Increasing Data Discovery and Re-Use: The Space Life Sciences Ontology

Daniel C. Berrios[1], Sara C. Jorgensen[2,4], Rachel K. Shoop[3,4], Carlos De Los Santos[4], Danielle K. Lopez[1,3], Alan E. Wood[1,3], Gregory D. Eley[3,4], Kathryn C. Richards[3,4], Olga V. Kieschnick[4], Sigrid S. Reinsch[1], John H. Dunn[4,5], Jessica A. Keune[4], Diedre Thomas[3,4]

[1]NASA Ames Research Center; [2]Aegis Aerospace; [3]KBR Wyle Services, LLC; [4]NASA Johnson Space Center; [5]Universities Space Research Association

INTRODUCTION

Two of the most important goals of adoption of the FAIR principles are:

- Increasing the ability of agents to find data
- Supporting data interoperability, facilitating their re-use.

Achieving these goals for space life sciences research is even more pressing, given the relatively expensive and scarce nature of these data. We have reported in the past on the progress made by exemplar life sciences data systems towards implementing FAIR, showing gaps particularly in the "interoperability area" of the principles [1]; the lack of common conceptual models for space life science research is one reason for this gap. There were few available resources that define, annotate, categorize or otherwise relate various kinds of metadata describing the acquisition, nature, and intent of investigational space life sciences data.

METHODS

NASA is working with the Open Biological and Biomedical Ontology Foundry (https://obofoundry.org/) to develop the Space Life Science Ontology (SLSO) that is intended to support archival and other kinds of systems that operate using these data. The scope of the ontology includes concepts regarding those aspects of investigation design and execution specific or unique to space environments, such as types of specialized equipment, operating organizations, and documentation. The ontology is continually being developed and published to the life science community (https://github.com/nasa/LSDAO/).

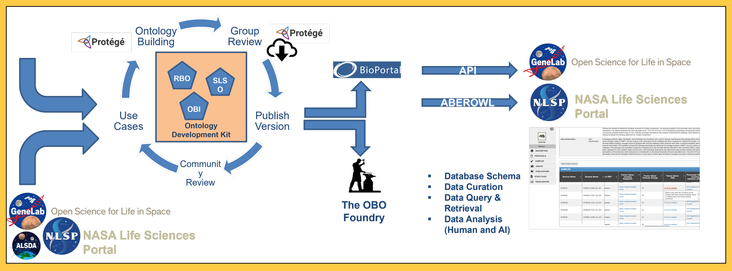

Fig. 1. Development cycle of the SLSO.

RESULTS

The SLSO newly and uniquely defines 30 types (classes), 90 properties, and 14 relations specific to space life sciences metadata. In addition, the SLSO reuses (imports) some 2,360 types (classes), 49 properties, and 393 relations from other ontologies that are relevant to these kinds of metadata.

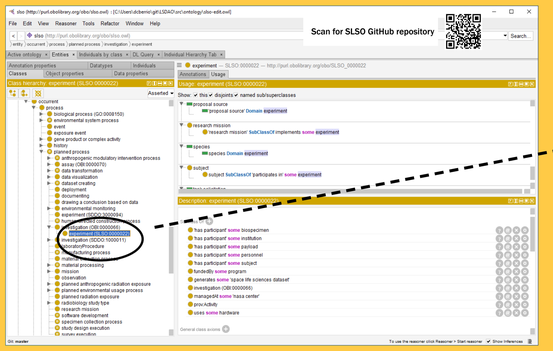

Fig. 2. The SLSO as viewed in the Protege Ontology Editor.

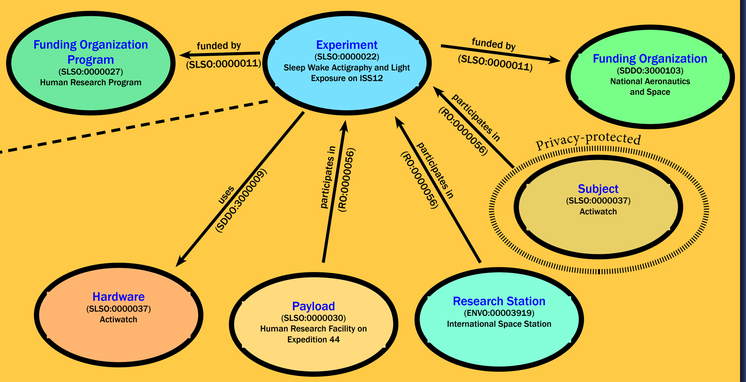

Fig. 3. An example knowledge graph constructed using types and relations from the SLSO and other OBO Foundry ontologies.

SUMMARY

In addition to its role as a common conceptualization for space biomedical research activities, the SLSO can also be used to provide automated support for traditionally difficult and expensive activities such as

- Data curation
- Cross-system data integration
- Data analysis
- Enhancing Large Language Models [2]

References

[1] Berrios DC, Beheshti A, Costes SV. FAIRness and Usability for Open-access Omics Data Systems. AMIA Annu Symp Proc. 2018 Dec 5;2018:232-241.
[2] Allemang D, Sequeda J. Increasing the LLM Accuracy for Question Answering: Ontologies to the Rescue! https://doi.org/10.48550/arXiv.2405.11706

IN MEMORIAM

John H. Dunn

Loved by all of us who worked with him over the last 8+ years, John was the manager and senior engineer (USRA) who led development of the Insight Data Science Platform underlying the NASA Life Science Portal. John's dedication to his team and to his NASA projects resulted in good friendships and great results. John received many awards during his time with USRA and NASA, including the SMA Crystal Ball Team award, the NASA Group Achievement Award, and the USRA President's Innovation Award.