

SSBD Ontology: Toward Global Interoperability and Harmonization of Bioimaging Metadata

Yuki Yamagata¹², Koji Kyoda³, Hiroya Itoga³, Emi Fujisawa³, Shuichi Onami¹³

¹ RIKEN R-IH, ² RIKEN BRC, ³ RIKEN BDR

Contact Us

yuki.yamagata@riken.jp

BACKGROUND

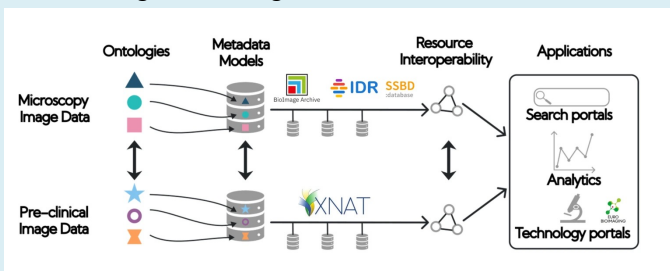
- Advanced bioimaging generates large datasets, but inconsistent metadata limits reuse.
- Repositories like BIA and IDR struggle with interoperability.

OBJECTIVE

To develop SSBD Ontology : a framework for improved metadata and supports global harmonization efforts of the foundingGIDE project

Role of ontology for metadata

- Ontology provides a controlled vocabulary and schema across different datasets and to facilitate data integration and analysis.
- Facilitating data integration and across diverse datasets.
- Enhancing data sharing and reuse.

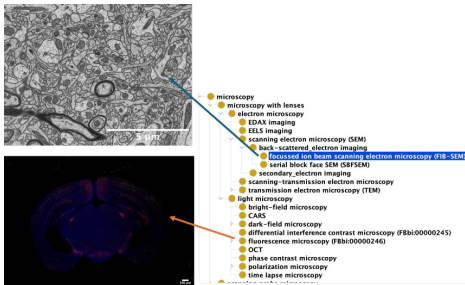


- Repository** : Minimal metadata for rapid publication
 - SSBD:repository, BioImage Archive
- Added-Value DB** : Rich metadata for enhanced data value.
 - SSBD:database, IDR

USE CASE

Cross-Modality Comparison

- Query same mouse strain (C57BL/6J) imaged by different microscopes.
- SPARQL → NGFF URLs → Vizarr viewer for side-by-side images.
- Ontology hierarchy links FIB-SEM vs fluorescence microscopy



This figure combines two distinct elements to demonstrate how the ontology bridges the schema and instance layers: (right) a sub-tree of the imaging-method class hierarchy for detailed annotation, and (left) a Vizarr snapshot of the actual imaging data instance retrieved by the query. This direct connection between abstract semantic descriptions and concrete datasets enables complex, cross-cutting queries. The specific SPARQL query that generates this result and retrieves the Vizarr URL of the dataset is provided in the GitHub repository (sparql/strain2zarr.rq).

```
PREFIX ssbd: <http://ssbd.riken.jp/ontology/#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?dataset ?title ?method ?label ?zarr ?vizarr
WHERE {
  # --- Biosample & Dataset ---
  ?dataset ssbd:has_about_strain ?strain
  ?dataset ssbd:has_biosample_information ?bs
  ssbd:bs dataset_title ?title

  # --- Imaging method ---
  ?dataset ssbd:has_imaging_method_total_info ?imNode
  ?imNode ssbd:has_imaging_method_recorded_type ?methodR
  OPTIONAL { ?methodR rdfs:label ?methodLabel }

  # --- NGFF endpoints (optional) ---
  OPTIONAL {
    ?dataset ssbd:ome_zarr_information ?imgff
    OPTIONAL { ?imgff ssbd:has_s3_endpoint ?s3 }
    OPTIONAL { ?imgff ssbd:has_vizarr_url ?vizarr }
  }
}
```

The SSBD graph can be queried live at

<https://knowledge.brc.riken.jp/bioresource/sparql>

How the SSBD Ontology Works

SSBD TWO-TIER ARCHITECTURE

Tier 1 – SSBD:repository → minimal core classes for fast publication.
Tier 2 – SSBD:database → enriched Biological & Imaging annotations.

SSBD Ontology

Core Layer: OWL classes

- Biological information: biosample, reagent

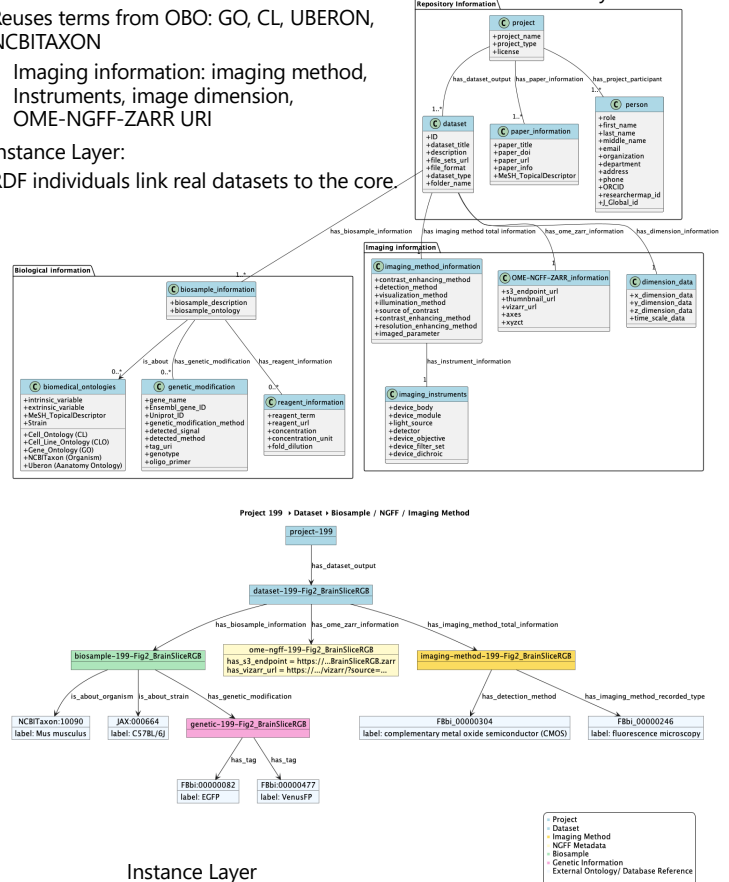
Reuses terms from OBO: GO, CL, UBERON, NCBITAXON

- Imaging information: imaging method, Instruments, image dimension, OME-NGFF-ZARR URI

Instance Layer:

RDF individuals link real datasets to the core

Overview of the SSBD Ontology's Core Layer



Instance Layer

Project → Dataset → Biosample → OME-Zarr relationships: Project 199 (Ichimura) illustrates the connection between a single Dataset, its corresponding Biosample, and the related OME-Zarr metadata.

CONCLUSION & FUTURE WORK

- ✓ A two-tier architecture enables FAIR data sharing.
- ✓ Approximately 500 datasets have been validated through complex queries.
- Next ▶ focus on expanding datasets, implementing LLM-assisted annotation.
- Aligning ontologies and boosting interoperability will establish a strong foundation for a global platform dedicated to bioimaging data sharing.



GitHub



BioPortal



ZENODO



Licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) ©2025 Yuki Yamagata (RIKEN)

