

Celebrating 30 years of access to NASA Space Life Sciences data

Lauren M. Sanders¹, Danielle K. Lopez², Alan E. Wood², Ryan T. Scott², Samrawit G. Gebre¹, Amanda M. Saravia-Butler², and Sylvain V. Costes^{1,*}

¹Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA 94035, USA

²KBR, Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA 94035, USA

*Correspondence address: Sylvain V. Costes, Space Biosciences Research Branch, NASA Ames Research Center, Moffett Field, CA 94035, USA.

E-mail: sylvain.v.costes@nasa.gov

Abstract

NASA's space life sciences research programs established a decades-long legacy of enhancing our ability to safely explore the cosmos. From Skylab and the Space Shuttle Program to the NASA Balloon Program and the International Space Station National Lab, these programs generated priceless data that continue to paint a vibrant picture of life in space. These data are available to the scientific community in various data repositories, including the NASA Ames Life Sciences Data Archive (ALSDA) and NASA GeneLab. Here we recognize the 30-year anniversary of data access through ALSDA and the 10-year anniversary of GeneLab.

Keywords: space biology, NASA, data, open science, data archive, FAIR, omics

Background

Space travel involves inherent hazards, including altered gravity, increased exposure to radiation, confinement, and distance from Earth [1]. Characterizing and mitigating these risks is the focus of the Thriving in Deep Space (TIDES) initiative, part of NASA's Space Biology Program within the Biological and Physical Sciences Division. NASA's upcoming Moon to Mars Missions will face increased health hazards [2]. TIDES aims to meet these challenges by investigating biological responses to the space environment and utilizing this knowledge to enhance biotechnological systems essential for human health and performance beyond low Earth orbit. Open and FAIR [3] (Findable, Accessible, Interoperable, Reusable) data are paramount for supporting these efforts by enabling scientific collaboration and gaining new knowledge. This view of data stewardship in the modern era is embraced by NASA and its Science Mission Directorate (SMD), which launched the Transform to Open Science (TOPS) initiative and released Science Policy Directive SPD-41a [4].

Data collection and archiving has been a key principle at NASA for decades. Here we recognize the 30-year anniversary in 2024 of the NASA Ames Life Sciences Data Archive (ALSDA), a database of experiment descriptions, payloads, missions, telemetry, results, and bioimaging data mostly from NASA-funded life sciences investigations. Similarly, 2024 marks the 10-year anniversary of NASA GeneLab, the omics repository for space biology and space-relevant experiments. In 2021, ALSDA and GeneLab merged to form the NASA Open Science Data Repository (OSDR), unifying these precious data resources under 1 platform [5].

History and 30-Year Anniversary of the NASA ALSDA

The work to create the NASA Life Sciences Data Archive (LSDA) began in 1990 by teams at Johnson Space Center (JSC), Ames Re-

search Center (ARC), and Goddard Space Flight Center (GSFC). It was established and fully funded in FY94 with archive centers at ARC, Kennedy Space Center (KSC), JSC, and GSFC. ARC focused on nonhuman data, JSC on human data, KSC on plant data, and GSFC provided Master Catalog functions. In 1995, LSDA was tasked with providing representation for Space Life Sciences to the Consultative Committee for Space Data Systems (CCSDS) and International Standards Organization (ISO) efforts to develop a reference model for the emerging challenge of archiving digital data. For 10 years, ALSDA led the Space Life Sciences effort for this ambitious endeavor. The resulting CCSDS/ISO product "The Open Archival Information System (OAIS) Reference Model" is the most widely adopted archival standards framework [6].

The repository was designed around experiments with relational links to personnel, hardware, payloads, missions, and so on. ALSDA was charged with collecting all ARC-funded life sciences experimental data and metadata to allow scientists to perform retrospective analyses across missions, experiments, disciplines, and research subjects/species. ALSDA began capturing hardware information, audio/visual media, images, slides, mission and payload information, experiment descriptions, raw experimental data, and raw telemetry downlink and ground control data in both analog and digital formats. Years later, ALSDA was tasked with creating and managing the biospecimen storage facility, capturing, cataloging, and disseminating biospecimens remaining from flight and ground experiments.

Faced with growing demands for open-access and high-quality data, ALSDA was unable to upgrade systems until it merged with the modern GeneLab data system [7] in 2021. As a result, this integration now maximizes science-return on all legacy and future data collections. In 2024, ALSDA remains responsible for curating, archiving, and making available space-relevant phenotypic, physiological, bioimaging, behavioral, and environmental telemetry datasets. Collection of scientific reports and publications was

streamlined to other program-level repositories such as Taskbook and PubSpace.

History and 10-Year Anniversary of NASA GeneLab

The early 2000s saw a paradigm shift in biological research, wherein costs of high-throughput sequencing dropped precipitously and large omics datasets were generated at an increasing rate. “Omics” refers to a group of biological disciplines that study comprehensive datasets of related molecules in an organism, such as genomics (DNA), proteomics (proteins), metabolomics (metabolites), and transcriptomics (RNA), to understand complex biological systems and processes.

In 2011, the National Academies’ Decadal Survey in Life and Physical Sciences Research emphasized the importance of omics, advocating for informatics technologies and a multilevel, systems biology approach for space biology research [8]. In keeping with the 2011 Decadal Survey recommendations, NASA initiated the GeneLab project in 2014. From the 2014 GeneLab Strategic Plan, the goal was to allow maximally open and unrestricted access to omics data generated from biological experiments in space, as well as associated experimental metadata.

A critical first step for GeneLab was to establish meticulous normalization standards to provide “rich metadata” using the ISA (Investigation, Study, Assay) model framework for data curation, thereby enhancing the FAIRness of these data [3]. In 2018, GeneLab established the Analysis Working Groups (AWGs), composed of data repository users and scientists. Together, they defined optimal analytical workflows for data, resulting in consensus bioinformatics pipelines to process and standardize all raw omics data, significantly enhancing dataset reusability by eliminating the need for downloading terabytes of raw sequencing data or having specialized bioinformatics expertise. Building on this foundation of standardized metadata and processed data, GeneLab recently released a visualization portal enabling users to effectively visualize experimental processed omics data.

Over the past decade, GeneLab has evolved into a sophisticated data system with a single point of entry for scientists to upload their data, expedite the curation of new datasets, search the database, download raw and processed data via a web browser or API, and visualize data using state-of-the-art tools. GeneLab is also fostering a vibrant scientific community through its AWGs, which regularly conduct meta-analyses, make new discoveries, and enhance the repository. These groups have been producing independent publications for many years through the reuse of data, culminating in the release of 2 major space biology paper collections [1, 9]. Additionally, GeneLab now serves as the omics repository for the European Space Agency’s SciSpacE research program, integrating European data through NASA’s portals.

The Open Science Data Repository: Best of Both Worlds

The integration of ALSDA with GeneLab [5] has resulted in the creation of OSDR, which is built on GeneLab’s robust framework and incorporates data and metadata from both databases. This integration has improved data accessibility and reusability through new tools and workflows, enabling the merging of multimodal and multihierarchical data from spaceflight experiments for enhanced knowledge discovery and mission support. The consolidation has standardized and made data from ALSDA searchable

and maximally openly accessible, while also enhancing GeneLab’s omics data by linking it transparently with non-omics data from the same experiments. Including over 20 years of ALSDA’s legacy data, along with valuable payload, mission, and hardware records, OSDR now hosts over 500 studies across all life forms, from microbes to animals, plants, and humans, each with curated metadata and accessible raw and processed data files. Additionally, the repository contains over 300 records related to payloads, missions, and hardware.

The AWG ecosystem has rapidly expanded, gaining over 300 new members in the past year and fostering the creation of new groups that contribute valuable insights to OSDR. Notably, subject matter experts from the ALSDA AWG have helped establish minimum data requirements and metadata standards for over 60 assay types on OSDR, many of which previously lacked standards. Additionally, the new artificial intelligence (AI)/machine learning AWG has been instrumental in enhancing the readiness of OSDR data for machine learning and AI applications. This group also played a key role in summarizing a NASA workshop focused on the use of AI to enable self-driving laboratories, automated science, and precision space health in future space missions [10].

Conclusions

A paradigm shift has occurred in modern space biosciences, where reliable data reuse is now essential, as well as science being interdisciplinary, inclusive, and maximally transparent. To enable large communities of scientists on Earth to participate in spaceflight discovery, NASA plays a leading role in this shift through a strong open science program. NASA OSDR is a growing star in this ecosystem, with the integration of 2 critical databases, and the scientific output shows that its whole is greater than the sum of its parts. The OSDR AWG is a thriving collaborative community that utilizes OSDR’s rich data and metadata to make new discoveries in space biology and health.

This merger marks a new era, epitomizing the maturity of open science in space biology. OSDR is committed to rapidly expanding its data through collaborations with other space agencies, institutions, and companies. By leveraging AI, OSDR also aims to enhance accessibility and reproducibility while streamlining the processes of data ingestion and interpretation.

Abbreviations

AI: artificial intelligence; ALSDA: Ames Life Sciences Data Archive; ARC: Ames Research Center; AWG: Analysis Working Groups; CCSDS: Consultative Committee for Space Data Systems; FAIR: Findable, Accessible, Interoperable, Reusable; GSFC: Goddard Flight Center; ISA: Investigation, Study, Assay; ISO: International Standards Organization; JSC: Johnson Space Center; KSC: Kennedy Space Center; LSDA: Life Sciences Data Archive; OAIS: Open Archival Information System; OSDR: Open Science Data Repository; SMD: Science Mission Directorate; TIDES: Thriving in Deep Space; TOPS: Transform to Open Science.

Acknowledgments

We would like to extend our sincerest thanks to both the previous and current teams, as well as to all AWG members and the broader scientific community around OSDR, GeneLab, and ALSDA. Their contributions, dedication, and unwavering commitment to our shared success are essential.

Author Contributions

Lauren M. Sanders (Writing—original draft [lead]), Danielle K. Lopez (Conceptualization [lead]), Alan E. Wood (Writing—review & editing [equal]), Ryan T. Scott (Writing—review & editing [equal]), Samrawit G. Gebre (Writing—review & editing [equal]), and Amanda M. Saravia-Butler (Writing—review & editing [equal]), Sylvain V. Costes (Conceptualization, Writing—review & editing [equal]).

Funding

OSDR is funded by the Biological and Physical Sciences (BPS) Space Biology Program within the NASA Science Mission Directorate (SMD) and the NASA Human Research Program (HRP).

Data Availability

All data are available on the NASA Open Science Data Repository: <https://osdr.nasa.gov/bio/>.

Competing Interests

The authors declare that they have no competing interests.

References

1. Afshinnekoo E, Scott RT, MacKay MJ, et al. Fundamental biological features of spaceflight: advancing the field to enable deep-space exploration. *Cell* 2020;183:1162–84. <https://doi.org/10.1016/j.cell.2020.10.050>.
2. Costes SV, Gentemann CL, Platts SH, et al. Biological horizons: pioneering open science in the cosmos. *Nat Commun* 2024;15:4780. <https://doi.org/10.1038/s41467-024-48633-2>.
3. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. <https://doi.org/10.1038/sdata.2016.18>.
4. NASA. Science information policy. 2021. <https://science.nasa.gov/researchers/science-information-policy/>. Accessed 1 July 2024.
5. Scott RT, Grigorev K, Mackintosh G, et al. Advancing the integration of biosciences data sharing to further enable space exploration. *Cell Rep* 2020;33:108441. <https://doi.org/10.1016/j.celrep.2020.108441>.
6. Reference Model for an Open Archival Information System (OAIS). The Consultative Committee for Space Data Systems. 2002. <https://public.ccsds.org/Pubs/650x0b1s.pdf>. Accessed 1 July 2024.
7. Berrios DC, Galazka J, Grigorev K, et al. NASA GeneLab: interfaces for the exploration of space omics data. *Nucleic Acids Res* 2021;49:D1515–22. <https://doi.org/10.1093/nar/gkaa887>.
8. National Research Council. Recapturing a future for space exploration: life and physical sciences research for a new era. Washington, DC: The National Academies Press; 2011. <https://doi.org/10.17226/13048>.
9. Nature Portfolio. Space Omics and Medical Atlas (SOMA) across orbits. 2024. <https://www.nature.com/immersive/d42859-024-00009-8/index.html>. Accessed 23 July 2024.
10. Scott RT, Sanders LM, Antonsen EL, et al. Biomonitoring and precision health in deep space supported by artificial intelligence. *Nat Mach Intell* 2023;5:196–207. <https://doi.org/10.1038/s42256-023-00617-5>.