SSBD update and RIKEN Open Science Platform

Shuichi Onami

RIKEN Center for Biosystems Dynamics Research

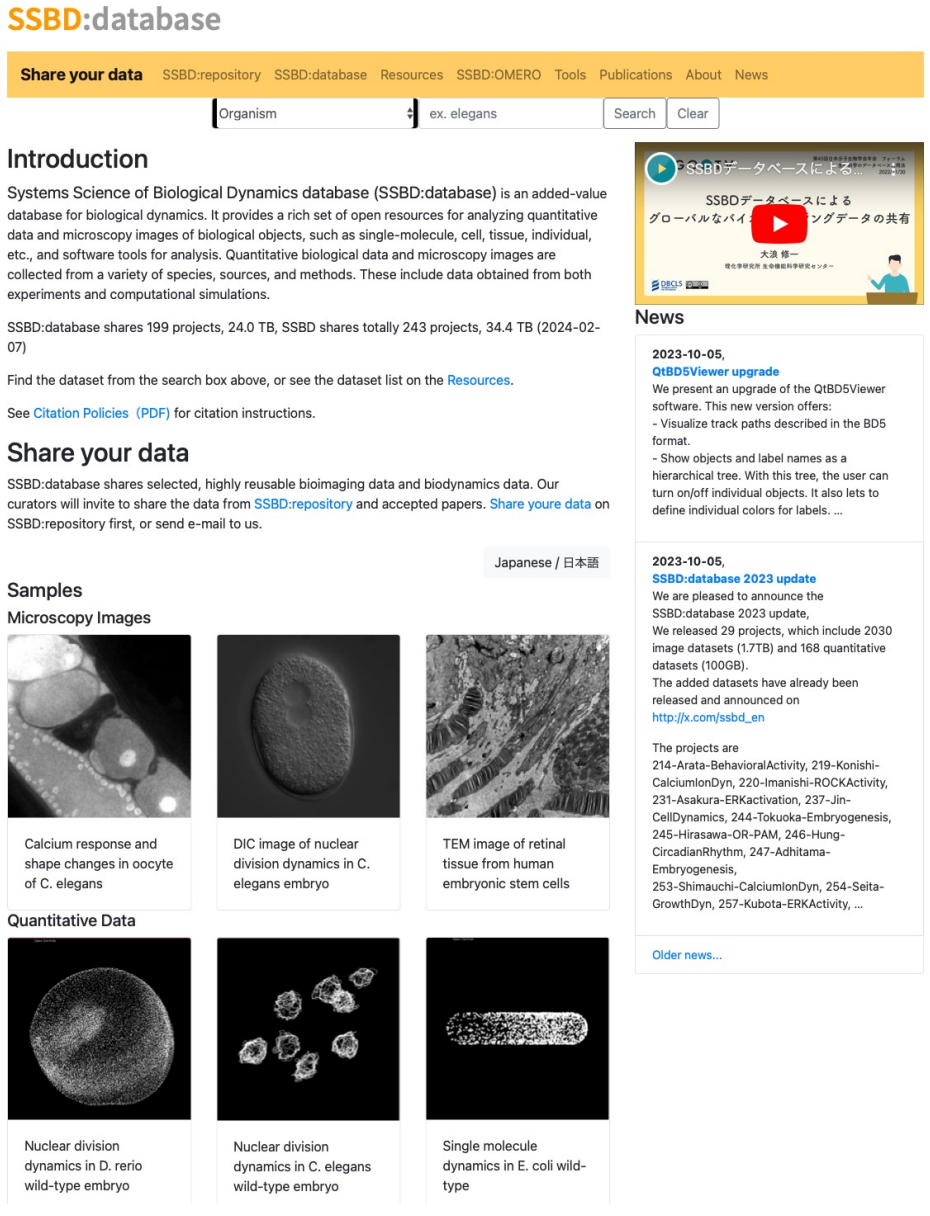
RIKEN, TRIP Headquarters

RIKEN Information R&D Headquarters

SSBD

a platform for Systems Science of Biological Dynamics

- Integrated database/public repository for sharing bioimaging data and biological dynamic data
- Comprehensive platform including tools and environment for utilization of bioimaging data and biological dynamics data



RIKEN National Science Institute is Japan's most comprehensive institute for the natural sciences, conducting cutting-edge research in a wide range of scientific fields, including physics, chemistry, brain science, energy, medicine, nuclear physics, sustainable resources, computer science, plant science, genetics, nanoscience, developmental biology, electronics, artificial intelligence, and much more.

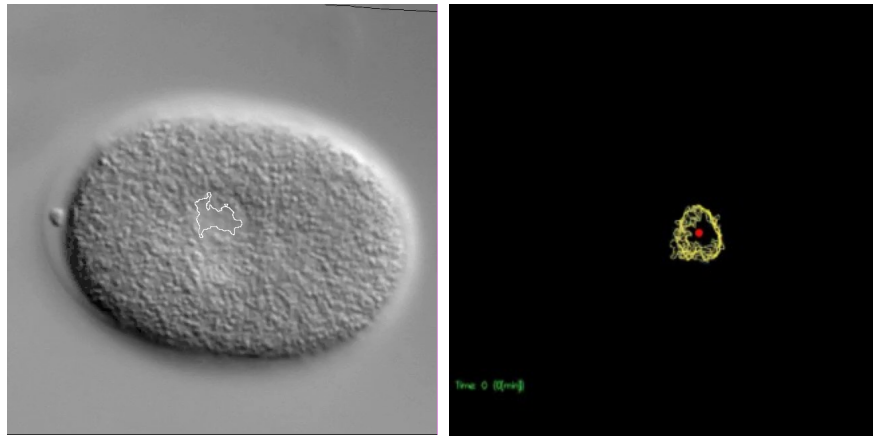
SSBD at the beginning

- Established with JST-NBDC funding as an integrated database of Systems Science of Biological dynamics in Japan (2013)
- Initially, it shares quantitative biosystems dynamics data and the image data used to obtain them in Japan.

Examples

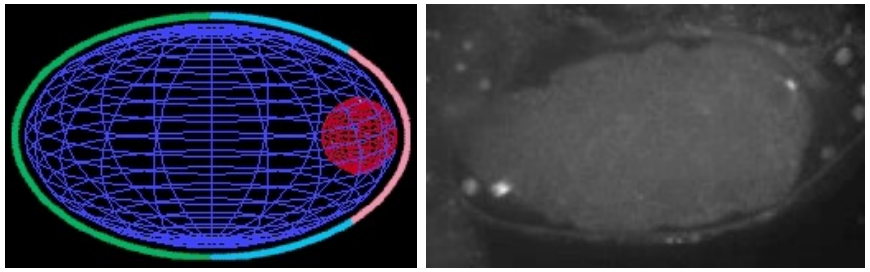
- **Quantitative measurement data** of biosystems dynamics under various perturbation conditions
- **Results of spatiotemporal dynamic simulation** of biological phenomena under various parameters

Measurement data of nuclear division dynamics in *C. elegans* embryos



Hamahashi et al. BMC Bioinformatics 2005

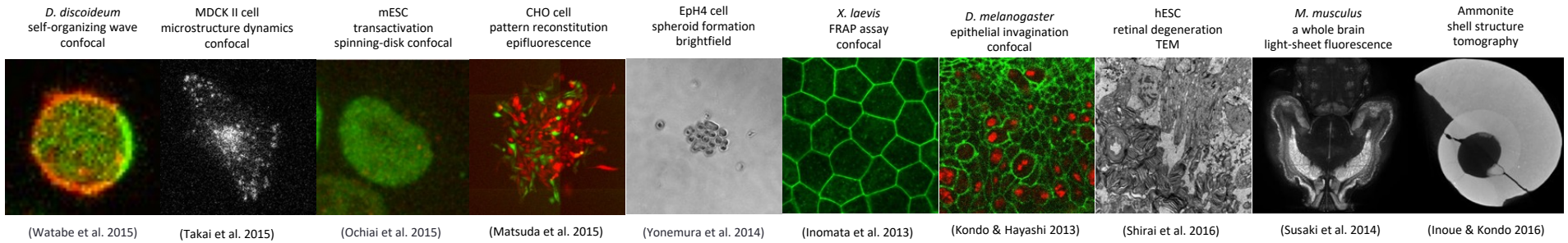
Simulation results of cell nuclear dynamics

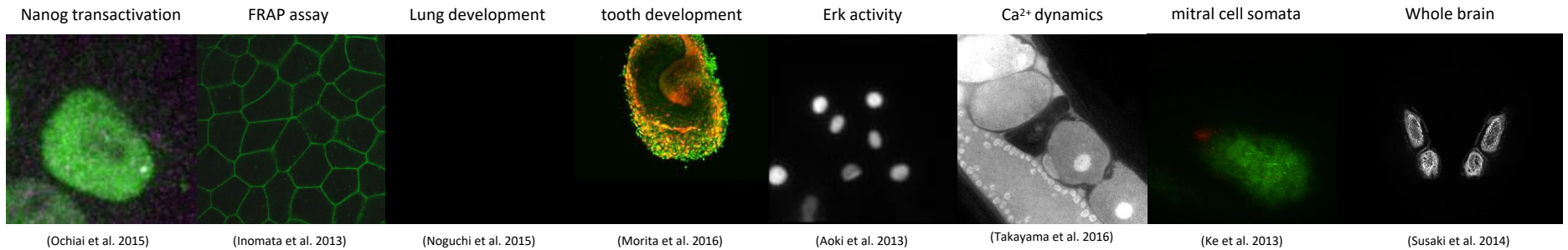


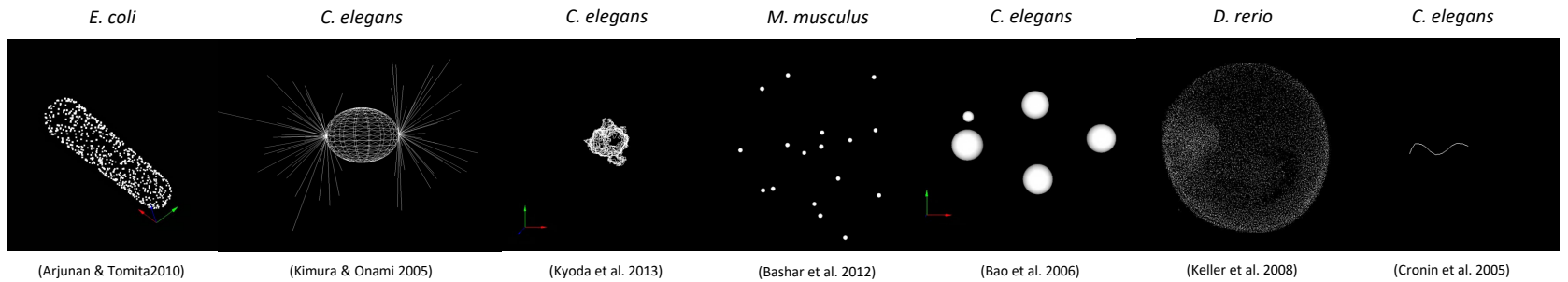
Kimura & Onami Dev. Cell 2005

SSBD as an integrated database of bioimaging and biosystems dynamic data

- Started sharing bioimaging data acquired with state-of-the-art technology and those from systematic experiments (2016).
- Launched repository service for life science image data (2016)
- Separated Added-value DB (SSBD:database) and public repository (SSBD:repository) (2019)



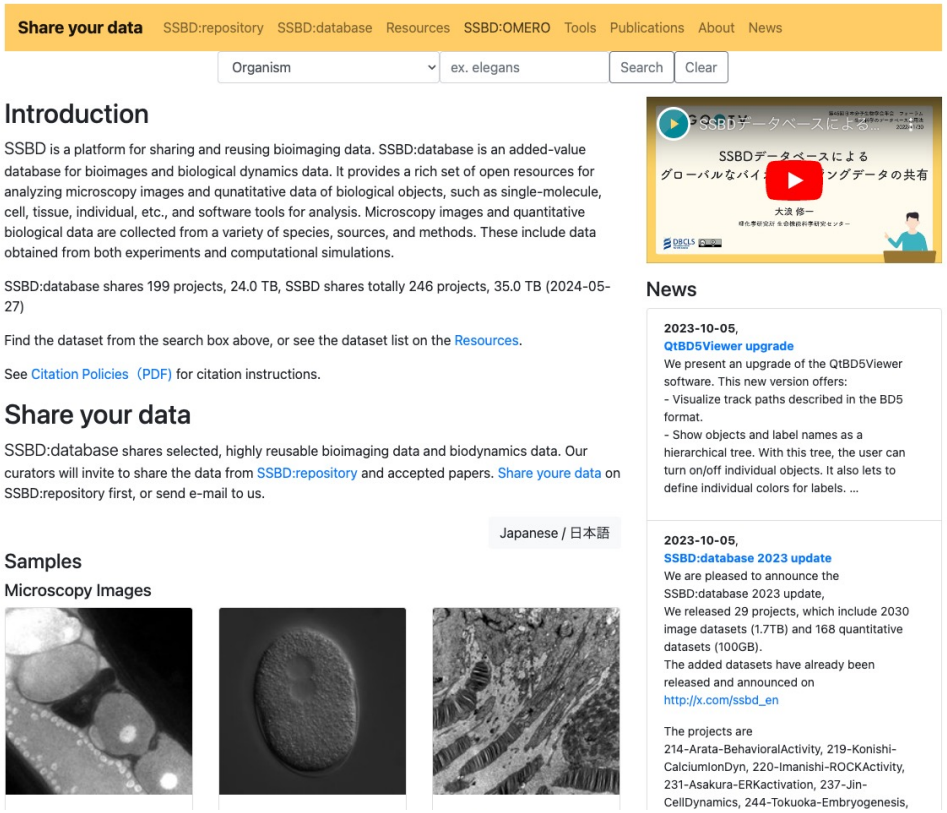




SSBD

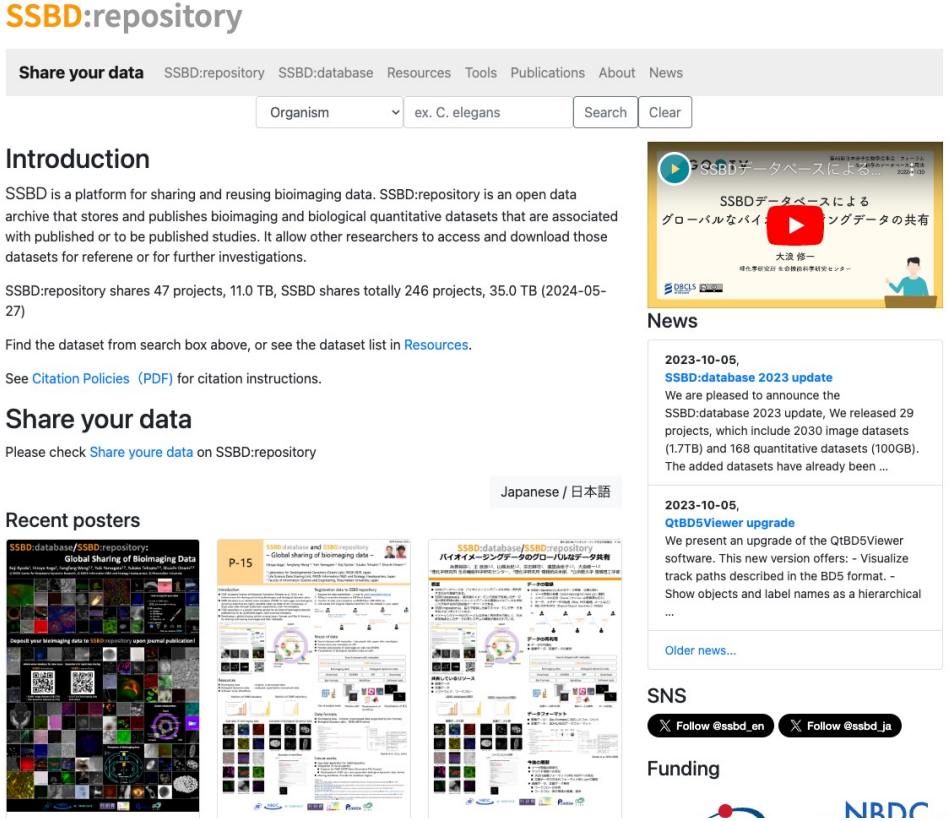
a platform for Systems Science of Biological Dynamics

SSBD:database



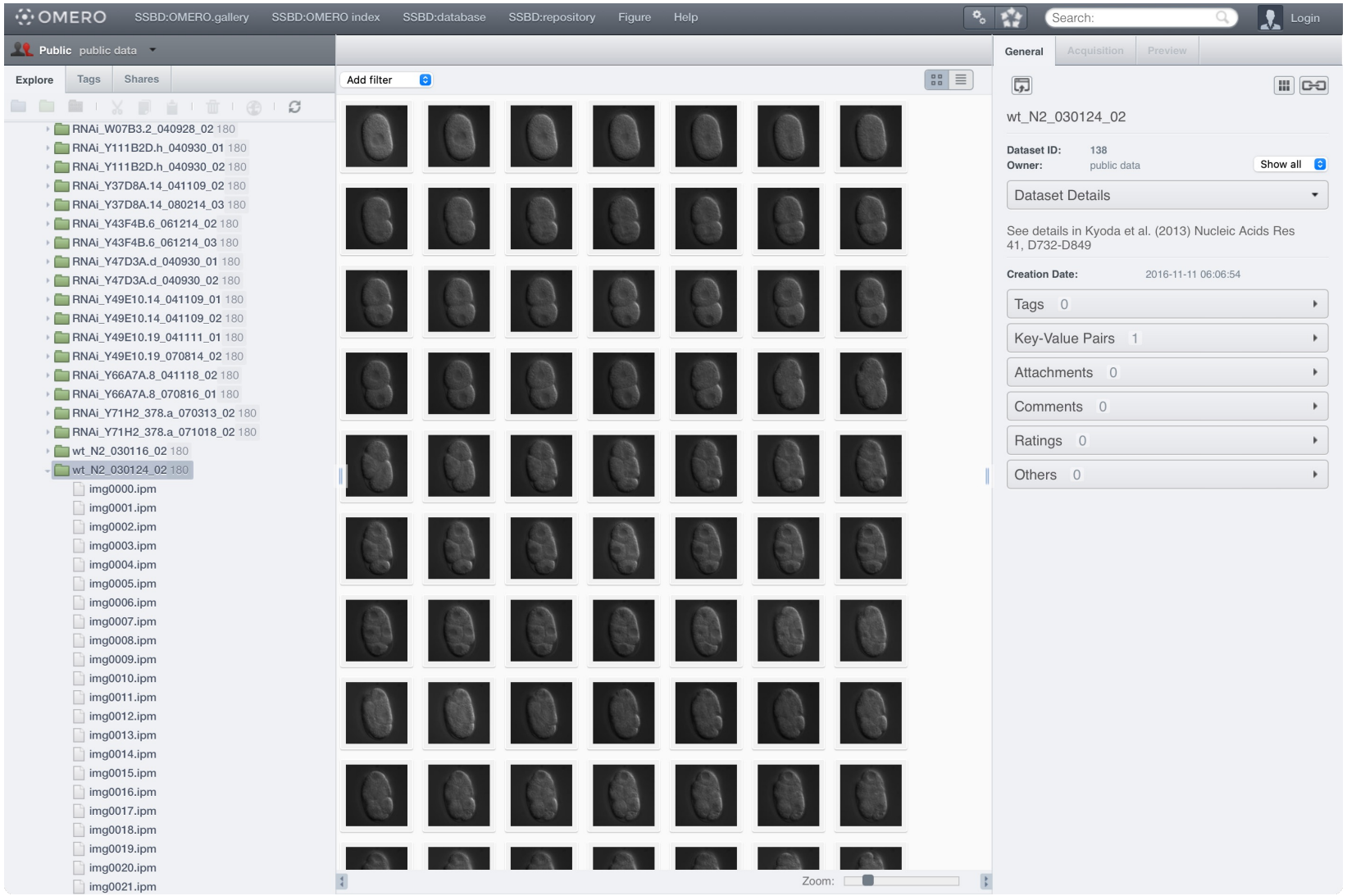
Added-value database that shares highly reusable bioimaging data and quantitative biosystems dynamics data with rich metadata.

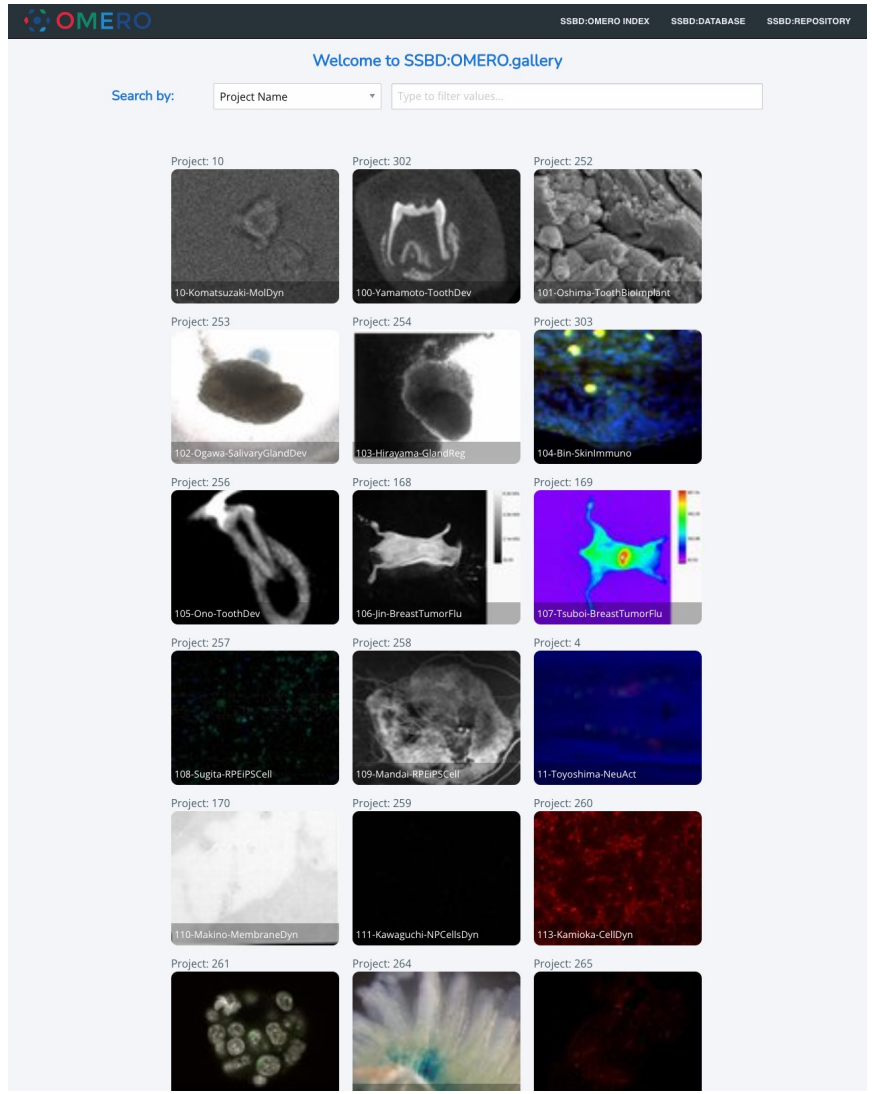
SSBD:repository



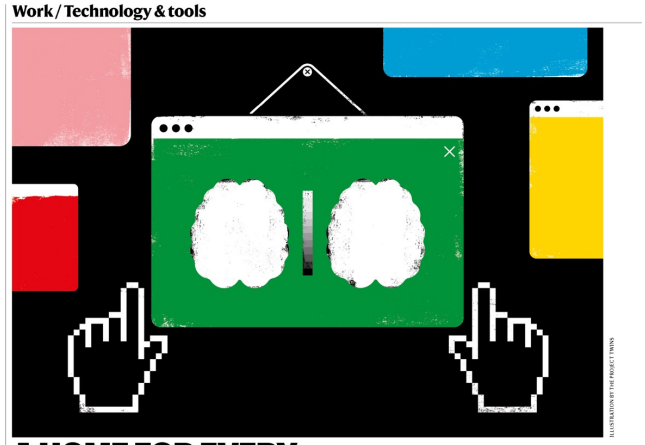
Public repository for sharing bioimaging data and quantitative biosystems dynamic data used in papers

SSBD uses OMERO for image data management





SSBD : a world's leading public repository for bioimaging data



Work / Technology & tools

A HOME FOR EVERY IMAGING DATA SET

Repositories let researchers store, share and access life-science images — and maybe even extract new findings. **By Amber Dance**

When Sjors Scheres set out to develop a tool to reverse flaws in cryo-electron microscopy images, he needed lots of data on which to test it. So Scheres, a structural biologist at the MRC Laboratory of Molecular Biology (LMB) in Cambridge, UK, turned to the Electron Microscopy Public Image Archive (EMPIAR), a database of raw images. There he downloaded, for free, data collected by the lab of Gabriel Lander, a structural biologist at Scripps Research in La Jolla, California.

Using his new technique, Scheres was able to squeeze sharper images from those data, improving[1] the resolution of one structure from 3.1 ångströms to 2.3 ångströms.

"That's precisely why we posted the data," says Lander. "We knew some brilliant people out there would be able to improve on our processing."

Services such as EMPIAR give researchers a central location in which to store, share and access a rapidly expanding corpus of biological images. "The data aren't just one picture any more," says Joshua Vogelstein, a neurostatistician at Johns Hopkins University in Baltimore, Maryland. Movies, 3D images and microscope-based screening data can take up gigabytes or terabytes of storage, and can't be e-mailed back and forth in the same way as individual TIFF or JPEG files. Moreover, grant agencies and journals increasingly require scientists to make their data available to all, but don't necessarily offer to host them. EMPIAR and its kin fill that gap, and often provide a digital object identifier or other citation so researchers can get credit for their data.

"Are you struggling to load your images?" asks Forrest Collman, a neuroscientist at the Allen Institute for Brain Science in Seattle, Washington. "Are you particularly struggling to share?" If so, he says, "looking into this kind of service makes sense for you".

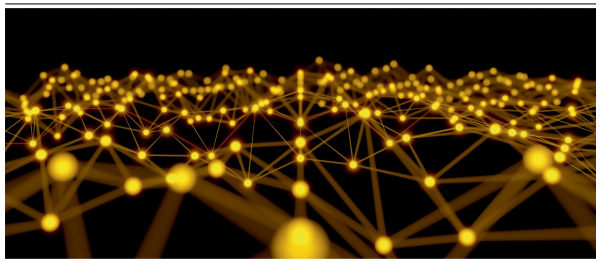
In 2019, when Collman spotted an odd-looking neuron in one of his electron-microscopy data sets[2], it was easy for him to send a colleague a link to that spot in the data repository, rather than a bulky file. She noticed another unique feature, and Collman identified a few similar cells. They might turn out to be a new type of neuron, Collman says.

There are a number of other image warehouses available, among them the Image Data Resource (IDR). Both it and EMPIAR are hosted by the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) in Hinxton, UK. Further options include,

Nature 579, 162-163 (2020)

Technology feature

https://doi.org/10.1038/s41592-023-01927-7



To share is to be a scientist

Check for updates

Wrangling big data is now part of being a biomedical scientist, and mandates on data sharing have entered the scene. Mandates can alter behavior, but data sharing also needs incentives and shifts in science culture.

By Vivien Marx

As data-spewing instruments spread across biomedical labs and multimodal approaches are embraced, data sharing must be powered up, too. Much has been achieved, say some researchers in genomics, proteomics, neuroscience and imaging, as do some big data producers, wranglers at repositories and shepherds of large-scale projects. Big biodata's next phase, they say, needs resources and shifts in science culture. Here are some views on how far things have come and what lies ahead.

With data as with pizza, it's considered good manners to share. Whereas pizza sharing is a private affair, data sharing is how good citizens of science give collaborators and strangers access to results generated mainly or entirely with public funds. As of January 2023, the US National Institutes of Health (NIH) mandates all who apply for funding must submit a Data Management and Sharing plan. The reaction has not been a chorus of hurrays.

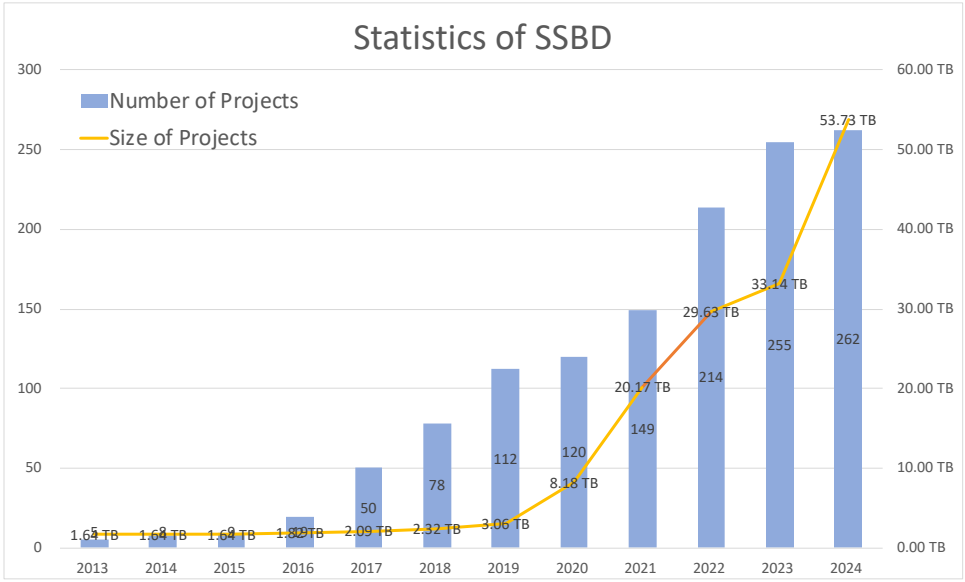
Some tense moments occurred at this year's American Association for Cancer Research (AACR) annual meeting, when program officers from the National Cancer Institute's (NCI) Office of Data Sharing presented the new NIH data sharing policies and held a question-and-answer session.

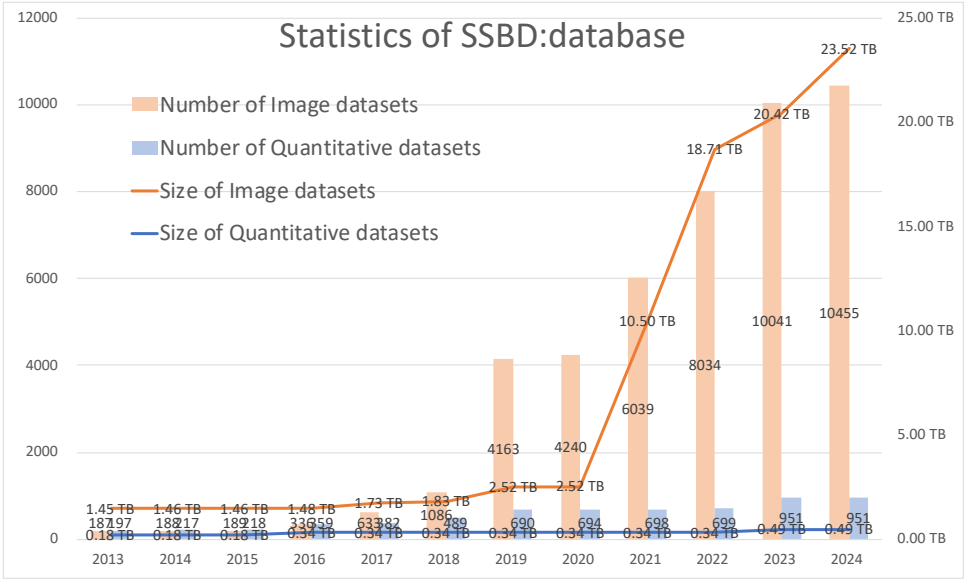
Some investigators said the mandate had been rolled out too suddenly. They asked how they are to find the time, skill and funding to set up a data management and sharing plan. Program officers directed attendees to guidance pages and offered personal conversations. Scientists can send questions to program officers or, in the case of NCI, to NCIOfficeOfDataSharing@mail.NCI.gov. The first grant proposals with data management and sharing plans have been submitted, the NCI Office of Data Sharing said in an unattributed statement after the conference. The plans are still in study section and review, so it's unclear how much back-and-forth will unfold. Says Emily Boja from the NCI Office of Data Sharing, who presented at the AACR session, it's understandable that the mandate can seem overwhelming for labs not used to organizing data for sharing, "especially when resources are limited."

The culture shift we are striving for is to weave data management and sharing into the conduct of science, says Heather Basehore.

CREDIT: S TOP — C JOHNNY / GETTY IMAGES; RIGHT: NIH/NCI OFFICE OF DATA SHARING

Nature Methods 20, 984-989 (2023)





Recent updates in SSBD

Metadata schema update

- We started using the SSBD Metadata template v3 for collecting data
  - The template has 6 sheets for describing *Project*, *Person*, *Datasets*, *Genetic Info*, *Treatment Info*, and *Extended Info for Datasets*.

Project

Project	SID	required	string
	name	required	text
	URL	required	string
	DOI	required	string
	type	recommended	text
	title	required	text
	description	required	text
	date of submission	required	date
	date of opened	recommended	date
	date of released	recommended	date
	date of updated	optional	date
	license	required	CC
	funding info.	optional	text
	version	required	int
	PubMed ID	optional	string
	PMC ID	optional	string
	DOI	recommended	string
	URL	recommended	string
	person	required	Person
	dataset	required	Dataset
	ext_dataset	optional	ExtDataset

Person

Person	role	required	text
	first name	required	text
	last name	required	text
	middle name	optional	text
	e-mail	required	string
	URL	optional	string
	organization	required	text
	department	optional	text
	laboratory	optional	text
	address	optional	text
	phone	optional	string
	ORCID	recommended	ORCID
	J-GLOBAL ID	optional	J-GLOBAL
	researchmap ID	optional	researchmap

Dataset

Dataset	ID	required	string
	title	required	text
	description	optional	text
	file format	recommended	string
	file size	recommended	text
	figure number	recommended	text
	dimensions	required	Dimensions
	biosample	required	Biosample
	imaging method	recommended	Imaging Method
	instruments	recommended	Instruments
	dimensions	required	string
	Xscale	recommended	float
	Xscale Unit	recommended	UO
	Yscale	recommended	float
	Yscale Unit	recommended	UO
	Zscale	recommended	float
	Zscale Unit	recommended	UO
	Tscale	recommended	float
	Tscale Unit	recommended	UO
	Channel	recommended	int
	description	required	text
	organism	required	NCBITaxon
	strain	recommended	text
	strain URL	optional	string
	cell	optional	CO
	cell line	recommended	CLO
	MeSH term (cell related)	optional	MeSH
	intrinsic variable	optional	text
	extrinsic variable	optional	text
	experimental variables	optional	text
	imaging method	recommended	FBbi
	detection method	recommended	FBbi
	visualization method	recommended	FBbi
	illumination method	recommended	FBbi
	source of contrast	recommended	FBbi
	contrast-enhancing method	recommended	FBbi
	resolution-enhancing method	recommended	FBbi
	imaged parameter	recommended	FBbi
	sample preparation method	recommended	FBbi
	MeSH term (Imaging Method)	optional	MeSH
	body	recommended	text
	module	recommended	text
	light source	recommended	text
	detector	recommended	text
	objective	recommended	text
	filter set	recommended	text
	dichroic	optional	text
	genetic	optional	Genetic
	treat	optional	Treat

Genetic Info

Genetic	gene name	optional	string
	Ensembl ID	optional	ENS
	protein name	optional	string
	UniProtKB ID	optional	UniProt
	method	optional	text
	method URI	optional	EFO
	detected signal/contrast	optional	string
	detection method	optional	text
	tag URI	optional	string
	genotype	optional	text
	oligo primer	optional	text

Treatment Info

Treat	reagent/compound term	optional	string
	reagent/compound URI	optional	ChEBI
	concentration	optional	string
	unit term (concentration)	optional	string
	UO term URI (concentration)	optional	UO
	fold dilution	optional	string
	unit term (fold dilution)	optional	string
	UO term URI (fold dilution)	optional	UO

ExtDataset

Ext Dataset	UBERON term	optional	UBERON
	anatomical entity (species-specific)	optional	text
	MeSH term (anatomy related)	optional	MeSH
	GO:Biological Process term	recommended	GO
	MeSH term (BP related)	optional	MeSH
	GO:Cellular Components	recommended	GO
	MeSH term (CC related)	optional	MeSH
	GO:Molecular Function	recommended	GO
	MeSH term (MF related)	optional	MeSH

- Fields in yellow are required.
- Fields in green can be described by controlled vocabulary/ontology terms.

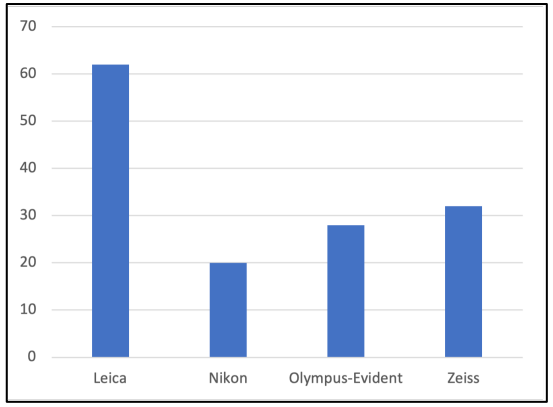
Metadata schema update

- We started using the SSBD metadata template v3 to collect and share

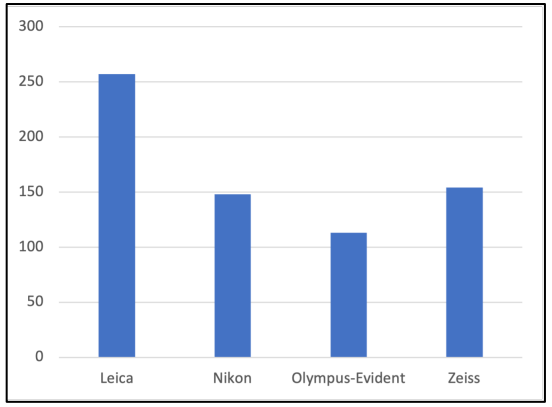
updated class	field	range	example
Genetic	gene protein method tag …	Ensemble ID UniProt ID string (EFO, etc.) FBbi …	rpl-23 (Ensembl:WBGene00004435) SOX2 (UniProtKB:P48432) transgenic (http://www.ebi.ac.uk/efo/EFO_0004020) GFP (http://purl.obolibrary.org/obo/FBbi_00000437) …
Treat	compound concentration unit …	ChEBI float, string UO (Unit Ontology) …	histamine (http://purl.obolibrary.org/obo/CHEBI_18295) 100.0, MOI-30, etc. micromolar (http://purl.obolibrary.org/obo/UO_0000064) …
Imaging method	imaging method detection method resolution-enhancing method …	FBbi FBbi FBbi …	fluorescence microscopy (http://purl.obolibrary.org/obo/FBbi_00000246) APD (http://purl.obolibrary.org/obo/FBbi_00000297) STED (http://purl.obolibrary.org/obo/FBbi_00000334) …
Instruments	body module objective detector …	*Select from a list of instruments provided by the vendors*	Leica DMi8 Zeiss PALM MicroBeam Olympus-Evident MPLAPON60x Nikon DS-Qi1Mc …

A list of instruments

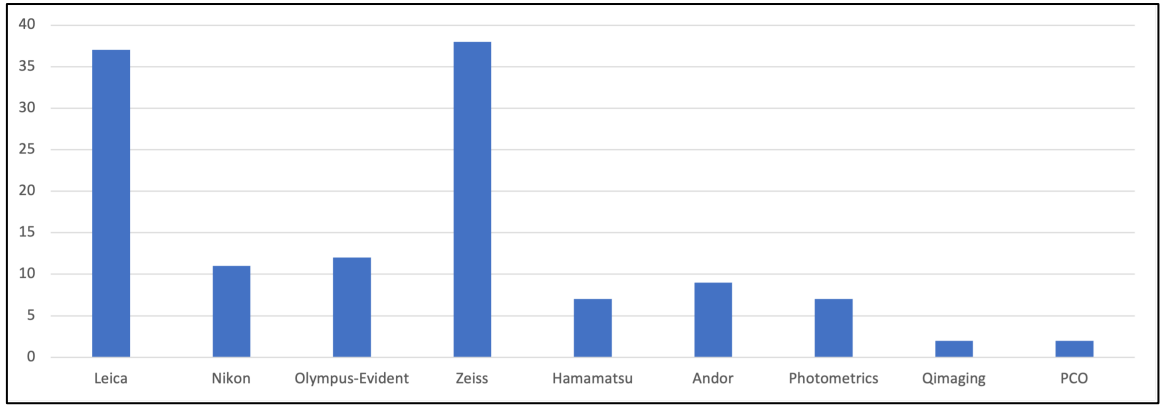
- The microscope vendors kindly helped us to collect a list of their products.
  - **Leica**, **Nikon**, **Olympus-Evident**, and **Zeiss** helped us to prepare the list.
  - We requested them to send a list of their products (Body, Module, Objectives, Detector, Light Source).
    - However, most vendors did not have digital files of their products.
  - We extracted the lists from vendor support web-pages and other sources.



Body (142)

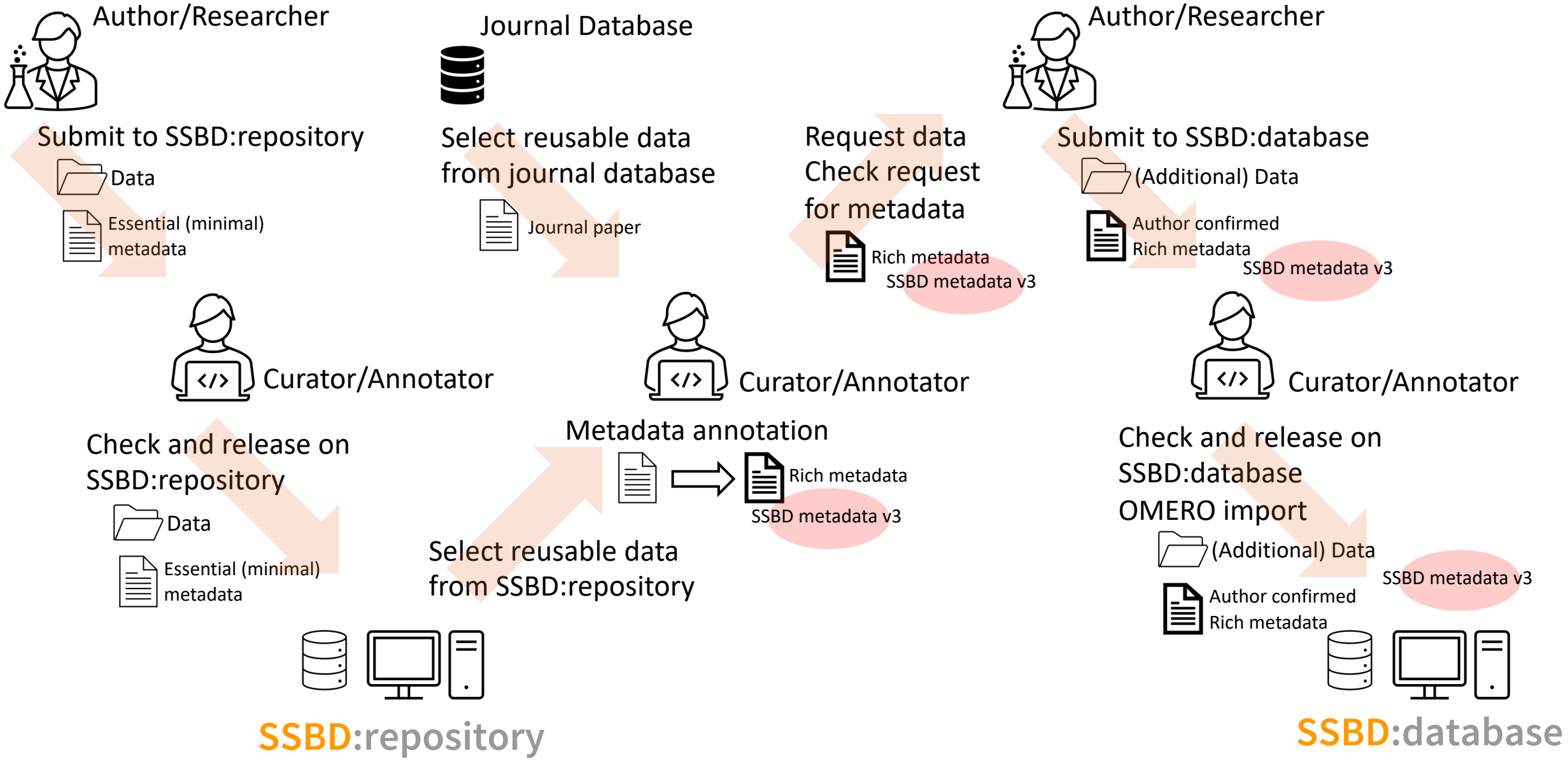


Objectives (672)



Detector (125)

SSBD data registration process



Re-annotation of already published data in SSBD:database

- 28% (55/199) of all projects (papers) have finished
  - 338 datasets
  - 430 gene/protein information was added
  - 233 treatment information was added

Schedule for SSBD metadata v3 update

- 2024/4/1 SSBD:repository started to accept metadata v2+
- 2024/4/29 SSBD:database started to use metadata v3 for annotation
- 2024/6/30 SSBD:repository metadata database v2+ update
  - SSBD minimal metadata v2+:
    pilot update for SSBD metadata v3,
    v2+ includes gene names/IDs and protein names/IDs, and
    minor updates for author and paper information

- 2024/10/5 SSBD:repository starts to accept metadata v3
- 2024/10/5 SSBD:database and SSBD:repository metadata database v3 update
  - SSBD metadata v3 for 55 projects (outsourced) and newly submitted projects

Started providing bioimaging data in OME-Zarr

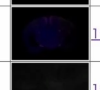
- SSBD:database has shared 12 samples of bioimaging data in OME-Zarr.
- Now we are converting all bioimaging data stored in SSBD:database to those in OME-Zarr.

SSBD ID	Dataset ID	License	Dimensions	Description
2-Kyoda-WormEmbryoRNAi	wt_N2_030303_01	CC BY-SA	600x600x66x1x180	DIC microscopy images of early *C. elegans* embryogenesis
98-Morita-ToothEpiCellDev	fig2ab_trajectory_epithelialcell	CC BY	862x855x43x2x271	Confocal microscopy images of tooth development with H2B-EGFP and Fucci indicator
48-Susaki-MouseBrainCUBIC	Fig3A_R26-pCAG-nuc-3xmKate2	CC BY	2560x2160x520x1x1	Light-sheet microscopy images of mouse brain cleared by CUBIC
48-Susaki-MouseBrainCUBIC	Fig3C_H2B-mCherry	CC BY	2560x2160x520x1x1	Light-sheet microscopy images of mouse brain cleared by CUBIC
48-Susaki-MouseBrainCUBIC	Fig3D_H2B-EGFP	CC BY	2560x2160x520x1x1	Light-sheet microscopy images of mouse brain cleared by CUBIC
48-Susaki-MouseBrainCUBIC	Fig3E_Thy1-YFP-H	CC BY	2560x2160x520x1x1	Light-sheet microscopy images of mouse brain cleared by CUBIC
199-Ichimura-MulticellularDyn	Fig2_BrainSliceRGB	CC BY-NC	13264x9180x1x3x1	100MP trans-scale scope (AMATERAS) image of brain slice
199-Ichimura-MulticellularDyn	Fig4_HeLaYC360_5spf	CC BY-NC	13264x9180x1x200	AMATERAS images of FRET-based calcium ion in HeLa cells
199-Ichimura-MulticellularDyn	Fig5-6_Flamindo2	CC BY-NC	13264x9180x1x1x1922	AMATERAS images of cell behaviors with cAMP indicator, Flamindo2
170-Tsutsui-HookBM	Airyscan_hook_BM	CC BY	844x844x58x3x1	High-resolution 3D microscopy images of the hook basement membrane
141-Sato-CellMorphology	Fig3a_FIB-SEM_synapse	CC BY	2048x1538x1x1x1	FIB-SEM images of spine synapse in C57BL/6J mice
163-Wang-MitochondrialDyn	Fig5AC_Mitochondrial_MitoPB	CC BY	1024x256x361x1x1	STED images of mitochondrial dynamics in MitoPB Yellow-labeled cells

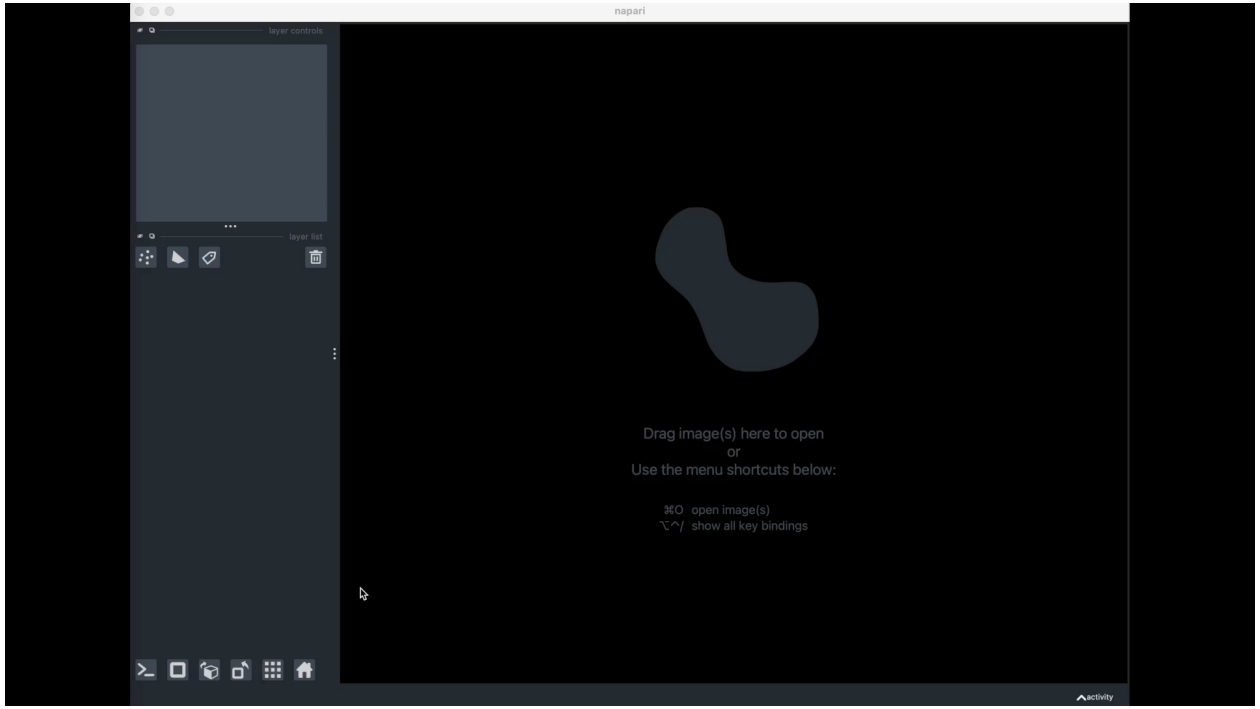
OME-Zarr sample data can be visualised in Vizarr and napari

SSBD OME-NGFF Samples

Sample SSBD image datasets for development and validation of NGFF software tools

OME-NGFF ver.	Thumbnail	SSBD Project ID	SSBD Dataset ID	license	sizeX	sizeY	sizeZ	sizeC	sizeT	Axes	view	S3-endpoiint	Date added
0.4		2-Kyoda-WormEmbryoRNAi	wt_N2_030303_01	CC BY-SA	600	600	66	1	180	XYZT	vizarr	https://ssbd.riken.jp/100 copy	2023-03-22
0.4		98-Morita-ToothEpiCellDev	fig2ab_trajectory_epithelialcell	CC BY	862	855	43	2	271	XYZCT	vizarr	https://ssbd.riken.jp/100 copy	2023-03-22
0.4		48-Susaki-MouseBrainCUBIC	Fig3A_R26-pCAG-nuc-3xmKate2	CC BY	2560	2160	520	1	1	XYZ	vizarr	https://ssbd.riken.jp/100 copy	2023-03-22
0.4		48-Susaki-MouseBrainCUBIC	Fig3C_H2B-mCherry	CC BY	2560	2160	709	1	1	XYZ	vizarr	https://ssbd.riken.jp/100 copy	2023-03-22
0.4		48-Susaki-MouseBrainCUBIC	Fig3D_H2B-EGFP	CC BY	2560	2160	665	1	1	XYZ	vizarr	https://ssbd.riken.jp/100 copy	2023-03-22
0.4		48-Susaki-MouseBrainCUBIC	Fig3E_Thy1-YFP-H	CC BY	2560	2160	696	1	1	XYZ	vizarr	https://ssbd.riken.jp/100 copy	2023-03-22
0.4		199-Ichimura-MulticellularDyn	Fig2_BrainSliceRGB	CC BY-NC	13264	9180	1	3	1	XYC	vizarr	https://ssbd.riken.jp/100 copy	2023-03-22
0.4		199-Ichimura-MulticellularDyn	Fig4_HeLaYC360_5spf	CC BY-NC	13264	9180	1	1	200	XYT	vizarr	https://ssbd.riken.jp/100 copy	2023-03-22
0.4		199-Ichimura-MulticellularDyn	Fig5-6_Flamindo2	CC BY-NC	13264	9180	1	1	1922	XYT	vizarr	https://ssbd.riken.jp/100 copy	2023-03-22

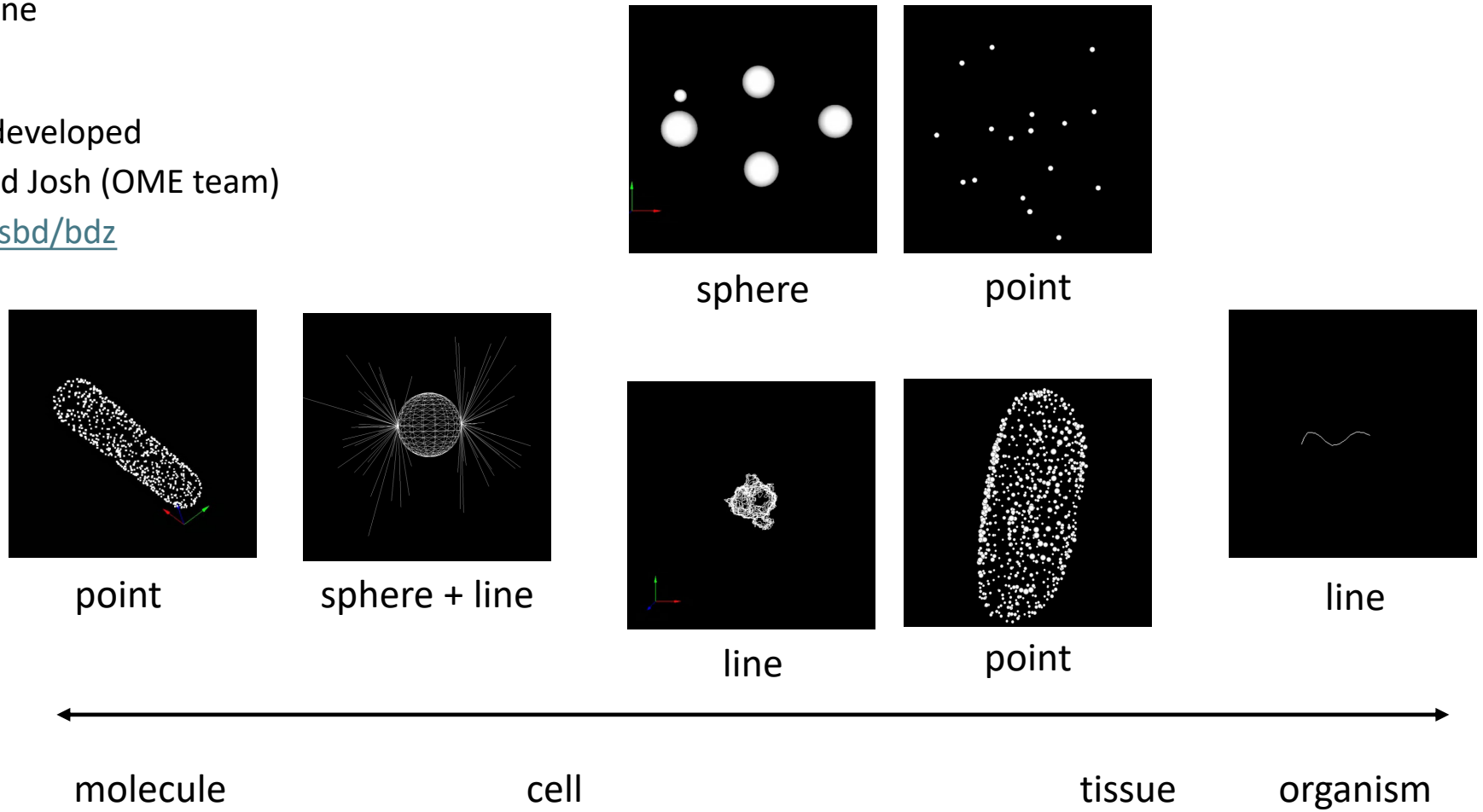
Visualisation of 100MP trans-scale scope images in Vizarr



Visualisation of light-sheet microscopy images of a whole mouse brain in napari

Data formats for spatiotemporal dynamics of biological objects

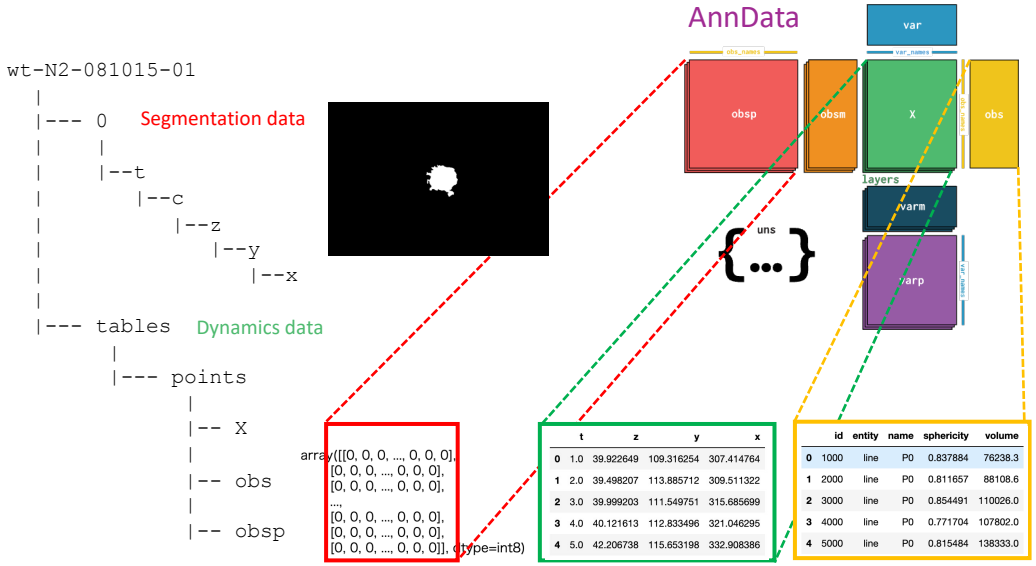
- BDML (Biological Dynamics Markup Language): an XML-based
  - Kyoda et al. (2015) Bioinformatics
- BD5: an HDF5-based
  - Kyoda et al. (2020) PLoS One
- BD-Zarr: a Zarr-based
  - α version of BD-Zarr was developed
    with the help of Will and Josh (OME team)
  - https://github.com/openssbd/bdz

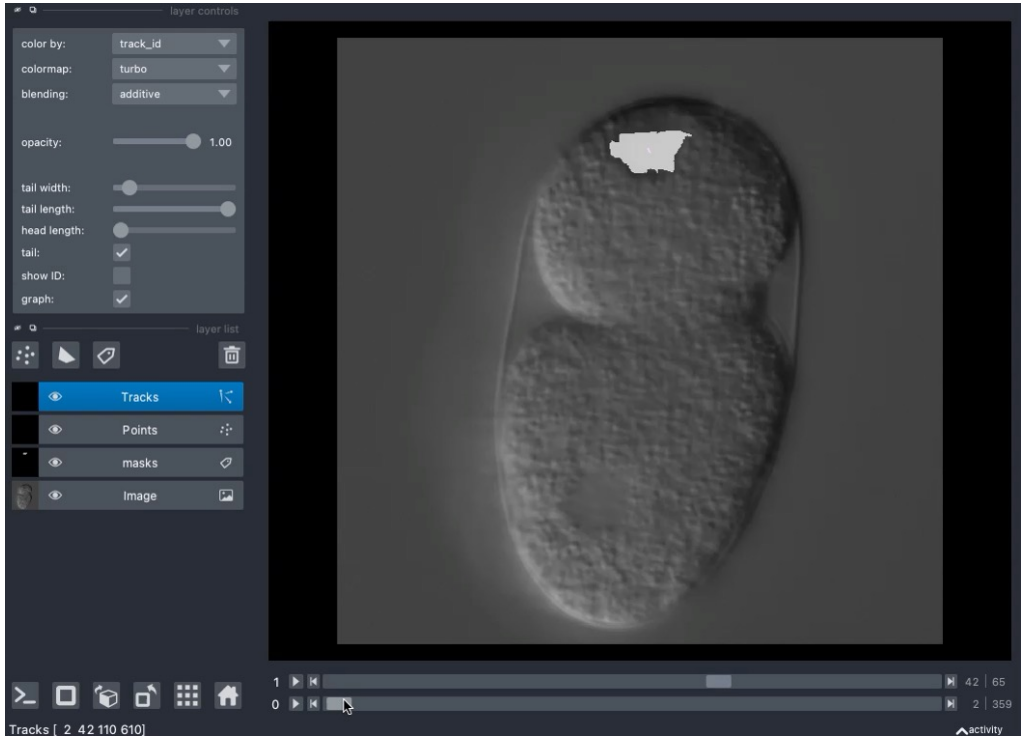


BD-Zarr development

- BD-Zarr uses OME-NGFF table to store
  - representative points of biological objects
  - trajectories/tracks of these points over time
  - features related to the biological objects

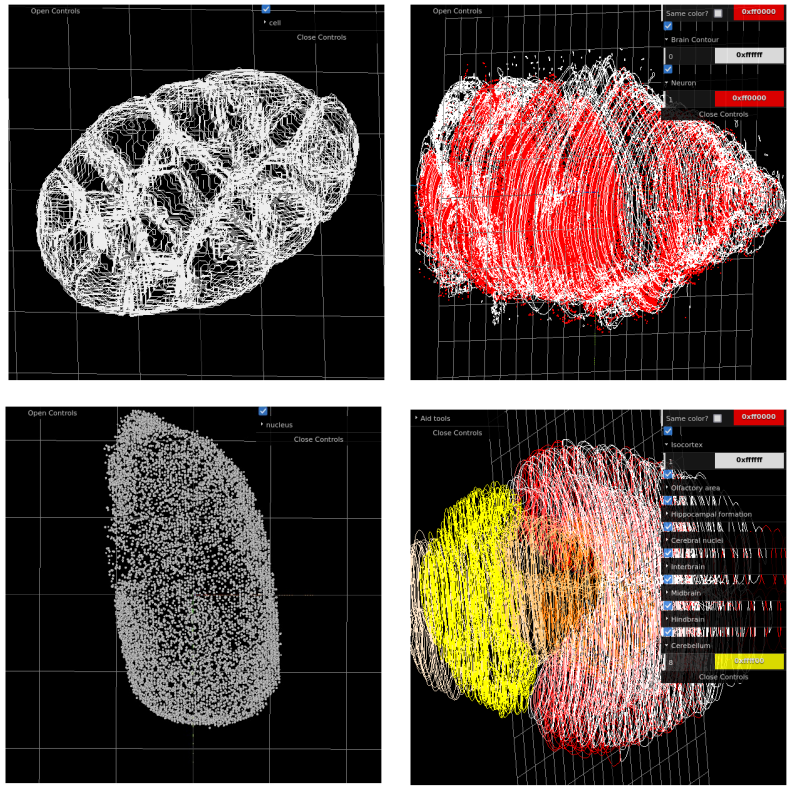
- Plan
  - Fix the specification [Aug, 2024]
  - I/O and visualization (napari) [Oct, 2024]
  - Converter (BD5 <-> BD-Zarr) [Oct, 2024]
  - Prepare some use cases [Dec, 2024]
  - Proposal to bioimaging community [Jan, 2025]

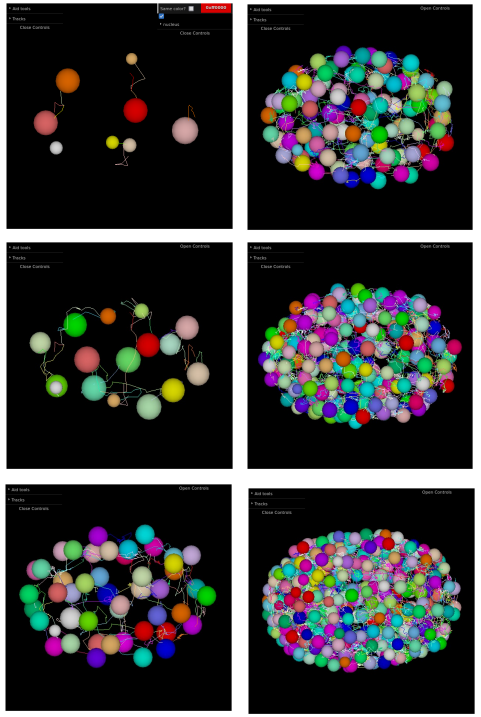


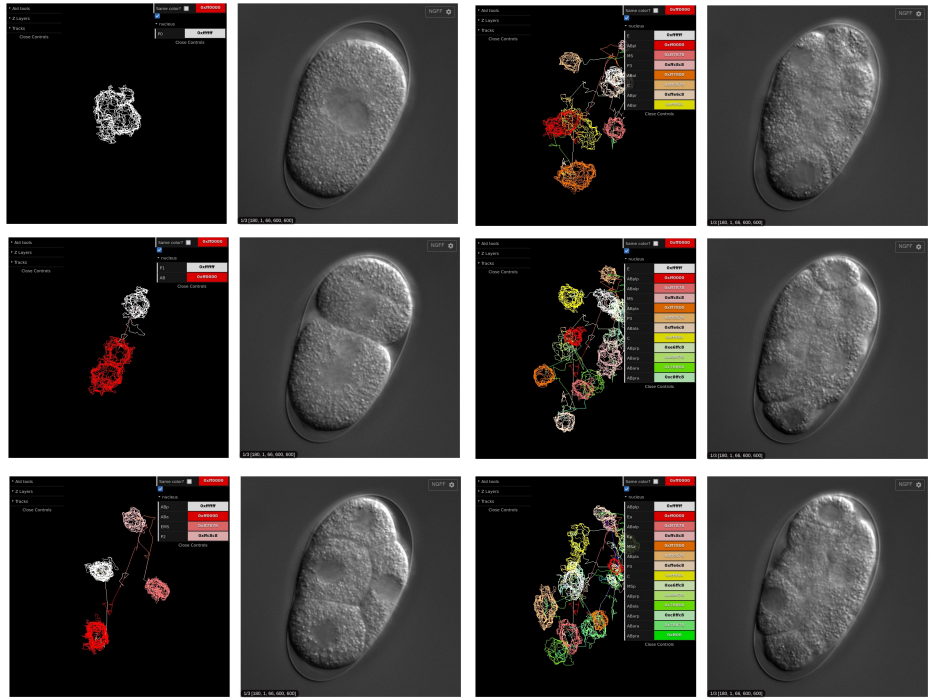


Quantitative data visualization system

- JavaScript-based BD5/BDML visualizer on SSBD web database
- More functional, more scalable, more network-effective
- Integrated visualization with OME-NGFF images



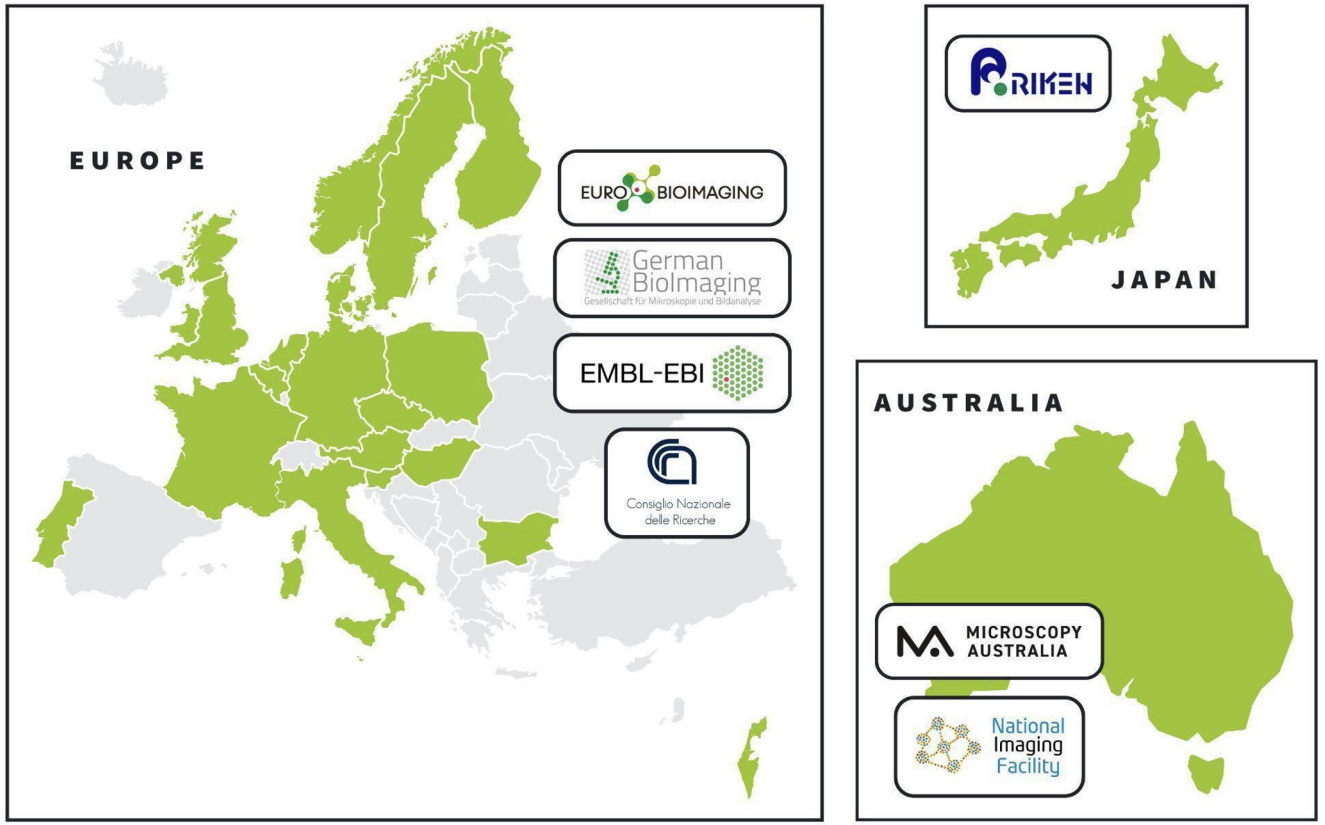




Founding**GIDE**: Founding a **Global Image Data Ecosystem**

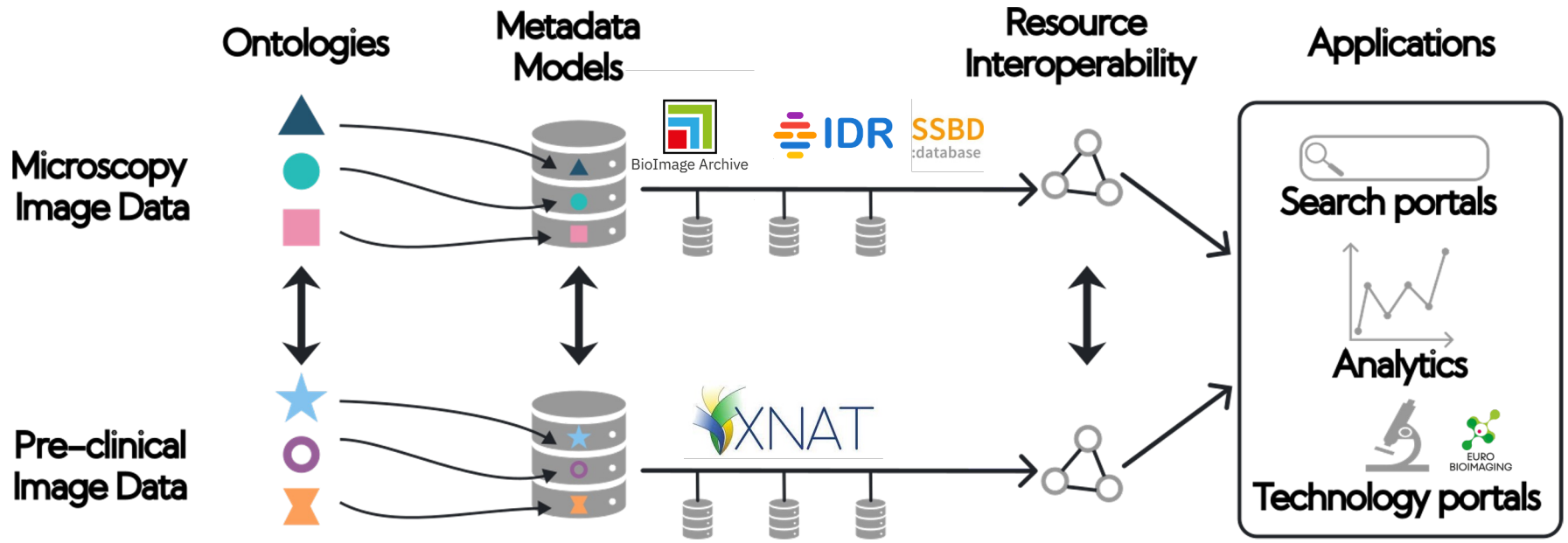
Laying strong foundation of an ecosystem for image data exchange based on global coordination of technical developments among data infrastructures and communities

- Global coordination among diverse imaging resources & communities
- Concerted development of Ontologies and Metadata models
- Adoption of outputs by global image data resources
- Interoperable solutions for microscopy and pre-clinical data
- Community recommendations FAIR image data management



Funded by the European Union

Integrating BioImage Archive, IDR and SSBD by harmonizing metadata models



Examination of the metadata models of BIA, IDR, and SSBD

- We have started collecting the metadata models used in BIA, IDR, and SSBD.
  - The comparison of the three metadata models will start in June.

class	item	type	multiplicity	range	description	examples
Project	SID	required	1	string	project id	2
	name	required	1	text	project name	2-Kyoda-WormEmbryoRNAi
	URL	required	1	string	URL for the project	https://ssbd.riken.jp/database/214/
	DOI	required	1	string	DOI for the project	https://doi.org/10.24631/ssbd.db.2023.05.296
	type	recommended	0..1	text	data types	image data, segmentation data, tracking data
	title	required	1	text	project title	Imaging data of centimeter-scale dynamics of multicellular systems observed with sub-cellular resolution
	description	required	1	text	description of the project	Optical images indicating optical performance of the imaging method, images of centimeter-scale dynamics of multicellular systems…
	date of submission	required	1	date	submission date	2021-09-23
	date of opened	recommended	0..1	date	open date (for reviewer)	2021-09-27
	date of released	recommended	0..1	date	release date (to the public)	2021-09-30
	date of updated	optional	0..1	date	update date	2021-10-08
	license	required	1	CC	license (Creative Commons)	CC BY 4.0
	funding info.	optional	0..1	text	funding information	The project is supported by …
	version	required	1	int	metadata version	5
	person	required	1..n	Person	people who contribute to the project	
	paper	required	1	Paper	paper of the project	
	dataset	required	1..n	Dataset	the sets of bioimaging data	
Person	role	required	1..n	text	role in the project	contact, imaging
	first name	required	1	text	first name	Shuichi
	last name	required	1	text	last/family name	Onami
	middle name	optional	0..1	text	middle name	
	e-mail	required	1	string	e-mail address	sonami@riken.jp
	URL	optional	0..1	string	URL for the laboratory	https://so.riken.jp
	organization	required	1	text	organization name	RIKEN Center for Biosystems Dynamics Research
	department	optional	0..1	text	department name	
	laboratory	optional	0..1	text	laboratory name	Laboratory for Developmental Dynamics
	address	optional	0..1	text	physical address	2-2-3 Minatojimaminamimachi, Chuo, Kobe …
	phone	optional	0..1	string	phone number	+81-xx-xxx-xxxx
	ORCID	recommended	0..1	ORCID	ORCID (https://orcid.org/)	0000-0002-8255-1724
	J-GLOBAL ID	optional	0..1	J-GLOBAL	J-GLOBAL ID (https://jglobal.jst.go.jp/en)	
	researchmap ID	optional	0..1	researchmap	researchmap ID (https://researchmap.jp/?lang=en)	
Paper	PubMed ID	optional	0..1	string	PubMed ID	23172286
	PMC ID	optional	0..1	string	PMC ID	
	DOI	recommended	0..1	string	DOI for the paper	10.1093/nar/gks1107
	URL	recommended	0..1	string	URL for the paper	https://academic.oup.com/nar/article/41/D1/D732/1059026
	paper info.	optional	0..1	text	paper information	Kyoda, K., et al (2003) …
Dataset	ID	required	1	string	ID of the dataset	RNAi_B0336.10_040518_01
	title	required	1	text	title of the dataset	3D time-lapse microscopy images about nuclear division dynamics in B0336.10(RNAi) embryo
	description	optional	0..1	text	description of the dataset	The image set was recoreded by differential interference contrast microscope …
	file format	recommended	0..1	string	file format of the dataset	ipm
	file size	recommended	0..1	text	file size of the dataset	8.0 GB
	figure number	recommended	0..1	text	figure number in which the dataset is used	Figure 2a; Figure 5b
	dimensions	required	1	Dimensions	dimensions of the dataset	
	biosample	required	1..n	Biosample	biosample used in the dataset	
	imaging method	recommended	1	Imaging Method	imaging method to obtain the dataset	
	instruments	recommended	1	Instruments	used instruments for the dataset	
Dimensions	dimensions	required	1	string	dimensions of the dataset (XYZTC)	600x600x66x180x1
	Xscale	recommended	0..1	float	physical size of a pixel	0.105
	Xscale Unit	recommended	0..1	UO	the Unit Ontology term for X	micrometer

metadata model of SSBD:database

Examination of the ontologies used in BIA, IDR, and SSBD

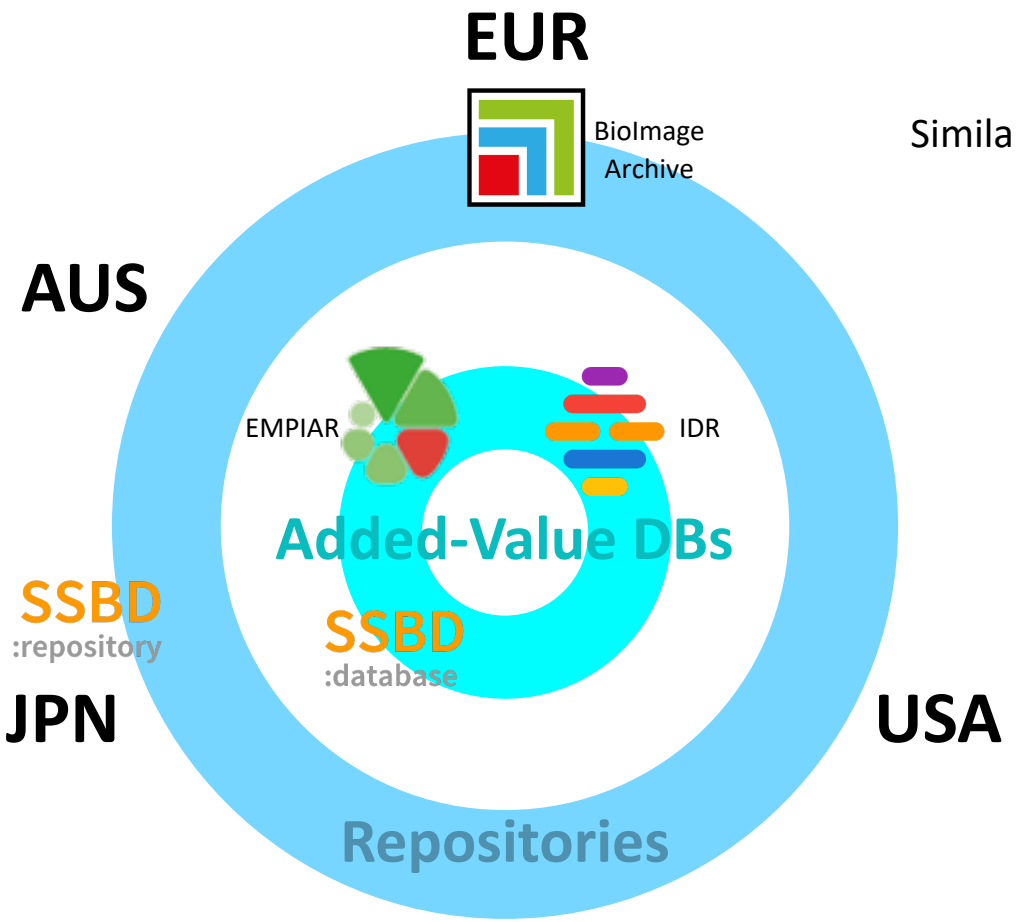
- We have started comparing the ontologies used in BIA, IDR, and SSBD.
  - Several vocabulary/ontology (NCBI Taxonomy, FBbi) are used in all repository/databases.
  - Different vocabulary/ontology (PubChem or CheBI, ENSEMBL or NCBI Gene) are used for some metadata.

Controlled vocabulary/ Ontology	Topic	BIA	IDR	SSBD
Gene Ontology (Biological Process)	Biologocal process	N	N	Y
Gene Ontology (Cellular Component)	Cellular component	N	N	Y
Gene Ontology (Molecular Function)	Molecular function, Molecular activity	N	N	Y
Cell Ontology	Cell	N	N	Y
Cell Line Ontology	Cell line	N	N	Y
Uberon multi-species anatomy ontology	Anatomy	N	N	Y
Biological Imaging Methods Ontology	Imaging	Y	Y	Y
Human Disease Ontology	Disease	N	N	N
Mondo Disease Ontology	Disease	N	N	N
SNOMED CT	Disease, Pathology	N	Y	N
Chemical Entities of Biological Interest Ontology	Chemical compound	N	N	Y
PubChem	Chemical compound	N	Y	N
NCBI organismal classification (NCBI Taxonomy)	Taxonomic classification, Organisms	Y	Y	Y
Ontology for Biomedical Investigations	Experimental conditions	N	N	Y
Experimental Factor Ontology	Experimental conditions	N	Y	Y
Human Phenotype Ontology	Phenotype	N	N	N
Mammalian Phenotype Ontology	Phenotype	N	N	N
Units of measurement ontology	Unit	N	N	Y
Cellular microscopy phenotype ontology	Cellular phenotype	N	N	N
ENSEMBL gene	Gene	N	Y	Y
NCBI Gene	Gene	N	Y	N
Uniprot	Protein	N	Y	Y
Medical Subject Headings (MeSH)	Controlled vocabulary	N	N	Y

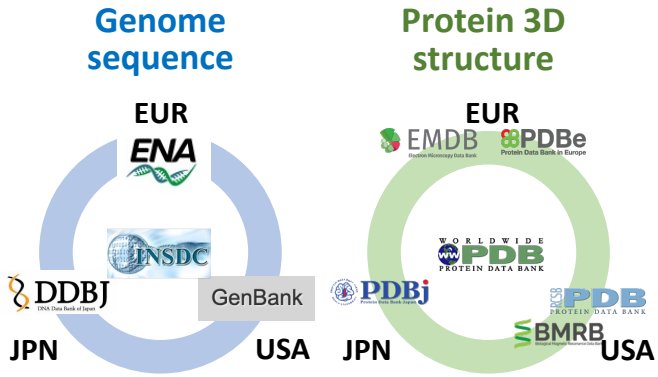
Integration of IDR and SSBD

- Both SSBD and IDR use the OMERO system – metadata is not the same
- The first targets
  - **Gene (name, ID)**
    - IDR – every dataset has gene name, id
    - SSBD – most datasets have gene names in the description (free text)
  - **Gene Ontology**
    - IDR – some datasets have Gene Ontology metadata
    - SSBD – every dataset has Gene Ontology metadata
- Toward integration using Gene and Gene Ontology
  - IDR – generate GO entries from gene names, ids
  - SSBD – Additional annotation of Gene for all datasets
    - New metadata template
    - Re-annotation of already published data

Global sharing of bioimaging data



Similarly to:



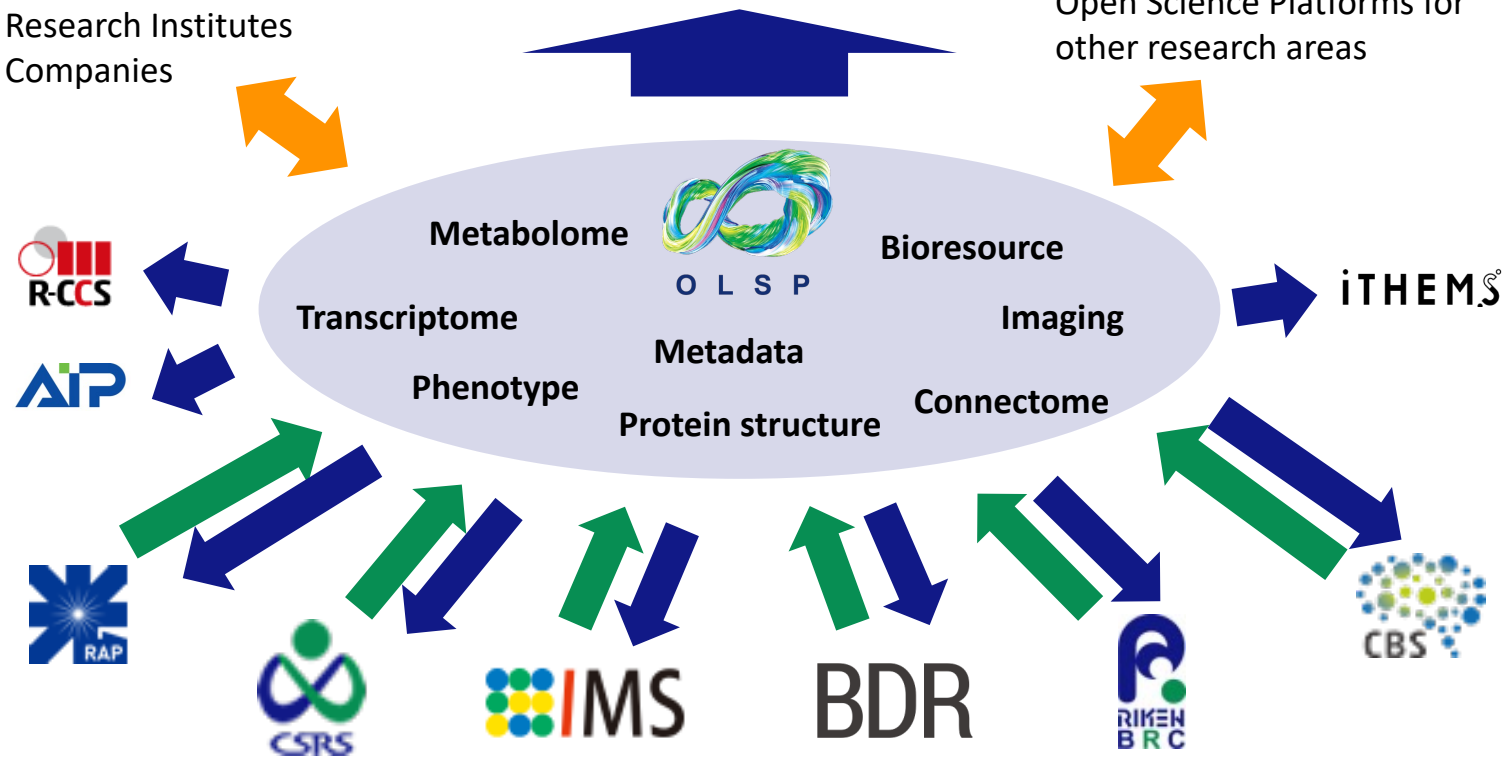
Biological data science and AI for Science in RIKEN

RIKEN Open Life Science Platform

Chair: Shuichi Onami

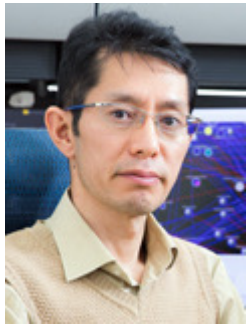
- Integrate and share RIKEN's cutting-edge life science data with the world
- Develop technologies and mechanisms necessary for open science
- Lead the world in open science

Disseminate RIKEN's life science data to the world
Facilitate reuse of the data



RIKEN started Open Science Promotion project in 2019
R-IH Life Science Data Sharing Unit (Onami UL) was established in 2020

Core member of RIKEN OLSP/R-IH Life Science Data Sharing Unit



Masanori Arita

CSRS

Metabolome
Sequencing

Shuichi Onami

BDR

BioImaging

Takeya Kasukawa

IMS

Transcriptome

Norio Kobayashi

R-IH

Metadata

Tomomi Shimogori

CBS

Brain science

Hiroshi Masuya

RIKEN BRC

BIoResource

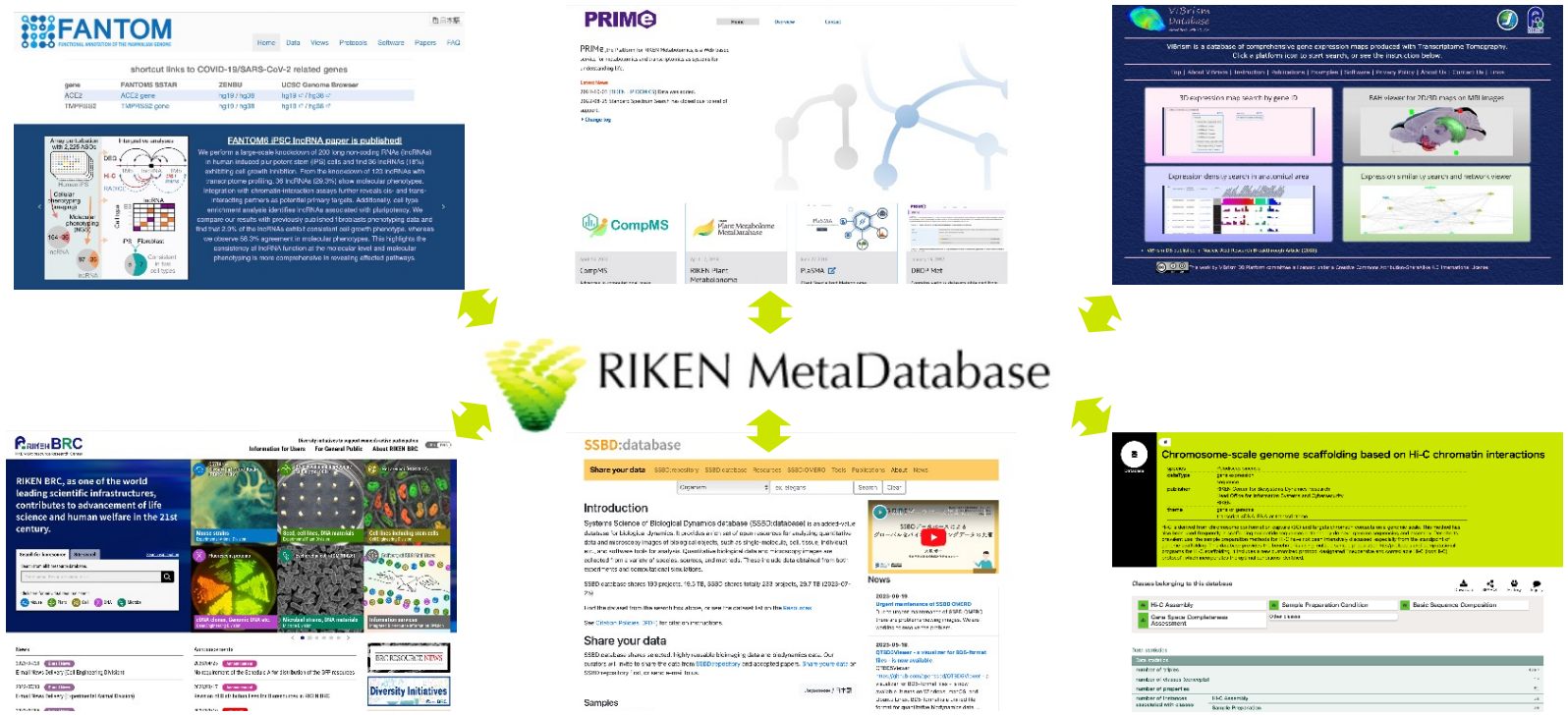
Hideo Yokota

RAP

Image
Processing

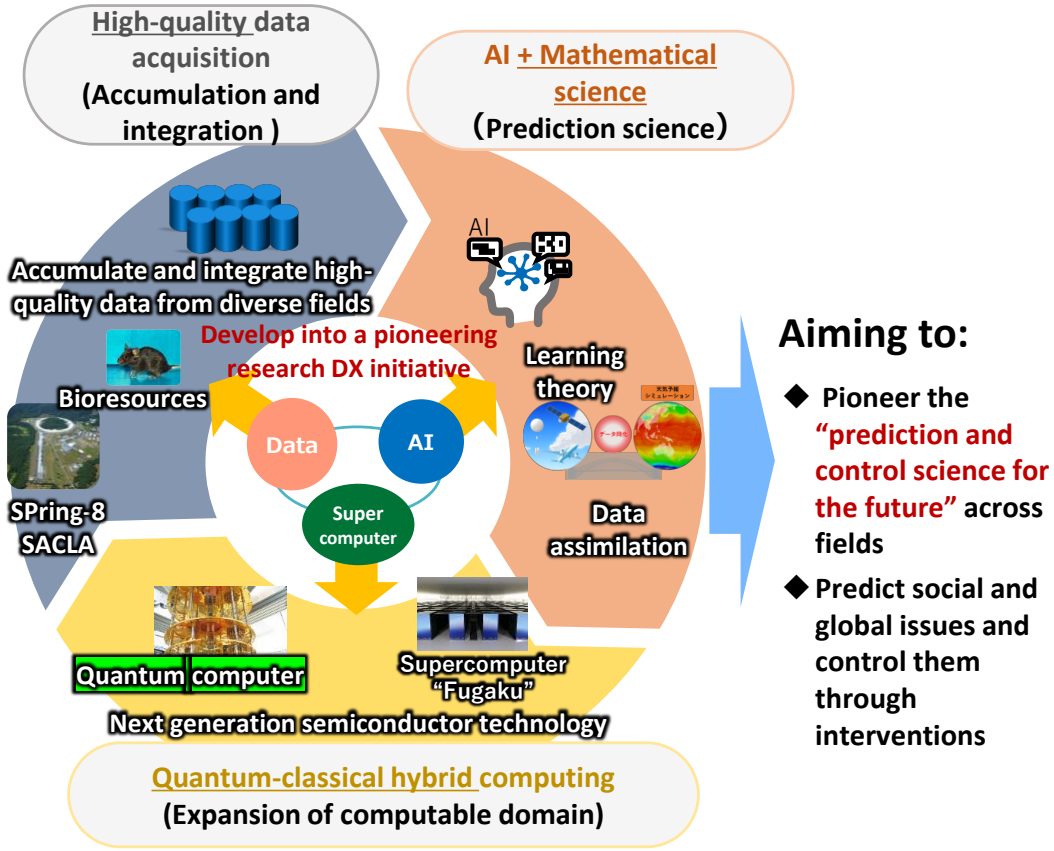
Integration of large-scale life science datasets in RIKEN

- Registered the metadata of large-scale life science databases in RIKEN in the RIKEN MetaDatabase
- Consolidated these databases at the metadata level
  - **bioresource data** from BRC; **FANTOM5 data** from IMS; **metabolome data** from CSRS; **bioimaging data**, **Hi-C data**, **scRNAseq data** from BDR; **ViBrism data** from RAP



Transformative Research Innovation Platform of RIKEN platforms (TRIP)

- Institution-wide project launched in 2024
- Aim to create new “prediction and control science” for solving social and global issues by combining
  - High-quality data acquisition and integration
  - AI + mathematical modeling
  - Quantum-classical hybrid computing
- Several pioneering directions
  - Transmutation of elements
  - Many-body electron system
  - Agriculture
  - Polymer chemistry
  - Drug discovery and medical technology
  - Foundation model for science



Source: RIKEN news

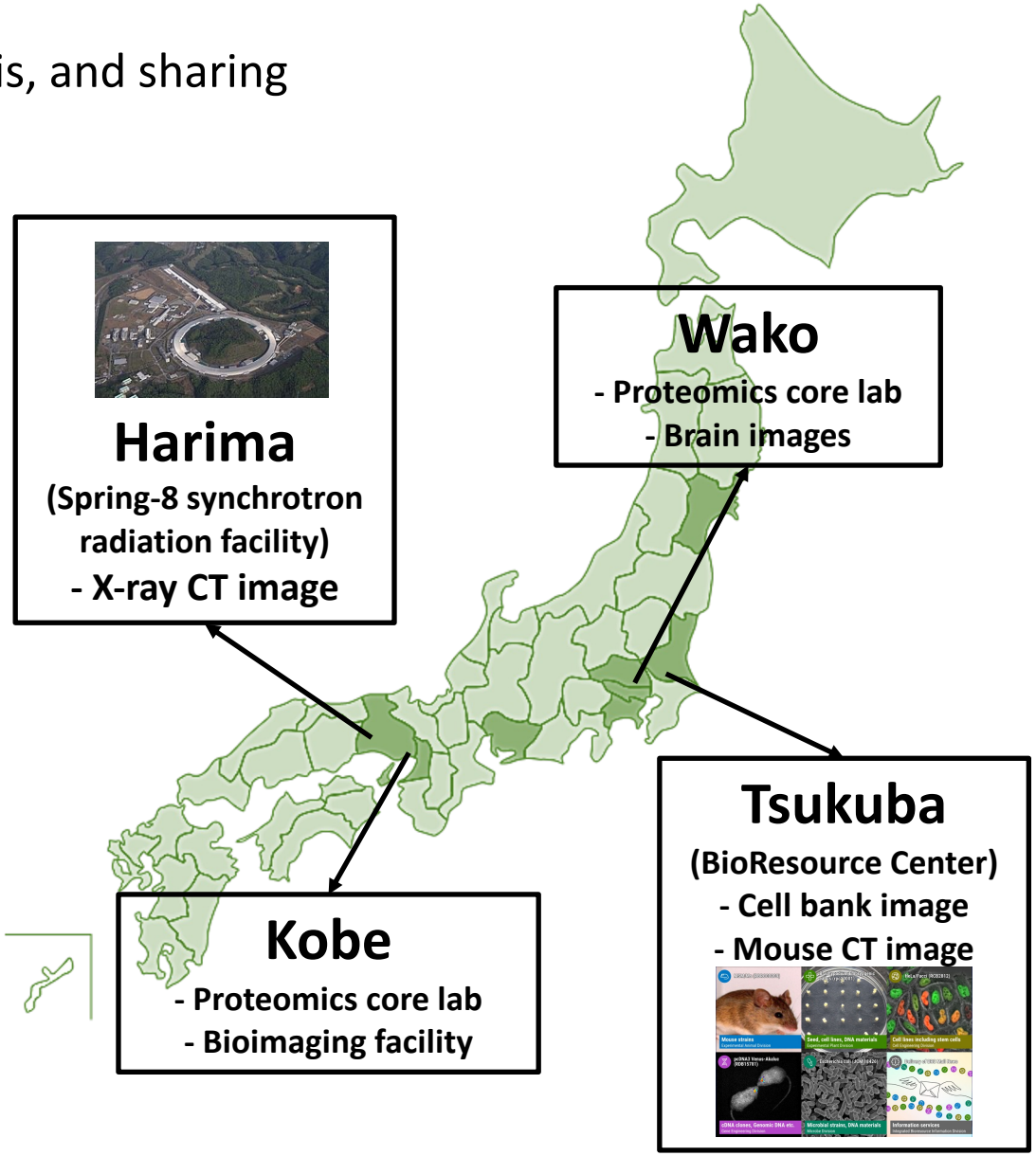
Research DX Foundation Team

- Started in 2023 to support the TRIP project
- Currently 10 members located in 4 campus with diverse scientific background
  - 5 biology
  - 4 physics
  - 1 chemistry
- Foster AI & data-driven science
  - Currently focusing on the data part
  - Future also move to analysis & application

- Now recruiting 10 more members!

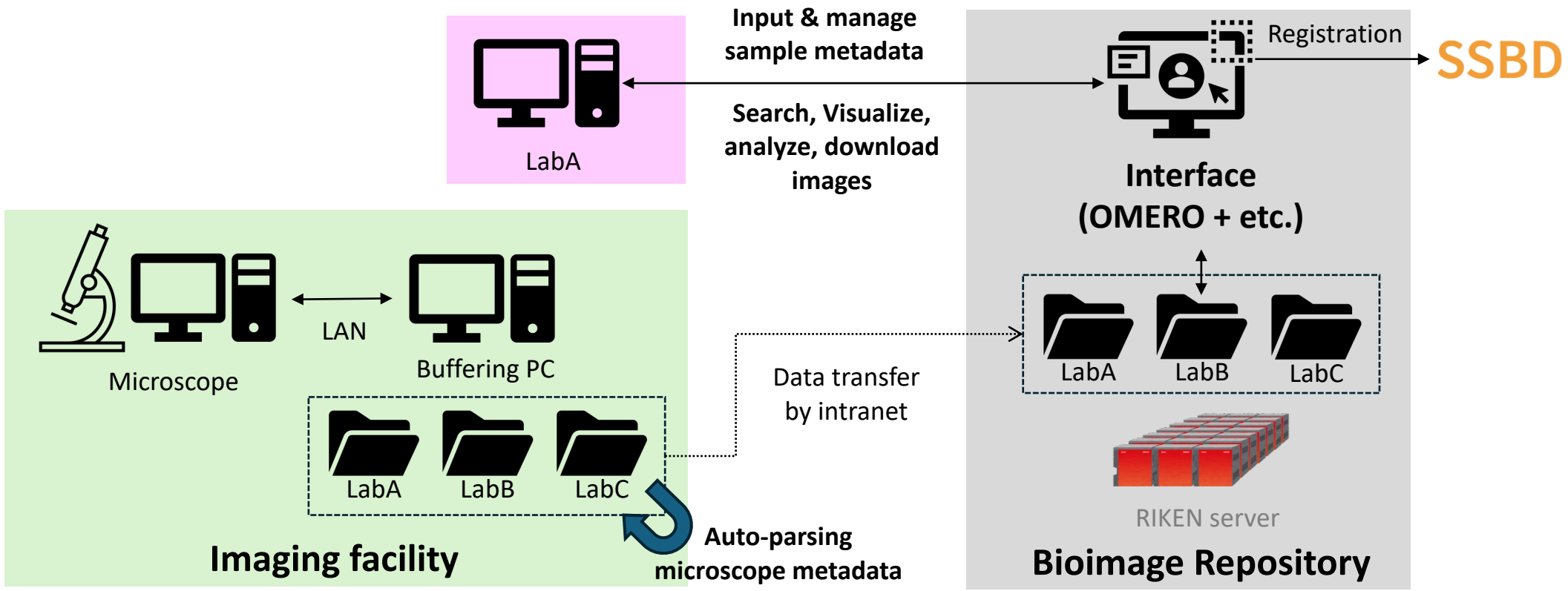
Towards an integrated data platform in RIKEN

- Integrated platform for data management, analysis, and sharing
- Data source
  - Core facility, National Infrastructure
  - Individual labs
- Data type
  - Bioimage (light microscope)
  - Animal/material CT image
  - Proteome/Metabolome (mass spectrometry)
  - Genome (DNA/RNA sequencing)
  - SAXS
  - NMR



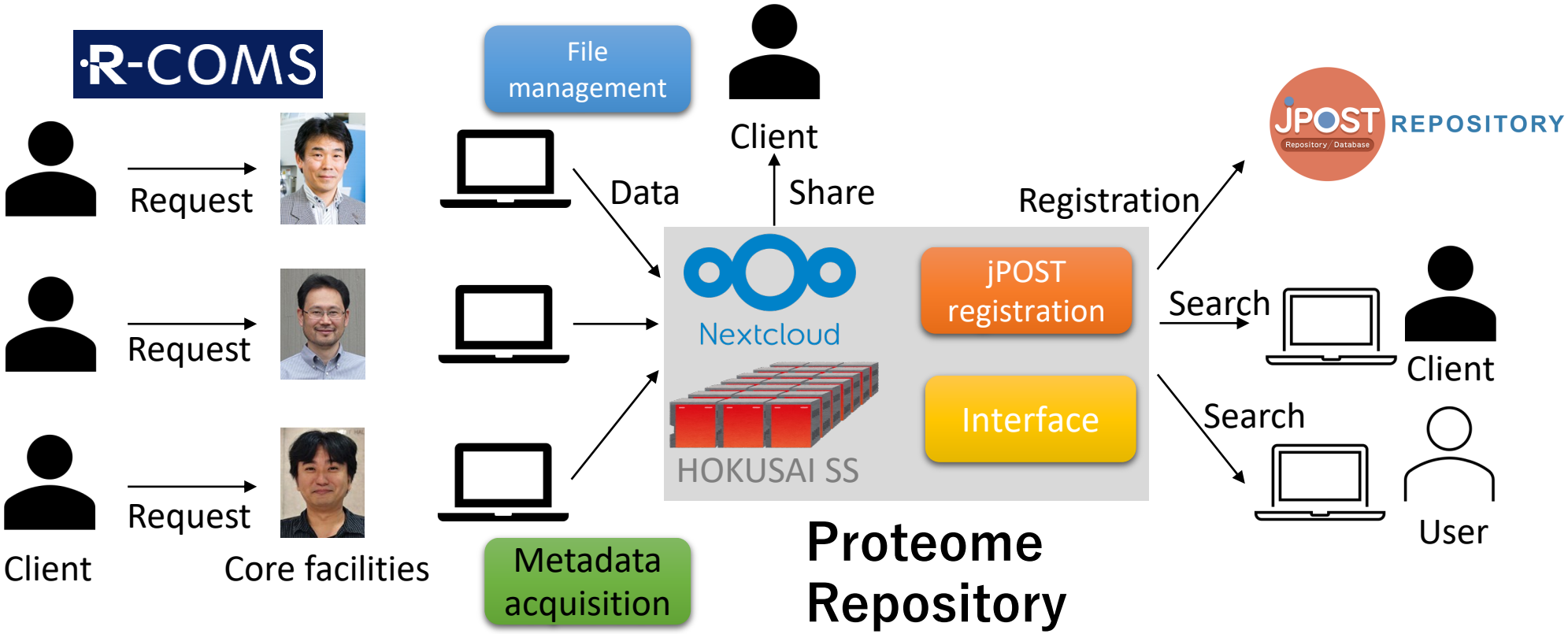
Bioimaging data

- Building a workflow for central management and access of bioimaging data
  - Bioimaging Facilities and Factory at Kobe
  - RIKEN cell bank located at Tsukuba
  - Parallelly brain imaging at Wako



Proteomics data

- Building an integrated workflow to streamline proteomics data acquisition, storage, analysis, and sharing across core facilities located across three campus
  - Standardize metadata input
  - Automatic deposit to external repository when ready for publication

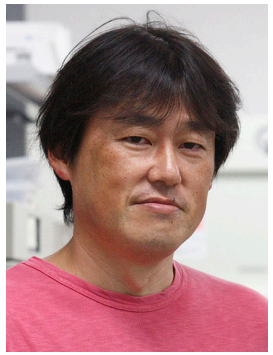


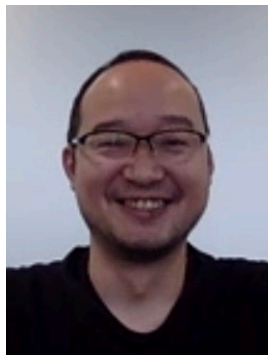
Share your data via SSBD !
Connect your database with SSBD !

Data registration to SSBD:repository: **SSBD-repos@ml.riken.jp**

Other questions: **SSBD@riken.jp**

SSBD team


S. Onami


K. Kyoda

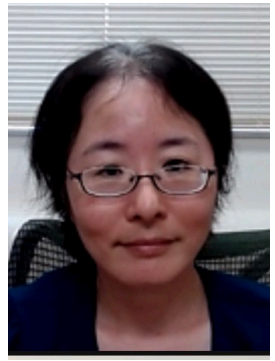

H. Itoga


E. Fujisawa


H. Yamamoto


M. Miranda


Y. Yamagata


Y. Tohsato

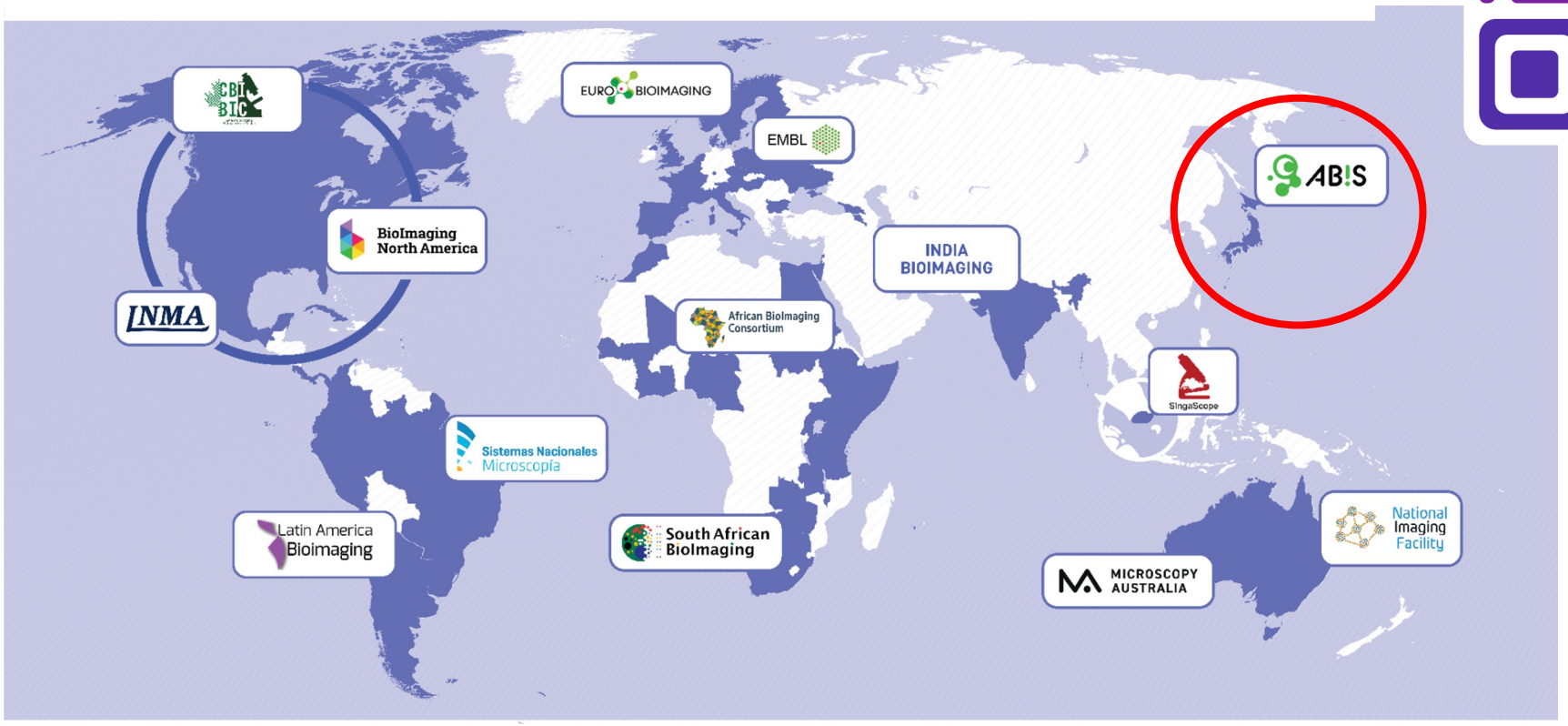
JST NBDC CREST ABiS foundingGIDE Funded by the European Union 科技ハブ産連本部 OLSP DECODE project Singularity 科研費 KAKENHI

Exchange of Experience 2024 - #GBI_EoE2024

DATES: 29th to 31st of October of 2024

LOCATION: Okazaki Conference Centre, Okazaki - Japan

TOPIC: Image Data Horizons - Global Strategies for Accessible Knowledge



GLOBAL BIOIMAGING
growing collaboration

foundingGIDE Community Event 2024

October 31 @ 13:00 – November 1 @ 17:00

foundingGIDE

Community Event 2024

We invite you to the first foundingGIDE Community Event 2024 that will bring together various stakeholders from the global imaging data community to Okazaki, Japan.

At **#foundingGIDE** we are dedicated to championing best practices in image data sharing and facilitating the exchange of bioimage data among major repositories.

A key step towards creating a global image data ecosystem is to **connect** the community.

Date: 31st of October – 1st of November

Location: Okazaki Conference Centre, **Okazaki**, **Japan**

Topic: Connecting communities in the **G**lobal **I**maging **D**ata **E**cosystem

Presentations available @

https://downloads.openmicroscopy.org/presentations/2024/Dundee