

SSBD: an integrated database of quantitative data and microscopy images of biological dynamics

Yukako Tohsato¹, Kenneth Ho H.L.¹, Koji Kyoda¹, Shuichi Onami^{1,2}
 1) RIKEN QBiC, Japan, 2) NBDC, JST, Japan

Overview

SSBD aims to provide an IT platform for accelerating data-driven approach for understanding biological phenomena as dynamical systems. We focus on **quantitative data of spatiotemporal dynamics** generated from microscopy image with image analysis and computer simulation (Fig. 1).

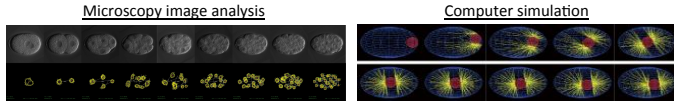


Fig. 1: Example of quantitative numerical data

Formats of such data are different (Fig. 2). Data and software tools are stored separately all over the Internet.

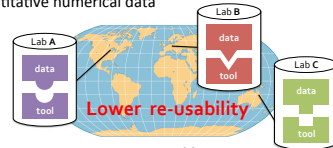


Fig. 2: Problem

Our solution:

- **BDML: Biological Dynamics Markup Language**
- **SSBD: Systems Science of Biological Dynamics database**

Biological Dynamics Markup Language (BDML)

BDML is a unified format for representing quantitative data of biological dynamics ranging from molecules to organisms (Fig. 2; Kyoda et al., Bioinformatics, 2015). The schema consists of five top-levels.

1. Info: Information of the BDML file, e.g. license, release date, etc.
 2. Summary: Short summary of the numerical information of biological dynamics
 3. Contact: Contact information of the author/owner/maintainer of the BDML files
 4. Methods: Sources files and extraction protocols
 5. Data: Description of numerical data
- } meta-data

Data are represented in five different types of features, point, line, face, circle, and sphere (Fig. 3).

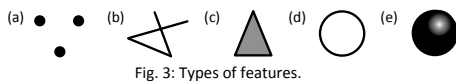


Fig. 3: Types of features.

For example, contours of nucleus in Z-sliced images can be represented as a set of lines (Fig. 4a). Nucleus labeled with GFP is represented as a sphere (Fig. 4b). Pronucleus and microtubules are represented as a set of sphere and lines, respectively (Fig. 4c).

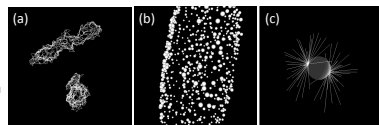


Fig. 4: Examples of visualization of various types of graphical representation of data for nuclei

Coordinates values of contour of nuclei can be described in BDML file (Figs. 5 and 6).

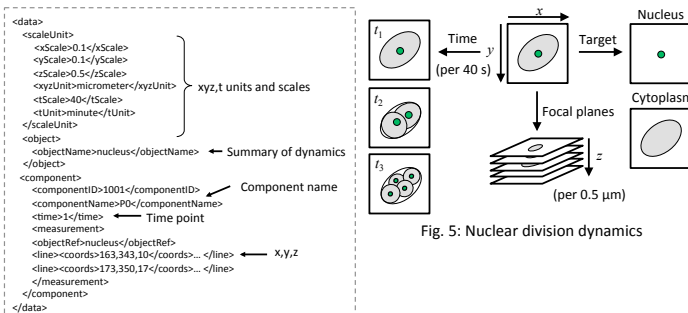


Fig. 5: Nuclear division dynamics

Fig. 6: Example of data section

An example of using BDML and SSBD in embryogenesis

By using BDML/SSBD, we analyzed the total number of nuclei over time during embryogenesis of three different types of organisms (Figs. 7 and 8). Stepwise increase indicates synchronization of nuclear division (for example C9 in Fig. 7b).

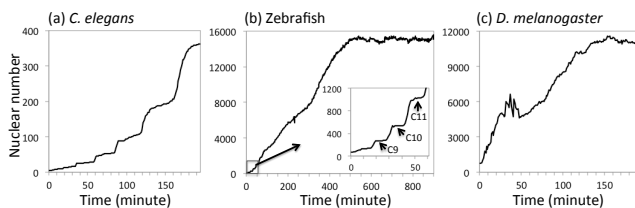


Fig. 7: Total number of nuclei over time in embryogenesis

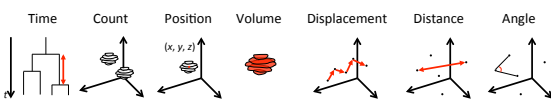


Fig. 8: Calculating phenotypic characters among components using Phenochar

SSBD database <http://ssbd.qbic.riken.jp/>

SSBD is a open repository for quantitative data and their corresponding microscopy image. It also includes software tools for visualization and analysis (Figs. 9 and 10). OMERO is integrated into SSBD for managing and displaying microscopy images.

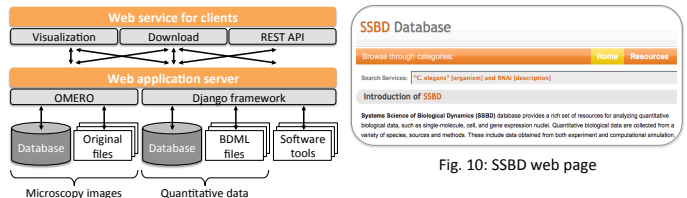


Fig. 9: Overview of SSBD database

Fig. 10: SSBD web page

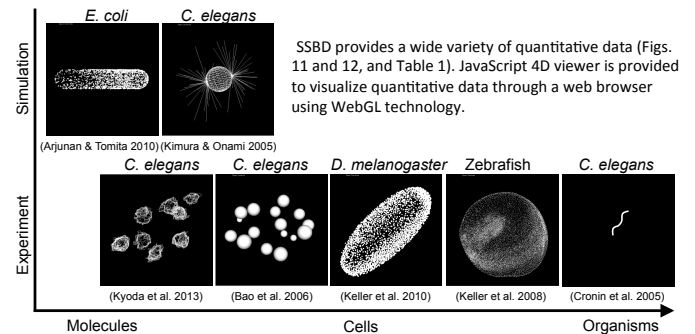


Fig. 11: Representative quantitative data and their visualization

Table 1: List of available resources in SSBD: ■ planned release, □ currently available

Organism	Dynamics	Source	Paper	#BDML	#images
<i>M. musculus</i>	nucleus	exp.	Bachar et al. 2012	1	2800
<i>M. musculus</i>	expression	exp.	Masumoto et al. 2010	8	
<i>M. musculus</i>	nucleus	exp.	Kurotaki et al. 2007	1	80
mammals	nucleosome	sim.	Hirata et al. 2012	1	
zebrafish	nucleus	exp.	Keller et al. 2008	7	
<i>D. melanogaster</i>	nucleus	exp.	Keller et al. 2010	2	
<i>D. melanogaster</i>	nucleus	exp.	Supatto et al. 2009	1	
<i>C. elegans</i>	nucleus	exp.	Kyoda et al. 2013	186	180x6x186
<i>C. elegans</i>	nucleus	exp.	Bao et al. 2005	2	
<i>C. elegans</i>	nucleus+microtubule	sim.	Kimura & Onami 2005	100	
<i>C. elegans</i>	behavior	exp.	Cronin et al. 2005	11	
<i>C. elegans</i>	nucleus+embryo	exp.	Tohsato et al.	1331	
<i>C. elegans</i>	nucleus	exp.	Kyoda et al.	259	360x6x259
<i>C. elegans</i>	nucleus	exp.	Takayama et al.	12	4899
<i>C. elegans</i>	expression	exp.	Sarov et al. 2012	273	
<i>D. discoideum</i>	molecule	exp.	Jin et al.	1	368
<i>E. coli</i>	molecule	sim.	Arjunan & Tomita 2010	1	

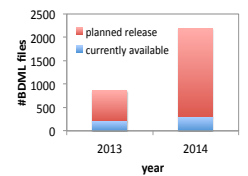


Fig. 12: Changes in the number of BDML files

ImageJ support

We are developing ImageJ plugins to support users to produce quantitative data from microscopy images using ImageJ (Fig. 13).

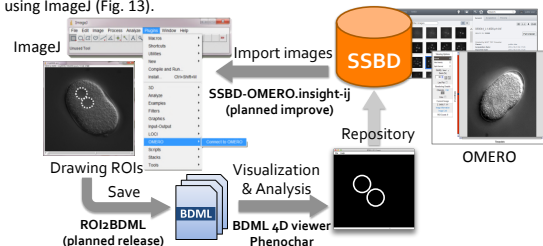


Fig. 13: Software tools related to ImageJ coordination

Towards RDF meta-data service

We will include RDF description on the meta-information of the data to all the BDML files (Fig. 14). We will also provide triplestores for querying RDF meta-data (Fig. 15).

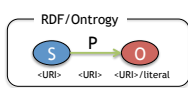


Fig. 14: Our strategy

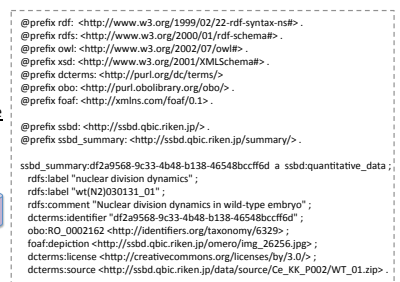


Fig. 15: Example of RDF triplestore for meta-data