

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

研究代表者 理化学研究所生命システム研究センター 大浪修一



©2014 大浪 修一（理化学研究所） licensed under CC表示2.1日本

2014年1月23日 JST東京本部

本研究開発の目標

- 我が国の生命動態システム科学分野の全てのデータベースを統合する体制と仕組みの構築
- 研究開発終了時における、我が国の生命動態システム科学分野の全てのデータベースの統合
- 生命動態システム科学分野の全世界のデータベースの統合

研究開発のスケジュール

研究項目	H24年度	H25年度	H26年度
1. DB統合の体制と仕組みの構築			
2. 理研QBiC内のDB統合			
3. 理研QBiC外のDB統合			

H24年度の進捗状況

- 研究項目1 DB統合の体制と仕組みの構築
 - プロジェクトグループの構築 (達成)
 - 統合データフォーマットの構築 (BDD-XML)
 - 統合データベースサーバの構築 (設計まで)
- 研究項目2 理研QBiC内のDB統合
 - 大浪研究室が保有する2つの大規模データの統合 (達成)

H25年度の目標、当初計画

- 研究項目 1 DB統合の体制と仕組みの構築
 - 理研QBiC内のデータの調査
 - 統合データフォーマットの改良
 - 統合データベースの構築
- 研究項目 2 理研QBiC内のDB統合
 - 理研QBiC内のデータベースの統合
- 研究項目 3 理研QBiC外のDB統合
 - 理研QBiC外のデータベースの統合の開始

H25年度の成果

- 研究項目 1 DB統合の体制と仕組みの構築
 - 理研QBiC内のデータの調査
 - 統合データフォーマットの改良
 - 統合データベースの構築
- 研究項目 2 理研QBiC内のDB統合
 - 理研QBiC内のデータベースの統合
- 研究項目 3 理研QBiC外のDB統合
 - 理研QBiC外のデータベースの統合の開始

理研QBiC内のデータの調査 (2013.04~05)

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

H25年度の成果

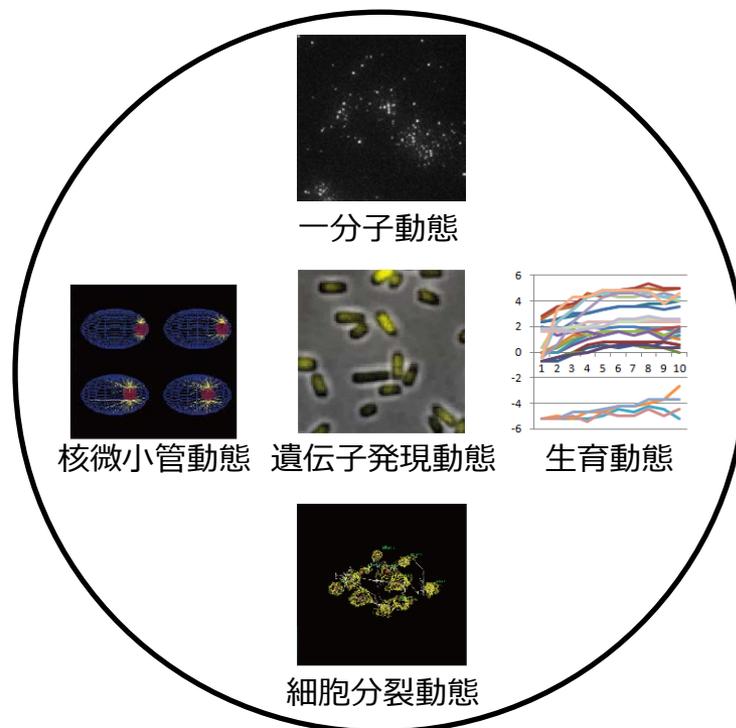
- 研究項目 1 DB統合の体制と仕組みの構築
 - 理研QBiC内のデータの調査
 - 統合データフォーマットの改良
 - 統合データベースの構築
- 研究項目 2 理研QBiC内のDB統合
 - 理研QBiC内のデータベースの統合
- 研究項目 3 理研QBiC外のDB統合
 - 理研QBiC外のデータベースの統合の開始

統合フォーマットの改良

- QBiCで産出される定量データに対応
 - BDMLの開発



BDD-XML (2013.01.21)



BDML ver. 0.15 (2013.10.4)

BDML

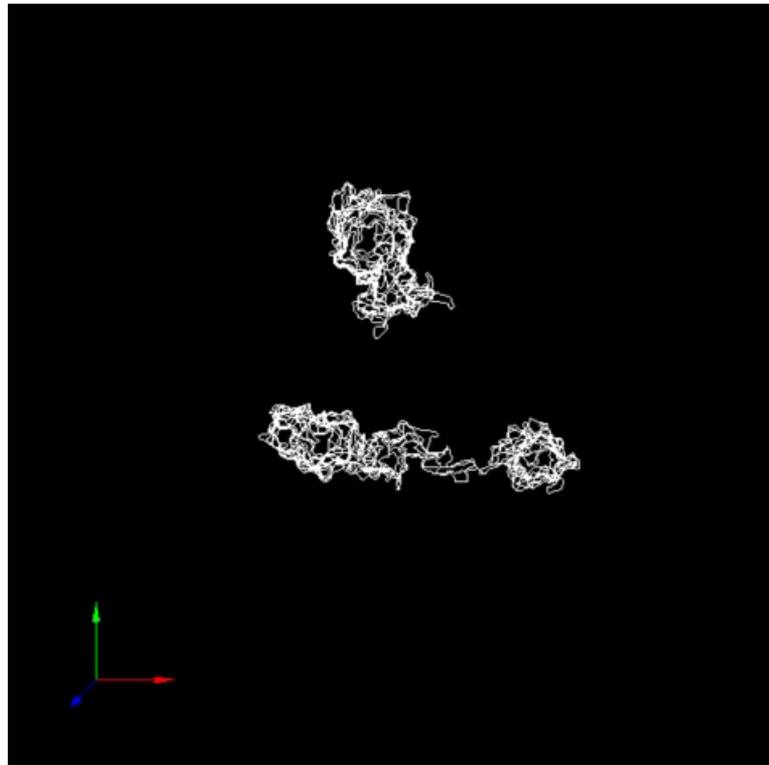
Biological Dynamics Markup Language

- XMLを基盤とした記述言語。
 - 利点：高い可読性・拡張性と計算機との親和性
 - 弱点：タグ多用によるオーバーヘッド

	tab-delimited	XML	JSON	BSON	HDF5
階層構造		○	○	○	○
可読性	○	◎	○		
拡張性		◎	○	○	○
多言語処理		○	○	○	○
ウェブ利用		○	○	○	
サーバ利用		○	○	○	○
ファイルサイズ				○	○
数値精度				○	○

BDMLで記述するデータ

- 空間情報 (XYZ) と時間情報 (T) を扱う。

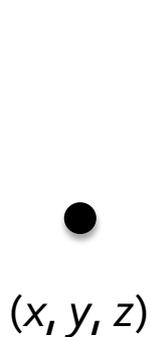


線虫の細胞分裂動態

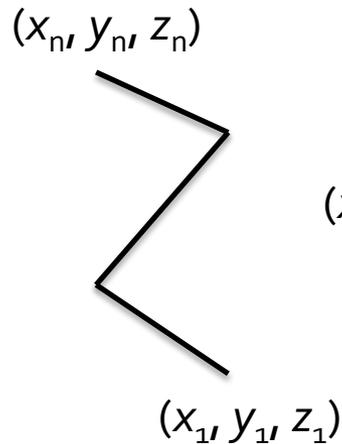
空間情報 (XYZ) の表現

- 位置や形状を表現するための5種類の表現法を準備した。
 - 特徴 (例えば遺伝子発現量) を付加することもできる。

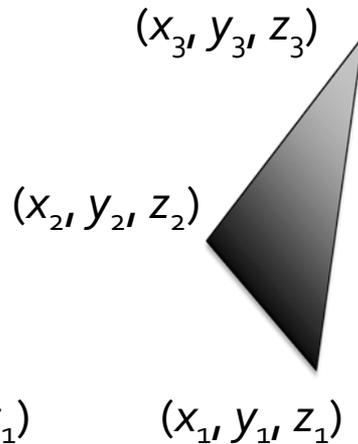
点



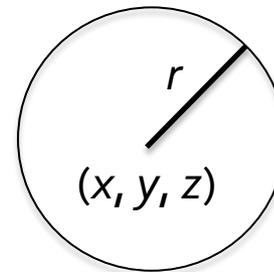
線



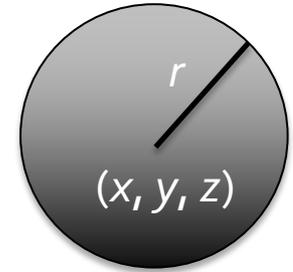
面



円

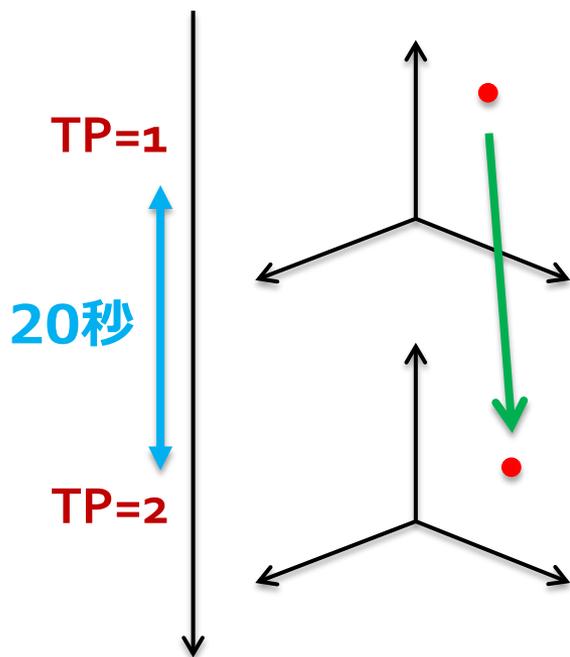


球



時間情報 (T) の表現

- 一定の間隔尺度を持つデータとして記述する。
- 対象物の時間的繋がりを表現する。



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

メタ情報の記述

- 画像またはモデルからの定量データ作成を再現可能にする。
 - 定量データの元になるファイルの在り処
 - 計測データ：顕微鏡画像
 - シミュレーション結果：モデル
 - 定量データ作成までの手順
 - PDPML: Procedures for Data Processing ML
 - 画像処理の手順（ソフトウェア、パラメータ等）
 - シミュレーションの手順（ソフトウェア、設定等）
- 詳細または簡易的な記述を可能にしている。

BDML ver. 0.15で記述可能な理研QBiCのデータ

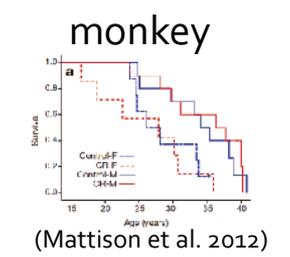
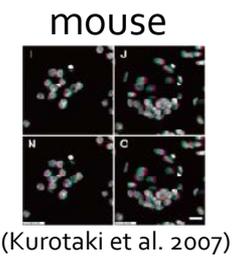
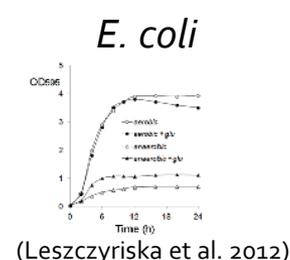
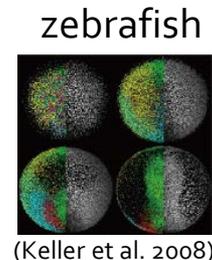
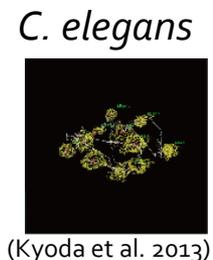
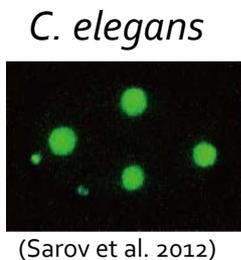
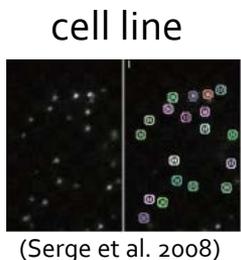
□ 記述可能

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

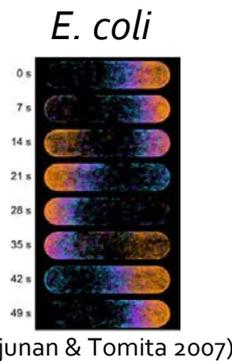
BDML ver. 0.15 で扱える定量データ

- 分子、細胞、個体から集団の動態まで。

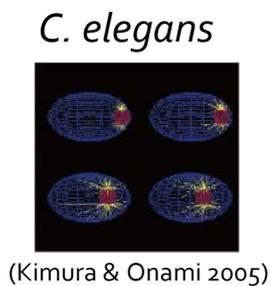
遺伝子発現



一分子



細胞核分裂



核・微小管

増殖・生存

計測データ

シミュレーション結果



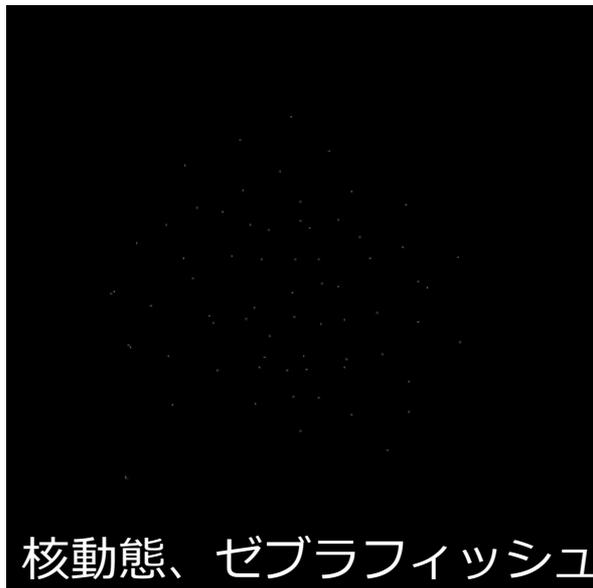
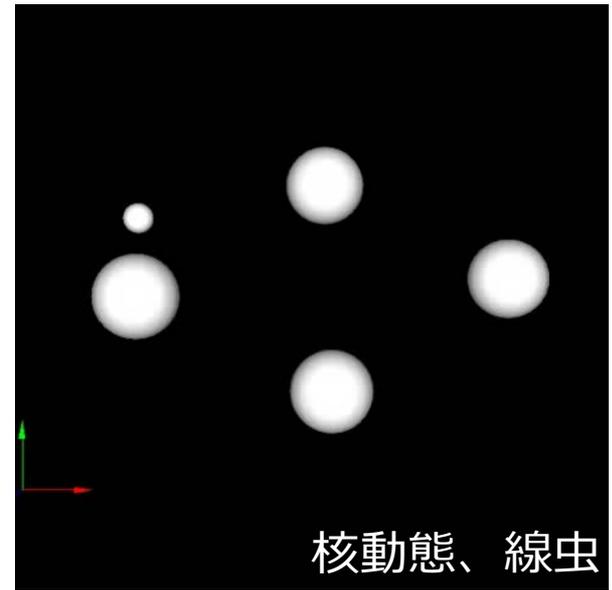
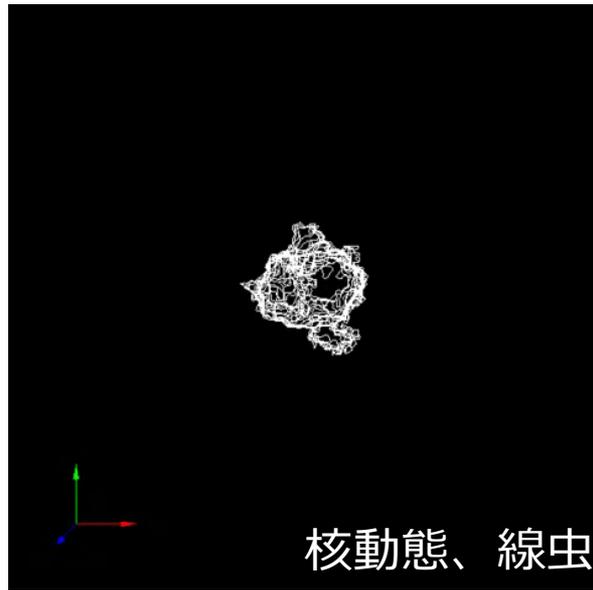
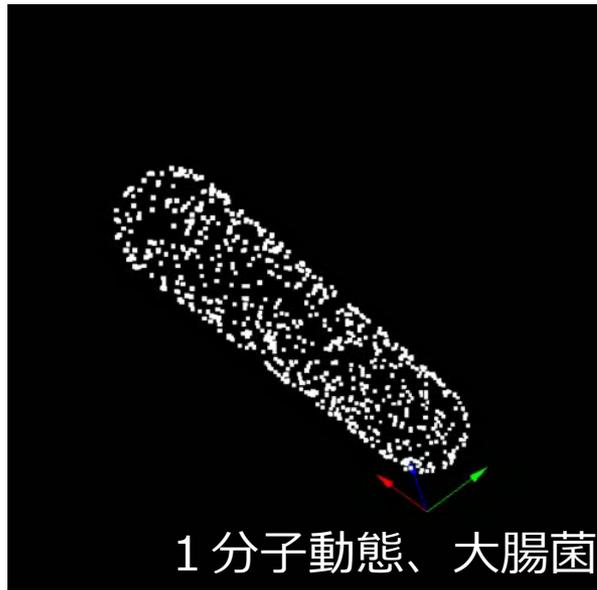
分子

細胞

個体

集団

BDMLデータの可視化



様々な生命動態の定量データを共通のソフトウェアで可視化できる。

H25年度の成果

- 研究項目 1 DB統合の体制と仕組みの構築
 - 理研QBiC内のデータの調査
 - 統合データフォーマットの改良
 - 統合データベースの構築
- 研究項目 2 理研QBiC内のDB統合
 - 理研QBiC内のデータベースの統合
- 研究項目 3 理研QBiC外のDB統合
 - 理研QBiC外のデータベースの統合の開始

統合データベースSSBDの構築 (2013.9.2公開)

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

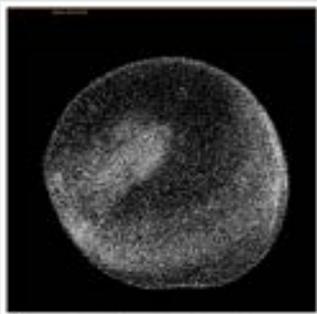
Browse through categories: [Home](#) [Resources](#) [Manuals](#) [Publicatons](#) [News](#) [Download](#)

Search Services: [organism] and wild-type [description] [Advanced](#) [Help](#)

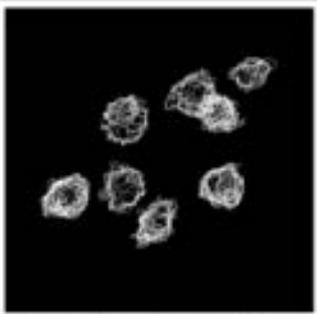
Introduction of SSBD

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.

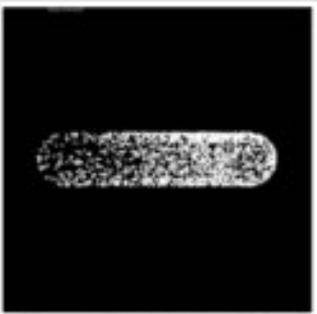
Sample Datasets



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).

News and Events

- Nov. 20, 2013: System maintenance notice (Date: Nov. 21 (JST))
- Nov. 11, 2013: System maintenance notice (Date: Nov. 12 (JST))
- Oct. 3, 2013: 17 BDML-files added.
- Oct. 2, 2013: System maintenance notice (Date: Oct. 9 (JST))
- Sept. 2, 2013: SSBD database and BDML schema released!

[Older news](#)

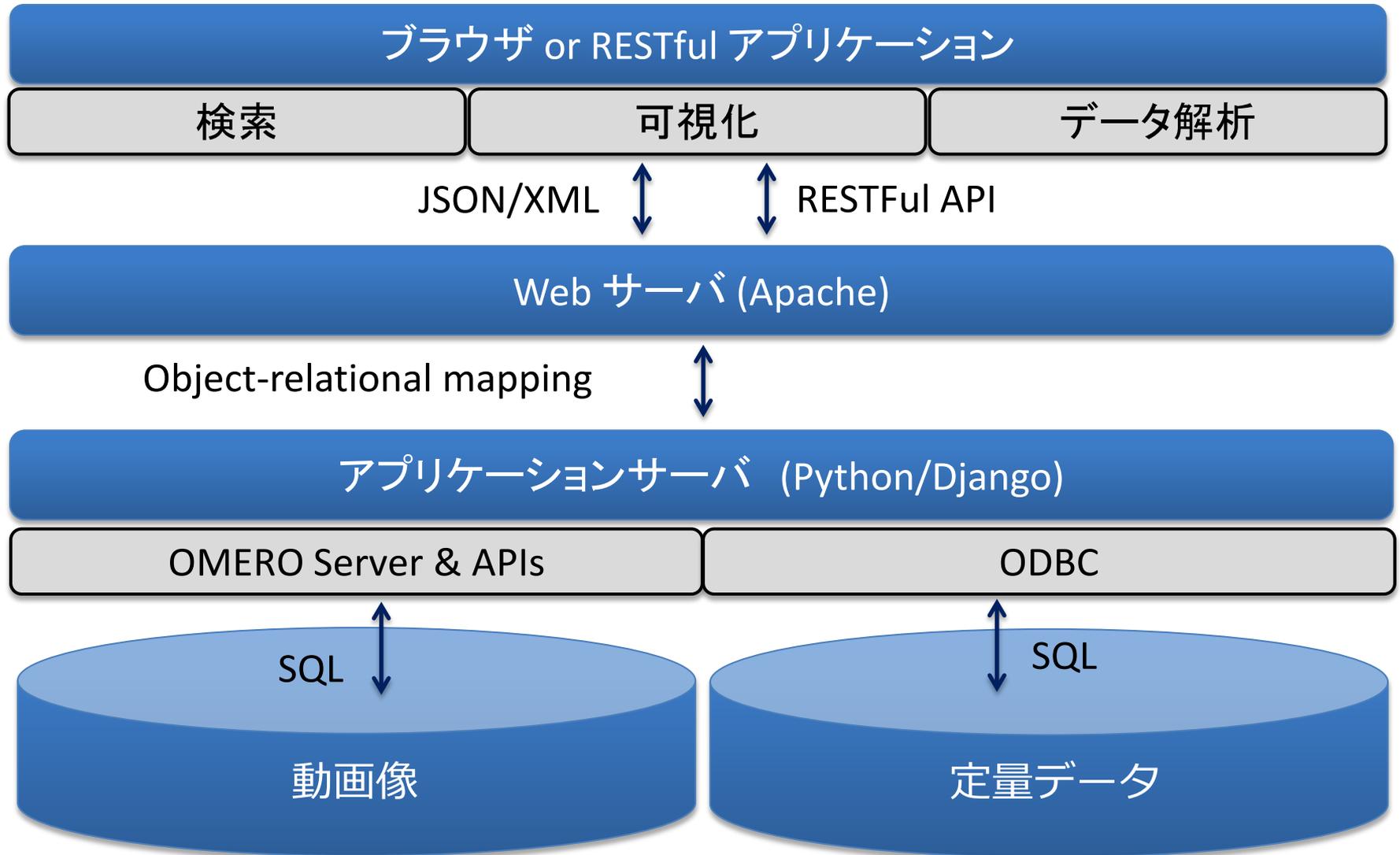
Copyright notice

Details can be found [here](#).

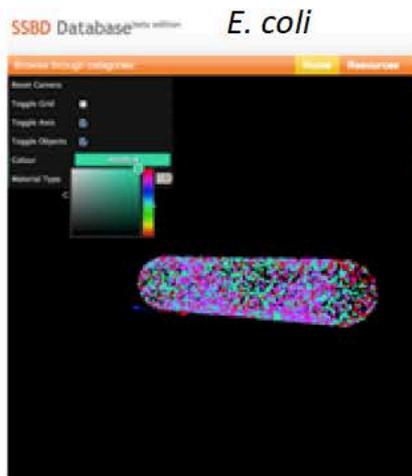
Links

[WDDD](#), [OME](#), [WormBase](#)

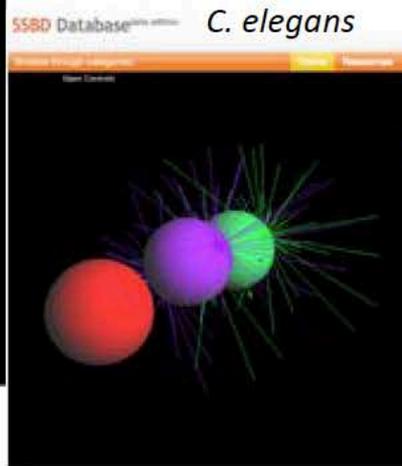
SSBDのシステム構成



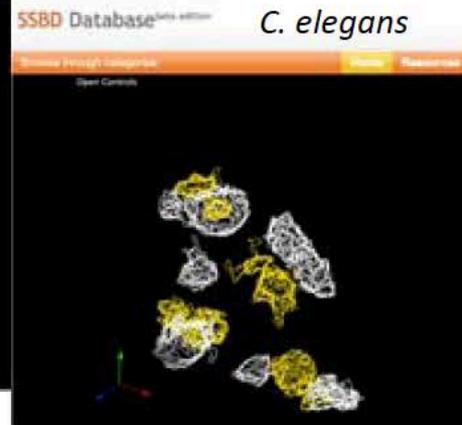
ブラウザ上でSSBDのデータを可視化



(Arjunan & Tomita 2007)



(Kimura & Onami 2005)



(Kyoda et al. 2013)

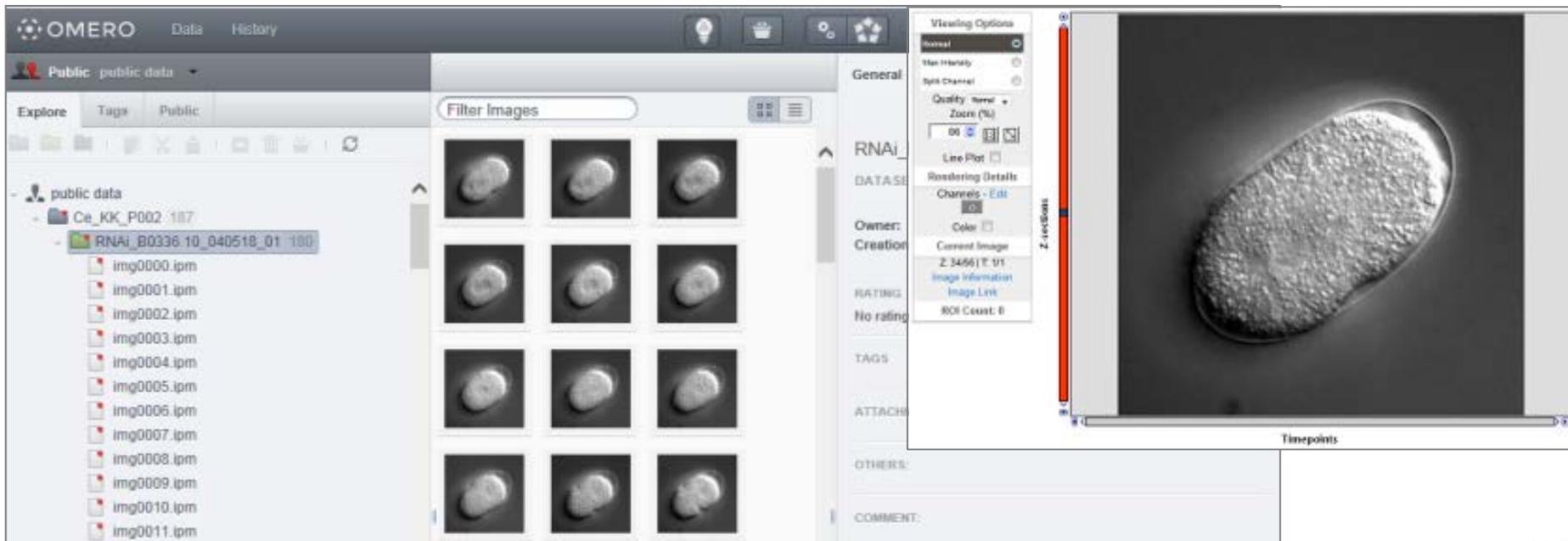


(Keller et al. 2008)

(Keller et al. 2010)

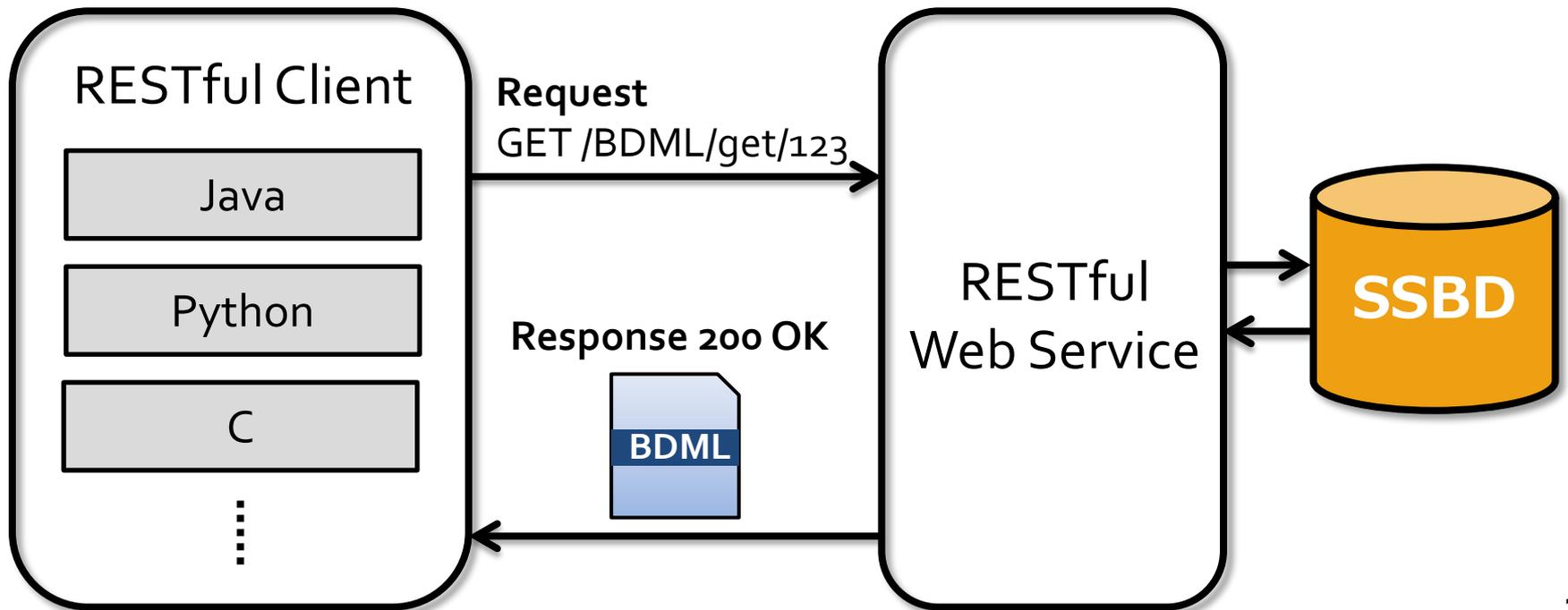
OMEROシステムによる動画像の管理

- 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。



H25年度の進捗状況

- 研究項目 1 DB統合の体制と仕組みの構築
 - 理研QBiC内のデータの調査
 - 統合データフォーマットの構築
 - 統合データベースの構築
- 研究項目 2 理研QBiC内のDB統合
 - 理研QBiC内のデータベースの統合
- 研究項目 3 理研QBiC外のDB統合
 - 理研QBiC外のデータベースの統合の開始

理研QBiC内のDB統合の H25年度の基本方針

- 時空間の定量情報を持つ大規模データをBDMLで記述し、SSBDに登録
- データ生産者の希望を優先してデータ登録、公開の日程を決定
 - 論文公開済みのデータはSSBDへの登録を推奨
- BDMLへの変換、SSBDへの登録は当プロジェクトが実行

理研QBiC内のデータの登録状況 (2014.1.22)

■ 公開中

生物種	対象	方法	研究室	BDML	Entries	画像
Hela細胞	一分子	計測	上田昌宏			
Hela細胞	一分子	モデル	高橋恒一			
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	186	75955	33480
線虫	核	計測	大浪修一	261	155873	256
線虫	核	計測	大浪修一	979	1913774	
線虫	核	計測	大浪修一	1	2988	
線虫	核	計測	大浪修一	12	20966	4899
線虫	核・微小管	モデル	大浪修一	100	2400100	
粘菌	一分子	計測	神 隆	1	987	1
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	1	721	

理研QBiCのDB統合の H25年度のまとめ

- 登録/公開可能な大規模データが予定よりも少なかった。
 - 未公開データが多かったため
- データの提供に関しては皆協力的である。
- H26年度中に多くのデータが登録/公開可能になることを期待する。

H25年度の進捗状況

- 研究項目 1 DB統合の体制と仕組みの構築
 - 理研QBiC内のデータの調査
 - 統合データフォーマットの構築
 - 統合データベースの構築
- 研究項目 2 理研QBiC内のDB統合
 - 理研QBiC内のデータベースの統合
- 研究項目 3 理研QBiC外のDB統合
 - 理研QBiC外のデータベースの統合の開始

理研QBiC外のDB統合の H25年度の基本方針

- 生命動態システム科学を代表する時空間定量情報の大規模データをBDMLで記述し、SSBDで公開
 - 論文発表済の公開データを対象
 - データは当プロジェクトが選択
- 公開に関しては、各データ生産者の同意を得る。
- BDMLへの変換、SSBDへの登録は当プロジェクトが実行

理研QBiC外のデータの登録状況

■ 公開中

生物種	対象	方法	論文	#BDML	#Entries	#画像
マウス	核	計測	Kurotaki et al. 2007	1	12096	80
マウス	核	計測	Bashar et al. 2012	1		
ゼブラフィッシュ	核	計測	Keller et al. 2008	7	56584840	
ショウジョウバエ	核	計測	Keller et al. 2010	2	5111828	
ショウジョウバエ	細胞	計測	Supatto et al. 2009	1	40534	
線虫	核	計測	Bao et al. 2005	2	24747	
線虫	発現	計測	Sarov et al. 2012	273	5713854	
線虫	行動	計測	Cronin et al. 2005	11	15822	

#Entries : 時間別にみた分子や核などの総登録数

理研QBiC外のDB統合の H25年度のまとめ

- 生命動態システム科学を代表する時空間定量情報データの登録/公開が実現した
- 多数のその他のデータについても公開の許可を得ている。
- H26年度に多くのデータが登録/公開されることが期待される。
 - CREST、推進拠点の協力も期待できる

SSBD登録済のデータ一覧

□ 公開待ち, ■ 公開中

生物種	対象	種類	論文	BDML	Entries	画像
マウス	核	計測	Kurotaki et al. 2007	1	12096	80
マウス	核	計測	Bashar et al. 2012	1		
ゼブラフィッシュ	核	計測	Keller et al. 2008	7	56584840	
ショウジョウバエ	核	計測	Keller et al. 2010	2	5111828	
ショウジョウバエ	核	計測	Supatto et al. 2009	1	40534	
線虫	核	計測	Kyoda et al. 2013	186	75955	33480
線虫	核	計測	Bao et al. 2005	2	24747	
線虫	核	モデル	Kimura & Onami 2005	100	2400100	
線虫	核	計測	Kyoda et al.	261	155873	256
線虫	核	計測	Sönnichsen et al. 2005	979	1913774	
線虫	核	計測	Gonczy et al. 2000	1	2988	
線虫	核	計測	Takayama et al.	12	20966	4899
線虫	発現	計測	Sarov et al. 2012	273	5713854	
線虫	行動	計測	Cronin et al. 2005	11	15822	
細胞性粘菌	一分子	計測	Jin et al.	1	987	1
大腸菌	一分子	モデル	Arjunan & Tomita 2010	1	721	

H25年度の成果のまとめ

- 研究項目 1 DB統合の体制と仕組みの構築
 - 理研QBiC内で産出されるデータを調査した
 - 理研QBiC内のデータに対応した統合データフォーマットを開発した (BDML ver. 0.15)
 - 生命動態システム科学の統合データベースSSBDを公開した
 - OMEMOによる統括的な画像の管理の実現
 - データ化可視化ソフトの導入
 - RESTful APIの構築
- 研究項目 2 理研QBiC内のDB統合
 - 理研QBiC内の登録可能なデータをSSBDへ登録した
 - 公開可能なデータの公開した
- 研究項目 3 理研QBiC外のDB統合
 - 理研QBiC外のデータのSSBDへの登録/公開を開始した
 - 代表的なデータを登録/公開した

プロジェクトグループ・ミーティング

(2014.1.7 神戸)

- これまでの開発内容についての了承
- 今後の開発計画についての了承
- H26年度にCREST、推進拠点の各研究組織が産出するデータの調査とSSBDへの登録への協力について了承

プロジェクトグループ

理研QBiC 

泰地真弘人 (理研) 上田昌宏 (理研)
上田泰己 (理研) 大浪修一 (理研)

JST CREST 

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)

文科省推進拠点 

松田道行 (京大) 井原茂男 (東大)
金子邦彦 (東大) 楯 真一 (広大)

研究開発のスケジュール

研究項目	H24年度	H25年度	H26年度
1. DB統合の体制と仕組みの構築			
2. 理研QBiC内のDB統合			
3. 理研QBiC外のDB統合			

H26年度の計画

- 理研QBiC外のDBの統合
 - 国内の全ての大規模データの統合
 - 国外の代表的な大規模データの統合
- 理研QBiC内のDBの統合の継続
 - H25年度までに未統合な大規模データの統合
 - H26年度に新たに登録/公開可能となった大規模データの統合
- DB統合の体制と仕組みの構築の継続
 - データフォーマットの拡張
 - 時空間オミクスデータへの対応など
 - BDML/SSBDに準拠した周辺ソフトウェアツール群の開発
 - スタンドアローンの可視化ソフトなど
 - 生命動態システム科学コミュニティーとの連携の強化
 - H25年度CREST採択課題との連携
 - 国際連携の体制や仕組みの構築の開始
 - OMEROとの連携の深化