THE NASA OPEN SCIENCE DATA REPOSITORY: BIOMEDICAL FAIR DATA, ANALYSIS TOOLS, USER COMMUNITIES, PUBLICATIONS, AND DISCOVERIES FOR DEEP SPACE MISSIONS

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 new knowledge discovery, new health countermeasures, and development of novel ecosystems, life support, crop production, and biomedical support capabilities. To meet NASA's Moon to Mars strategic program goals for Human and Biological Sciences, findable, accessible, interoperable, reusable (FAIR), and maximally open-access data is going to be required to enable humanity to thrive in deep space. Indeed, this cornerstone perspective on FAIR and maximally open access data was also recommended in the recent 2023-2032 Decadal Survey from the National Academies of Sciences, Engineering, and Medicine.

The NASA Open Science Data Repository (OSDR) is a maximally open access and FAIR database, and meets various scientific, technical, and operational spaceflight needs. It offers public users and submitters the ability to upload, download, search, share, analyze, and visualize data across 'omics, physiological, phenotypic, behavioral, bioimaging, video, and environmental monitoring telemetry datasets. OSDR includes NASA GeneLab, NASA Ames Life Sciences Data Archive, and the NASA Biological Institutional Scientific Collection. OSDR has >455 studies with datasets from model organisms and 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 have raw FASTQ and FASTA files relating to the epitranscriptome, and a new request feature is live in OSDR (with a backend review process established) which was developed based on industry norms. OSDR also recently began a collaboration with the European Space Agency (ESA) to scientifically curate and make available >200 terabytes of human and model organism space-relevant data. The OSDR submission portal is designed to ingest and curate ~25 'omics assay data types, and ~50 physiological-phenotypic-imaging assay data types, spanning ultrasonography, micro-computed tomography, histology, morphometric photography, rebound tonometry, gait analysis, optical coherence tomography, novel object recognition, flow cytometry, and immunohistochemistry.

A suite of analysis tools are available for OSDR users including: 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, which compiles radiation measurements relevant to human spaceflight and provides tools for accessing and manipulating the data, and 3) a Multi-study visualization tool which enables users to look across and combine GeneLab's omics datasets across different experiments and missions.

There are ~600 volunteer OSDR Analysis Working Group (AWG) members who: 1) provide feedback on scientific standards for reuse (subject and assay metadata; processing pipelines; dataset formats and uniformed structures for machine-readability), and 2) collaborate to mine-reuse OSDR data conducting scientific analysis. OSDR has enabled 60 publications as of September 2023, many directly from AWG collaborations most notably the Cell Press package in 2020. Lastly, there are at least 15 articles which mine OSDR data part of a package of ~50 articles across Nature Portfolio with research stemming from I4, the Japan Aerospace Exploration Agency, NASA Space Biology, and the NASA Human Research Program.