NASA’s Biological and Physical Sciences Division Enable Open Science

National Aeronautics and Space Administration

AGU 12/13/2024

A Vision for What's Next: NASA’s Evolving Data, Software, and Science

Amanda Saravia-Butler, Ph.D.

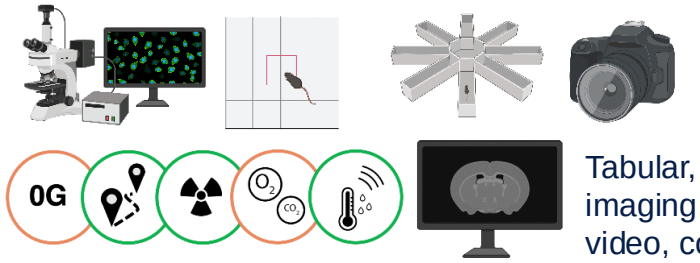
NASA GeneLab Science Lead

Contractor: Amentum

Biological & **P**hysical **S**ciences

NASA Open Science Data Repository Overview

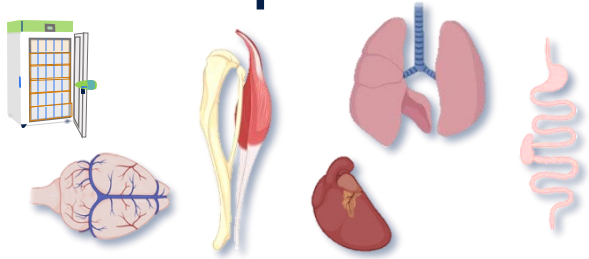
Physiological/Phenotypic/Imaging/ Environmental Telemetry Data



Tabular, text, imaging, video, code

Molecular/Omics Data

Biospecimens



ALSDA

GeneLab

NBISC

NASA Open Science Data Repository (OSDR)

nasa.gov/osdr/

- **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
  - Transparency, Inclusivity, and Reproducibility

NASA OSDR – By The Numbers

522 Studies

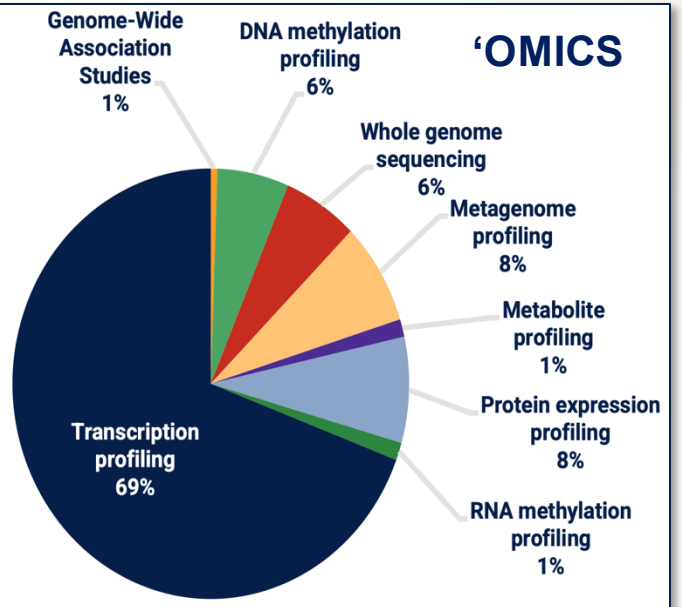
973 Datasets

45 Species

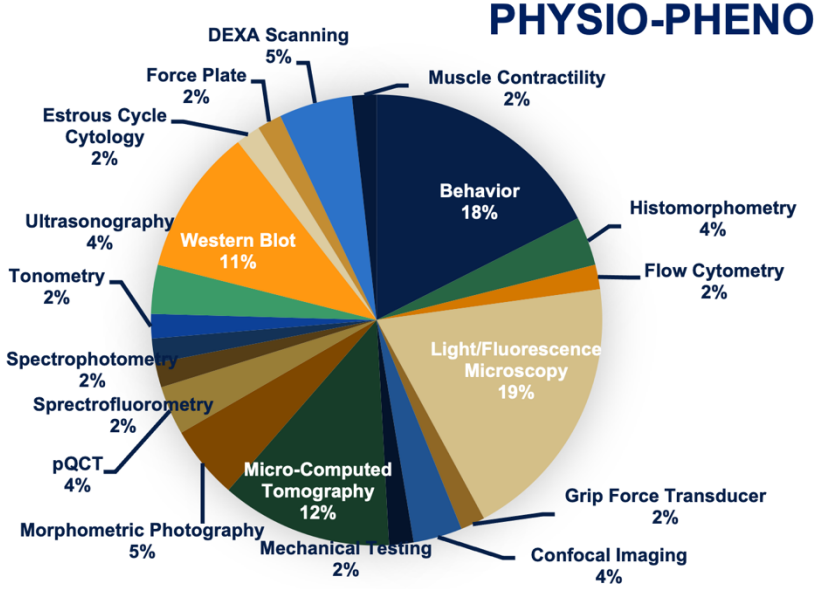
>80 Assays

>210TB Data

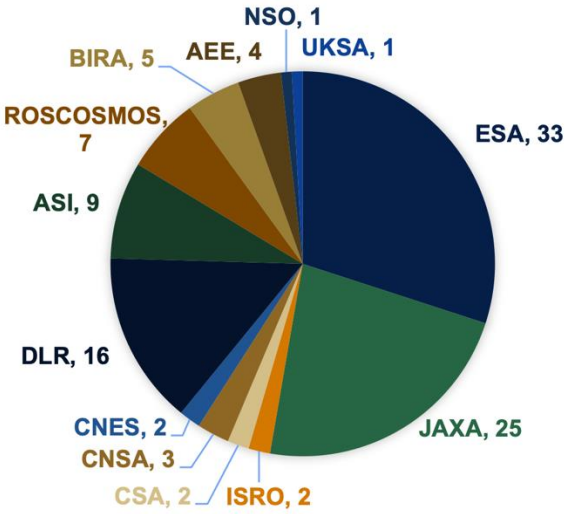
'OMICS



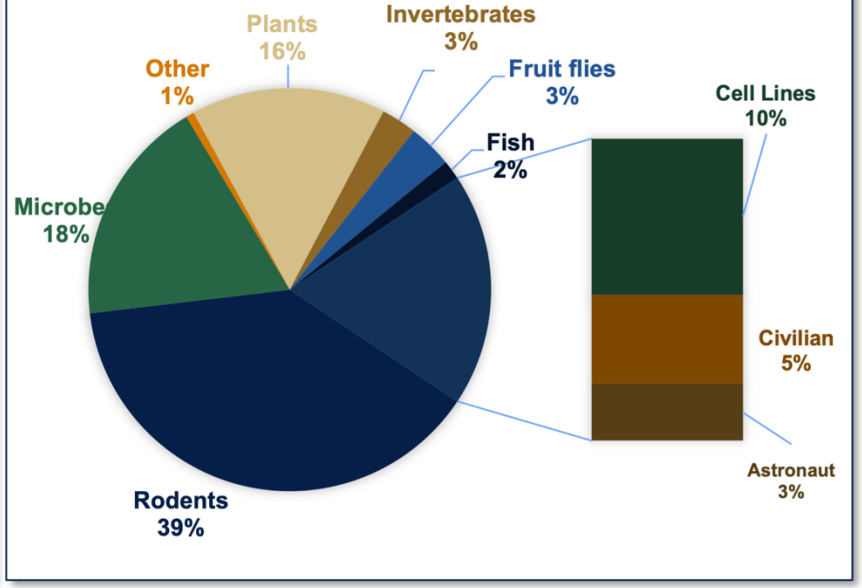
PHYSIO-PHENO



INTERNATIONAL STUDIES



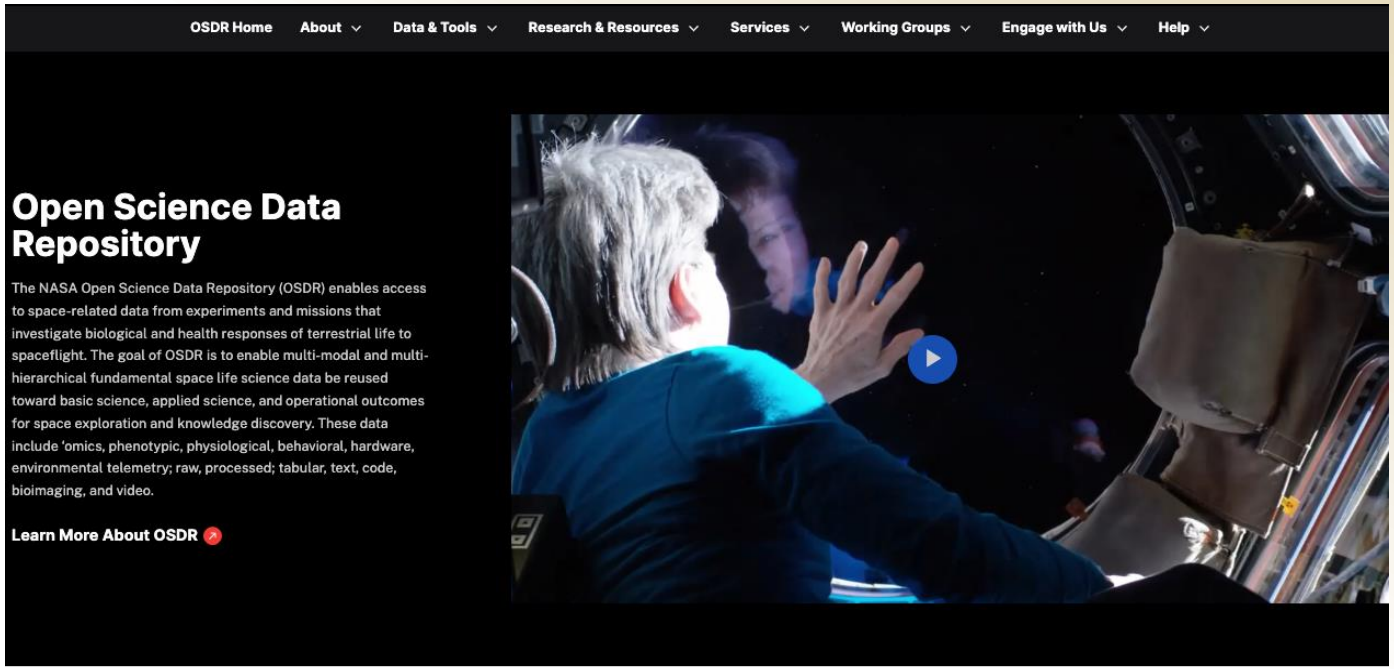
ORGANISM



USAGE METRICS (12 MO)

COUNTRY	ACTIVE USERS	
United States	48K	↑1,90...
India	1.7K	↑654....
China	1.2K	↑910....
United Kingdom	593	↑329....
Canada	542	↑530....
Germany	472	↑604....
South Korea	470	↑571....

Open Science Data Repository (nasa.gov/osdr)



Explore and Contribute

Open Science Data Repository

Search and upload spaceflight datasets

Explore the Data Repository

Submit Data to OSDR

Have space-relevant data to submit?

Contribute via the Submission Portal

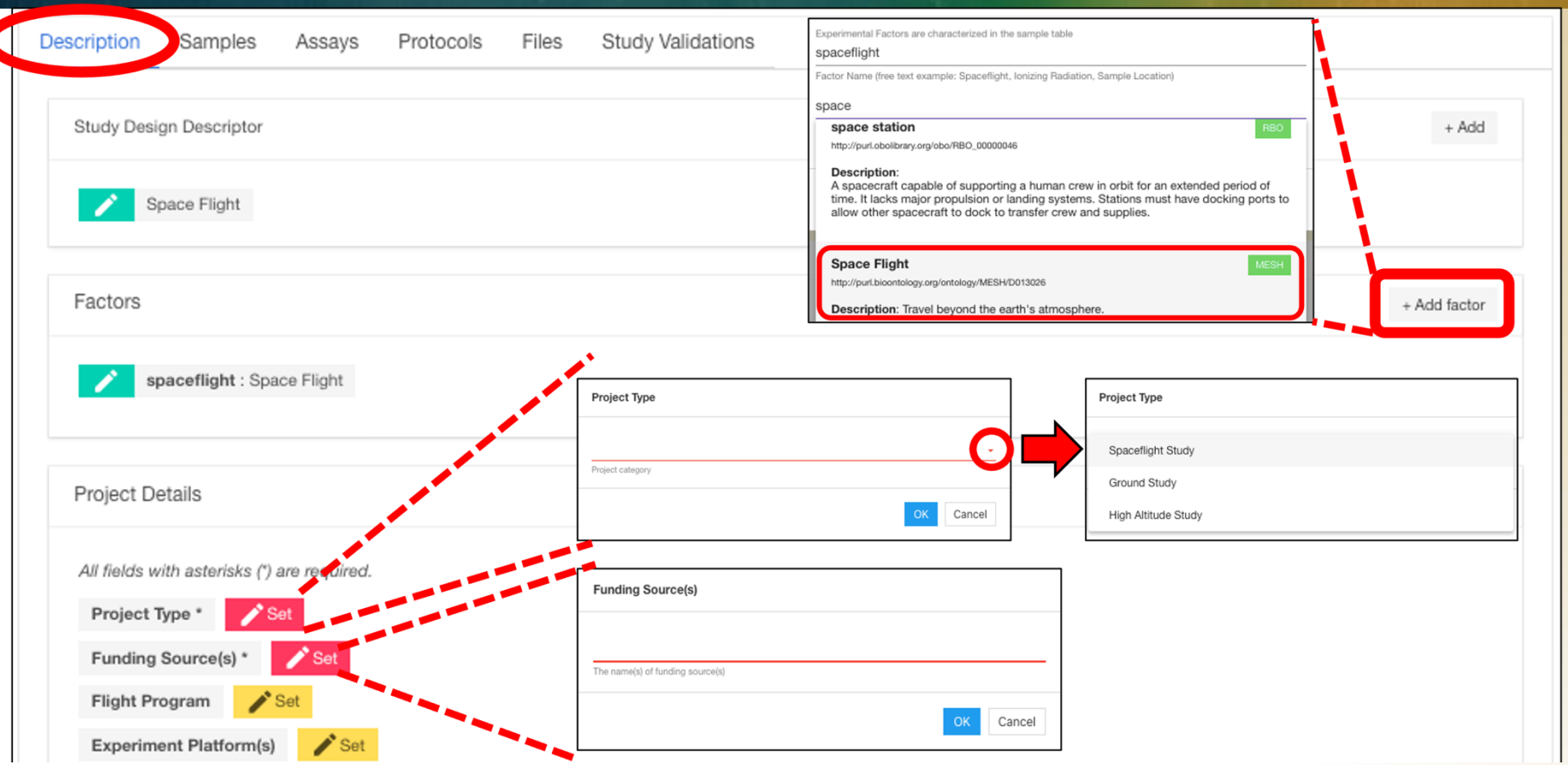
Join OSDR AWG's

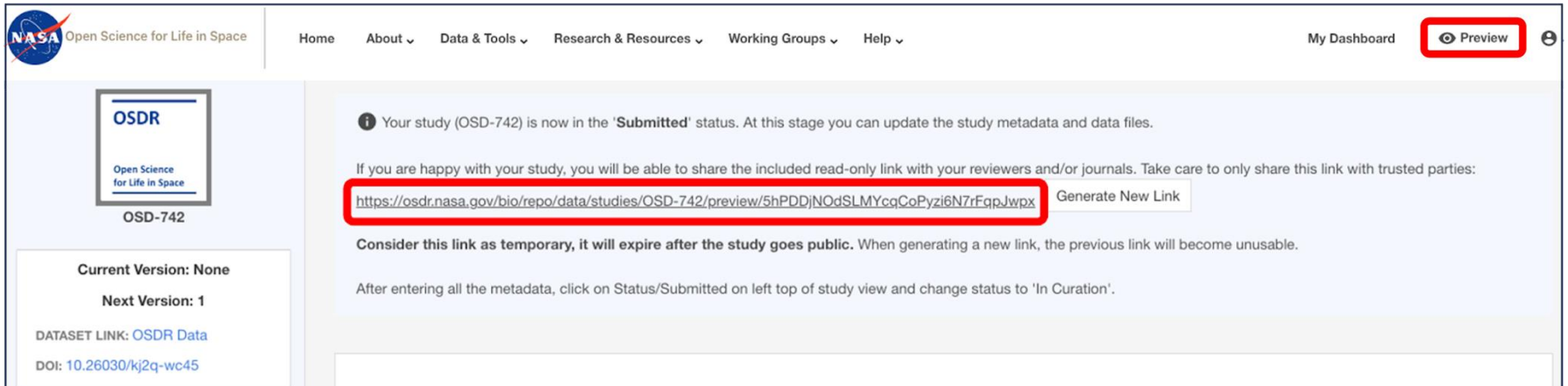
We look forward to collaborating with you!

Interested in participating in an AWG?

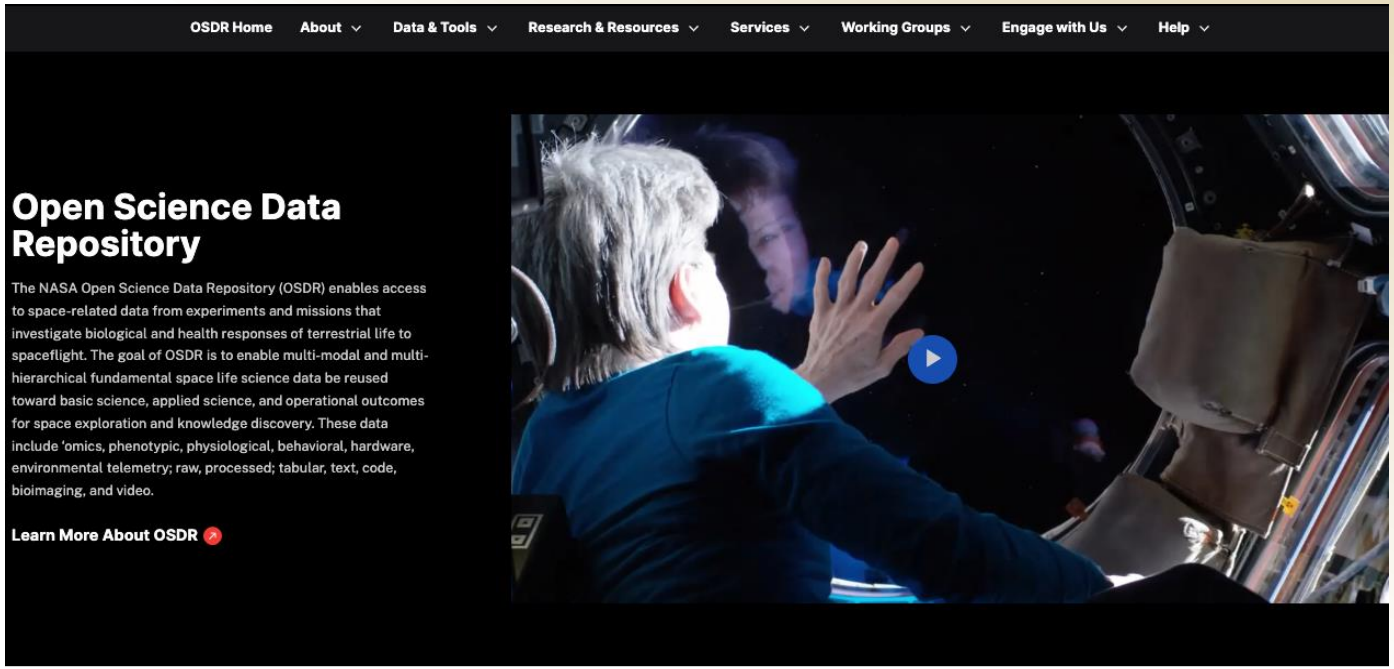
Data Submission Portal: Biological Data Management Environment (BDME)

- PIs can submit data through BDME.
- Allows efficient input of sample and assay level metadata alongside data files.
- Metadata inputs connected to controlled and normalized ontology.
- Tailored to specialized needs of spaceflight samples.
- Used for both omics (GeneLab) and non-omics (ALSDA) data submissions.

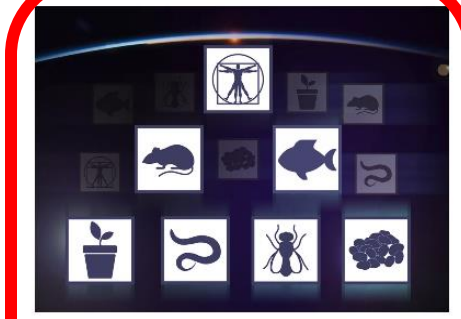




Open Science Data Repository (nasa.gov/osdr)



Explore and Contribute



Open Science Data Repository

Search and upload spaceflight datasets

Explore the Data Repository

Submit Data to OSDR

Have space-relevant data to submit?

Contribute via the Submission Portal

Join OSDR AWG's

We look forward to collaborating with you!

Interested in participating in an AWG?

Open Science for Life in Space

Home | About | Data & Tools | Working Groups | Help

Open Science Data Repository Search

Search Datasets

Sort By: Release Date

Items per page: 25 | 1 – 25 of 435

General Search Filters

Data Source

- [x] GeneLab
- [x] ALSDA
- [ ] NIH GEO
- [ ] EBI PRIDE
- [ ] ANL MG-RAST

Data Type

- [x] Study
- [ ] Experiment
- [ ] Subject
- [ ] Biospecimen
- [ ] Payload

Show more

Study Search Filters

Project Type

- [ ] Ground
- [ ] Spaceflight
- [ ] High Altitude

Assay Type

- [ ] Amplicon Sequencing Assay
- [ ] Bisulfite Sequencing
- [ ] ChIP-Seq
- [ ] Behavior (Gait)
- [ ] Gel Electrophoresis

Show more

Organism

Study

OSD-385

Persistence of Escherichia coli in the microbiomes of red Romaine lettuce (Lactuca sativa cv. 'Outredgeous')- does seed sanitization matter?

Organisms	Factors	Assay Types	Release Date	Description
Microbiota	Treatment Seed Sanitization Tissue	Amplicon Sequencing	19-Apr-2024	Seed sanitization via chemical processes removes/reduces microbes from the external surfaces of the seed and thereby could have an impact on the plants,health or productivity. To determine the impact ...

Highlights: *cgene*

Study

OSD-580

Transcriptional profiling of heart tissue from mice flown on the RRRM-2 mission

Organisms	Factors	Assay Types	Release Date	Description
Mus musculus	Spaceflight Age Euthanasia Location	transcription profiling	03-Jan-2024	In the Rodent Research Reference Mission (RRRM-2), forty female C57BL/6NTac mice were flown on the International Space Station. To assess differences in outcomes due to age, twenty 12 week-old and twe...

Highlights: *cgene*

Study

OSD-576

Transcriptional profiling of tibialis anterior muscle from mice flown on the RR-23 mission

Organisms	Factors	Assay Types	Release Date	Description
Mus musculus	Spaceflight	transcription profiling	12-Dec-2023	The objective of the Rodent Research-23 mission (RR-23) was to better understand the effects of spaceflight on the eyes, specifically on the structure and function of the arteries, veins, and lymphati...

Highlights: *cgene*

Study

OSD-524

Ionizing radiation induces transgenerational effects of DNA methylation in zebrafish

Organisms	Factors	Assay Types	Release Date	Description
Danio rerio	Ionizing Radiation Generation	DNA methylation profiling	31-Aug-2023	Ionizing radiation is known to cause DNA damage, yet the mechanisms underlying potential transgenerational effects of exposure have been scarcely studied. Previously, we observed effects in offspring ...

Enabling *FAIR* Science

- Description
- Experiments
- Payloads
- Missions
- Protocols
- Samples
- Assays
- Publications
- Files
- Version History
- Visualization

OSD-666 Version 3

Transcriptional profiling of right quadriceps femoris muscle from mice flown on the RR-23 mission

Study

420.51 GB

Submitted Date: 26-Sep-2023
Initial Release Date: 18-Oct-2023

GeneLab ID: GLDS-605
DOI: 10.26030/7c15-jm96

Cite this Study

Description

Description

The objective of the Rodent Research-23 mission (RR-23) was to better understand the effects of spaceflight on the eyes, specifically on the structure and function of the arteries, veins, and lymphatic vessels that are needed to maintain vision. To this end, twenty male, C57BL/6J, 16-17 weeks-old mice were delivered to the International Space Station (ISS) on SpaceX-21 in a single transporter, transferred to two rodent habitats, and maintained in microgravity for 38 days. Flight mice were then returned to Earth alive (January 13th, 2021). After splashdown in the Atlantic Ocean, mice were transported to Kennedy Space Center via helicopter. The 20 Flight, 20 Habitat Ground Control (HGC), and 20 Vivarium Ground Control (VGC) mice were removed from Rodent Transporters (Flight and HGC) or vivarium cages (VGC), placed into shipping containers, and flown to Texas A and M University. There, mice underwent post-flight procedures, before euthanasia and tissue collection. Flight, HGC and VGC animals were euthanized and dissected on Jan 14th, 17th or 20th of 2021, respectively. Right quadriceps femoris muscle samples were preserved by immersion in liquid nitrogen and stored at -80C until RNA was extracted, and libraries generated and sequenced (target 60 M clusters per sample, PE 150 bp). This dataset features 9 samples from the Flight group, 9 samples from the Habitat Ground Control group, and 9 samples from the Vivarium Ground Control group.

Factor(s)

Factor	Ontology: Concept
Spaceflight	Space Flight

Organism(s)

Mus musculus

Assay(s)

Measurement	Technology	Device Platform
transcription profiling	RNA Sequencing (RNA-Seq)	Illumina

Project

Payload Identifier	RR-23
Project Title	Effects of Microgravity on Ocular Vascular Hydrodynamics
Project Type	Spaceflight Study
Flight Program	International Space Station (ISS)
Experiment Platform	Rodent Flight Hardware (Transporter and Habitat)

Enabling *FAIR* Science: Metadata Curation

Samples

Controlled, Standardized Ontology / Terminology

Select Export Columns

Source Name	Sample Name	Characteristics: Organism	Characteristics: Strain	Characteristics: Genotype	Characteristics: Animal Source	Characteristics: Material Type	Factor Value: Spaceflight	Characteristics: Sex	Characteristics: Age at Launch	Protocol REF	Parameter Value: habitat	Parameter Value: duration	Parameter Value: light cycle	Parameter Value: Enrichment material
RR-23_F2	RR23_R-Quad_FLT_F2	Mus musculus	C57BL/6J	Wild Type	Jackson Laboratory	Right quadriceps femoris	Space Flight	Male	16-17 week	Animal Husbandry	Rodent Flight Hardware (Transporter and Habitat)	38 day	12 h light: 12 h dark. Lights on at 7:00 GMT	Hut
RR-23_F3	RR23_R-Quad_FLT_F3	Mus musculus	C57BL/6J	Wild Type	Jackson Laboratory	Right quadriceps femoris	Space Flight	Male	16-17 week	Animal Husbandry	Rodent Flight Hardware (Transporter and Habitat)	38 day	12 h light: 12 h dark. Lights on at 7:00 GMT	Hut
RR-23_F4	RR23_R-Quad_FLT_F4	Mus musculus	C57BL/6J	Wild Type	Jackson Laboratory	Right quadriceps femoris	Space Flight	Male	16-17 week	Animal Husbandry	Rodent Flight Hardware (Transporter and Habitat)	38 day	12 h light: 12 h dark. Lights on at 7:00 GMT	Hut
RR-23_F5	RR23_R-Quad_FLT_F5	Mus musculus	C57BL/6J	Wild Type	Jackson Laboratory	Right quadriceps femoris	Space Flight	Male	16-17 week	Animal Husbandry	Rodent Flight Hardware (Transporter and Habitat)	38 day	12 h light: 12 h dark. Lights on at 7:00 GMT	Hut

Assays

Controlled, Standardized Ontology / Terminology

Assay Name: transcription profiling - RNA Sequencing (RNA-Seq)

Technology Type: RNA Sequencing (RNA-Seq)

Technology Platform: Illumina

Select Export Columns

Sample Name	Protocol REF	Parameter Value: QA Instrument	Parameter Value: QA Assay	Parameter Value: QA Score	Extract Name	Protocol REF	Parameter Value: Spike-in Quality Control	Parameter Value: Spike-in Mix Number	Protocol REF	Parameter Value: Library Selection	Parameter Value: Library Layout	Protocol REF	Parameter Value: Read Depth	Parameter Value: rRNA Contamination
RR23_R-Quad_FLT_F2	Nucleic Acid Extraction	Agilent 4200 TapeStation	Agilent RNA ScreenTape Assay	8.4 RINe	RR23_R-Quad_FLT_F2	Spike-in Protocol	ERCC ExFold RNA Spike-In Mix	Mix 1	Library Construction	Ribo-depletion	PAIRED	GeneLab raw data processing protocol	101335022 read	0.61 percent
RR23_R-Quad_FLT_F3	Nucleic Acid Extraction	Agilent 4200 TapeStation	Agilent RNA ScreenTape Assay	8.3 RINe	RR23_R-Quad_FLT_F3	Spike-in Protocol	ERCC ExFold RNA Spike-In Mix	Mix 1	Library Construction	Ribo-depletion	PAIRED	GeneLab raw data processing protocol	98135600 read	0.5 percent
RR23_R-Quad_FLT_F4	Nucleic Acid Extraction	Agilent 4200 TapeStation	Agilent RNA ScreenTape Assay	7.7 RINe	RR23_R-Quad_FLT_F4	Spike-in Protocol	ERCC ExFold RNA Spike-In Mix	Mix 1	Library Construction	Ribo-depletion	PAIRED	GeneLab raw data processing protocol	106203766 read	0.52 percent
RR23_R-Quad_FLT_F5	Nucleic Acid Extraction	Agilent 4200 TapeStation	Agilent RNA ScreenTape Assay	7.9 RINe	RR23_R-Quad_FLT_F5	Spike-in Protocol	ERCC ExFold RNA Spike-In Mix	Mix 1	Library Construction	Ribo-depletion	PAIRED	GeneLab raw data processing protocol	105949962 read	0.69 percent

Enabling *FAIR* Science: Omics Data Processing

Build consensus data processing pipelines with the scientific community

Raw Sequence Data

```
@A00654:12:HFY5YDSXX:1:1101:2121:1000_CTCGCGTTT 1:N:0:NGGAGAGT+TCCTACCT
AATATCAGTGATATTTAGAAACCACATAGTAAGCTAACAACTAATAATGGAATGGTTTTAATATCCTGTGACAAGTTAATGTGGATACTATGCGGTCTTCCTTAAAA
+
FFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FF
@A00654:12:HFY5YDSXX:1:1101:5466:1000_CCTCCCTCG 1:N:0:NGGAGAGT+TCCTACCT
GGCGGGTCTGAGCAAGAACACTCTGCCATCGCCGGCTCGGGACGGGAGGCCCACCCGTGGCCGAGCACGCAGACCTGGATGAACCGAGGCCTCGATGGGCCTGGACG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00654:12:HFY5YDSXX:1:1101:14109:1000_TGATCTACG 1:N:0:NGGAGAGT+TCCTACCT
TGTCACTACAACCAGCTGTGCCTGTGCTATTGCAGTTACACAGTGTCACTACAACCAACTGTGCCTGTGCTATTGCAGTTACACAGTGTCACTACAACCAGCAGATC
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,:FFFFFFF:FF:F:FF
```

Processed Amplicon Sequence Data: Differential Abundance Analysis

ASV_ID	family	genus	Group.Mean_(FLT)	Group.Mean_(GC)	log2FC_(FLT)v(GC)	stat_(FLT)v(GC)	padj_(FLT)v(GC)
ASV_16S_1	Tannerellaceae	Parabacteroides	38674.17	33866.03	0.19	1.16	0.51
ASV_16S_10	Lachnospiraceae	NA	382.22	1622.88	-2.09	-2.60	0.05
ASV_16S_100	Tannerellaceae	Parabacteroides	0.00	8.47	-5.62	-1.85	NA
ASV_16S_101	Erysipelotrichaceae	Holdemania	0.00	12.07	-6.13	-2.02	NA
ASV_16S_102	Tannerellaceae	Parabacteroides	8.23	0.00	5.53	1.82	NA
ASV_16S_103	Peptostreptococcaceae	NA	0.00	7.48	-5.44	-1.79	0.24

Amplicon Sequencing Data

Raw data fastQC (fastqc/multiqc) → Primer removal (dada2) → Quality filter/trim (dada2) → Filtered data fastQC (fastqc/multiqc) → Resolve ASVs / Remove Chimeras (dada2) → Assign Taxonomy (DECIPHER SILVA DB for rRNA)

Primary outputs:
- Recovered ASV sequences (fasta)
- Counts of sequences per sample
- Taxonomic classifications of sequences

Data Visualizations:
- Beta and alpha diversity plots
- Taxonomic bar plots
- Differential abundance analysis and volcano plots

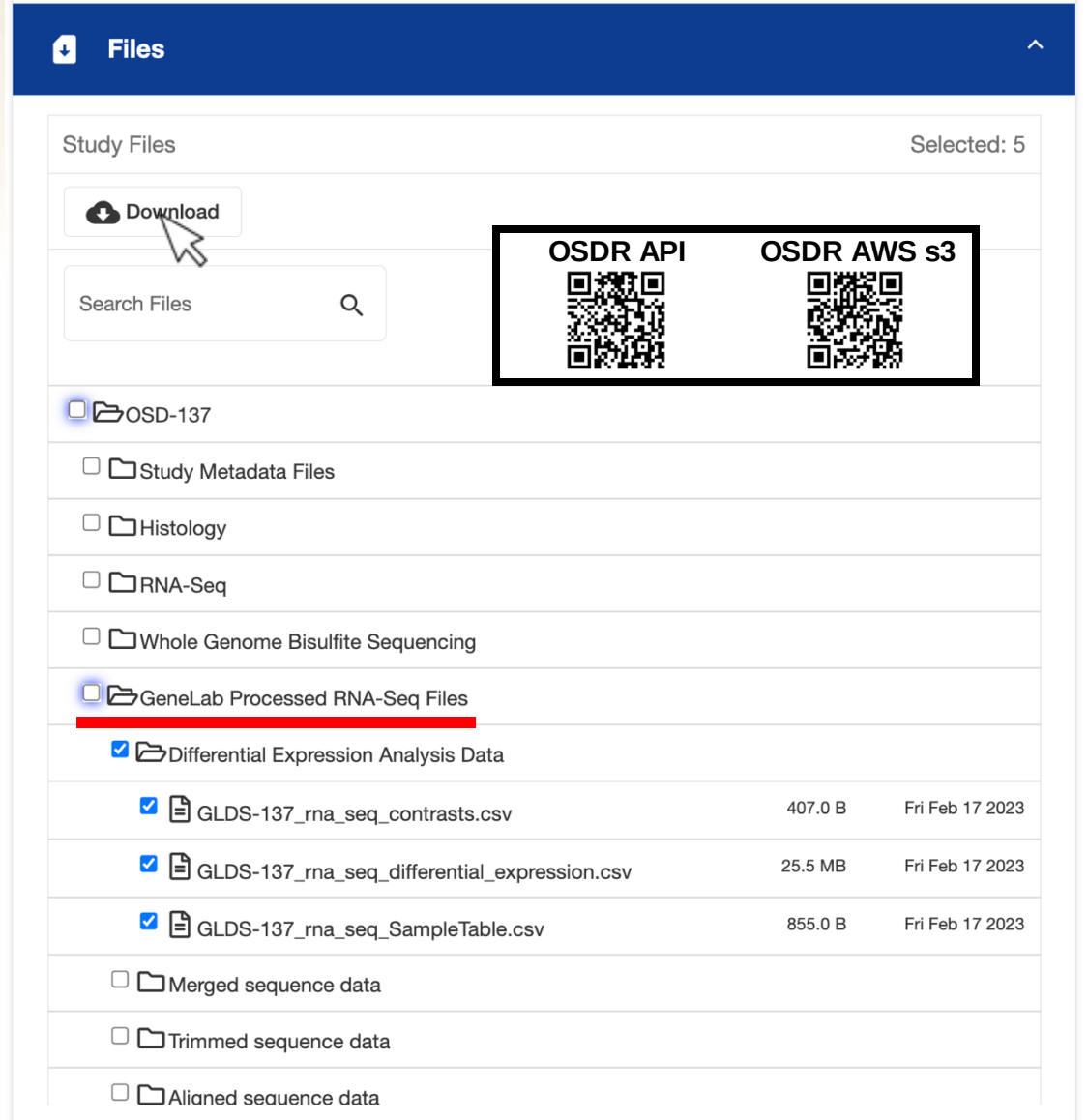
```
>ASV_1
AGTTTGATCA...
>ASV_2
TTTATGGAGA...
...
```

	Sample_A	Sample_B	...
ASV_1	975	537	...
ASV_2	40	480	...
...	...	...	...

	domain	phylum	...
ASV_1	Bacteria	Proteobacteria	...
ASV_2	Bacteria	Firmicutes	...
...	...	...	...

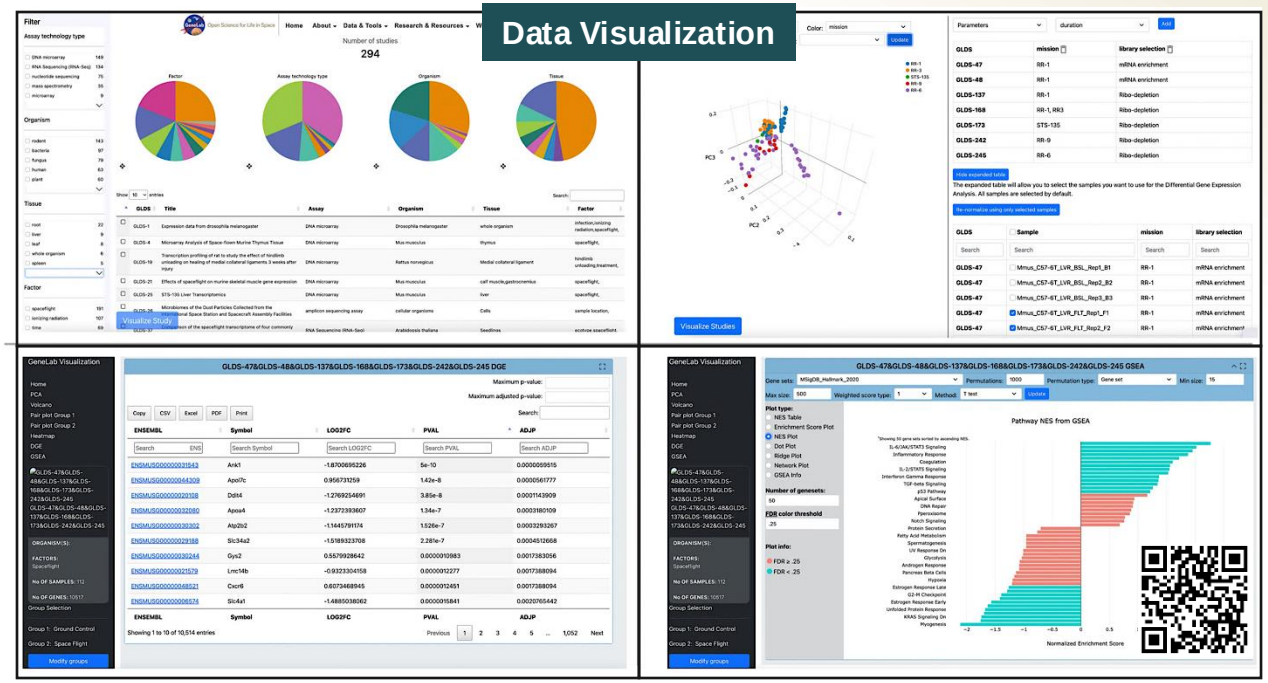
Data Visualization (stats / phyloseq / vegan)

Differential Abundance (DESeq2)

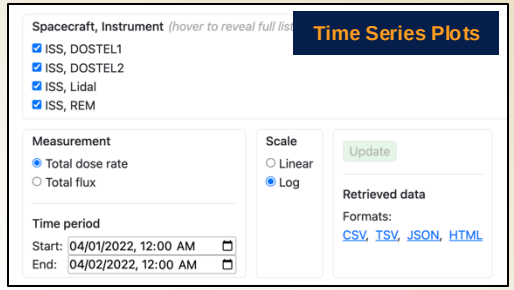


OSDR Visualization Platforms

Data Visualization

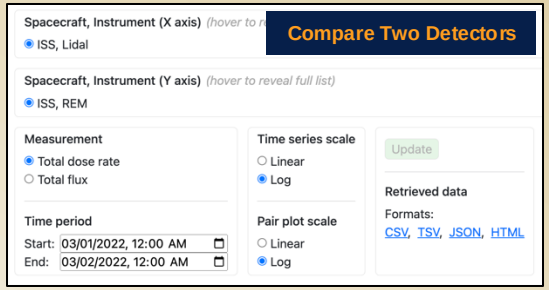


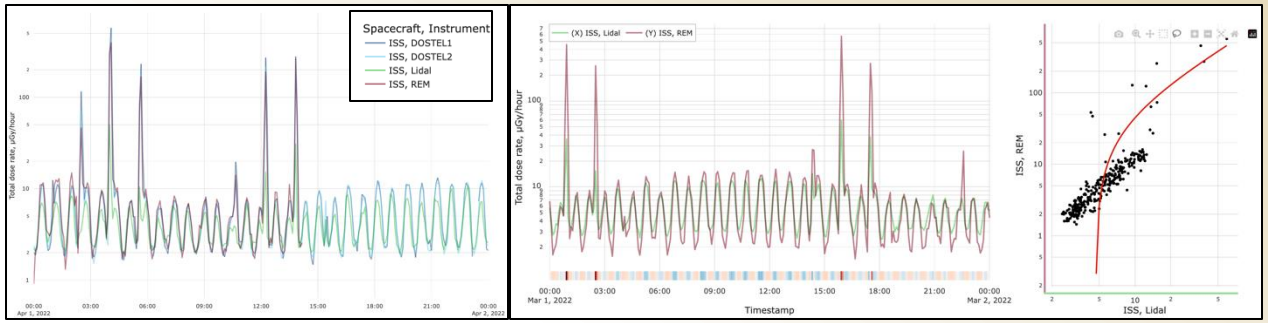
Time Series Plots



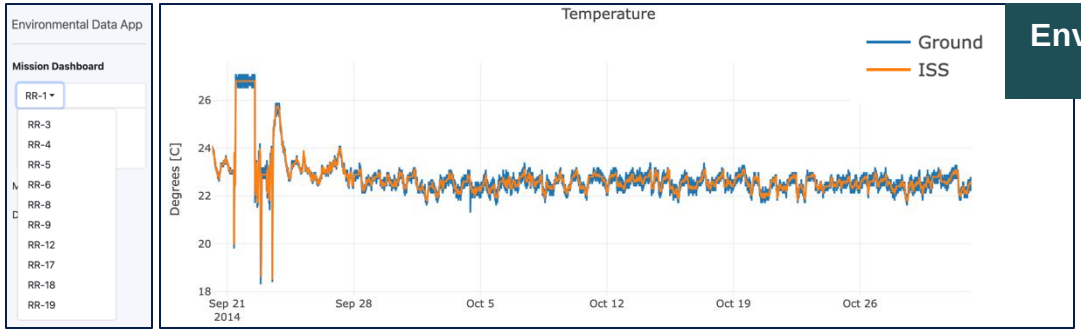
RADLAB

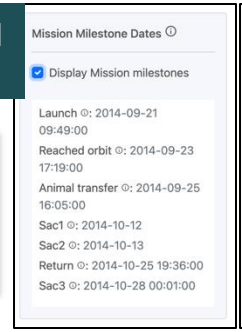
Compare Two Detectors

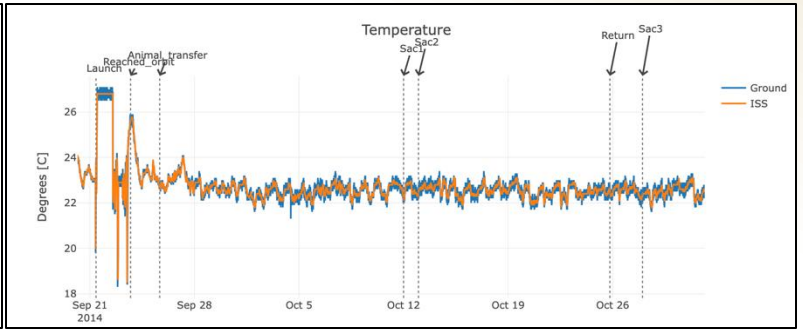


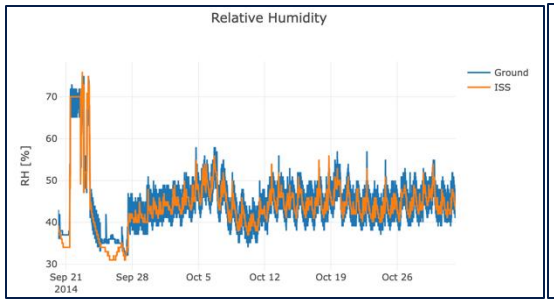


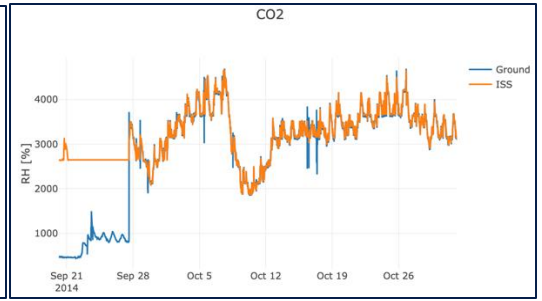
Environmental Data

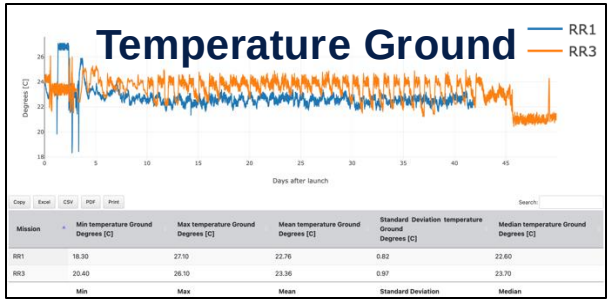


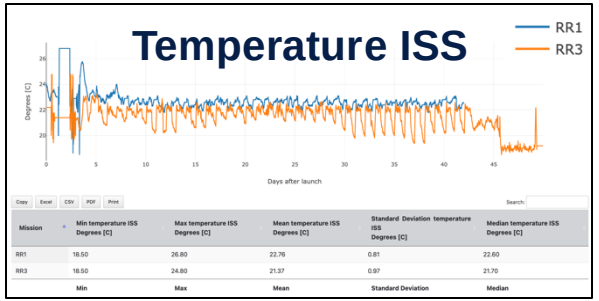












Open Science Analysis Working Groups (AWGs)

ANIMAL

226 members

MULTI-OMICS

351 members

MICROBIAL

216 members

Join an AWG!

PLANTS

176 members

AI/ML

323 members

ALSDA

(Physiological/BioMedical)

224 members

FEMALE REPRO

88 members

HUMAN

39 members

RADLAB

67 members

92

Enabled Publications linked to OSDR

109

Original Publications linked to OSDR

150+

Datasets used in enabled publications

Collaborate on Data Mining/Publications

Cell Press Package 2020:
The Biology of Spaceflight
https://www.cell.com/c/the-biology-of-spaceflight

Nature Portfolio Collection 2024:
Space Omics and Medical Atlas across orbits (SOMA)
https://www.nature.com/collections/ebdbcahdgc

Physical Sciences Informatics (PSI)

https://nasa.gov/psi

The new PSI Data Repository and Submission Portal was released in September 2024

Key updates to the PSI Repository include:

- PSI utilized NASA OSDR infrastructure to develop a new data repository and user submission portal
- New look and layout, available investigations are visible as a scrollable list.
- The Submission Portal is a new tool to streamline the process of publishing new investigations.
- Improved data repository interface and new search capabilities
- To support the sharing of scientific information, each dataset has an assigned DOI.
- Version history is available for datasets

PSI Focuses on 6 Research Areas

Fluid Physics

The study of fluid motion and the associated transport of mass, momentum and energy to address critical space exploration needs for advanced life support systems, propulsion, and thermal control systems.

Biophysics

Growing larger and more perfect protein macromolecule crystals for analysis, biomaterials/biofilms, and biofluids to sustain long duration human presence in space and improve quality of life on earth.

Combustion Science

Obtain benchmark microgravity data to advance computational models for spacecraft fire safety and improved efficiency & reduced pollutant emission in practical terrestrial combustion.

Fundamental Physics

Quantum Coherence & entanglement, quantum interferometry & precision measurement quantum matter & many-body systems.

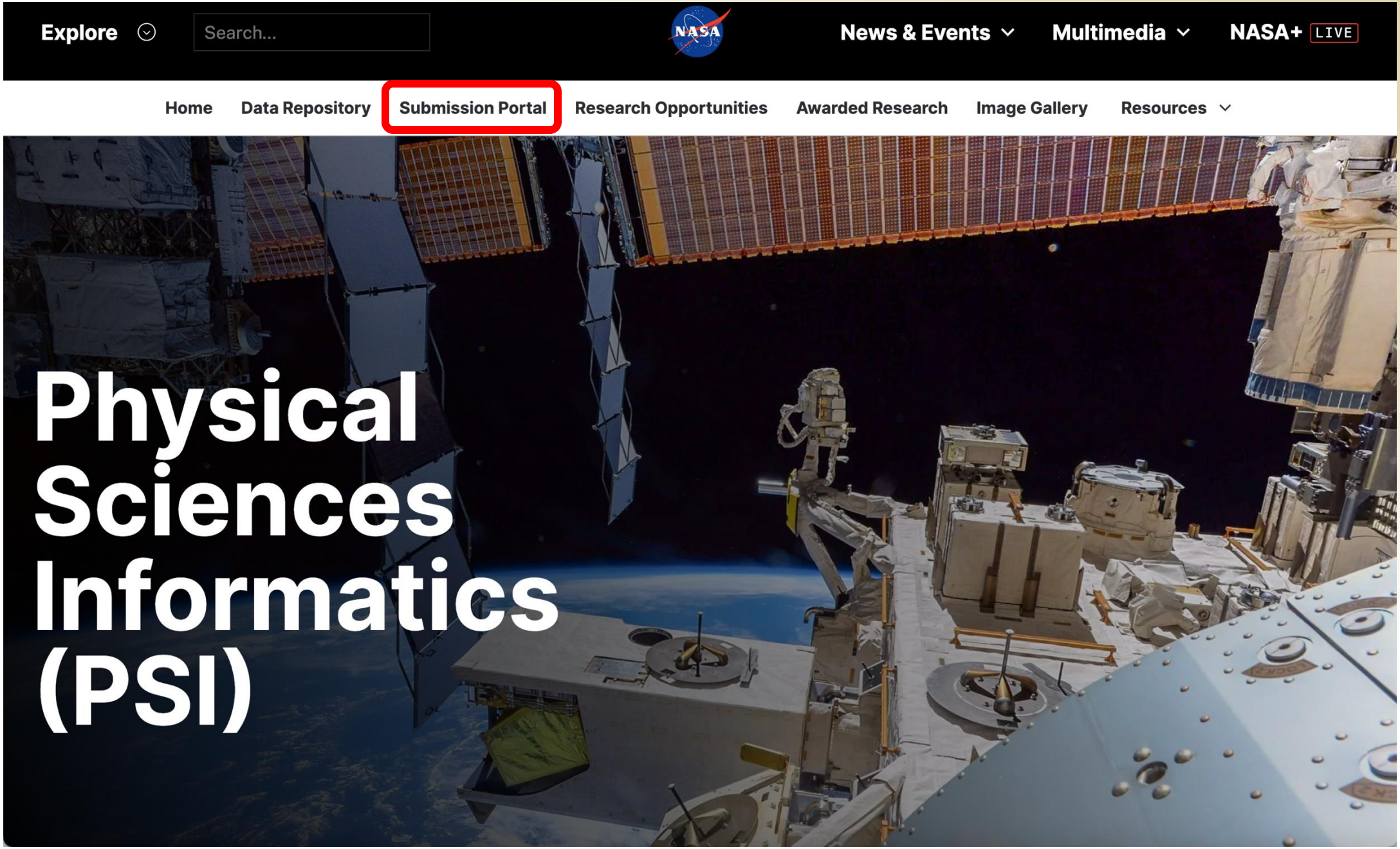
Materials Science

Microstructural evolution, phase coarsening, and quantifying thermophysical properties to improve computational models and enhance material properties terrestrially and in space - Glasses & Ceramics, Granular Materials, Metals, Polymers & Organics & Semiconductors.

Complex Fluids Soft Matter

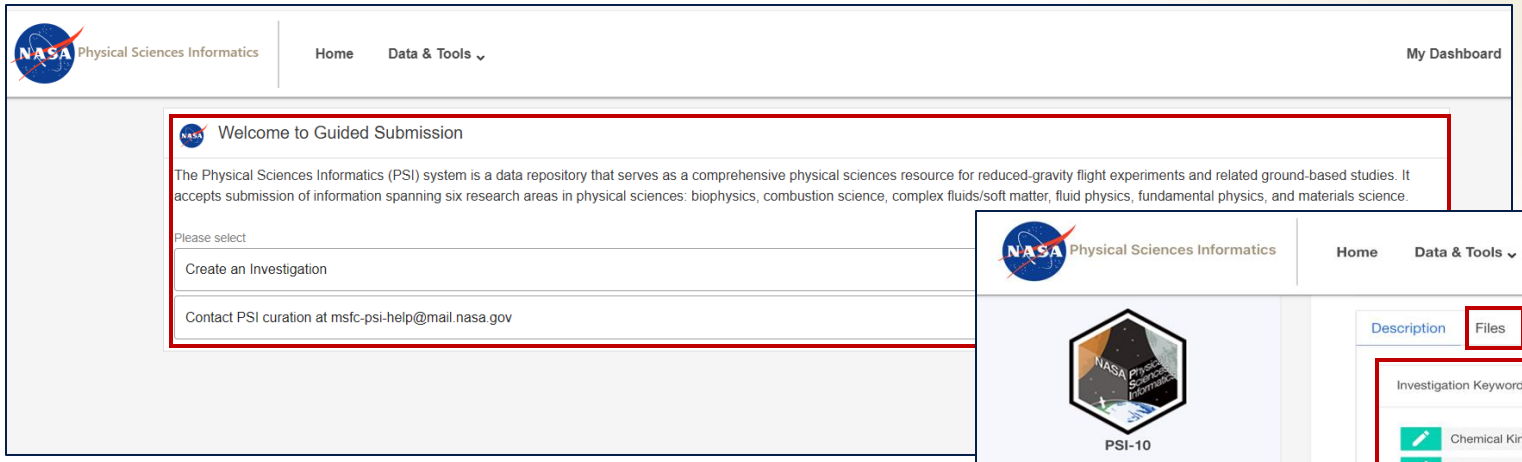
Study of binary mixtures such as colloidal systems, liquid crystals, foams, gels, emulsions, and granular flows.

Physical Sciences Informatics (nasa.gov/psi)

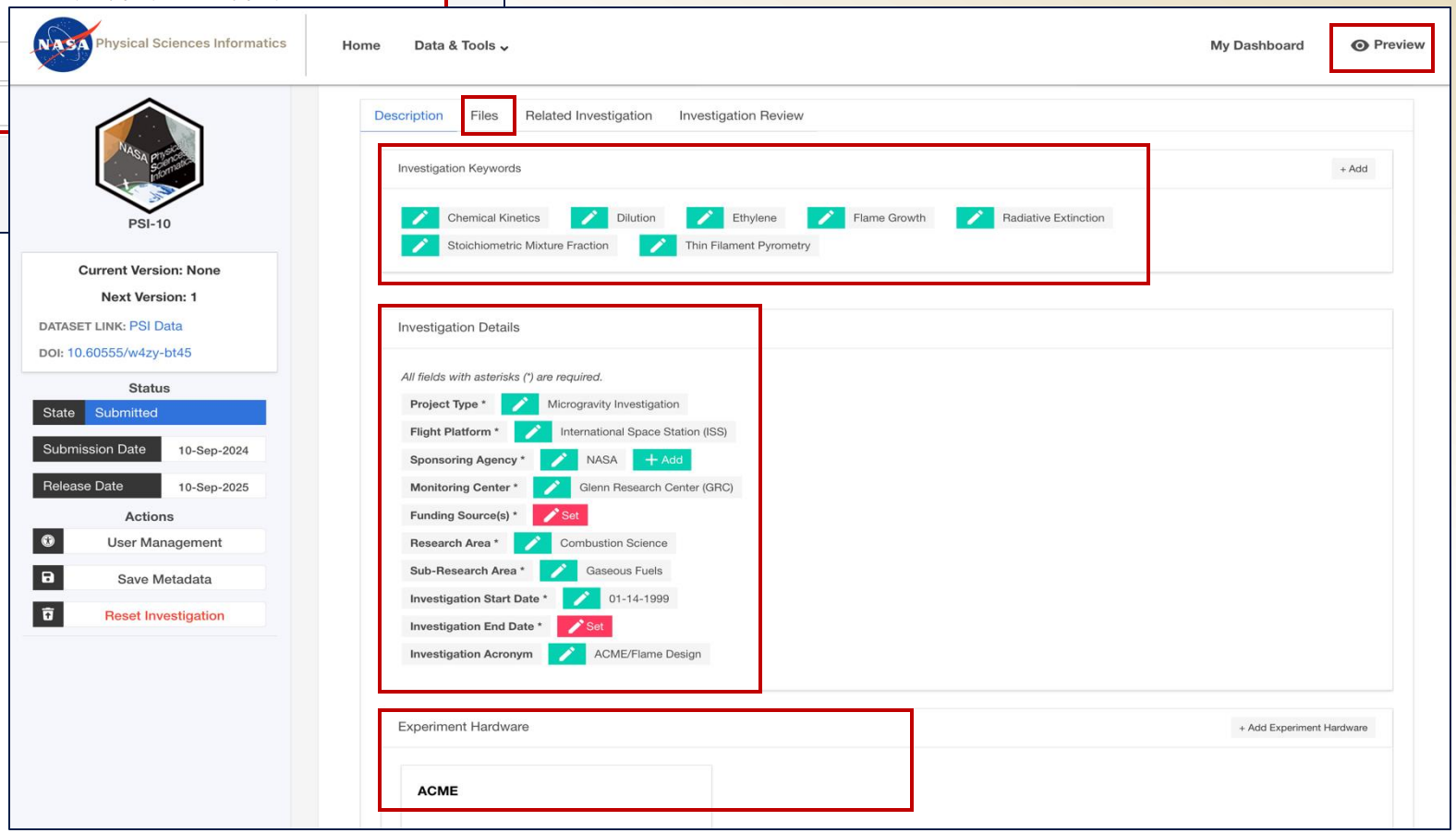


Submission Portal

The Guided Submission directs Data Submitters through the process of creating a new investigation and offers the option of contacting a PSI Curator for more help.



Preview provides a view of the Investigation page before publishing.



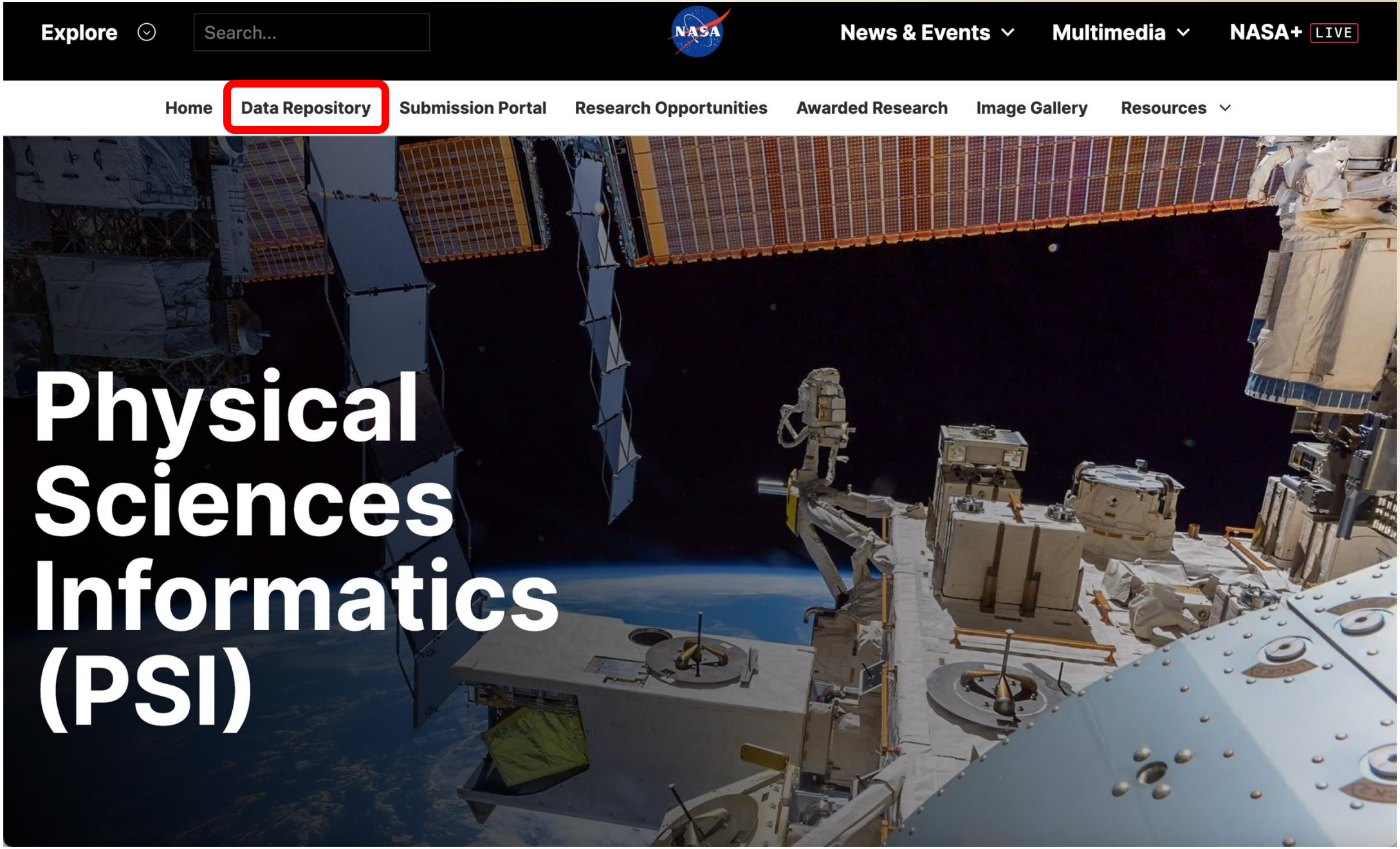
Investigation Description captures:

- Keywords
- Investigation Details
- Experiment Hardware

Investigation Files include:

- Raw and Analyzed Data
- Science Documents
- Reports and Presentations

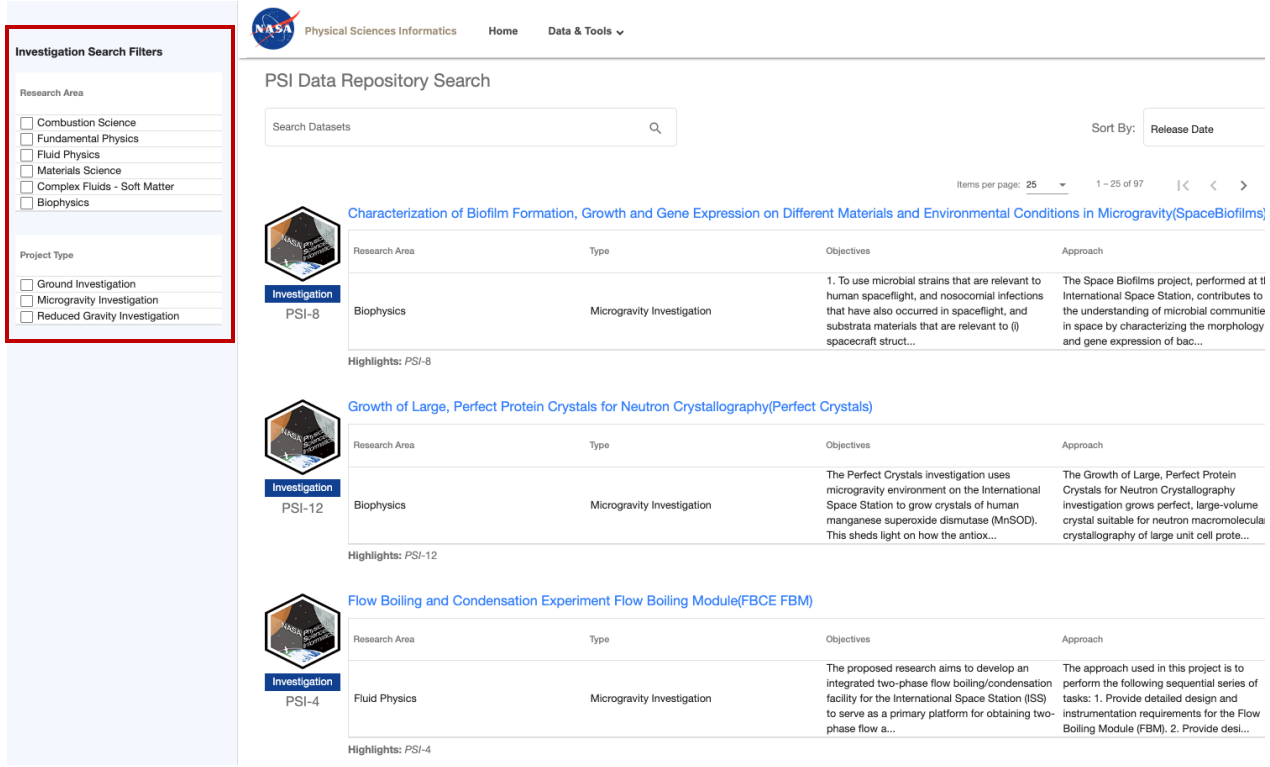
Physical Sciences Informatics (nasa.gov/psi)



New Data Repository User Interface

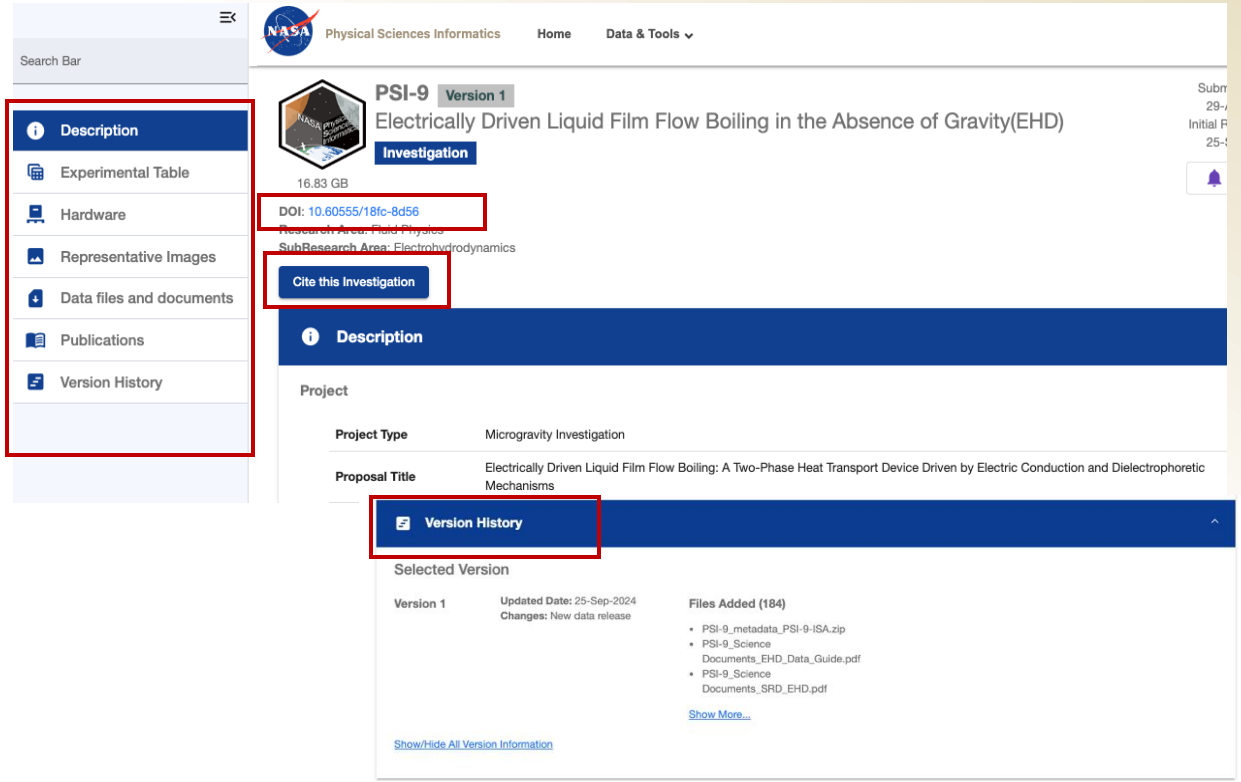
Researchers can filter the investigations by

- Research Area
- Project Type
- Keywords



New Investigation Records include:

- Navigation Menu
- Digital object identifier (DOI) upon creation
- Citation
- Version History



Acknowledgements

PSI Team

Open Science for Life in Space Teams

Open Science Analysis Working Group Members

Support

NASA Space Biology Program
NASA Biological and Physical Sciences
NASA Science Mission Directorate
NASA Human Research Program