Report of NASA Workshop on Artificial Intelligence & Modeling for Space Biology

By Workshop Organizers: Ryan Scott, Lauren Sanders, Sylvain Costes
Date: July 13, 2021
Contact: lauren.m.sanders@nasa.gov

The NASA Workshop on Artificial Intelligence (AI) and Modeling for Space Biology was held virtually on June 24-25, 2021. The workshop was supported by the NASA Space Biology Program within the Biological and Physical Science (NASA-BPS) Division part of the NASA Science Mission Directorate (NASA-SMD).

The purpose of the workshop was to gather a cohort of NASA-external AI and Modeling subject matter experts (SMEs), and develop a vision of AI and Modeling recommendations for space biology in the next decade. As part of the overall purpose, the content gained from the workshop is now being prepared for a journal review-workshop summary article, as well as a package of White Papers to be submitted to the NASEM-CBPSS for consideration in their 'Decadal Survey' 2023-32.

A total of 111 participants attended the workshop between the two days. A numerical summary of participants across both days is shown in Table 1. Attendees included NASA-related space biology and human space health researchers, along with the AI-Modeling SME cohort from industry and academia. For the great majority of SMEs, this was their first exposure to spaceflight science. The breadth of institutions represented by the SMEs was expansive, including Caltech, Johns Hopkins University, Google, Mayo Clinic, Stanford University, Lawrence Berkeley National Laboratory, UC San Francisco, University of Colorado-Anschutz Medical Campus, San José State University, University of Texas San Antonio, Amazon Web Services, Michigan State University, Rutgers New Jersey Medical School, Harvard University, Ohio State University, Weill Cornell Medicine, University of Minnesota Medical School, Salk Institute for Biological Studies, Baylor College of Medicine, Icahn School of Medicine at Mount Sinai, University of Alabama Birmingham, and the University of North Florida.

Day	Participants		
	NASA-related	External SMEs	Total
Thursday, June 24	54	48	105
Friday, June 25	39	42	83

Table 1. Workshop participant summary.

The first day of the workshop consisted of talks and breakout sessions primarily meant to educate the AI-Modeling SME cohort regarding: (1) the long-term required types of biological and health capabilities needed for Lunar, Martian, and deep space missions, (2) a status of relevant data repositories, their content-structure, and overall workflow of the current data resources to be mined, (3) current space-relevant biological data science and AI projects, (4) the unique statistical, data volume, cross-comparison, and logistical challenges of space biology science and data.

The workshop discussion focused on key 'central domain topics' which included:
• AI and Modeling for Knowledge Discovery: 'Omics and other Space Biological Data
• AI Applications in Imaging Space Biology Research Data (including Behavioral Analysis AI Tools for Space Data)
• Precision Medicine Utilization of AI
• Data Collection through Wearables, Sensors, Monitoring Hardware Systems and Integration with AI and Modeling Power
• Space Health Risk Predictions through AI, Modeling, Network Analyses
• Spaceflight Countermeasure Predictions Utilizing AI, Modeling, and Network Analyses
• AI Applications for Microbiology and Synthetic Biology
• AI Techniques and Translational Science Across Model Organisms and Species Towards Human Health

The workshop started with keynote talks given by Dr. Mike Snyder, a 'Twins Study' Investigator at the intersection of health monitoring and spaceflight multi-omics analyses (Stanford), and NASA Astronaut Dr. Kate Rubins, a molecular biologist who was the first person to sequence DNA in space.

The first day also included four general-AI and biological-health educational talks given by Drs. David Van Valen, Greg Corrado, Amina Qutub, and Casey Greene. These were organized to inform the NASA-related participants of the changing landscapes of biological sciences and health with the advances in AI, platforms, computer sciences, and statistics.

A concurrent session was then held for the external cohort of SMEs to hear from NASA leaders and PIs regarding the big vision of space biology requirements towards the Moon and beyond, as well as the statuses of data repositories, the available data to be mined, along with current space-related data science projects.

Participants then were assigned to a first Breakout Session, with the rooms discussing knowledge and technology gaps in space biology AI research by answering questions such as: *"What is required (and how to go about) to ensure that the quality and structure of data collected from NASA's science community PIs will enable the application of AI technologies?"* and *"How best should NASA leverage the power of AI in large volumes of big-data, particularly image data, to accelerate scientific knowledge discovery?"*

Discussion centered around the need for standardized and automated data acquisition systems to generate AI-ready data, data sharing incentives and guidelines, and data

analytical tools to incorporate multiple data types. A data life cycle was developed with common themes across all sessions and specific areas of needed development, as shown in Figure 1. The findings from this discussion were formalized into a slide deck at the end of the first day.

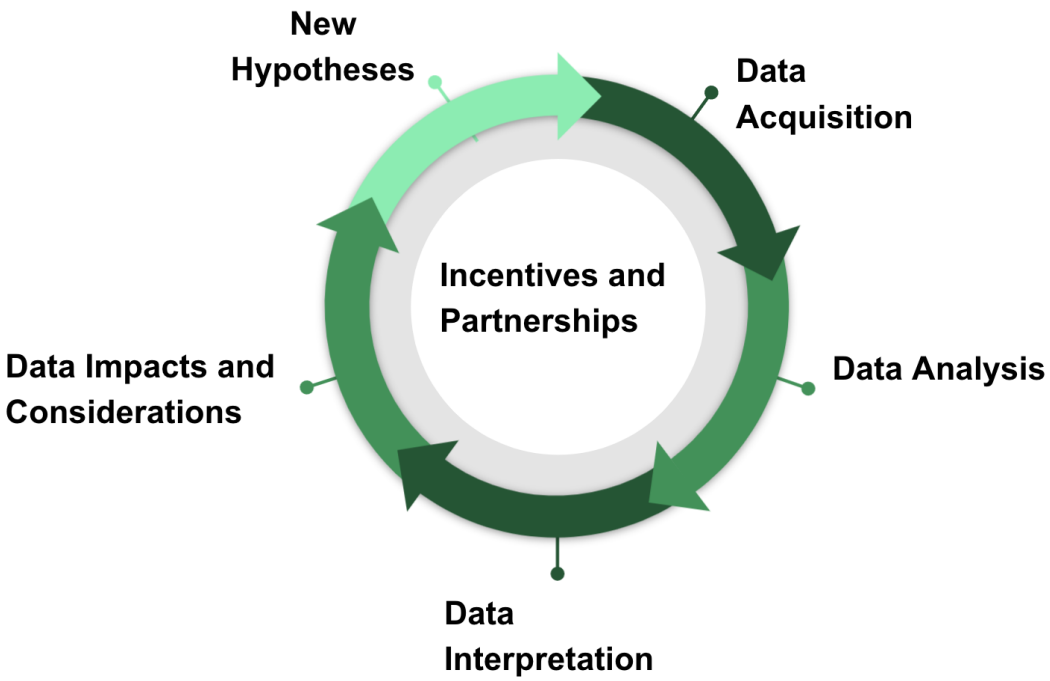

Figure 1. Data life cycle and areas of needed development.

The second day began with a closed session in which the AI SMEs gathered to develop a specific roadmap for AI in space biosciences for the next decade. This session was attended by approximately 35 AI and Modeling experts from industry and academia. Several concrete recommendations emerged from this discussion, including the implementation of AI-biolabs with AI-assisted experimental design and automation, emphasis on longitudinal data collection, investment in an AI model "zoo" and benchmark datasets for space biology challenges, and development of an automated data sharing manager and workspace. The data life cycle was updated with recommendations and ideas as shown in Figure 2.

A second Breakout Session followed, attended by ~60 participants, in which each group worked to identify the top 5 priorities for next-decade development from the SME recommendations.

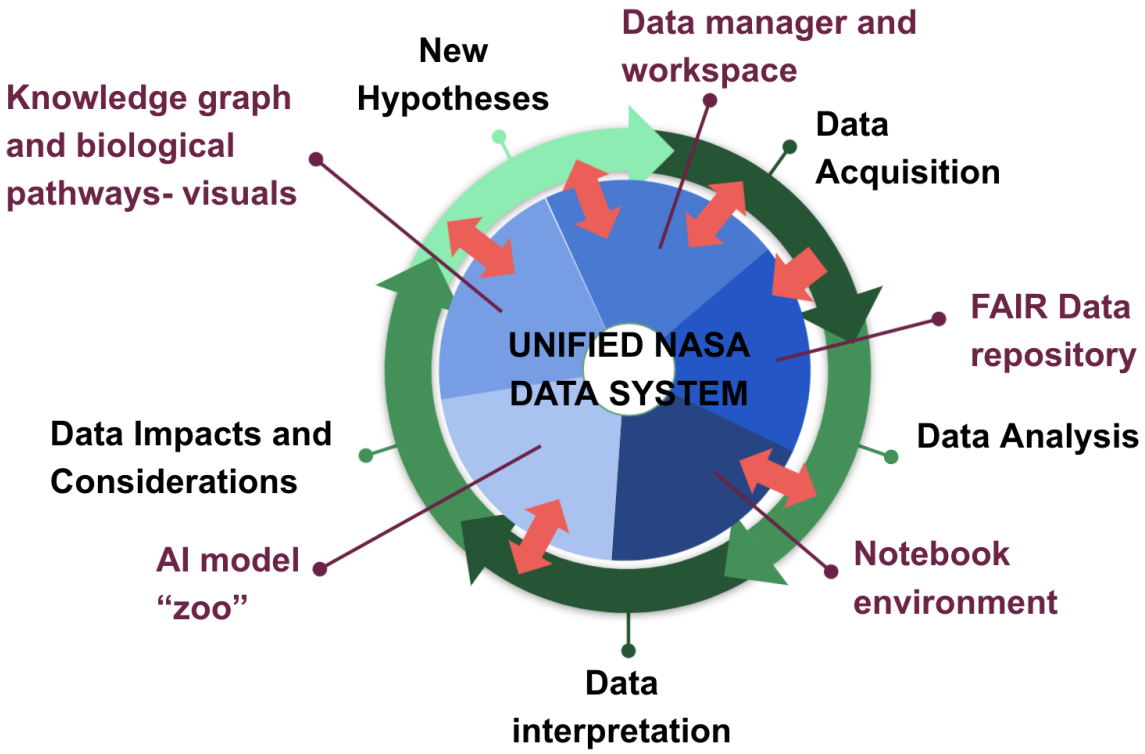

Figure 2. Future NASA Open-AI data life cycle.

The second day closed with an extended writing session with approximately 55 participants, resulting in a comprehensive draft manuscript summarizing findings from the workshop. This manuscript is being reviewed by the organizing committee and prepared for submission to Nature Machine Intelligence as a review-workshop recommendations article. Concurrently, the content is being collated into a group of White Papers to be submitted to the NASEM-CBPSS for the Decadal Survey. More information on the Decadal Survey can be found here. Drafts of both the White Papers and Journal Article will be sent to authors for review prior to submission.