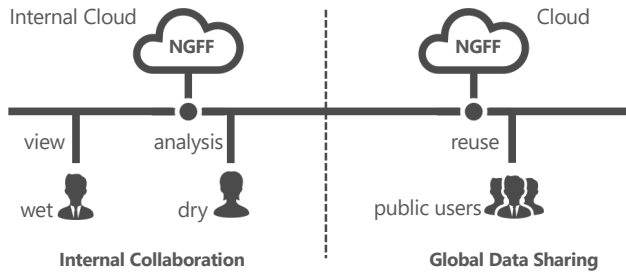


## Summary

- With the bioimaging community, we are developing the cloud-native **next generation file formats (NGFFs)** of bioimaging and biosystems dynamics data.
- SSBD** is a platform for sharing and reusing **bioimaging and biosystems dynamics data** globally.



## Toward Sharing Bioimaging Data

### Issues:

- Fast and parallel access to bioimaging, segmentation and tracking data stored in internal or public storages
- Building of a global sharing system of bioimaging and biosystems dynamics data

### How to resolve the issues:

- Use of the NGFFs such as OME-Zarr and BD-Zarr instead of TIFF or the other conventional formats
- Collaboration with BioImage Archive and Image Data Resource on metadata harmonization, cross-database searching, and API development



## Metadata and File Formats for Global Sharing of Bioimaging Data

Koji Kyoda<sup>1</sup>, Hiroya Itoga<sup>1</sup>, Yuki Yamagata<sup>2,3</sup>, Emi Fujisawa<sup>1</sup>, Miguel Miranda<sup>1</sup>, and Shuichi Onami<sup>1,2</sup>

<sup>1</sup>RIKEN BDR, <sup>2</sup>RIKEN R-IH, <sup>3</sup>RIKEN BRC

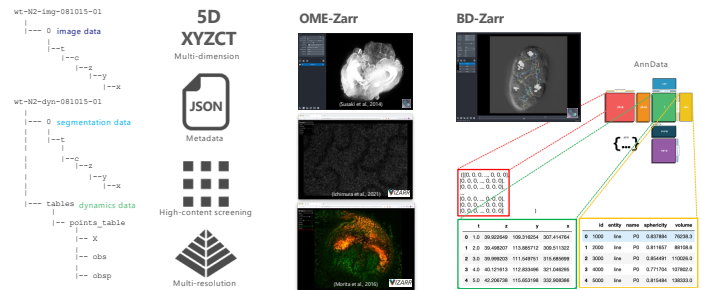
## File Formats

### OME-Zarr

- Next generation file format (NGFF) for storing bioimaging data

### BD-Zarr

- NGFF for storing biosystems dynamics data (e.g., tracking data, feature data)

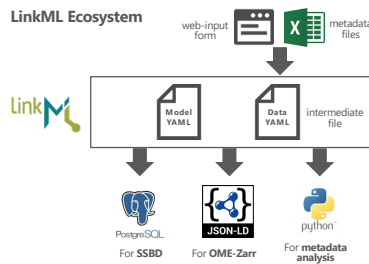


## Metadata

**SSBD metadata has been improved to align with the other data resources.**

- REMBI (Sarkans et al. 2021)
  - A guideline for metadata of optical and electron microscopy imaging data
- QUAREP-LiMi (Nelson et al. 2021)
  - Specifications for quality assurance and quality control of optical microscopies

A part of metadata extracted from SSBD metadata



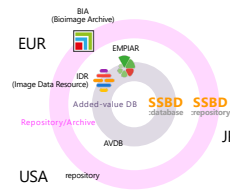
## Global Sharing

### SSBD:repository

- An archive for quick sharing of bioimaging data for paper submission/publication

### SSBD:database

- An added-value database of highly reusable bioimaging data with rich metadata



### Issues

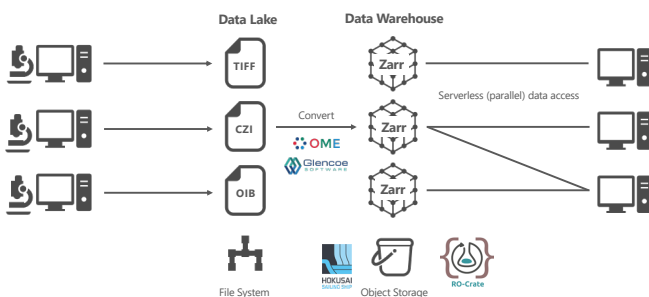
- Harmonization of metadata
- Cross-database search
- NGFF datasets (in OME-Zarr)
- Common APIs

**SSBD is positioned as a core repository and database within the founding GIDE.**

## Data Reuse and Analysis

### For reuse and analysis of bioimaging data:

- We need an execution environment capable of performing workflow analysis.
- In SSBD, data should be shared together with image-analysis workflows and analysis results.



## Plans

- Sharing data in OME-Zarr and BD-Zarr formats in internal and public storages
- Developing tools and libraries for reading, writing and reusing biosystems dynamics data in BD-Zarr
- Harmonizing metadata of SSBD, BioImage Archive (BIA) and Image Data Resource (IDR) to enable cross-search across these data resources

### References

- Moore et al. (2023) OME-Zarr: a cloud-optimized bioimaging file format with international community support. *Histochem. Cell Biol.* **160**, 223-251.
- Kyoda et al. (2020) BD5: An open HDF5-based data format to represent quantitative biological dynamics data. *PLoS One* **15**, e0237468.
- Kyoda et al. (2015) Biological Dynamics Markup Language (BDML): an open format for representing quantitative biological dynamics data. *Bioinformatics* **31**, 1044-1052.

### Acknowledgment

