THE NASA OPEN SCIENCE DATA REPOSITORY: BIOMEDICAL DATA, ANALYSIS TOOLS, AND INFORMATIC COLLABORATIONS

Ryan T. Scott[1], Danielle K. Lopez[1], Amanda Saravia-Butler[1], Lauren M. Sanders[2], Samrawit G. Gebre[3], Sylvain V. Costes[3]
[1]KBR, Moffett Field, CA; [2]Blue Marble Space Institute of Science, Moffett Field, CA; [3]Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA

Increased biomedical risks and challenges associated with deep space missions require knowledge discovery, health countermeasures, and biomedical support capabilities. Maximally open-access and reusable data is needed by developers, scientists, and engineers to develop these systems. The NASA Open Science Data Repository (OSDR) is a maximally open access and FAIR database (ie., findable, accessible, interoperable, and reusable), and meets various scientific, technical, and operational needs. It offers users and submitters the ability to upload, download, search, share, analyze, cite, and visualize data across 'omics, physiological, phenotypic, payload, hardware, behavioral, bioimaging, video, and environmental monitoring telemetry datasets. OSDR is an expanded database, based upon the successes of NASA GeneLab.

OSDR has >460 studies with datasets covering model organisms to non-NASA human astronauts. There are ~12 datasets from the Inspiration 4 (I4) mission, spanning metagenomics, comprehensive metabolic panels, clonal hematopoiesis, spatial transcriptomics, proteomics, and cytokine panels. In the interest of data privacy, two I4 datasets with raw files relating to the epitranscriptome, and a new request feature is live in OSDR (with a backend review process established) developed from industry norms. OSDR is collecting and curating biomedical human data from a new sub-orbital research flight and is open to more space life science/biomedical submissions from the international and commercial sectors. OSDR also recently began a collaboration with the European Space Agency (ESA) to collect and curate >200 terabytes of human and model organism data. The OSDR submission portal is designed to ingest and curate ~25 'omics and ~50 physiological-phenotypic-imaging assay data types.

Tools available for OSDR users include: 1) an Environmental Data Application to compare radiation, CO2, relative humidity, temperature, and other telemetry across missions and subjects, 2) the RadLab database, a collaboration between NASA, ESA, the German and Italian Space Agencies, and the Bulgarian Academy of Sciences, and 3) a Multi-study visualization tool which enables users to look across and combine 'omics datasets.

There are ~600 volunteer OSDR Analysis Working Group (AWG) members providing feedback on scientific data/metadata standards and collaborating to mine-reuse OSDR in research. OSDR/GeneLab has enabled ~60 publications reusing data as of October 2023.