

JST-NBDC

平成24年度ライフサイエンスデータベース統合推進事業

統合化推進プログラム研究開発課題（平成24-26年度）

生命動態システム科学のデータベースの統合化

生命動態システム科学のデータベース 統合化の現状と今後の展開

大浪修一

理化学研究所生命システム研究センター



背景

- 生命動態システム科学
 - 複雑な生命現象の動態を時空間を有する先端定量計測と高精度モデリングをもとに、in silicoとin vitroで再構成
- 生命動態システム科学のデータ
 - 時空間情報を数値として含む新しい様式の生命科学の研究データが解析の中心
 - 計測データ
 - シミュレーションデータ
- 生命動態システム科学のDB
 - 公開中のDBの数はまだ多くない
 - データ生産が急速に増加中
 - 今なら、DBの乱立に先行して統合化のスキームを構築できる。

生命動態システム科学のDBの統合化

目標

- 我が国の生命動態システム科学分野の全てのDBを統合する体制と仕組みの構築
- 我が国の生命動態システム科学の全てのDBの統合
- 生命動態システム科学分野の全世界のDBの統合

謝辞

プロジェクト・グループ

JST CREST研究代表者

飯野雄一（東大）
影山龍一郎（京大）
黒田真也（東大）
洪 実（慶大）
近藤 滋（阪大）

理研QBiCグループリーダ

上田泰己（理研）
上田昌宏（理研）
泰地真弘人（理研）

文科省推進拠点研究代表者

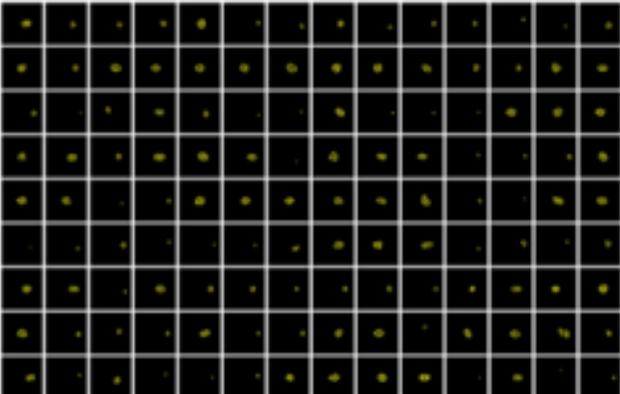
井原茂男（東大）
金子邦彦（東大）
楯 真一（広大）
松田道行（京大）

統合フォーマットの構築

生命動態システム科学のデータの課題点

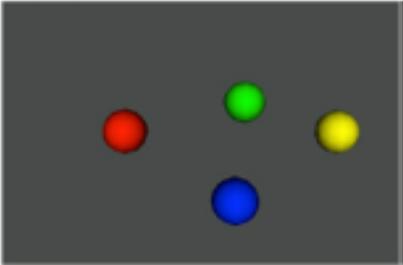
異なるデータ・フォーマットの使用

Onami et al. 2001, Hamahashi et al. 2007, Kyoda et al. 2013



```
0000, 1000, 51.8232, 37.7117, 27.1820, 1, P0, 17  
477, 342, 23, 211111111210111212112.....  
486, 321, 24, 212112222100121222102.....  
.....  
1000, 2000, 33.3146, 32.1893, 25.2702, 2, P0, 17  
320, 257, 23, 211111211111211111112.....  
.....
```

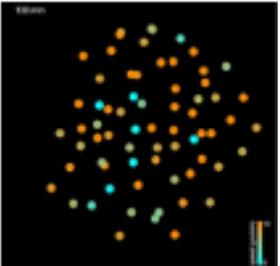
Bao et al. 2006



t001-nuclei
t002-nuclei
.....

```
1, 1, -1, 4, -1, 380, 366, 16.1, 80, EMS, 21281015,  
2, 1, -1, 2, -1, 387, 153, 16.6, 86, ABp, 2836266,  
3, 1, -1, 3, -1, 189, 251, 17.2, 88, ABa, 2850348,  
4, 1, -1, 5, -1, 562, 269, 18.1, 80, P2, 2168825
```

Keller et al. 2008



<68x17 double>
<69x17 double>
.....

```
611.2904, 563.0863, 80.3444, 13.8991, 14.7172, 606, 551, ....  
485.5862, 546.2255, 91.1758, 14.0556, 18.5152, 480, 534, ....  
562.8646, 459.1969, 95.8143, 14.2670, 17.0606, 556, 450, ....  
.....
```

BDML: Biological Dynamics Markup Language

生命動態システム科学の定量データを記述する言語

- XMLを基盤
 - 高い可読性
 - 高い拡張性

```
<scaleUnit>
  <tScale>20</tScale>
  <tUnit>second</tUnit>
</scaleUnit>
<component>
  <componentID>100</componentID>
  <time>1</time>
  <measurement>
    <line><coords>10.32,30.42,18.32</coords></point>
  </measurement>
</component>
<component>
  <componentID>101</componentID>
  <time>2</time>
  <prevID>100</prevID>
  <measurement>
    <line><coords>9.57,32.05,14.91</coords></point>
  </measurement>
</component>
```

BDMLの開発の歴史

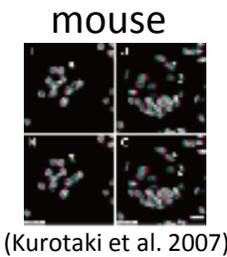
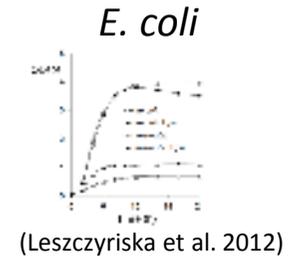
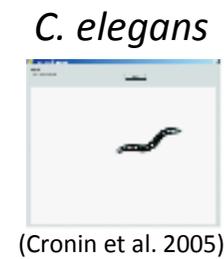
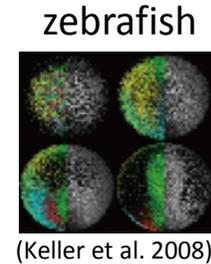
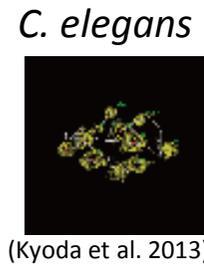
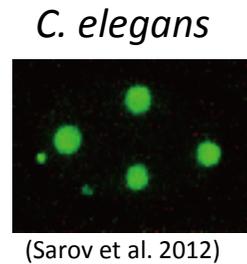
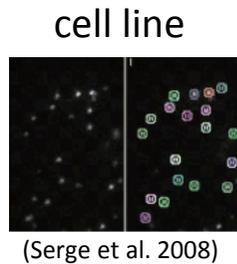
- BDD-XML Jan 21 2013
 - 細胞核分裂動態のみ対応。
- BDML ver. 0.15 Sep 2 2013
 - 初代公開version。一分子から個体、集団、遺伝子発現動態に対応。
- BDML ver. 0.18 Jun 18 2014
 - Series, Setの導入。
- BDML ver. 0.2 Aug 28 2014
 - RDF, オントロジーへの対応。

BDMLで記述可能な定量データ

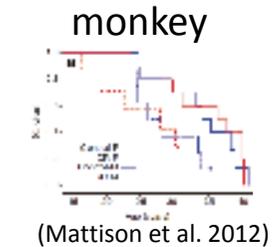
- 分子、細胞、個体から集団の動態まで。

計測データ

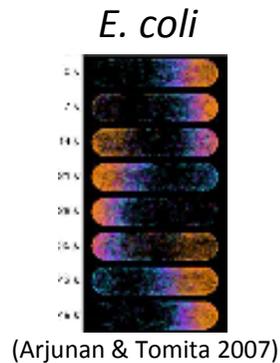
遺伝子発現



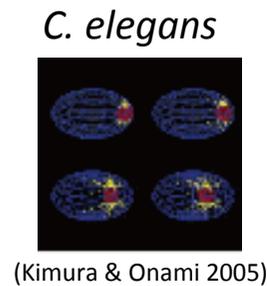
行動



一分子



細胞核分裂



増殖・生存

核・微小管

BDML ver. 2.0

シミュレーション結果

分子

細胞

個体

集団

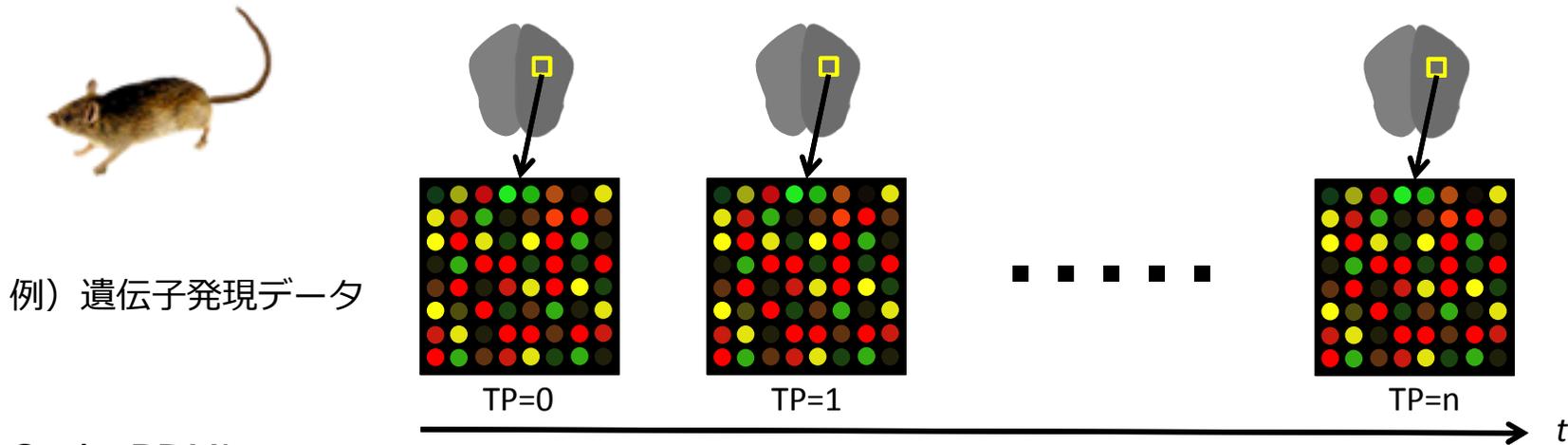
BDMLで記述可能な理研QBiCのデータ

BDML ver. 2.0 記述可能

生物種	対象	方法	研究室
Hela細胞	一分子	計測	上田昌宏
Hela細胞	FCS	計測	神 隆
Hela細胞	一分子	モデル	高橋恒一
肝臓細胞	MS	計測	升島 勉
マウス	一分子	計測	岡田康志
マウス	発現	計測	上田泰己
ゼブラフィッシュ	核	計測	岡田康志
線虫	核	計測	大浪修一
線虫	核	計測	大浪修一
線虫	核・微小管	モデル	大浪修一
粘菌	一分子	計測	神 隆
出芽酵母	発現	計測	谷口雄一
大腸菌	生育	計測	古澤 力
大腸菌	一分子	モデル	高橋恒一

OmicxBDML

- 時空間情報を持つオミクスデータセットのためのフォーマット。
 - 既存のフォーマットは時空間情報を定量的に表現できない。
 - 既存フォーマット: MiNiML, mzMLなど
 - BDMLには効率的に格納できない。
- 既存のフォーマットを取り込んで再利用。



OmicxBDML

```

<xml version=" " />
<omicxbdml ...>
  <MINiML:Series>
    <MINiML:Status />
    <MINiML:Title />
    ...
  </MINiML:Sample>
  <time>0</time>
  <target>PT</target>
  <MINiML:Sample>
    <MINiML:Status />
    <MINiML:Title />
    ...
  </MINiML:Sample>
  <time>1</time>
  <target>PT</target>
  <MINiML:Sample>
    <MINiML:Status />
    <MINiML:Title />
    ...
  </MINiML:Sample>
  ...
  <time>n</time>
  <target>PT</target>
  <MINiML:Sample>
    <MINiML:Status />
    <MINiML:Title />
    ...
  </MINiML:Sample>
</omicxbdml>
    
```

統合データベースの構築

統合データベースSSBD

http://ssbd.qbic.riken.jp

2013年9月公開

SSBD Database

Sign in to SSBD

Browse through categories:

Home Resources Manuals Publications News Download

Search Services: "C. elegans" [organism] and wild-type [description]

Search Advanced Help

Introduction of SSBD

Systems Science of Biological Dynamics (SSBD) database provides a rich set of resources for analyzing quantitative biological data, such as single-molecule, cell, and gene expression nuclei. Quantitative biological data are collected from a variety of species, sources and methods. These include data obtained from both experiment and computational simulation. These quantitative numerical data are represented in a new **Biological Dynamics Markup Language (BDML)**. The new data format allows users to exchange, store, compare and analyze data through the SSBD database. Users can download quantitative biological dynamics data directly in BDML format from the SSBD database. The system utilizes OMERO server to manage image data and experimental conditions. A range of software tools and applications for visualizing and analyzing quantitative biological dynamical data are being developed through a set of SSBD APIs.

News and Events

June 18, 2014: BDML schema 0.18 released!

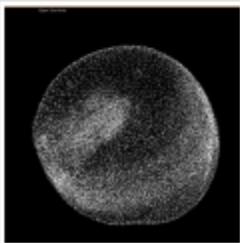
BDML schema version 0.18 has been released. All BDML-files and software have been updated.

March 17, 2014: System maintenance notice (Date: Mar. 19 (JST))

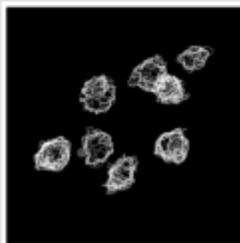
Due to system maintenance, SSBD database will be unavailable Mar. 19, 2014 10:00 am to 13:00 pm (About 4 hours, JST).

[Older news ...](#)

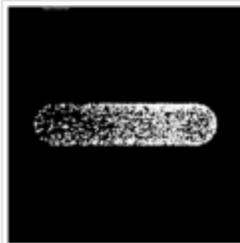
Sample Datasets



Nuclear division dynamics in zebrafish wild-type embryo



Nuclear division dynamics in C. elegans wild-type embryo



Single molecule dynamics in E. coli wild-type

Menu

BDML and PDPML schemas can be found [here](#).

OMERO web: Some images can be viewed on [OMERO.web](#). If you have problem viewing the images on the website, please click on the drop-down arrow on the right of 'public data' on the bar above the data tree, select 'Public' group and 'public data' to view the images (click [here](#) for more details).

Introducing SSBD Database



Copyright notice

Details can be found [here](#).

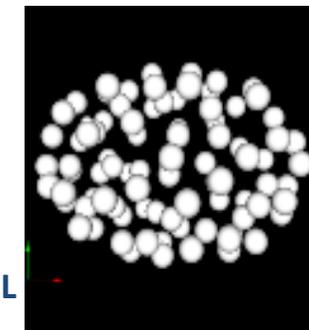
Links

[OME](#) [WDDO](#) [WormBase](#)

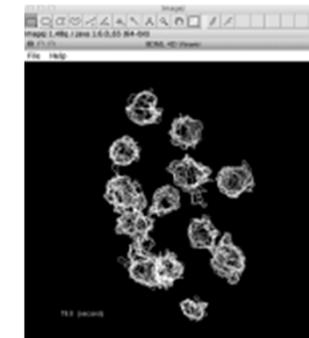
画像データ



定量データ



ツール



BDML
OmicsBDML



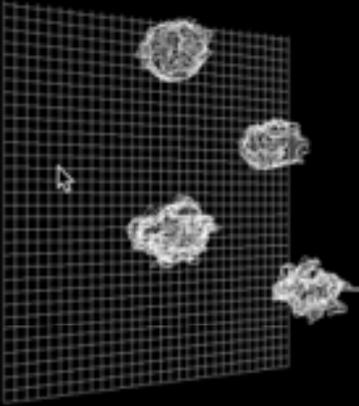
BDML

ブラウザ上でデータを可視化

SSBD Database beta edition Sign in to [SSRD](#)

Browse through categories: [Home](#) [Resources](#) [Manuals](#) [Publications](#) [News](#) [Download](#)

Open Controls



The visualization shows a 3D grid plane with several cell-like structures. A mouse cursor is visible over the grid. A 3D coordinate system with red, green, and blue axes is shown in the bottom left corner.

Time point: 42

BOML ID: 48eb0bbb-8c18-43de-a05c-154a24cb3710

Title: BOML file for quantitative information about nuclear division dynamics of wild-type embryo

Version: Version 1, updated on May 18, 2012, CC BY-NC-SA license

Description: quantitative information about nuclear division dynamics in wild-type embryo

Organism: *C. elegans*

Datatype: nuclear division dynamics

Identifier: wt(NZ)030210_01

Based on: Measurement

Contributor: Kyoda, K., Furukawa, M., Arai, R., Onami, S.

PubMed ID: [23172286](#)

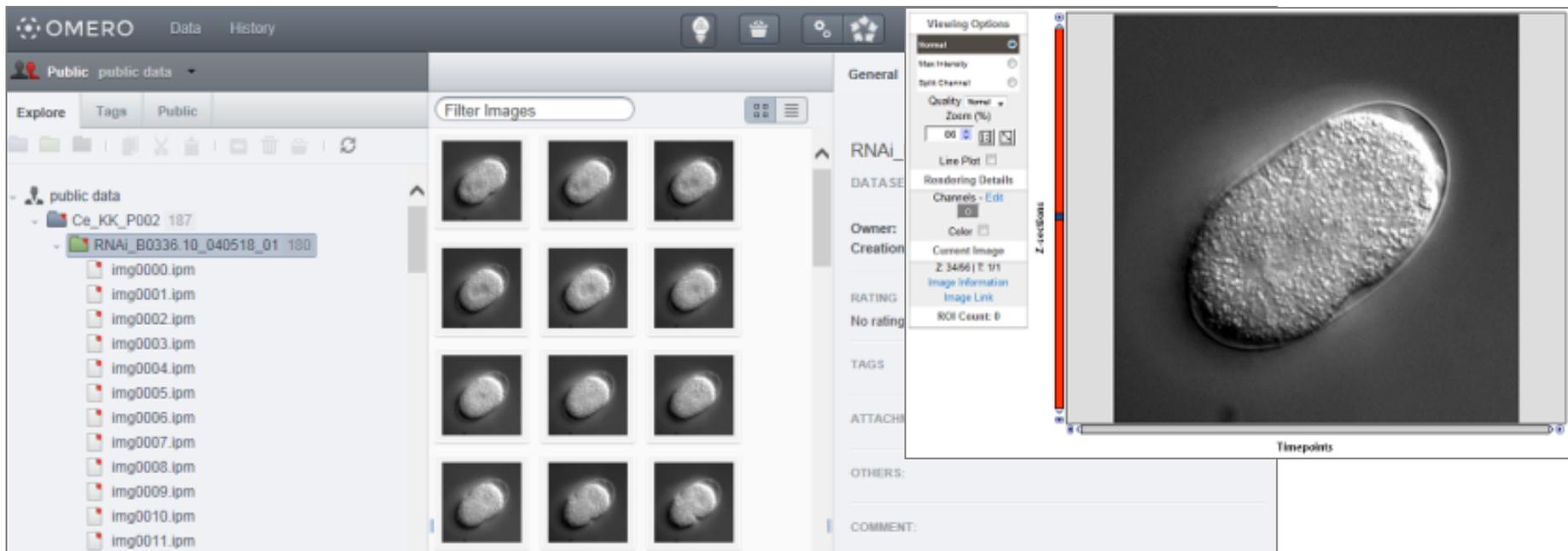
Contact Information: Shuichi Onami, RIKEN, Quantitative Biology Center, Laboratory for Developmental Dynamics

Summary of methods: See details in Kyoda et al. (2013) *Nucleic Acids Res* 41, D732-D737.

Data Scale (XYZ, T): 0.105 x 0.105 x 0.5 (micrometer), 40.0 (second)

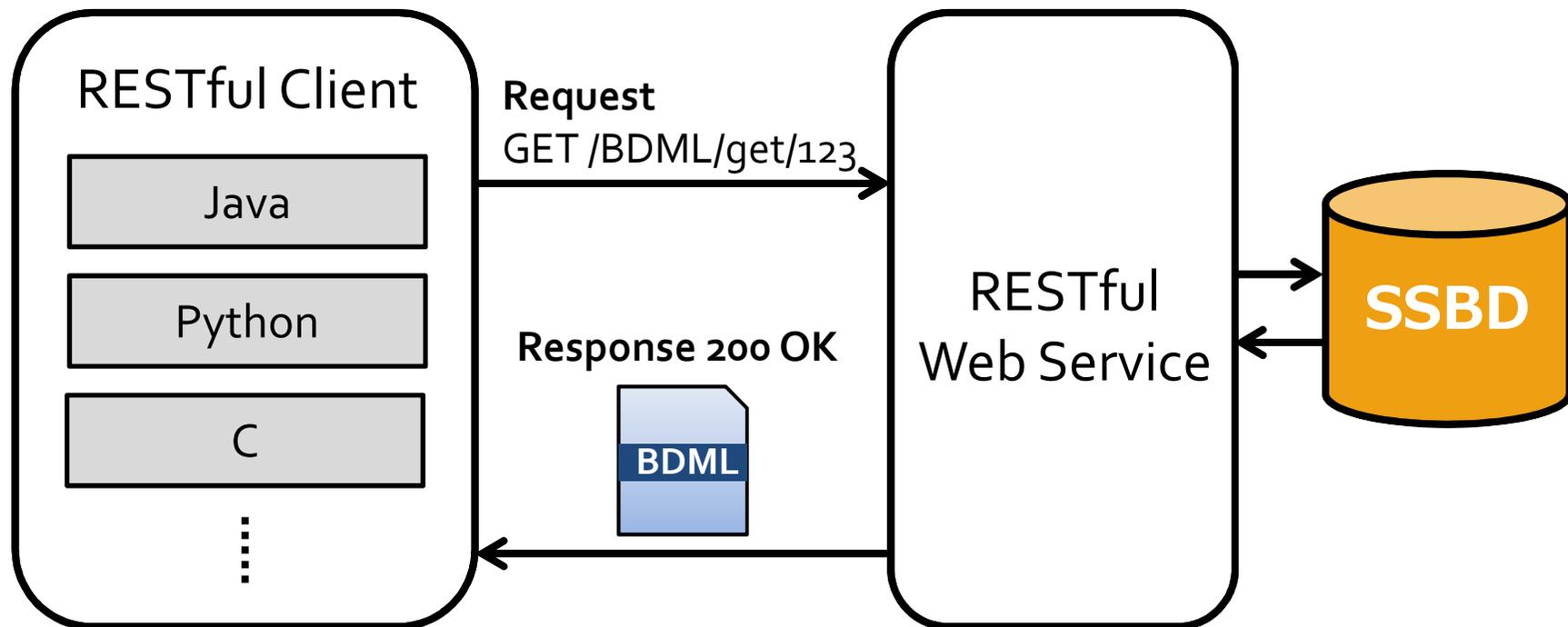
OMEROシステムによる動画像の管理

- 顕微鏡から得られるテラバイト単位の動画像の管理を目的に開発されたフリーのソフトウェア
 - 127種類の動画像のフォーマットに対応
 - ImageJ, MATLAB, Java, C++, Python によるクライアント・アプリケーションの開発が可能



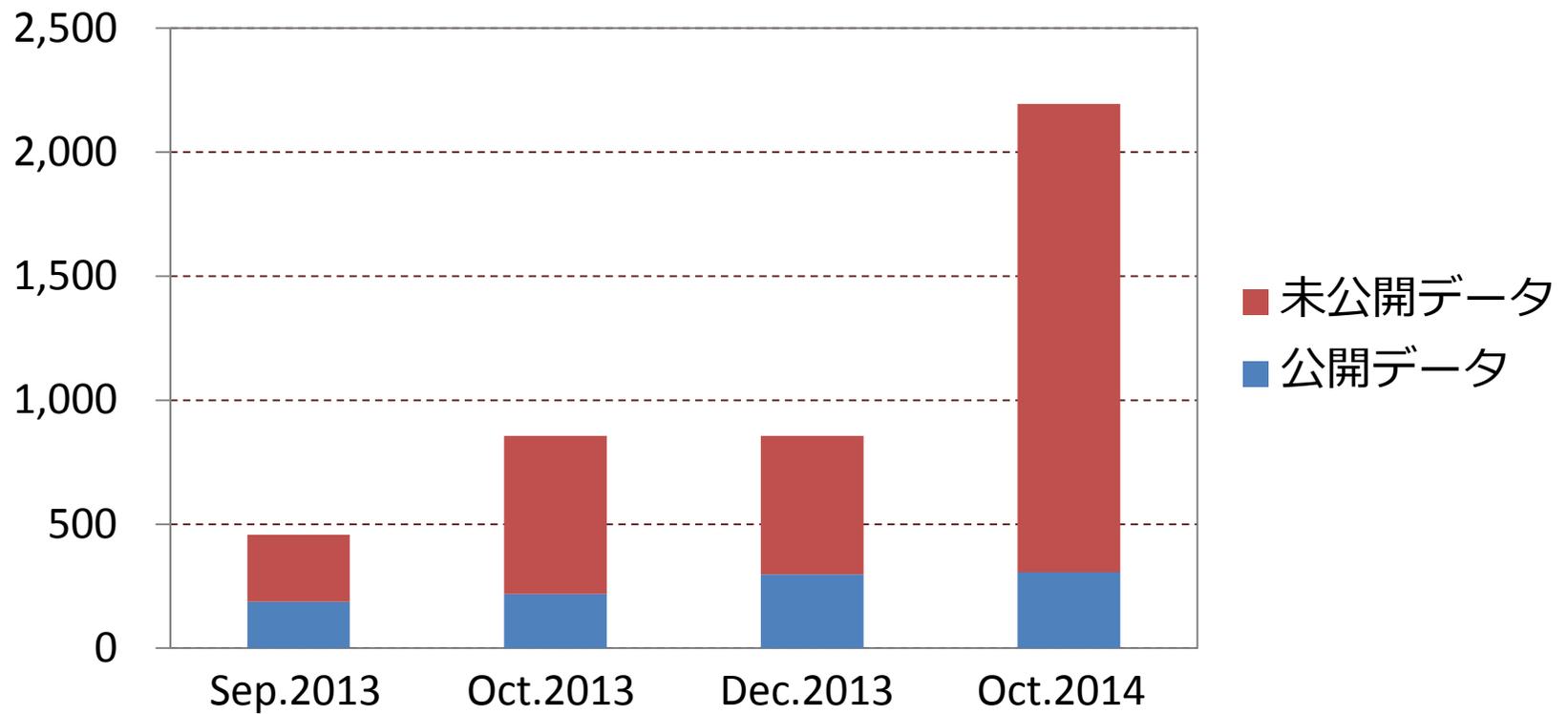
RESTful API を構築

- HTTPのURI (Uniform Resource Identifier) でデータの読み取りができる。
 - 第三者がクライアントのアプリケーションをプログラミング言語やプラットフォームに制限されることなく開発できる。



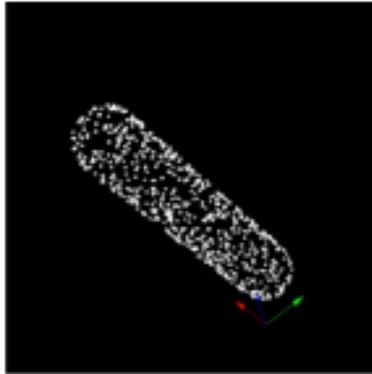
SSBD登録データ

BDMLファイル数



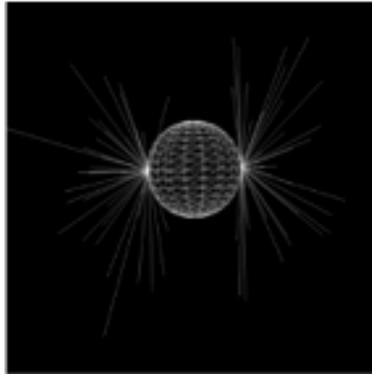
SSBDのデータ

E. coli



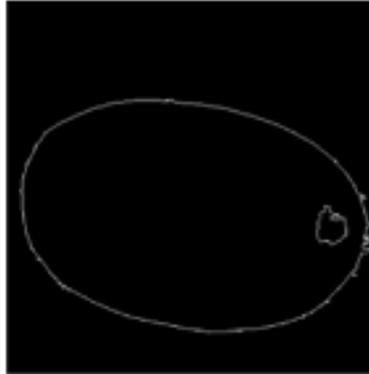
(Arjunan & Tomita 2007)

C. elegans



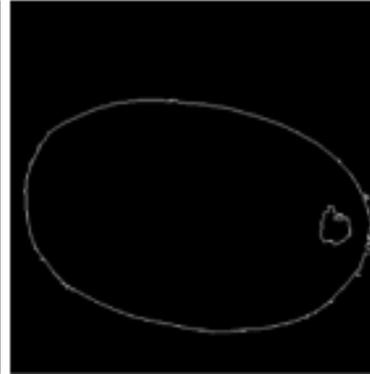
(Kimura & Onami 2005)

C. elegans



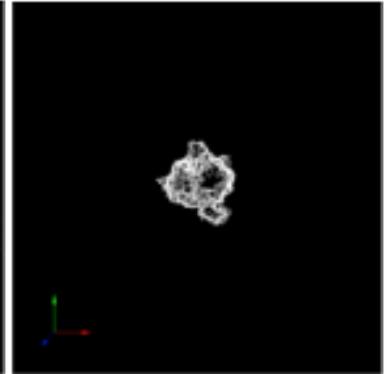
(Tohsato et al., submitted)

C. elegans



(Kyoda et al. 2013)

C. elegans



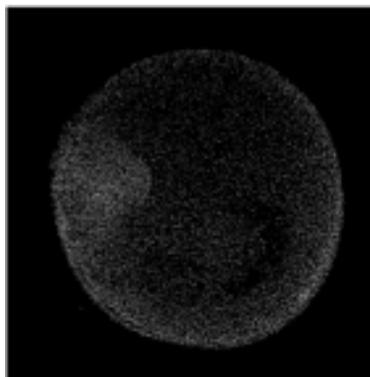
(Bao et al. 2006)

D. melanogaster



(Keller et al. 2010)

Zebrafish



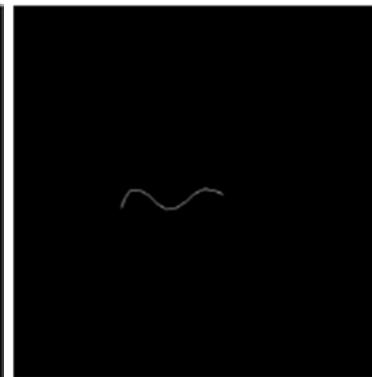
(Keller et al. 2008)

mouse



(Bashar et al. 2012)

C. elegans



(Cronin et al. 2005)

公開予定・公開中のデータ一覧

2014年10月5日

生物種	対象	種類	文献	#BDML	#component	#images
mouse	核	計測	Bashar et al. 2012	1	2055	2800
mouse	発現 (MINiML)	計測	Masumoto et al. 2010	8	48	
mouse	核	計測	Kurotaki et al. 2007	1	12096	80
mammals	ヌクレオソーム	モデル	Hihata et al. 2012	1		
zebrafish	核	計測	Keller et al. 2008	7	56584840	
<i>D. melanogaster</i>	核	計測	Keller et al. 2010	2	5111828	
<i>D. melanogaster</i>	核	計測	Supatto et al. 2009	1	40534	
<i>C. elegans</i>	核	計測	Kyoda et al. 2013	186	75955	180x66x186
<i>C. elegans</i>	核	計測	Bao et al. 2005	2	24747	
<i>C. elegans</i>	前核+微小管	モデル	Kimura & Onami 2005	100	2400100	
<i>C. elegans</i>	行動	計測	Cronin et al. 2005	11	15822	
<i>C. elegans</i>	前核+核+胚	計測	Tohsato et al.	1330	1872137	
<i>C. elegans</i>	核	計測	Kyoda et al.	259	155873	360x66x259
<i>C. elegans</i>	核	計測	Takayama et al.	12	20966	4899
<i>C. elegans</i>	発現	計測	Sarov et al. 2012	273	5713854	
<i>D. discoideum</i>	一分子	計測	Jin et al.	1	987	368
<i>E. coli</i>	一分子	モデル	Arjunan & Tomita 2010	1	721	

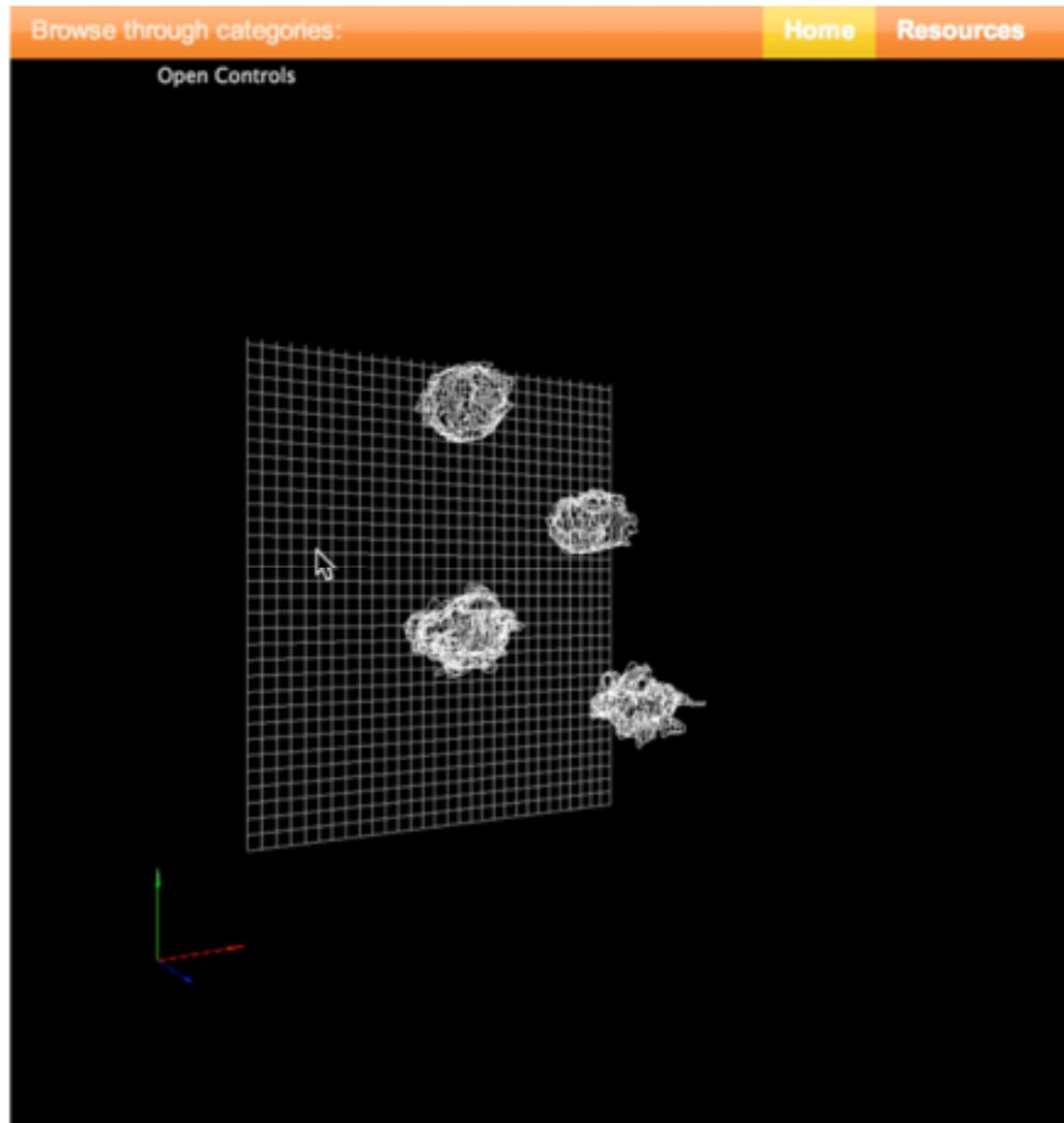
BDML、SSBDの活用への取り組み

生命動態の時空間定量データの活用

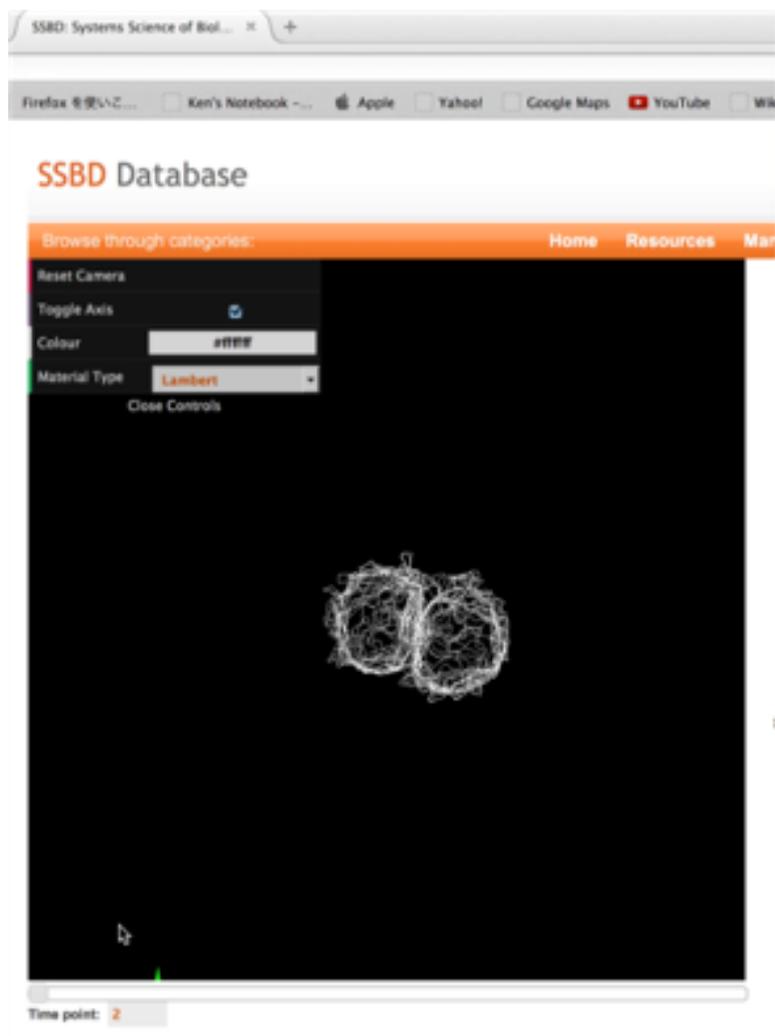
- 可視化
- 従来の解析の自動化、高速化
- データ駆動型の新手法による解析

ツール群の開発/公開、サービスの開始

可視化



ブラウザでの可視化の速度の向上



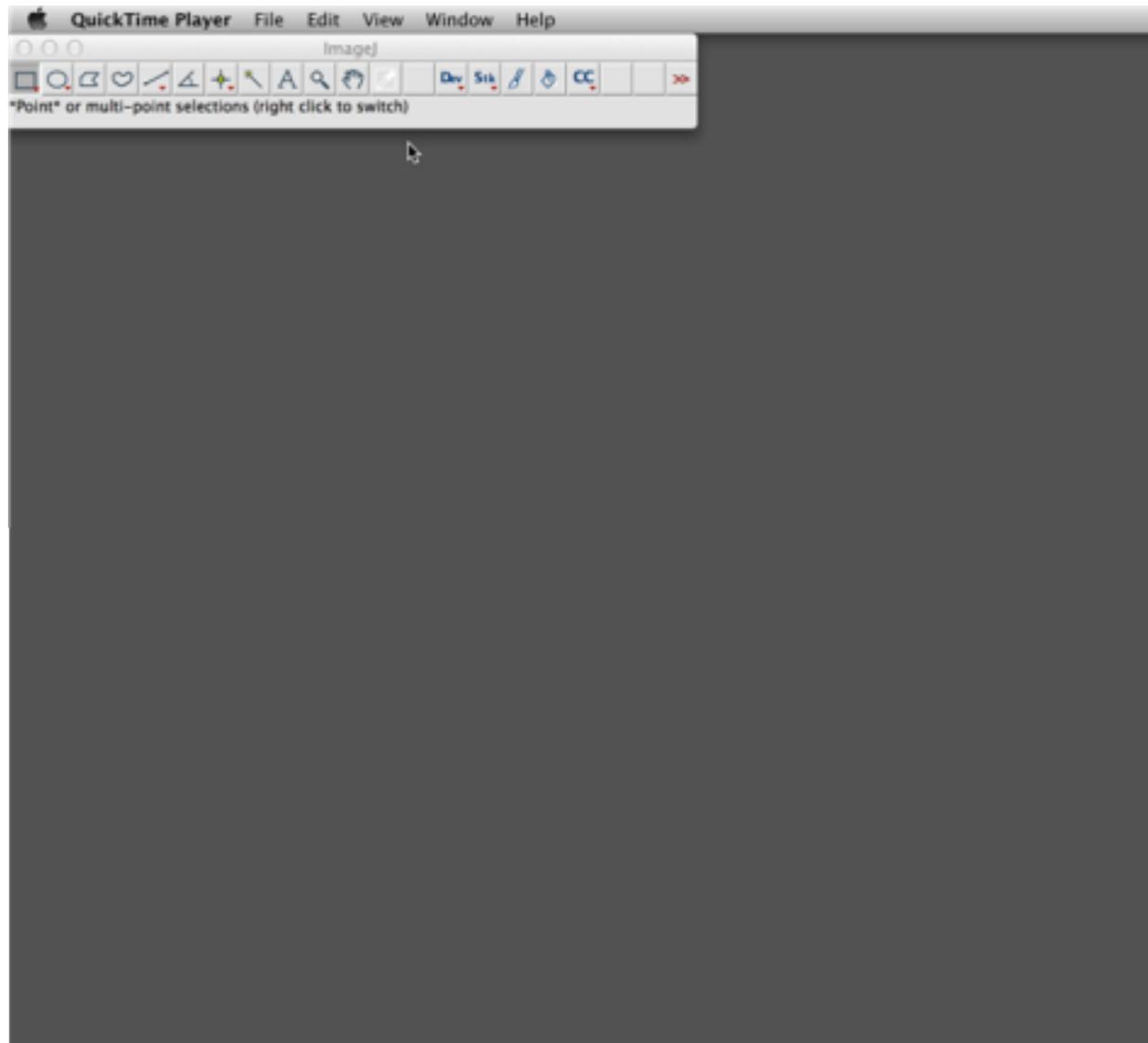
Old Viewer



New Viewer

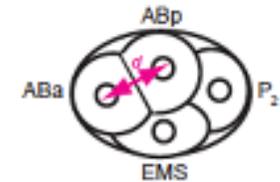
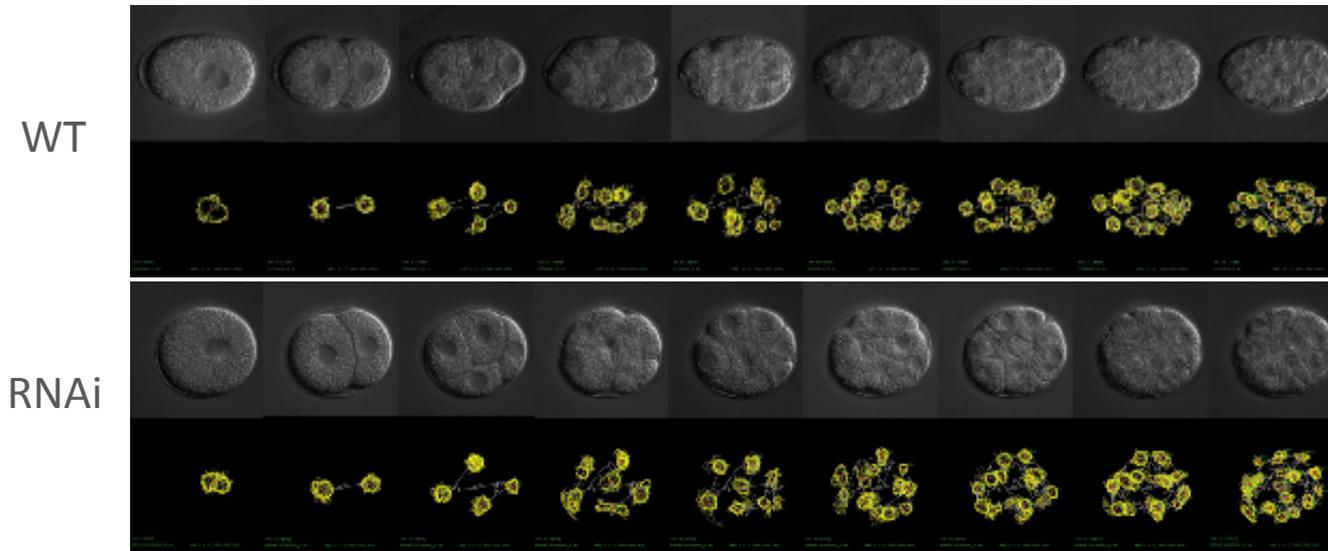
BDMLの可視化ソフトウェア

ImageJのプラグインとして開発



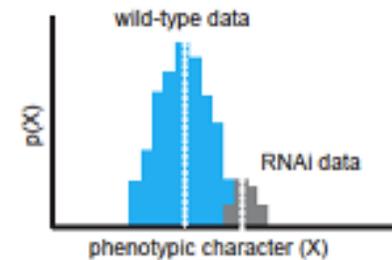
従来法の自動化、高速化の典型例

計算表現型解析

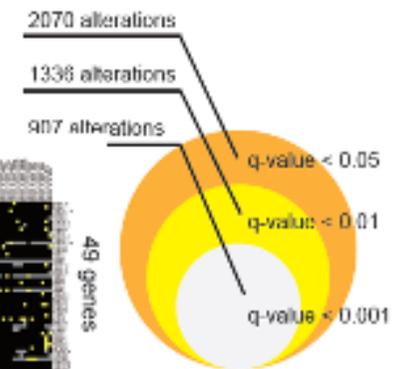
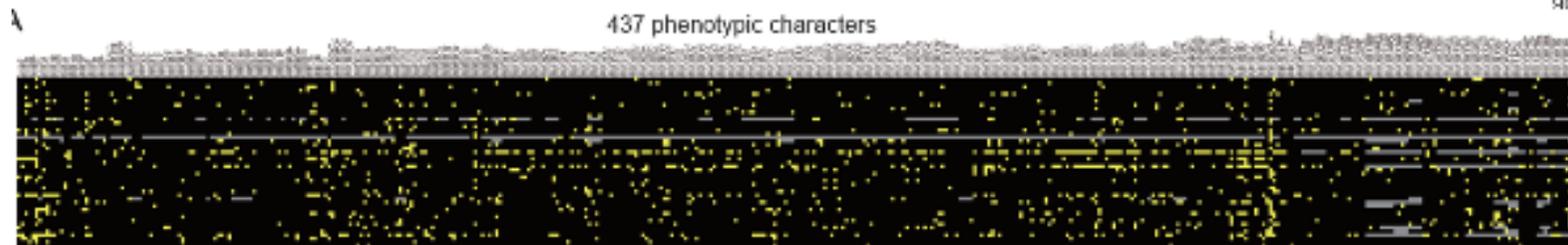


Distance between ABa nucleus and ABp nucleus

$$d = \sqrt{(x_{ABa} - x_{ABp})^2 + (y_{ABa} - y_{ABp})^2 + (z_{ABa} - z_{ABp})^2}$$

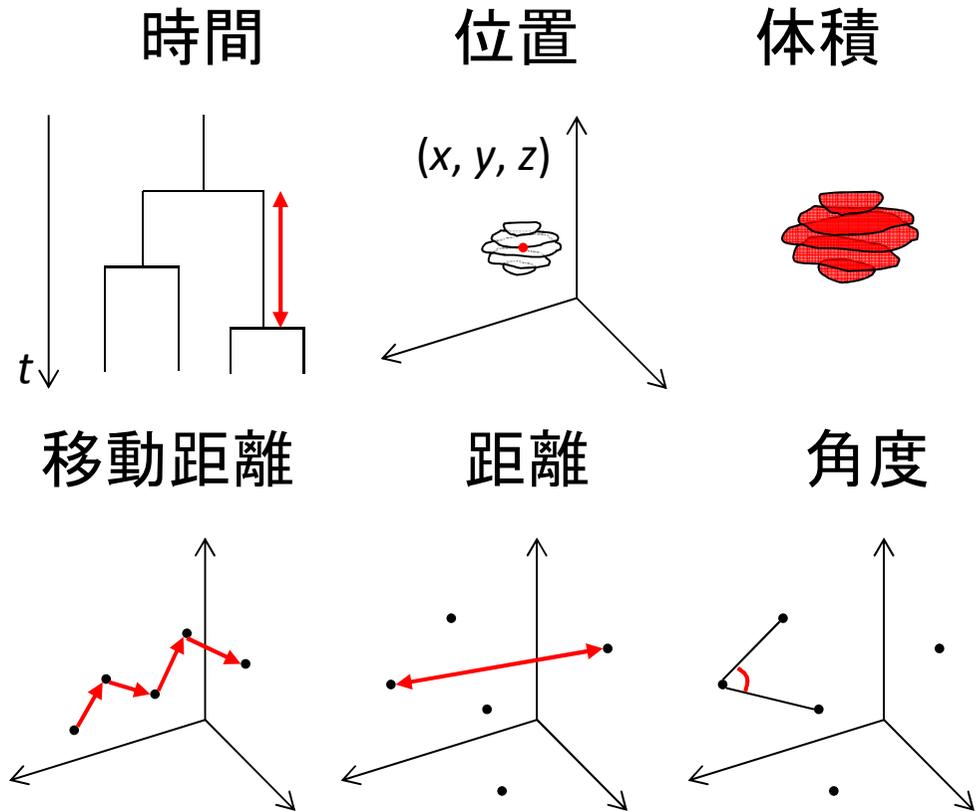
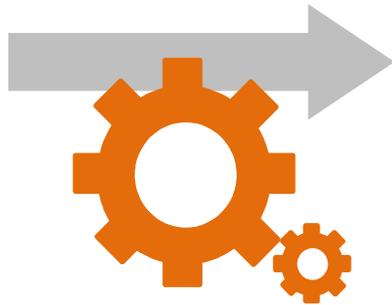


- All essential embryonic genes (97 genes) on Chromosome III
- 437 phenotypic characters between 1- and 8-cell stage
- 1336 RNAi induced phenotypic alterations



表現型解析ツールの開発/公開

- Phenochar (ver. 1.0)
 - 表現型の特徴量を計算するツール



計算表現型解析サービス

- 国内の線虫研究者に試験的に開始



東北大学



国立
遺伝学研究所



関西学院大学



理化学研究所



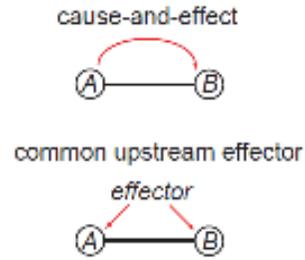
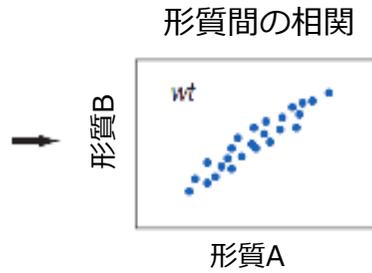
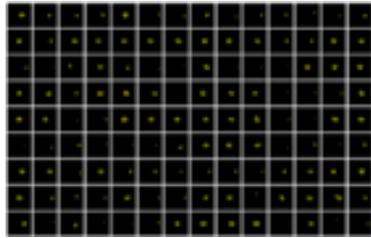
大阪大学



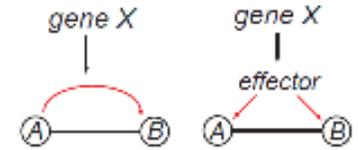
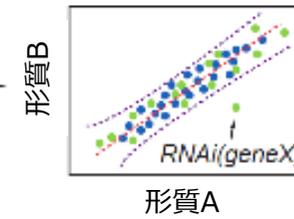
データ駆動型の新手法による解析

時空間データを利用した発生の粗視化モデルの導出

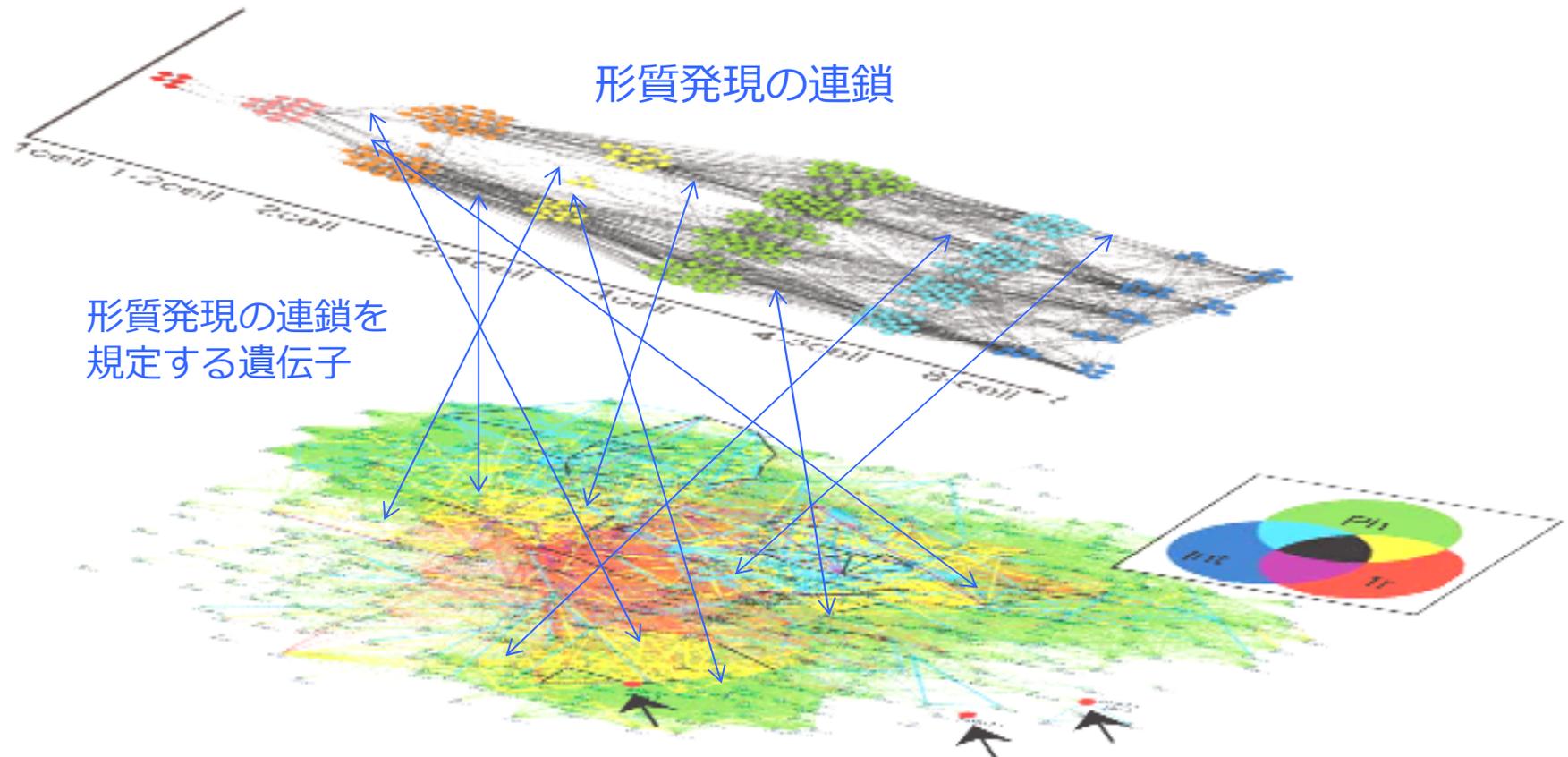
遺伝子ノックダウン胚の
細胞分裂動態データ



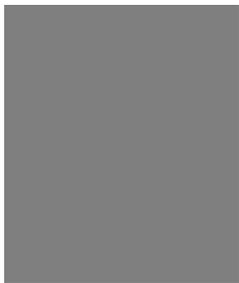
相関のはずれ値



Kyoda et al., submitted



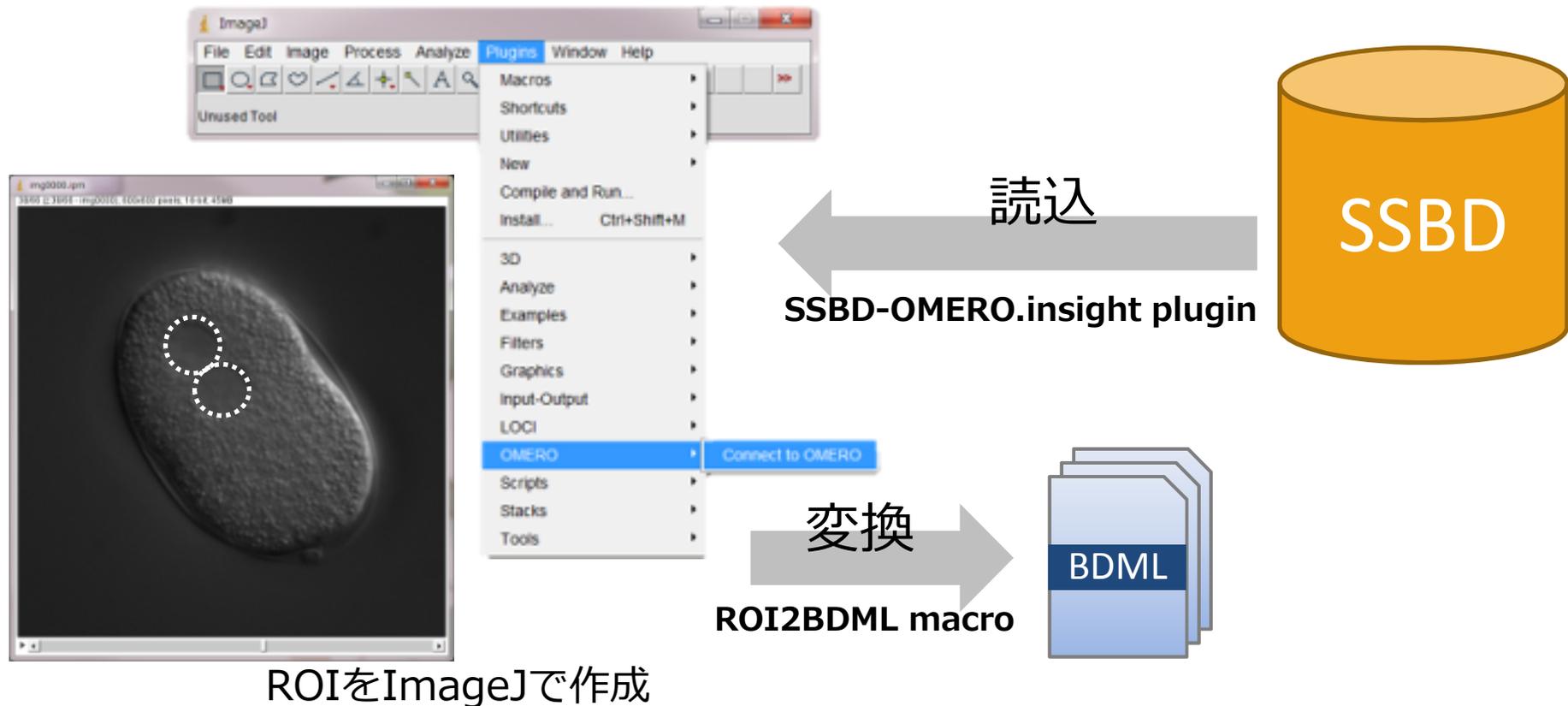
情報科学研究者との共同研究



より多くの実験生物学者へ

SSBD/OMEROからのBDML生成

- SSBDからの画像の読み込みを可能にするImageJ pluginを公開し、BDML形式のROI情報の保存を容易にした。



今後の展望

- 生命動態システム科学のDB統合化の国内での推進を継続
 - プロジェクト・グループとの連携の継続、強化
 - 当分野の主要な研究者との連携の拡大
- BDML/OmicsBDMLの国際標準化を推進
 - COMBINEとの連携
 - Bioimage Informatics communityとの連携
- SSBDの国際連携の拡大、推進
 - 生命動態データのDB統合化の国際的な連携組織を構築
 - 当分野の主要な研究者との連携の拡大
 - OMEROとの連携
- BDML/SSBDの活用
 - データから知識を抽出する手法、ソフトウェアの開発・公開
 - バイオインフォマティクス、数理生物学、ビッグデータ解析分野との連携
 - 類似動態、類似動画画像の検索機能の開発
- 画像データの拡大
 - より一般的な画像データへの取り組み

まとめ

- 生命動態システム科学のデータを統一的に記述するフォーマットを開発した。
 - BDML
 - OmicsBDML
- 生命動態システム科学の統合データベースを構築した。
 - SSBD
- 国内外の生命動態システム科学の定量データをSSBDに格納した。
 - QBiCの公開可能なデータ全て
 - 国内の公開可能なデータをほぼ全て（継続中）
 - 国外の生命動態システム科学の代表的なデータ（継続中）
- BDML/OmicsBDML, SSBDを活用するツール群の開発/公開を開始した。
 - Phenochar
 - SSBD-OMERO.insight plugin
 - ROI2BDML macro
 - ImageJ Plugin - BDML Viewer
- DB統合化の国際連携についての議論を開始した。