



Enabling Model Organism and Commercial Astronaut Data Access Through the NASA Open Science Data Repository



Amanda M. Saravia-Butler¹, Lauren M. Sanders², Ryan T. Scott¹, San-huei Lai Polo¹, Danielle K. Lopez¹, Samrawit G. Gebre², and Sylvain V. Costes²

¹KBR, Moffett Field, CA, USA; ²Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA, USA

ABSTRACT

NASA's Open Science Data Repository (OSDR) brings together omics data from NASA's GeneLab project and non-omics data, including physiological, phenotypic, imaging, and behavioral data from NASA's Ames Life Sciences Data Archive (ALSDA) collected from decades of space biology research, providing open and FAIR (findable, accessible, interoperable, and reusable) access of these precious data to scientists worldwide. This rich source of meticulously curated metadata and data from spaceflight and analog studies has been mined by the scientific community resulting in dozens of high impact scientific publications that reveals a complex network of molecular and physiological effects of spaceflight across living systems, from microbes to plants, to mammals. Understanding how these effects translate to the human condition is critical as we move deeper into the era of commercial space travel. However, the integration of data, specifically omics data, from astronauts is particularly challenging due to their sensitive nature. OSDR has risen to this challenge by developing a mechanism to control access to identifiable levels of omics data, such as raw sequence data, while enabling public access to processed, unidentifiable, data and associated metadata that will allow the scientific community to interrogate human astronaut data alongside data from model organisms to begin answering these critical questions.

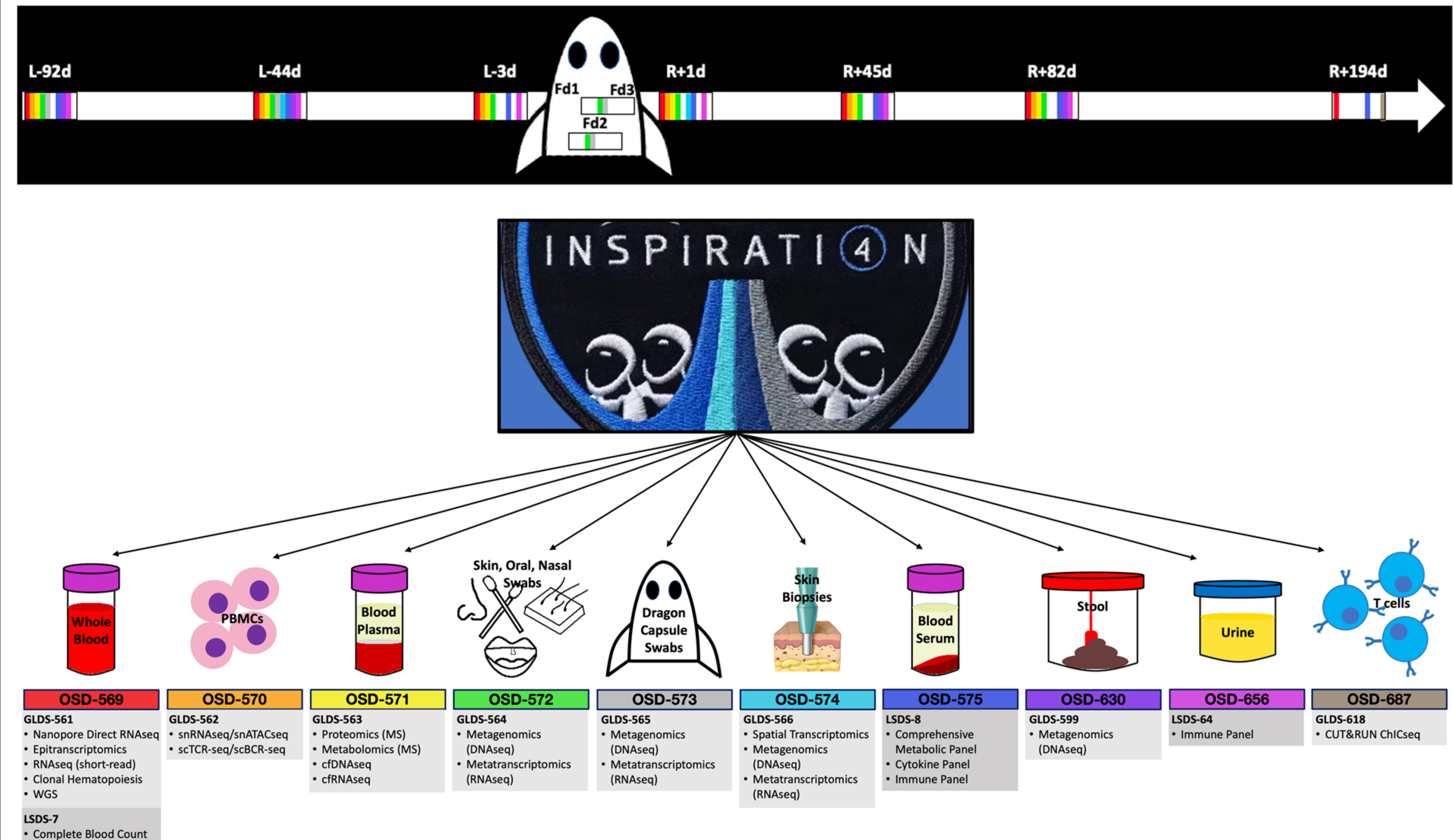
The 2021 SpaceX Inspiration4 (I4) mission collected a comprehensive atlas of biological measurements from four civilian astronauts, providing a wealth of data to characterize the effects of spaceflight on the human body. These data include both non-omics and omics assays such as direct RNA sequencing (RNA-seq), single nuclei ATAC-seq and RNA-seq, metagenomics, proteomics, and comprehensive metabolic and cytokine panels, all of which have been integrated into the OSDR system across no less than 10 studies. Each study has been carefully curated using community-backed OSDR standards for sample and assay level metadata ensuring these data are findable and accessible. In addition to hosting both raw and processed data from the principal investigator team for each assay type, the GeneLab team plans to re-process the I4 omics data using GeneLab's standard processing pipelines. The GeneLab processed data outputs will allow for comparisons across studies on OSDR and enable visualization of these data through the OSDR data visualization platform thereby enabling data reusability and interoperability.

Here we describe the robust privacy and security protocols implemented by OSDR to safeguard sensitive health data from astronauts while facilitating metadata and processed data sharing for research purposes. We further provide a road map for navigating the vast amount of data provided for each I4 study on the OSDR, including experimental design, associated experiments, payloads, and missions, data generation and analysis protocols, and associated scientific articles. Additionally, we illustrate how to interrogate the standardized metadata provided in the sample and assay tables as well as instruction for how to download and access the data. The I4 datasets described here represent the first ever comprehensive collection of commercial astronaut data.

KEEPING PRIVATE DATA SECURE

The workflow for data access involves several steps: 1. A user requests access to restricted files in the OSDR portal. 2. A Data Access Committee reviews the request. 3. A Data Use Agreement (DUA) is generated and approved. 4. A Data Access Request (DAR) form and IRB are submitted. 5. Once approved, data access information is available on the DAR portal, allowing users to download files.

INSPIRATION 4 MISSION DATA ON OSDR



I4 Data Organization On OSDR: I4 mission data are separated into 10 OSD studies based on sample/tissue type as shown above. For each OSD, the associated omics and non-omics data are listed under the GLDS-ID (light gray) and LSDS-ID (dark gray), respectively. The time points samples were collected for each OSD are color-coded in the I4 mission timeline at the top. L=Launch, F=Flight, R=Return, d=Days.

I4 OSDR STATS

- 5 New Configs Created
- 799 Astronaut-Derived Samples
- 1358 Sample/Assay Measurements
- 14.85 Terabytes of Data Added to OSDR

ENABLING FAIR SCIENCE

The OSDR Meta/Data Model (ISA format) is structured as follows:

- investigation:** Study overview, including: Study Description, Factors, Organism(s), Assay(s), Project Info, Acknowledgements, Contact(s).
- sample:** Table containing individual sample-level info including: Source & Sample Names, Characteristics, Parameters, Factors, Protocol References, Uses standardized ontology.
- assay:** Table containing assay-level info for each sample including: Technology Type & Platform, Assay Name, Parameters (e.g. Reagents, Kits, Instruments, Settings, etc.), Associated Data Files, Protocol References, Uses standardized ontology.
- data:** Downloadable study data files: ISA.zip, Molecular-omics, Physiological, Phenotypic, Behavior, Imaging, etc.; Tabular, Text, Images, Video, Sequence, etc.; Raw, Processed, Transformed; Uses standardized formats.

 Screenshots show the OSDR portal interface, including the 'Request Access to Restricted Files' button, the 'Description' page for a specific study, and the 'Assays' table listing various data types and their associated metadata.

OSDR

NEW AVENUES FOR BPS/HRP COLLABORATION



The open access of datasets in NASA's OSDR provides a unique opportunity for the scientific community, as well as citizen scientists and students, to continue using OSDR resources to further unlock profound insights into the consequences of space travel on the human body. Through implementation of security measures to protect sensitive human data, the OSDR seeks to strengthen the science exchange between the Biological and Physical Sciences Program and the Human Research Program, per recommendation 4-1 of the 2023-2032 Decadal Survey, and encourage further sharing and dissemination of astronaut data to provide the scientific community with the resources needed to lay the groundwork for developing targeted mitigation strategies to help withstand the rigors of long-duration spaceflight.

ACKNOWLEDGEMENTS

OSDR AWG Members: ALSDA, GeneLab, NBISC

I4 Team: Weil Cornell Medicine, Lead by: Drs. Elish Overbey and Chris Mason

Funding: This work was funded by the NASA Space Biology Program within the NASA Science Mission Directorate's (SMD) Biological and Physical Sciences (BPS) Division.