Open Science for Life in Space: Data Sharing and Tools for Knowledge Discovery

Ryan T. Scott[1], Lovorka Degoricija[1], Lauren M. Sanders[2], Danielle K. Lopez[1], Samrawit G. Gebre[1], and 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

The fast-growing array of space biological data, which in the past was simply archived after minimal analysis, holds great potential if it can be reorganized and formatted for Open Science. Organizing the data for such analysis is a challenge because of its diverse nature (molecular, cellular, tissue, whole organism, behavior; tabular, imagery).

Open Science is the concept that the more people have access to scientifically curated data, the more knowledge will be gained. This led NASA to start the development of GeneLab in 2015. GeneLab houses spaceflight and space-analog multi-omics datasets from plant, rodent, small animal, and microbial experiments. The success and knowledge gained from GeneLab led to a new alliance of NASA "Open Science Data Repositories" (OSDR), which include the Ames Life Sciences Data Archive (ALSDA) and the NASA Biological Institutional Scientific Collection (NBISC). Both are adopting the GeneLab data system, so data are more findable, accessible, interoperable, and reusable (FAIR). OSDR systems provide users the ability to upload, download, search, share, analyze, and visualize.

Open Science also needs strong confidence in the data, which is gained through building science communities. With ~400 current members, GeneLab and ALSDA formed Analysis Working Groups (AWGs) to provide feedback on processing pipelines, metadata curation standards (for 'omics and phenotypic-physiological-behavioral assays), and to collaborate in effectively reusing data. The AWG also led to the development of the Radiation Biology Ontology (RBO), ensuring radiation metadata are efficiently captured, connected, and interoperable. Feedback from the AWG provided design input toward the new single point-of-entry data submission portal for all investigators to submit, curate, and share their research data.

Space biological data is now maximally open access, collected-curated with rich metadata, and formatted for interoperability to enable systems biology, meta-analysis, knowledge graphs, machine learning, modeling, and other reuse approaches. With potential for further federation of OSDR for data mining with traditional biological and medical databases (NIH, NCI, EBI, etc.), a new era for space biology has begun to support the knowledge discovery necessary for Lunar and Martian missions.