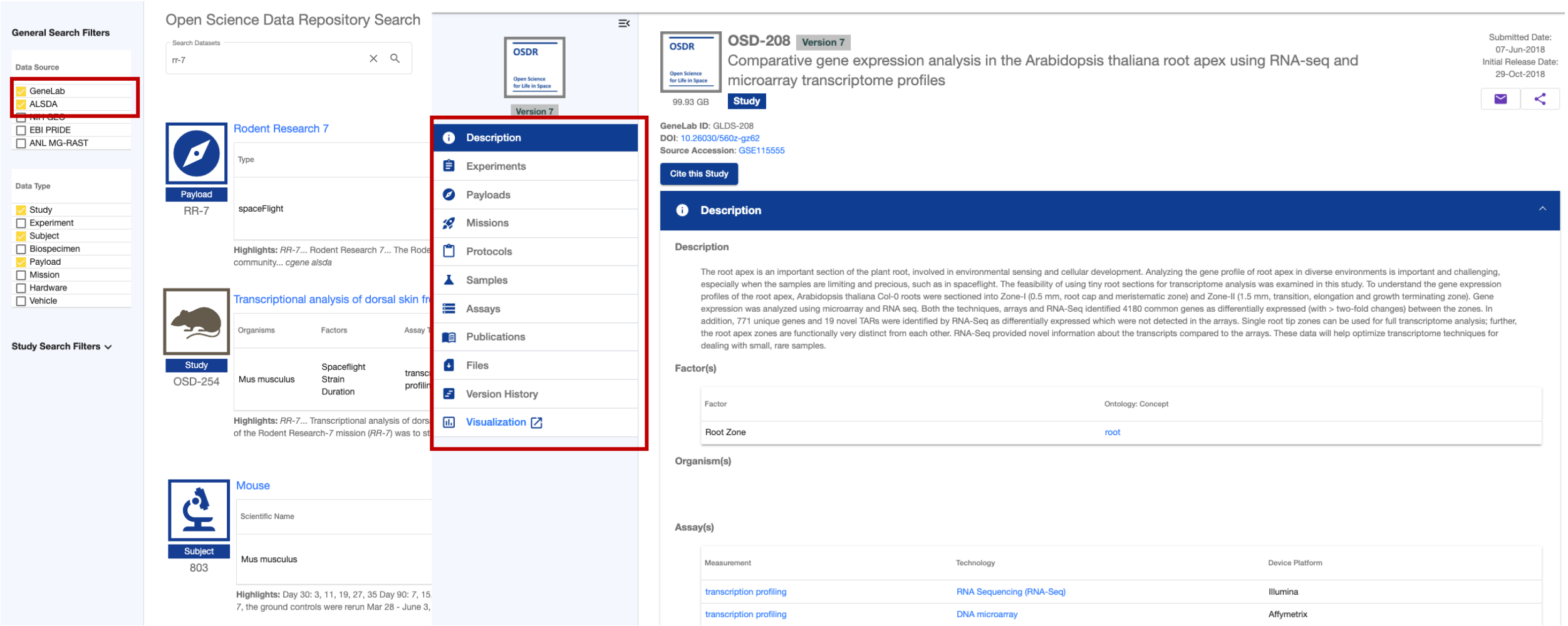




Single Submission Portal for GeneLab and ALSDA – Easy to use web portal to enter and publish research data (omics and phenotypic data).



Integrated Data Systems - Search



Going beyond the data.. connecting the scientific community

In January 2018, we kicked-off the **Analysis Working Groups (AWGs)** with over 200 scientists from multiple space agencies, international institutions, and industry.

To date, we have **600+ scientists** that meet monthly with each group to provide feedback, develop standards, and analyze data.

ANIMAL

79 members

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

PLANTS

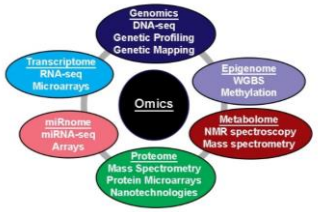
66 members

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

MULTI-OMICS

259 members

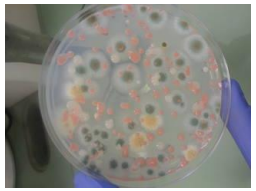
Interactions between the different omics to provide a better understanding of the systemic response.



MICROBES

61 members

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



ALSDA

110 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.

AI/ML

56 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.

BPS data to Open AWS Registry – *coming soon*

Open access directly from the cloud!

- High value research data from Space Biology and Physical Sciences Programs
- Simplify access to data by PIs and their teams in academia while under analysis, through a public S3 bucket
- Lower cost of duplication of data storage
- Reduce download times for large datasets
- Enable analysis on AWS compute nodes without any data downloads

Registry of Open Data on AWS

NASA Space Act Agreement

Amazon Web Services and the National Aeronautics and Space Administration (NASA) have entered into a Space Act Agreement to explore best practices around discovery, access, and use of high-value NASA science datasets. Making analytics-optimized data stores available to the science community will minimize the need for data wrangling and preprocessing within the community, leading to a faster time to insight and quicker innovation.

New BPS Data Policy

To increase scientific research throughput and the transparency of government funded research, NASA and the Science Mission Directorate (SMD) has issued new policies as described in the Scientific Information Policy.

To adhere to the new guidance, we recently released the **BPS Science Data Management Policy**, which defines policies and provides guidelines for managing scientific data by its programs, projects, investigators, and repositories.

Download Policy here!

Key highlights:

- The policy apply to scientific information from all BPS-funded activities including **Research Data, Operations Data, and Software.**
- Research data shall become publicly available in BPS designated OSDR no later than the publication of the investigation results.
- Research software developed using BPS funding and used in support of a scientific, peer-reviewed publication shall be released as open-source software no later than the publication date.
- Operations data: BPS shall commit to full and open sharing of information produced by BPS Mission Projects. This includes environmental data from flight and ground-control analogs, animal husbandry data from flight and ground, and logs of flight operations activities.

Event: BPS Townhall, coming in mid-December

BPS Data Policy – For Researchers

Guidelines for BPS Funded Investigators:

1. Complete a Research Data Submission Agreement (RDSA).
2. Must have a PI unique identifier, for example ORCID
3. Are urged to submit data throughout grant.
4. Submit all data to designated BPS Open Science Data Repository (OSDR)
5. All scientific data supporting any publication must be submitted to OSDR at time of publication.
6. Any data not used to support publication must be submitted by end of period of performance (grant). No-costs extension on grants may be requested.
7. Funded research software to support published results must be reported and released as open-source.
8. Investigators sponsored with BPS funding to attend science events open to the public, including but not limited to conferences, workshops, and symposia shall publish any public presentations through a designated NASA repository.

BPS Open-Source Science Initiatives - Knowledge

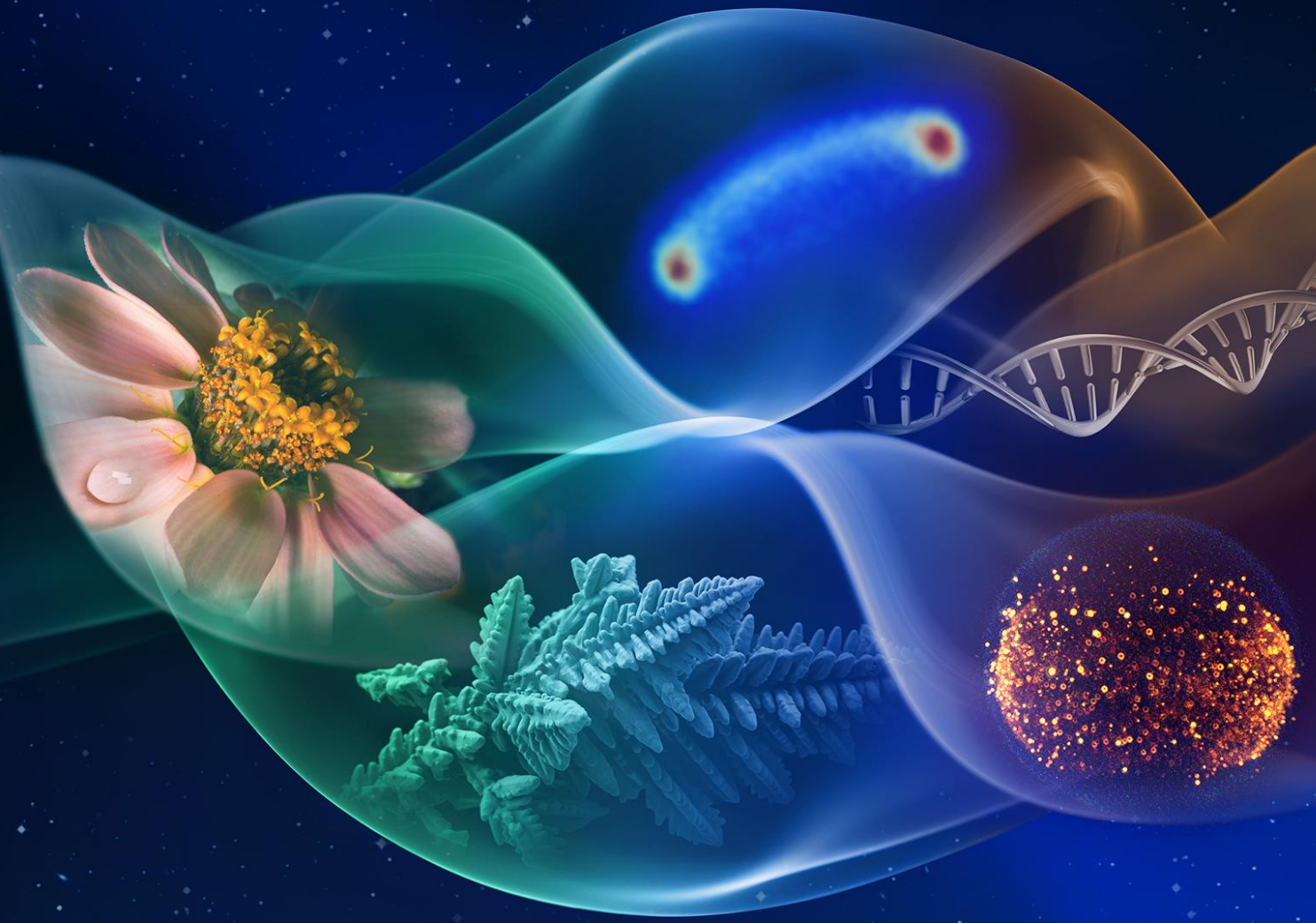
Training the next-generation of scientists

- GeneLab for High School (GL4HS) is a summer training program providing students an opportunity to immerse themselves in space life sciences with a specific focus on omics-based bioinformatics research.
- GeneLab for Universities (GL4U) was able to conduct a bioinformatics bootcamp for educators from 4 HBCUs and MSIs institutions. NASA scientific compute resources will be available for each educator to train their students through the 2022-2023 school year.

Enabling AI/ML Activities for BPS

- Creation of benchmark Space Biology datasets to train AI/ML algorithms
- Adapting AI/ML tools from other disciplines for biological data

SMD Open Science Opportunities



Leading the Path to Open-Source Science

NASA's Transform to Open Science (TOPS) is a $40 million* 5-year NASA Science Mission Directorate mission

Goals:

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

Metrics:

- ★ 20K earn Open Science Badge
- ★ 5+ major discoveries
- ★ Increase participation of underrepresented groups by 2x

A NASA OPEN-SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

New Funding Opportunity!

F.14 Transform to OPen Science Training (TOPST) solicits proposals to advance open science literacy. This element will support three types of activities:

1. development of **ScienceCore materials**, discipline specific curricula modules around NASA Science Mission Directorate (SMD) divisions,
2. implementation of **Summer Schools** for SMD science teams, and
3. **Virtual Cohorts** participating in OpenCore curricula training.

Dec 8, 2022 – Proposal Due

Apply Today!

THANK YOU!

Open Science Projects are funded by the Biological and Physical Sciences Division.

https://www.nasa.gov/PSI

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

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

Twitter: @NASAGeneLab

https://genelab.nasa.gov