

BEYOND FAIR: ENGAGEMENT, DATA USABILITY, AND OPEN COMMUNITY PRODUCTIVITY THROUGH THE NASA OPEN SCIENCE DATA REPOSITORY

Ryan T. Scott¹, Amanda M. Saravia-Butler¹, Lauren M. Sanders², 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

The FAIR principle (findable, accessible, interoperable, and reusable) governs the storage and sharing of NASA space biology and health data[1]. These guiding principles maximize reuse of data and the reproducibility of scientific findings. The NASA Open Science Data Repository (OSDR; an expansion of NASA GeneLab) was built on the FAIR principles and houses over 500 studies and close to 1000 datasets from decades of space life sciences experiments. OSDR embodies the FAIR principles through data governance that includes mediated, embargoed, and fully open access data.

The FAIR data governance principles were recently proposed to be expanded to encompass a FAIREST framework for assessing research data repositories (FAIR + Engagement, Social connections, and Trust)[2]. FAIREST emphasizes the importance of data repositories engaging with the scientific community and gaining the trust of researchers regarding data quality. Trust also refers to the TRUST principles developed for assessment of digital repositories: Transparency, Responsibility, User Focus, Sustainability, Technology[3].

We present the “Open Science for Life in Space” Analysis Working Groups (AWGs) as evidence regarding the power of engagement, social connections, and trust which has enhanced OSDR’s capabilities and productivity. AWG members engage in two main activities. One, members provide feedback on OSDR scientific standards for data ingestion, curation, and reuse (study, subject and assay metadata; processing pipelines; dataset formats and uniformed structures for machine-readability). Two, AWG members collaborate to mine-reuse OSDR data to conduct scientific analysis. With nearly 800 active members, the AWGs have resulted in 32 publications re-using OSDR data and contributed many papers in two major special issues in *Cell* (2020) and *Nature* (2024). AWGs also serve as networking groups, facilitate social connections between researchers at all levels of experience, and also have a social online ‘Forum’ used to keep members informed on projects and opportunities. This community-centric, productive, and trustworthy data culture has resulted in a broader effect with international space agencies, academics, and the commercial space sector wanting to submit their data to OSDR. Ten studies of Inspiration 4 data were recently publicly released by OSDR, as were some JAXA human data. Coming up soon in OSDR are data submissions from the European Space Agency, Virgin Galactic PIs, and SpaceX Polaris Dawn.

A major benefit of OSDR is the array of standardized and uniformly formatted data (which was developed through AWG member consensus), from which visualization tools, analysis tools, and machine learning models can be built or trained. This talk will cover the Multi-Study Visualization Tool, the Environmental Data Application, RadLab, and a UCSF-NSF funded knowledge graph biomedical health discovery tool ‘SPOKE’ currently being integrated with OSDR. OSDR also provides training programs in bioinformatics and machine learning to improve the scientific community’s awareness of data availability and to boost their ability to perform data analysis.

The increasing engagement of the scientific community and the public with technologies powered by artificial intelligence (AI) heightens the need for data analysis to be transparent. The AI for Life in Space initiative leverages the data products provided in OSDR to train AI models, with an emphasis on explainable and trustworthy AI, which would not be possible without FAIR data and metadata.

Overall, here we will demonstrate the importance for NASA life sciences data repositories to adhere to the FAIREST framework, by providing examples and success stories from different aspects of OSDR.

REFERENCES

[1] Stall, S. et al. (2019) *Nature* 570, 27–29. [2] d’Aquin, M. et al (2023). *Data Intell.* 5, 202–241. [3] Lin, D. et al. (2020) *Sci Data* 7, 144.