



Ontology Development to Support Open Science and Data Interoperability in Space

Dan Berrios MD PhD MPH
NASA Ames Research Center



Open Science



- Science cannot be termed “open” unless its conduct is transparent
 - Metadata transparency means conveying what was done clearly and uniformly
 - Unambiguous and richly-annotated attributes, values
 - Community-developed and –maintained (open-source) terminology models
 - Data transparency means using open standards for data whenever possible
- Transparency enables scientific reproducibility
 - Data cannot be reproduced if the context in which it is generated is not well understood

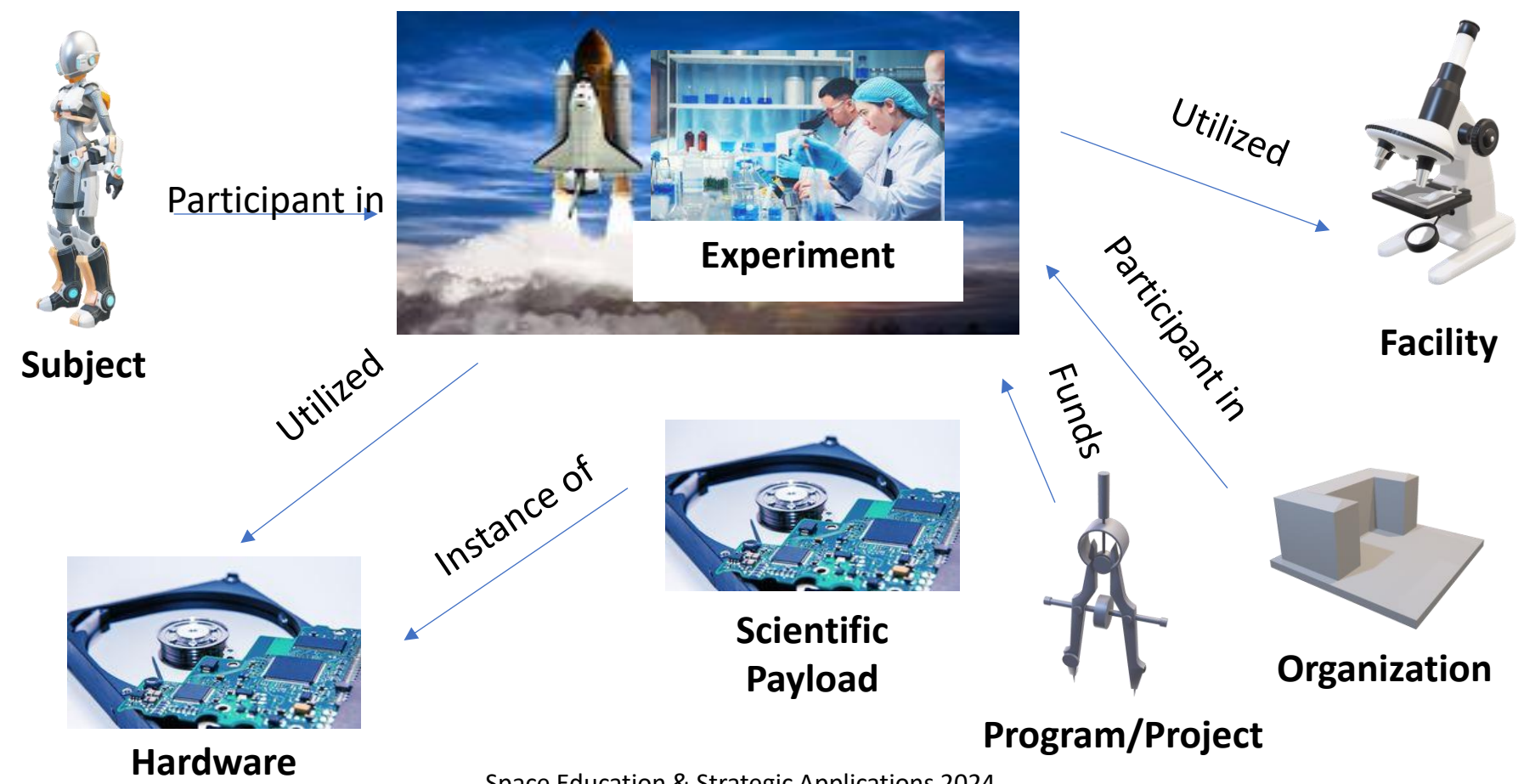


Metadata Transparency: The Problem



- Lack of uniformity of metadata across biomedical data system (**metadata harmonization**)
- Failure to capture correspondences between metadata (i.e., so-called **linked data**)
 - “This specimen in this experiment is a sample of that organism in that experiment”
 - “This instrument used in this experiment is the same as that instrument used in that experiment”
- Extensively linked metadata constitute a **Knowledge Graph**
- Metadata harmonization and linked data are **key features** of **FAIR systems** and **both are evaluated in various nascent FAIR Interoperability Metrics**
- Metadata harmonization and linked data are both foundations for **Open Science**, in that they enhance transparency of investigations, support data interoperability, and facilitate scientific reproducibility.

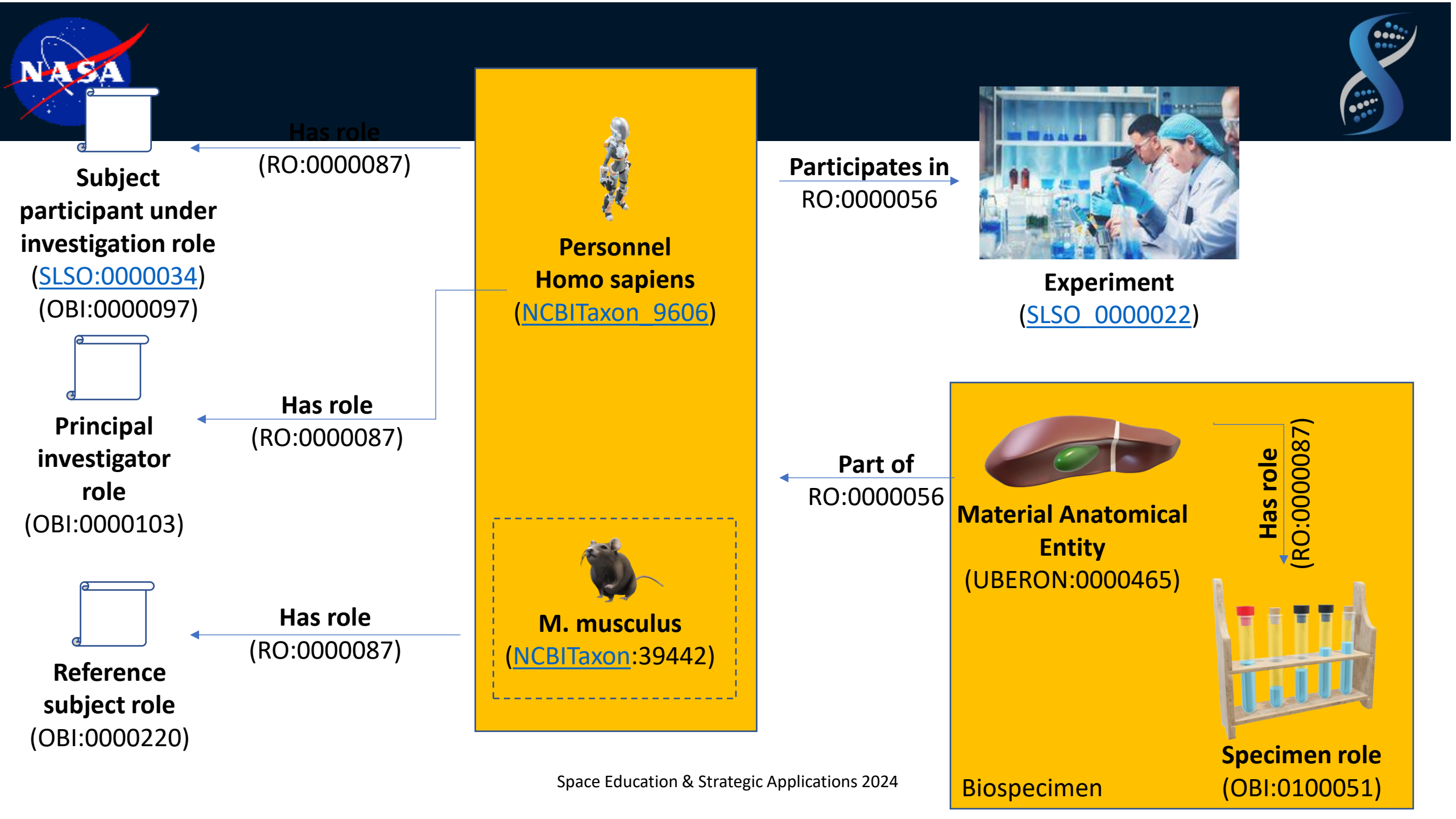
- A collection of concept instances organized as a graph with instance properties (relationships) forming the edges





What is a FAIR-compliant Knowledge Graph?

- A Knowledge Graph in which nodes and edges are each drawn from a “broadly applicable language for knowledge representation” (**ontologies**)
- **Biomedical Ontology** development efforts have been accelerating in the last 10 or so years, **enabling KG implementation**
- **OBO Foundry (~ 200 ontologies)**
 - OBI Ontology for Biomedical Investigations
 - GO Gene Ontology
 - ENVO Environment Ontology
 - **RBO Radiation Biology Ontology (2021)**
 - CHEBI (Chemical Entities of Biological Interest)
 - ENVO (updates in progress for space environment concepts)
 - TaxRank (updates in progress for missing definitions of strain, cultivar, other subspecies)
 - **SLSO: Space Life Sciences Ontology (SLSO) (2024)**
- The next slide shows an exemplar FAIR-compliant Knowledge (sub-)Graph

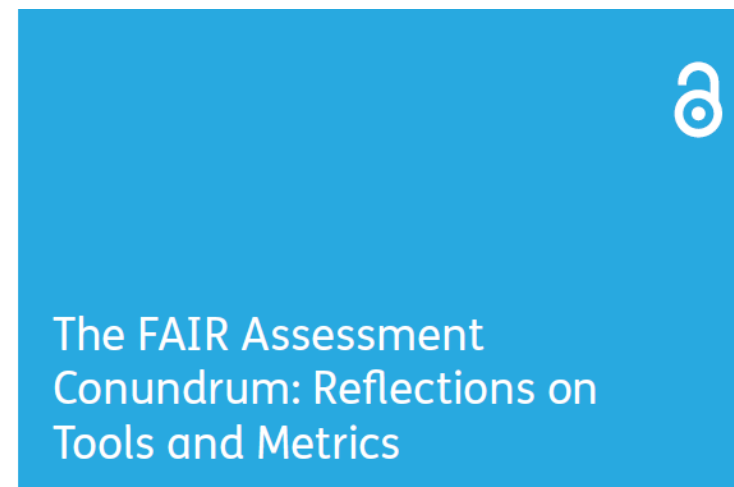




Measuring Data Interoperability per FAIR



- FAIR: Interoperability
 - I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
 - I2: (Meta)data use vocabularies that follow the FAIR principles
- Of 1180 FAIR metrics studied:
- 100 measure interoperability
- Many of these 100 metrics imply or require the use of **formal ontologies for metadata**



ESSAY

ubiquity press

LEONARDO CANDELA

DARIO MANGIONE

GINA PAVONE

*Author affiliations can be found in the back matter of this article

<https://doi.org/10.5334/dsj-2024-033>

Integrated Ontology/Schema Development for NASA Data Repositories: Modularized, with Limited Scope & Specified Governance

Scope

Governance

Space Sciences
ES,AP,BPS,HR,
PS,HP

NASA & Partners

Space Life Sciences

NASA HRP, BPSD, OBO

Radiation Biology

Consortium

NASA, EU (STOREDB),
Univ. Wash., OBO

