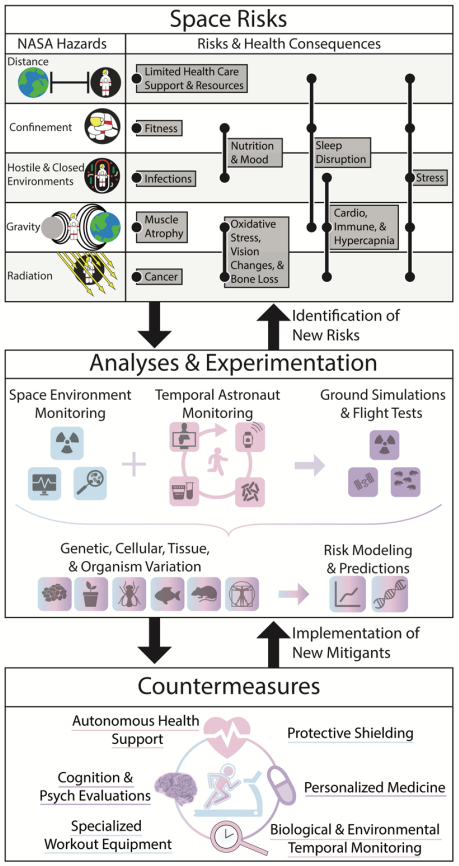
NASA Open Science Data Repository: Maximizing Spaceflight Bioscience Data

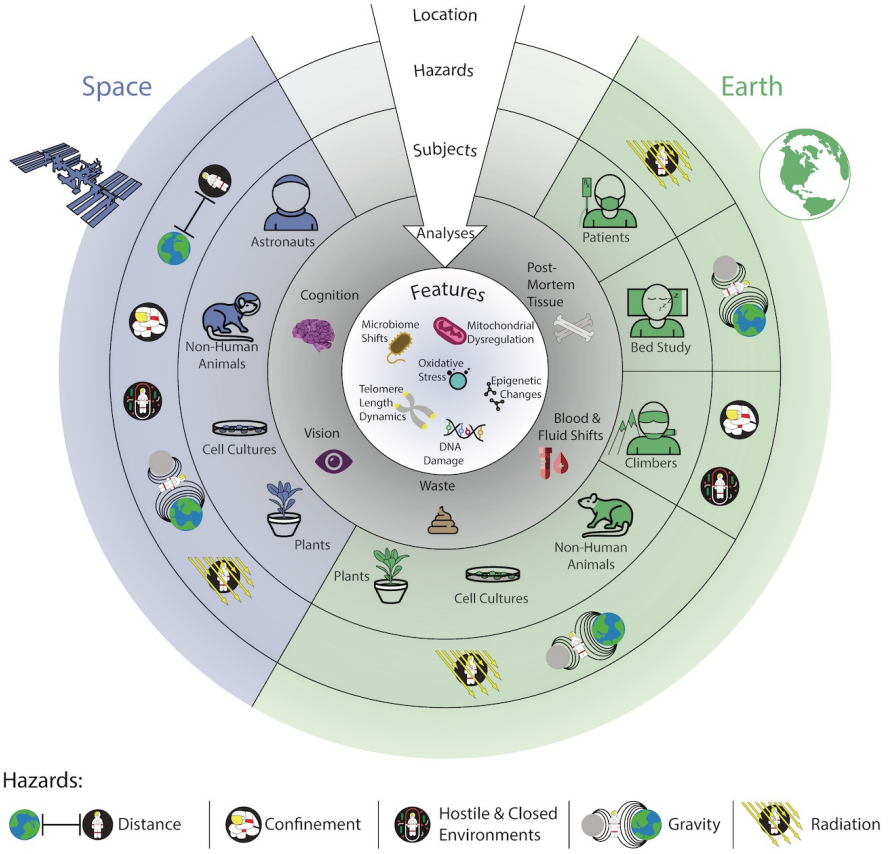
National Aeronautics and Space Administration



Ryan T. Scott
KBR | Science Lead
NASA Ames Life Sciences Data Archive
NASA OSDR: Open Science for Life in Space
NASA Ames Research Center, Moffett Field, CA
ryan.t.scott@nasa.gov

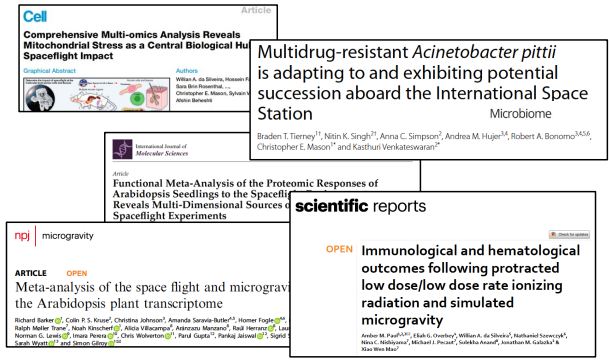
Field of Spaceflight Biology and Health Research





Reusable Data is Essential for Life and Humans to Thrive in Deep Space

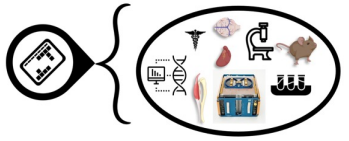
Knowledge Discovery



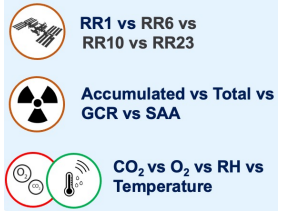
Enable Informatics, AI/ML, KG, DAG, AOP, Omic-Phenotypic Conclusions

Model development
Active and adaptive learning
Causal and network inference
Explainable AI
Federated learning
Few- or one-shot learning
Generative learning (synthetic datasets)
Knowledge graphs
Lifelong learning (living lab)
Out-of-distribution learning
Transfer learning

Telemetry for Biomedical Context



Analysis Across Missions

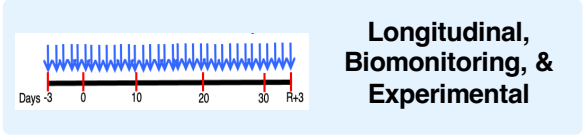




Across Hazards

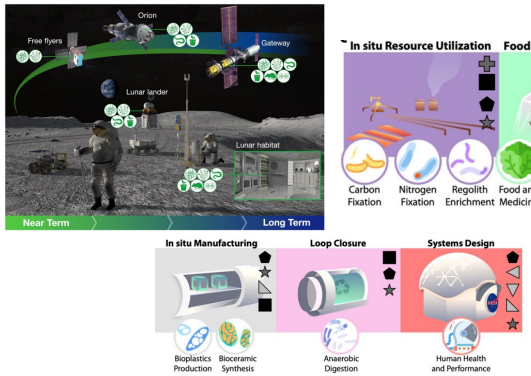


Across Species, Multimodal Data Types

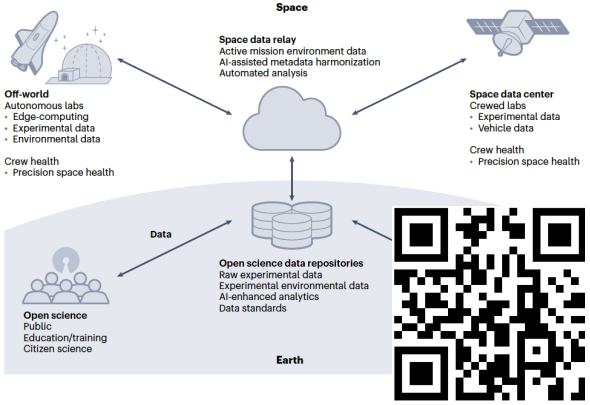


Longitudinal, Biomonitoring, & Experimental

Novel Ecosystem, Platform, Countermeasure Development, Health Risk Quantification



Involve the World in Spaceflight



All Relevant Data for Biological, Biomedical, Health, Telemetry Analytics

Physiological/Phenotypic/Imaging/ Environmental Telemetry Data

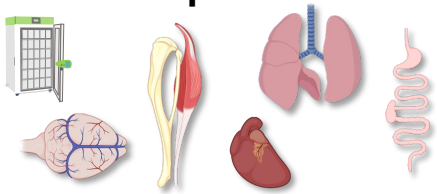


Tabular, text, imaging, telemetry, video, code

Molecular/Omics Data



Biospecimens



ALSDA

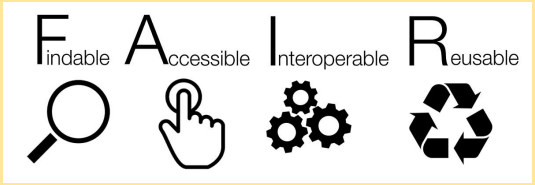
GeneLab

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)**
- **Maximally Open Access with Necessary Controls for Sensitive Data**
- **Data Maximally FAIR**



NASA Open Science Data Repository

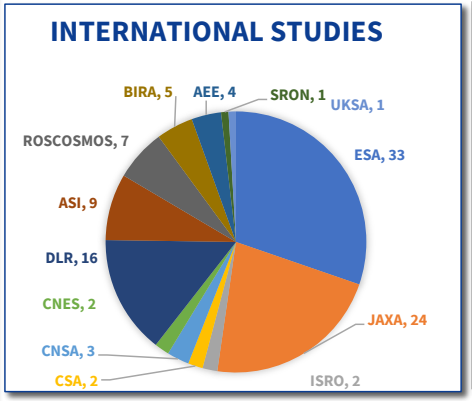
453
Studies

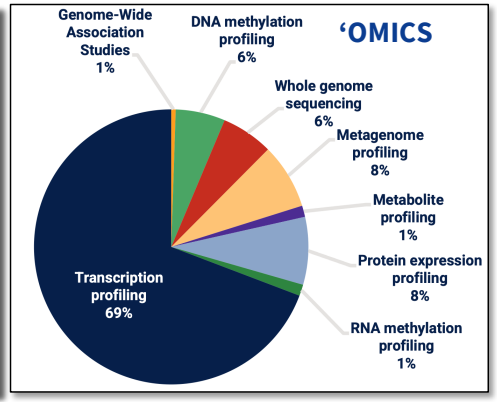
880
Datasets

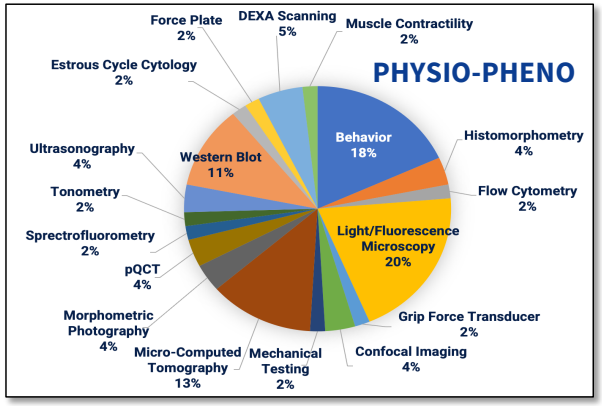
45
Species

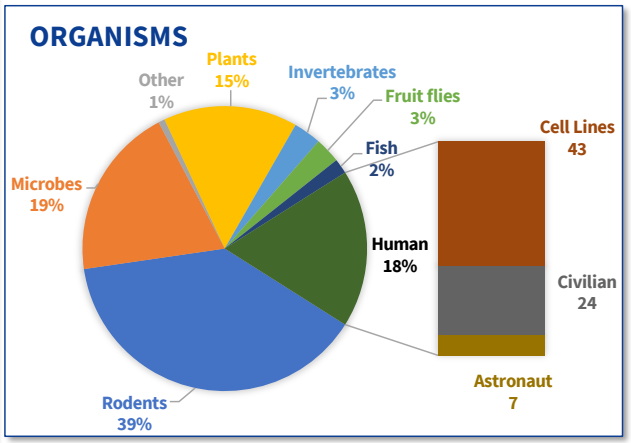
>55
Assays

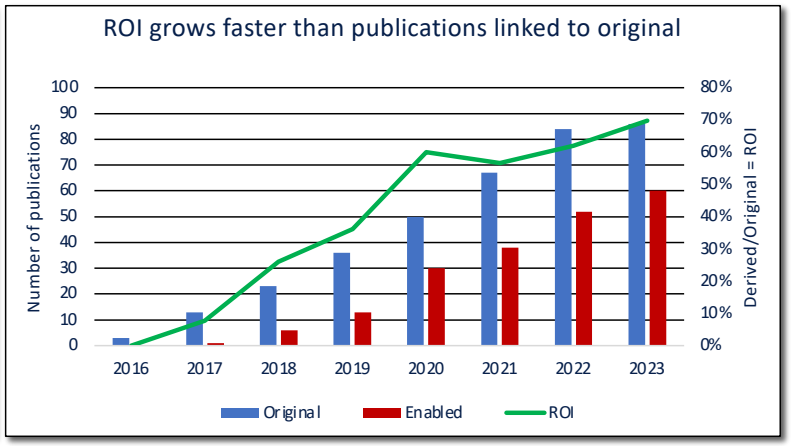
>150TB
Data











91
Original Publications Have Data Submitted to OSDR

60
Publications Enabled by OSDR Data Mining

142+
Datasets Used in Enabled Publications

ryan.t.scott@nasa.gov

A Buffet of Data: From Raw Sequence Data to Imaging to Videos

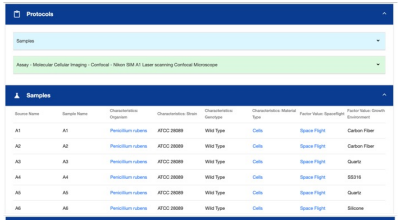
Study
OSD-628

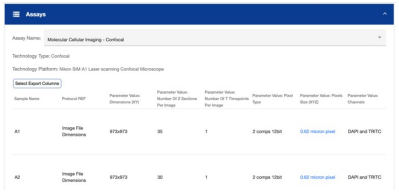
Characterization of Biofilm Formation, Growth, and Gene Expression on Different Materials and Environmental Conditions in Microgravity (Morphology of Penicillium rubens biofilms)

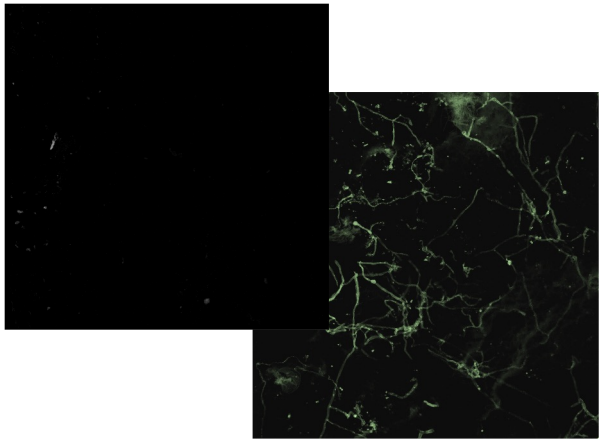
Organisms	Factors	Assay Types	Release Date	Description
Penicillium rubens	Spaceflight Growth Environment Time	Molecular Cellular Imaging	12-Sep-2023	Microorganisms' natural ability to live as organized multicellular communities – also known as biofilms – provides them with unique survival advantages. For instance, biofilms are protected against en...









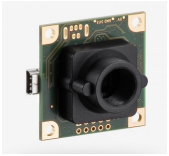


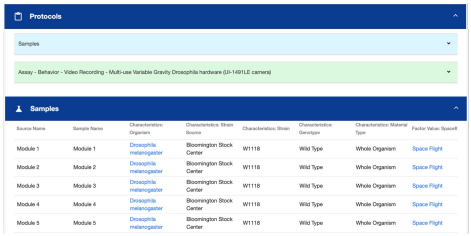
Study
OSD-595

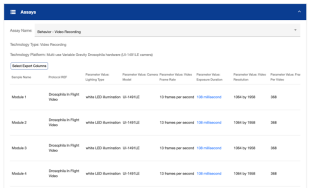
Artificial gravity partially protects space-induced neurological deficits in Drosophila melanogaster

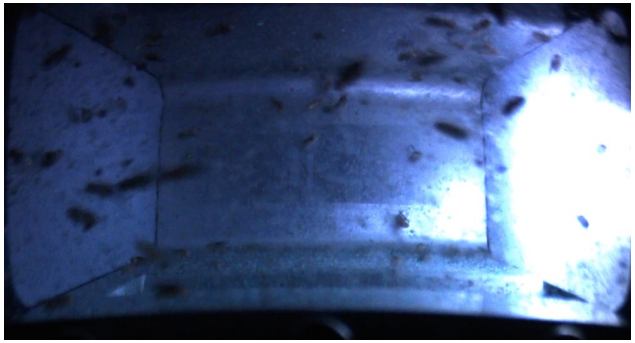
Organisms	Factors	Assay Types	Release Date	Description
Drosophila melanogaster	Altered Gravity Spaceflight	Behavior	25-May-2023	Spaceflight poses risks to the central nervous system (CNS), and understanding neurological responses is important for future missions. We report CNS changes in Drosophila aboard the International Spa...











Open Science OSDR Analysis Working Group: ~600 Members Worldwide

ANIMAL

128 members

MULTI-OMICS

342 members

MICROBIAL

130 members

PLANTS

121 members

AI/ML

138 members

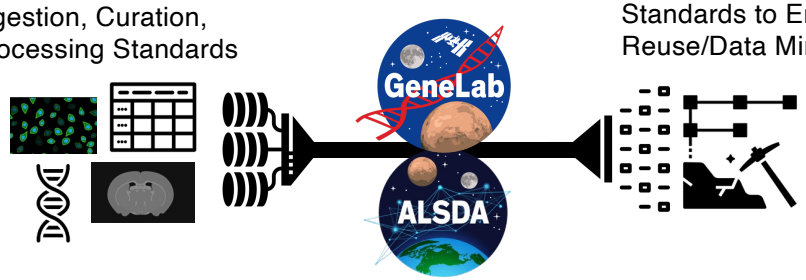
ALSDA

(Physiological/BioMed)

249 members

Feedback on Ingress/Egress Standards

Ingestion, Curation, Processing Standards



Standards to Enable Reuse/Data Mining

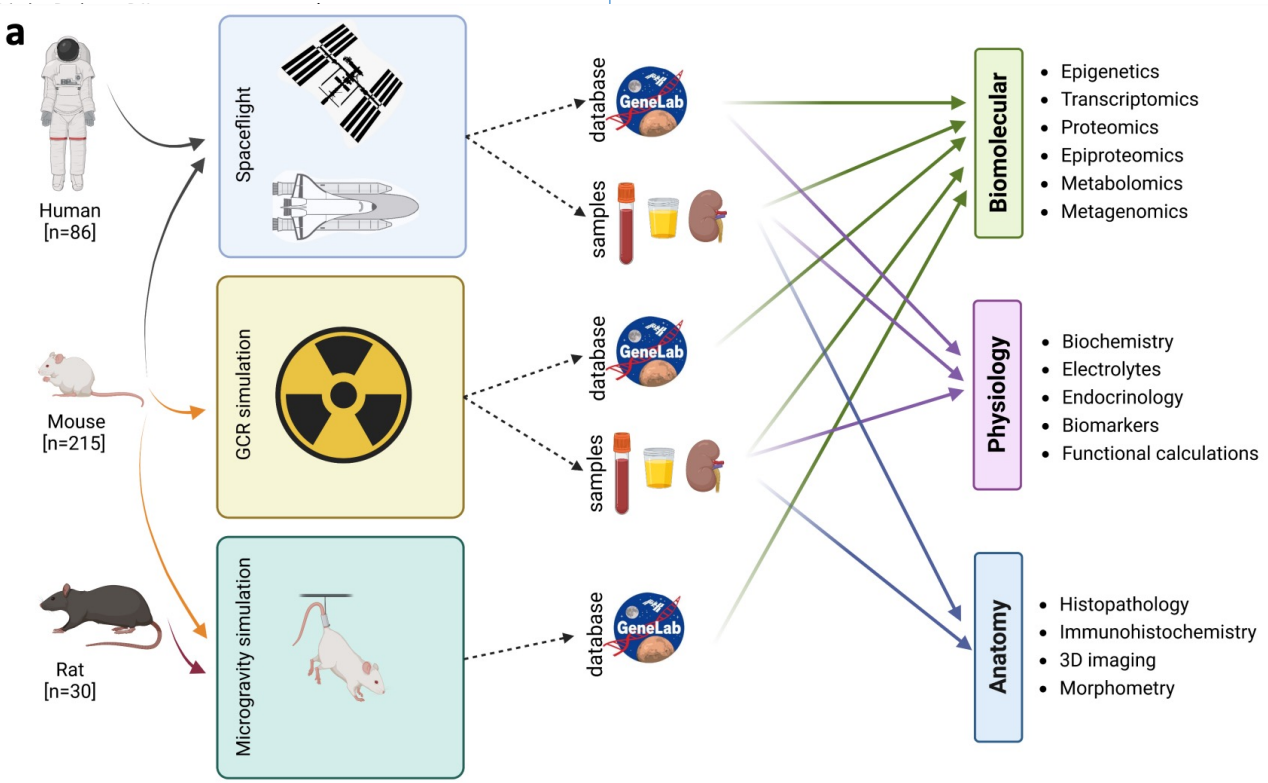
Collaborate on New Studies/Data Mining/Publications

https://www.cell.com/c/the-biology-of-spaceflight

https://osdr.nasa.gov/bio/data/publications.html

Example of Open Science AWG Collaboration and Publication:

Mission	Experimental Condition	Species	Sex	Background	GCR (mG)	Mission Duration (d)	Sample Collection	n	Control	Data Source	Tissue	DNA Meth	RNA	Protein	PTMs	Metab	Clin Chem	Morpho	Microbes
Inspiration4	Spaceflight	Human	Mix	USA	Unknown	3	R+0	4	Paired	Original	P, U			✓		✓	✓		
JAXA	Spaceflight	Human	♂	Japan	Unknown	120	L+120	6	Paired	Original	P		✓						
NASA	Spaceflight	Human	Mix	USA	Unknown	180	L+120	66	Paired	Database	P, U						✓		
Roscosmos	Spaceflight	Human	♂	Russia	Unknown	199	R+1	10											
BNL-1	simGCRsim	Mouse	♀	C57BL/6	500.0	~1.5yr sim	PE+1	10											
	MGsim				N/A	14	PE+1	10											
	simGCRsim + Mgsim				500.0	~1.5yr sim	PE+1	10											
BNL-2	simGCRsim	Mouse	♀	C57BL/6	500.0	~1.5yr sim	PE+1	10											
	MGsim				N/A	14	PE+1	10											
	simGCRsim + Mgsim				500.0	~1.5yr sim	PE+1	10											
BNL-3	simGCRsim	Mouse	♀	C57BL/6	500.0	~1.5yr sim	PE+1	10											
	MGsim				N/A	14	PE+1	10											
	simGCRsim + Mgsim				500.0	~1.5yr sim	PE+1	10											
NSRL-22A	GCRsim	Mouse	♀	C57BL/6	750.0	~2.5yr sim	PE+180	12											
	GCRsim	Mouse	♂	C57BL/6	750.0	~2.5yr sim	PE+180	12											
STS-135	Spaceflight	Mouse	♀	C57BL/6CR	4.7	13	R+0	7											
RR-1	Spaceflight	Mouse	♀	C57BL/6J	5.0	38	L+38	6											
RR-3	Spaceflight						L+41	6											
RR-6	Spaceflight						R+4	9											
							L+75	9											
RR-7	Spaceflight						L+25	5											
							L+75	5											
							L+25	5											
							L+75	5											
RR-9	Spaceflight						R+1	5											
RR-10	Spaceflight						L+28	10											
RR-19	Spaceflight						R+0	8											
RR-23	Spaceflight						R+1	9											
MHU-1	Spaceflight						R+2	6											
MHU-3	Spaceflight						R+0	6											
CNSA	MGsim						R+0	30											



- Renal/Kidney Research
  - 53 datasets
  - 5 continents of collaborators
  - 20 spaceflight missions & ground simulations

(Credit: Keith Siew, Stephen Walsh, et al., 2024)

Engaging With Commercial Entities for Human Data into OSDR

INSPIRATION4

OSDR Number	Samples	Assays
OSD-569	whole blood	Whole Genome Sequencing
		Clonal Hematopoiesis
		Complete Blood Count
		Long & Short Read RNA-seq
		Epitranscriptomics
OSD-570	PBMCs	Single-Nuclei RNA-seq
		Single-Nuclei ATAC-seq
		Single-Cell TCR-seq
		Single-Cell BCR-seq
OSD-571	blood plasma	Proteomics
		Proteomics (EVPs)
		Proteomics (metabolome)
		cfDNA-seq
		cfRNA-seq
OSD-572	skin, oral and nasal swab	Metagenomics
		Metatranscriptomics
OSD-573	dragon capsule	Metagenomics
		Metatranscriptomics
OSD-574	skin biopsies, swab	Spatial Transcriptomics
		Metagenomics
		Metatranscriptomics
OSD-575	blood serum	Comprehensive Metabolic Panel
		Cytokine Biomarker Panel
		Cardiovascular Biomarker Panel
		Immune Biomarker Panel
OSD-630	stool	Metagenomics
OSD-656	urine	Immune Biomarker Panel

- Upcoming Nature Portfolio 'I4/JAXA/SB Package'
- Credit: Chris Mason's Lab; Overbey et al., 2024
- In Peer Review at *Nature*

Just Confirmed!

IIAS International Institute for Astronautical Sciences

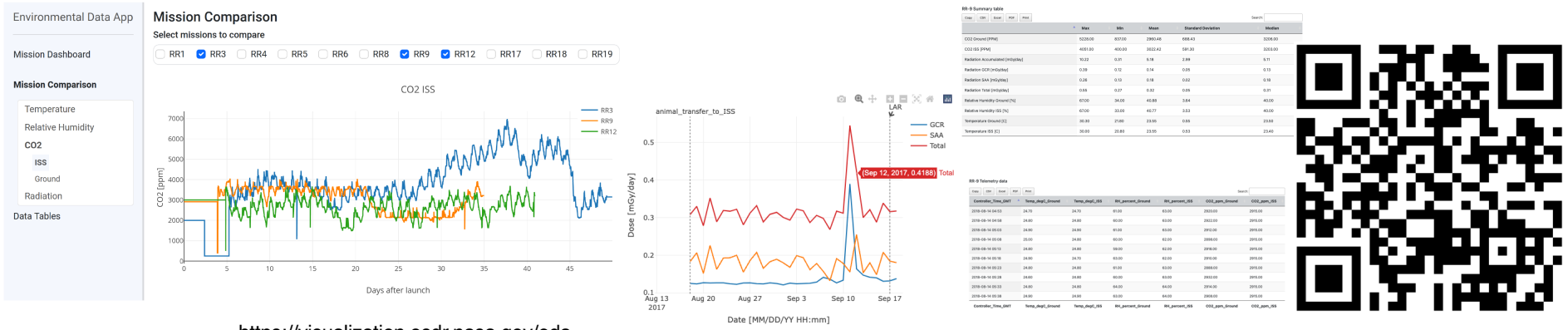


- **Commercial Sub-Orbital**
- **Galactic-05 – Flew Nov. 2**
  - **Astroskin (ECG, systolic BP, heart rate (HR), breathing rate (BR), skin temperature, activity levels (provided by 3-axis accelerometer) and oxygen saturation in blood (SpO2)**
  - **Blood Glucose Monitoring Device**

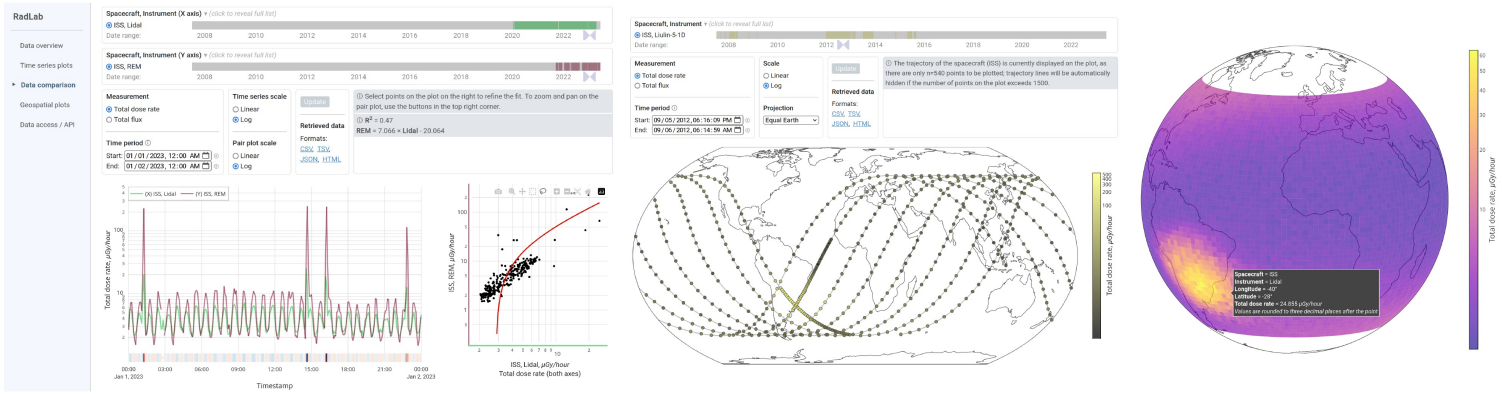
Soon!

- **Credit: Dr. Shawna Pandya**

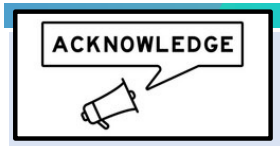
Environmental Data App and RadLab



https://visualization.osdr.nasa.gov/eda



https://visualization.osdr.nasa.gov/radlab



Open Science Data Repository

ALSDA
- Danielle Lopez
- Rachel Gilbert
- Xavier-Lewis Palmer
- Alan Wood
- Sylvain Costes
- Christina Johnson

GeneLab Team
- Amanda Saravia-Butler
- Lauren Sanders
- San-huei Lai Polo
- Kristen Peach
- Sam Gebre
- Ana Uriarte Acuna
- Jamie Bales
- Sylvain Costes
- Kirill Grigorev

AI for Life in Space
- James Casaletto
- Lauren Sanders
- Sylvain Costes

NBISC
- Joey Varelas
- Sigrid Reinsch
- Sam Gebre











Collaborators

ALSDA AWG
249 Members

AI/ML AWG
96 Members

2021 AI/ML Workshop Participants
~100 Participants

I4 Team
C. Mason, E. Overbey

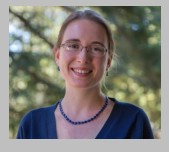
Stakeholders-Management

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

Tuesday, 4:00-5:30p
Location: 2014 - West (Level 2, West, MC)
Title: IN24A-03: Developing Open-Source Training Materials for AI/ML and Space Biological Sciences Using NASA Cloud-Based Data


James Casaletto

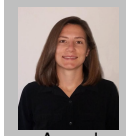
Tuesday, 4:00-5:30p
Location: 2014 - West (Level 2, West, MC)
Title: IN24A-02: Large Pretrained Models for Space Life Sciences


Lauren Sanders

Wednesday, 2:15-2:30p
Location: Exhibit Hall Booth #531
Title: Hyperwall Presentation on the Open Science Data Repository


Sylvain Costes

Friday, 8:30-12p
Location: 160 - South (Upper Mezzanine, South, MC)
Workshop Title: INV51B - FAIR and Beyond: Ensuring the Future of Findable, Accessible, Interoperable, and Reusable (FAIR) Multiomics Data for all Stages of the Sample and Data Lifecycle to Enable Earth and Space Science Innovation and Discovery


Amanda Saravia-Butler


James Casaletto


Lauren Sanders


Sylvain Costes


Ryan T Scott