

Integrating OMERO with SSBD to create a database of quantitative data and microscopy images of biological dynamics

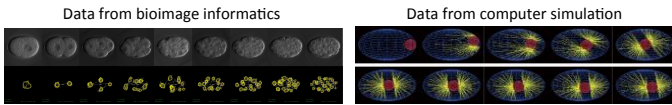
Yukako Tohsato^{1,2}, Kenneth H.L. Ho¹, Koji Kyoda¹, Shuichi Onami¹

1) RIKEN Quantitative Biology Center, Kobe, Japan, 2) Osaka Electro-Communication University, Osaka, Japan

Motivation

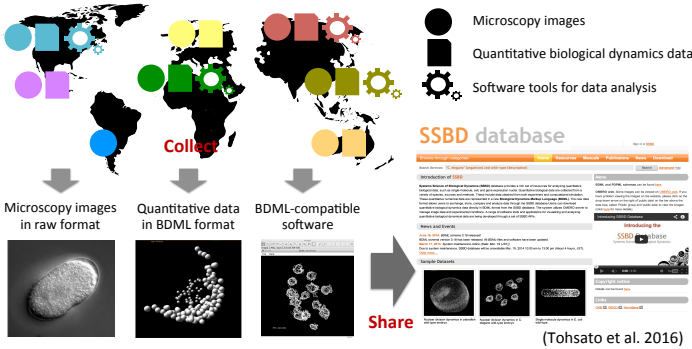
Recent progress in bioimage informatics and computer simulation produced a large amount of quantitative data on spatiotemporal dynamics of various kinds of biological objects.

Example of quantitative biological dynamics data



There is a need to bring these data centrally in an open and unified data format. We developed **SSBD (Systems Science of Biological Dynamics database)**, with the support from National Bioscience Database Center of Japan. SSBD collects and shares quantitative data and microscopy images. It also includes software tools for visualization and analysis.

Concept of SSBD



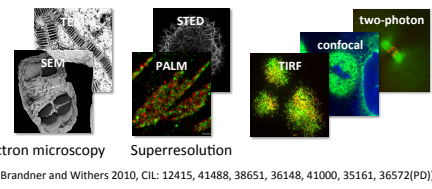
Cooperation with Japanese societies

Microscopy images are managed by using **OMERO platform** internally in SSBD.

Since last year, SSBD begins to store and share raw microscopy images with no quantitative data. We focus on, but not limited to, time-lapse image data that are obtained by using state-of-the-arts

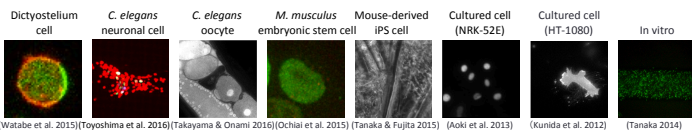
imaging systems. We plan to collect further images by cooperating with Japanese societies for cell biology and developmental biology.

Microscopy image in cell and developmental biology



(Brandner and Withers 2010, CIL: 12415, 41488, 38651, 36148, 41000, 35161, 36572(PD))

Summary of new images



SSBD stores microscopy images and quantitative data ranging from molecules to cells to organisms.

List of available data resources in SSBD

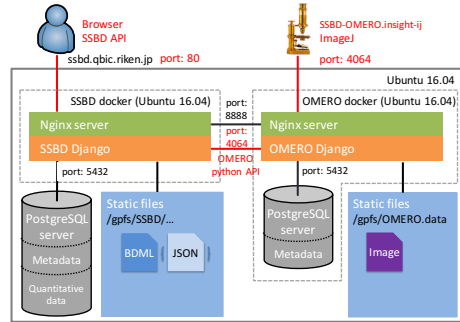
Species	Type	BasedOn	Reference	#BDML	#entities	#Images	Microscopy
<i>E. coli</i>	single molecule	Simulation	Arjunan & Tomita 2010	1	721	0	-
<i>D. discoideum</i>	single molecule	Experiment	Komatsuzaki et al. 2014	1	987	1,800	TIRF
<i>D. discoideum</i> , PC12, NA	single molecule	Simulation / Experiment	Watabe et al. 2015	4	745,846,573	41	Confocal
<i>C. elegans</i>	nucleus	Experiment	Bao et al. 2006	2	24,747	0	-
<i>C. elegans</i>	nucleus	Experiment	Kyoda et al. 2013	186	75,955	26,640	DIC
<i>C. elegans</i>	nucleus	Experiment	Tohsato et al. 2016	14	125,426	13,954	Confocal, SDC, EF
<i>C. elegans</i>	nucleus	Experiment	Tohsato et al.	1,582	2,430,536	4,512	DIC
<i>C. elegans</i>	nucleus	Experiment	Kyoda et al.	1,147	360x66x1147	-	DIC
<i>C. elegans</i>	cell	Experiment	Takayama & Onami 2016	119	24,608	77,306	SDC
<i>C. elegans</i>	pronucleus/microtubule	Simulation	Kimura & Onami 2005	100	2,400,100	0	-
<i>C. elegans</i>	locomotion	Experiment	Cronin et al. 2005	11	15,822	0	-
<i>D. melanogaster</i>	nucleus	Experiment	Keller et al. 2010	2	5,111,828	0	-
<i>D. rerio</i>	nucleus	Experiment	Keller et al. 2008	7	56,584,840	0	-
<i>M. musculus</i>	nucleus	Experiment	Bashar et al. 2012	1	2,054	2,800	SDC
<i>M. musculus</i>	gene expression	Experiment	Harima et al. 2013	2	146	2	BF
<i>M. musculus</i>	molecule	Experiment	Ochiai et al. 2015	0	0	61	SDC
<i>M. musculus</i> (iPS), NA	particle	Experiment	Tanaka & Fujita 2015	2	0	14,573	Phase
<i>M. musculus</i>	gene expression	Experiment	Masumoto et al. 2010	8	48	0	-
NRK-52E	cell	Experiment	Aoki et al. 2013	0	0	856	EF
HT-1080	cell	Experiment	Kunida et al. 2012	0	0	211	FRET
NA	particle	Experiment	Tanaka 2014	0	0	200	EF

http://ssbd.qbic.riken.jp

System update into Docker containers

OMERO platform and SSBD core platform were individually rebuilt using Docker containers. This enables us to minimize interferences between the two platforms and also makes it easier for us to manage libraries and software tools compatibility. The OMERO is incorporated into SSBD by using OMERO Python API and Nginx proxy allowing SSBD to interact with OMERO.

Overview of SSBD database system



Hardware

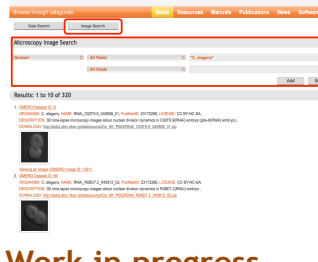
- Production environment:
- IBM System x3550 M4
 - Xeon Processor E5-2620 2.00 GHz
 - 32 GB RAM
 - 10 GbE connectivity
 - 200 TB GPFS RAID space
- Testing environment:
- Dell PowerEdge R620, R720xd
 - Xeon Processor E5-2650 2.6 GHz
 - 128 GB RAM
 - 10 GbE connectivity
 - Dell PowerVault MD3260, 3060e, 3460
 - 300 TB + 200 TB RAID space

We also released a Dockerfile to build a ready-to-use OMERO in auto-login mode at GitHub (<https://github.com/openssbd/omero-web/>).

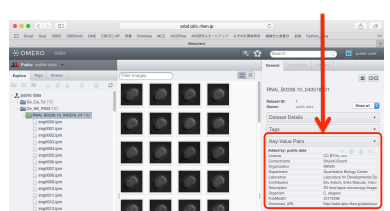
Django app. development to search images in OMERO

An image search web-service becomes available by developing a Django application of the OMERO. This enables users to search for images based on the description using metadata.

SSBD Database



Metadata are annotated as list of key-value pairs for Projects, Datasets, and Images.

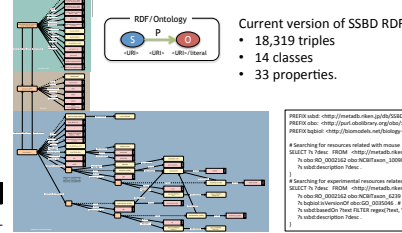


Work in progress

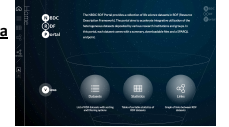
RDF-formatted metadata based on OME-XML

To accelerate integrative use of the heterogeneous datasets

Meta-information about image data and quantitative data



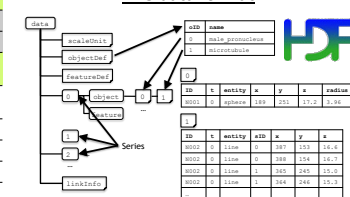
NBDC RDF Portal (<http://integbio.jp/rdf/>)



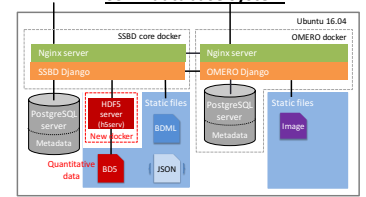
SSBD - adopting HDF5 for representing quantitative data

To allow more efficient access to quantitative data and allowing a more flexible architecture

BD5 data format



SSBD database system



Ref.

- Tohsato, Y. et al. (2016) SSBD: a database of quantitative data of spatiotemporal dynamics of biological phenomena. *Bioinformatics* **32**, 3471-3479.
- Kyoda, K. et al. (2015) Biological Dynamics Markup Language (BDML): an open format for representing quantitative biological dynamics data. *Bioinformatics* **31**, 1044-1052.

Funding

This work was supported in part by the National Bioscience Database Center (NBDC) of the Japan Science and Technology Agency (JST); the Strategic Programs for R&D (President's Discretionary Fund) of RIKEN; and a Grant-in-Aid for Scientific Research (C) (No. 16K00414).