平成27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

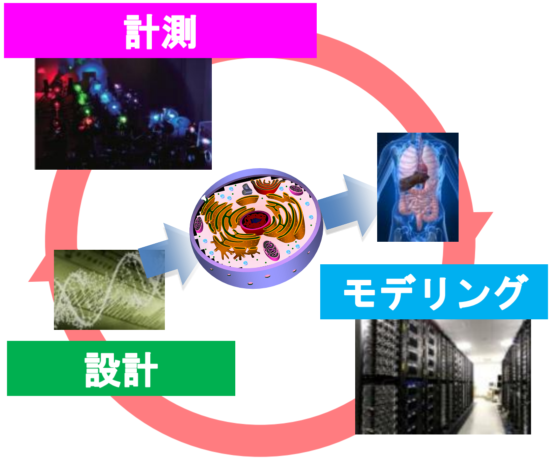
生命動態情報と細胞・発生画像情報の統合データベース

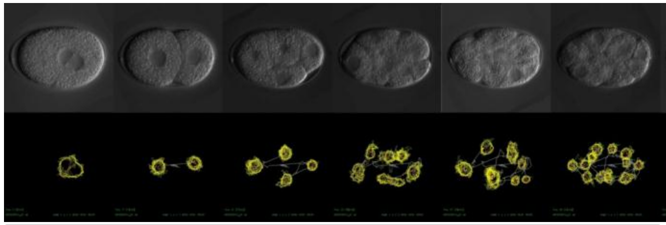
理化学研究所生命システム研究センター
大浪修一

©2015 大浪 修一(理化学研究所) licensed under CC表示2.1日本

生命動態システム科学

- 生命を動的システムとして理解し、操作する新しい生命科学
- 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密モデリングをもとに、in silicoとin vitroで再構成
- 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





0000, 1000, 51.8232, 37.7117, 27.1820, 1, P0, 17
477, 342, 23, 2111111112101112121112..........
486, 321, 24, 2121122221001212221O2..........
....
1000, 2000, 33.3146, 32.1893, 25.2702, 2, P0, 17
320, 257, 23, 21111121111121111111112..........
....

生命動態システム科学のデータベースの統合化

- 研究開始時に公開・運用されているデータベースは少ない
- 分野の成長に伴い、その後の急速な増加が見込まれる

↓

研究開始時なら、データベースの乱立に先行して統合化の体制と仕組みを構築し、当分野の全てのデータベースがその体制と仕組みの下で開発される体制の構築が可能

↓

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
- 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

データベース統合の体制の構築

生命動態システム科学のコミュニティーの合意の下でDB統合を実施するためにプロジェクトグループを構築した。

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

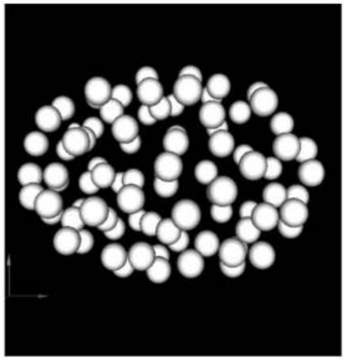
松田道行（京大）　井原茂男（東大）
金子邦彦（東大）　楯　真一（広大）

生命動態システム科学の定量データを記述する言語を構築

BDML: Biological Dynamics Markup Language

- XMLを基盤
  - 高い計算機可読性
    - 既存のXML用のライブラリやAPIの活用により、効率的なソフトウェア開発が可能
  - 高い拡張性
    - 新しいタイプのデータにも柔軟に対応可能

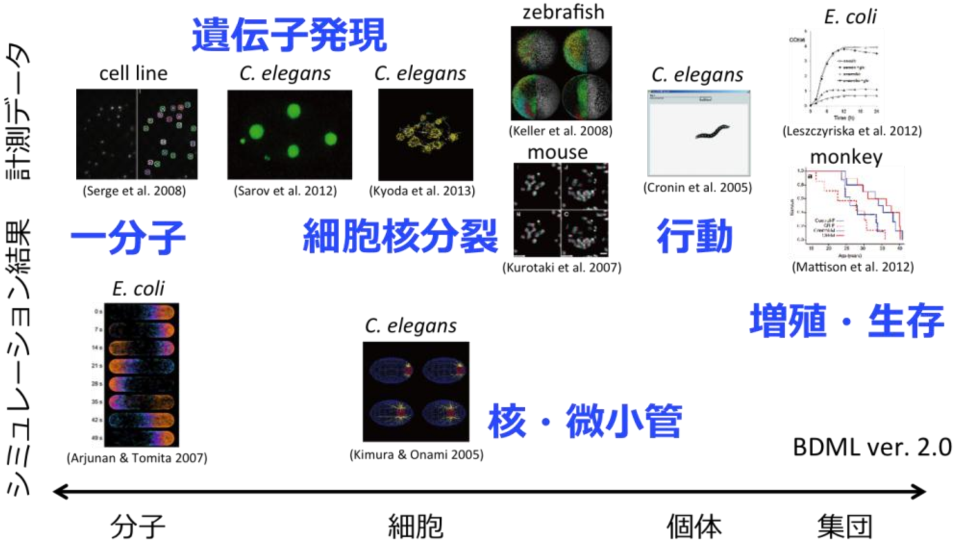
BDMLの記述例



```
<measurement>
  <objectRef>nucleus</objectRef>
  <sphere>
    <xyz>
      <x>261</x>
      <y>243</y>
      <z>8.1</z>
    </xyz>
    <radius>1.98</radius>
  </sphere>
  <property>
    <featureRef>total
       GFP signal</featureRef>
    <featureVal>438689</featureVal>
  </property>
</measurement>
```

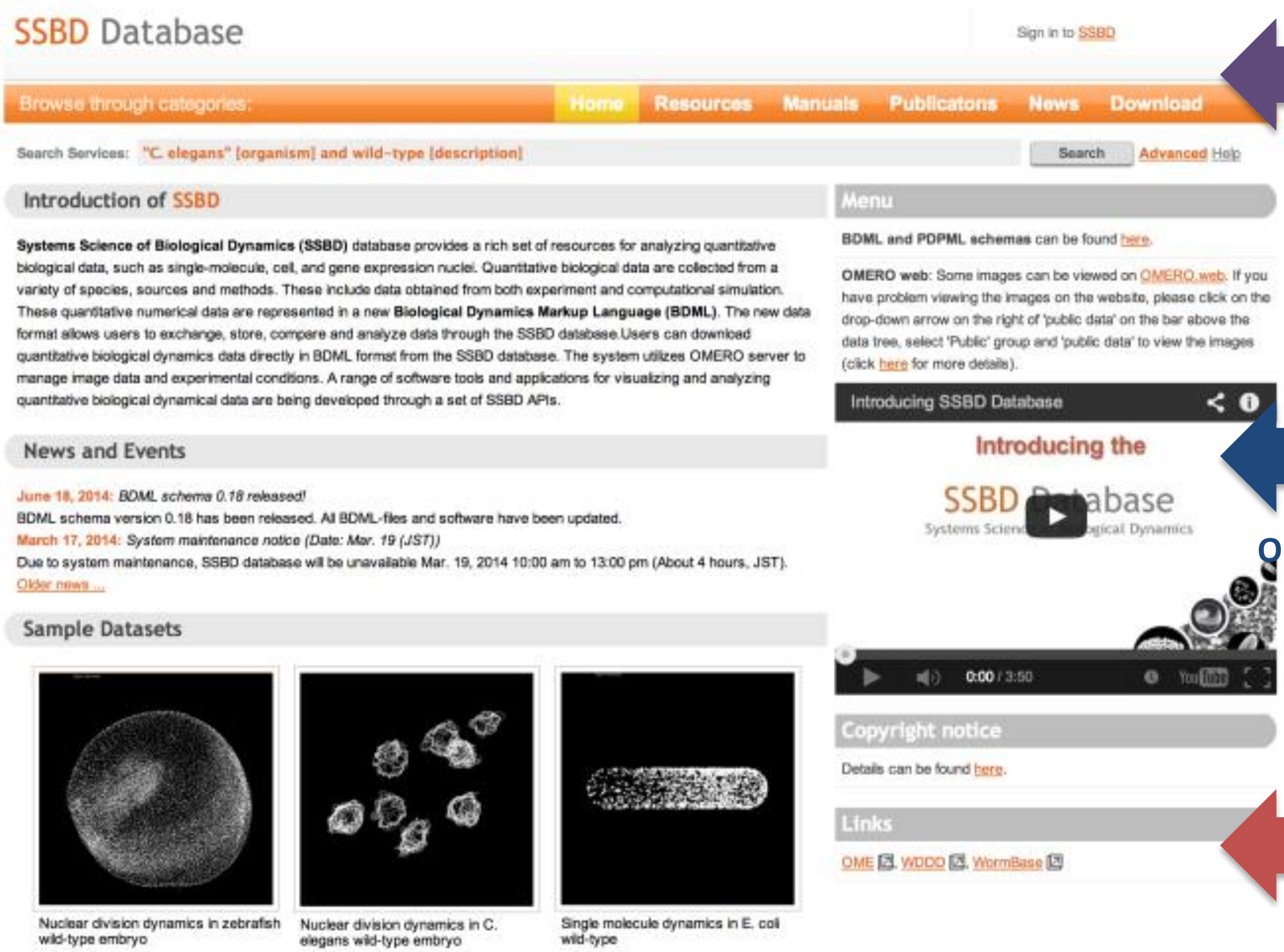
Kyoda et al., Bioinformatics, 2015

BDMLで記述可能なデータ



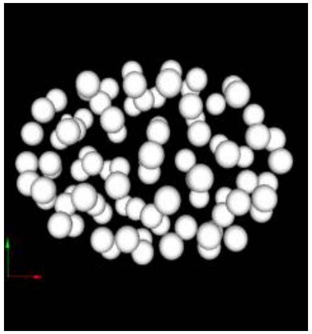
生命動態システム科学の統合データベースを構築

SSBD Database



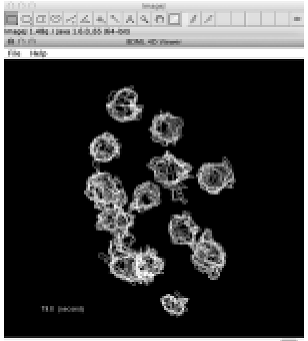
画像データ

定量データ



BDML
OmicsBDML

ツール



BDML

2013年9月公開

http://ssbd.qbic.riken.jp

国内の全ての生命動態システム科学のDBを統合

生物種	対象	種類	文献	#BDML	#component	#images
mouse	核	計測	Bashar et al. 2012	1	2,055	2,800
mouse	発現（MINiML）	計測	Masumoto et al. 2010	8	48	
mouse	核	計測	Kurotaki et al. 2007	1	12,096	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	56,584,840	
D. melanogaster	核	計測	Keller et al. 2010	2	5,111,828	
D. melanogaster	核	計測	Supatto et al. 2009	1	40,534	
C. elegans	核	計測	Kyoda et al. 2013	186	75,955	2,209,680
C. elegans	核	計測	Bao et al. 2005	2	24,747	
C. elegans	前核＋微小管	モデル	Kimura & Onami 2005	100	2,400,100	
C. elegans	行動	計測	Cronin et al. 2005	11	15,822	
C. elegans	前核＋核＋胚	計測	Tohsato et al.	1330	1,872,137	
C. elegans	核	計測	Kyoda et al.	259	155,873	6,153,840
C. elegans	核	計測	Takayama et al.	12	20,966	4,899
C. elegans	発現	計測	Sarov et al. 2012	273	5,713,854	
D. discoideum	一分子	計測	Jin et al.	1	987	368
E. coli	一分子	モデル	Arjunan & Tomita 2010	1	721	

今後の課題

- データの急増、大規模化への対応
  - 生命動態システム科学の３推進策が成果発表の時期に突入
- 持続可能な体制と仕組みの構築
- データベース活用の促進
- 生命科学分野全体への拡大
- 国際連携体制の構築

細胞生物学・発生生物学分野の画像データ

- 近年、最先端のイメージング技術の開発が加速
  - 超解像顕微鏡、レーザーシート顕微鏡etc.
- 最先端の画像が研究の競争力を決定づける

The Nobel Prize in Chemistry 2014

Photo: A. Mahmoud
Eric Betzig
Prize share: 1/3

Photo: A. Mahmoud
Stefan W. Hell
Prize share: 1/3

Photo: A. Mahmoud
William E. Moerner
Prize share: 1/3

(Nobelprize.org)

- 最先端の画像の活用は一部の研究室に限られる
  - 最先端のイメージング技術は高価or未市販

- 最先端の画像の共有化が望まれる
  - 一般的な細胞生物学・発生生物学研究への再利用
  - 動態の定量化 → SSBD → 生命動態研究

画像データは定量化を前提に管理するのが効果的

本研究課題の目的

- 我が国の生命動態システム科学のデータベースの統合化の発展的な継続
- 我が国の生命動態システム科学のデータベース統合の中長期的な継続を可能にする体制と仕組みの構築
- 我が国の細胞生物学および発生生物学の画像データベースと生命動態システム科学のデータベースとの統合

研究構想

（１） 我が国の生命動態システム科学のデータベースの統合化の発展的継続

1a. 生命動態システム科学の最新データの継続的な統合
1b. RDF/オントロジーを利用したデータベース連携の実現
1c. データベースツールおよび解析ソフトウェアの充実

（２） 我が国の生命動態システム科学のデータベース統合の中長期的な継続を可能にする体制と仕組みの構築

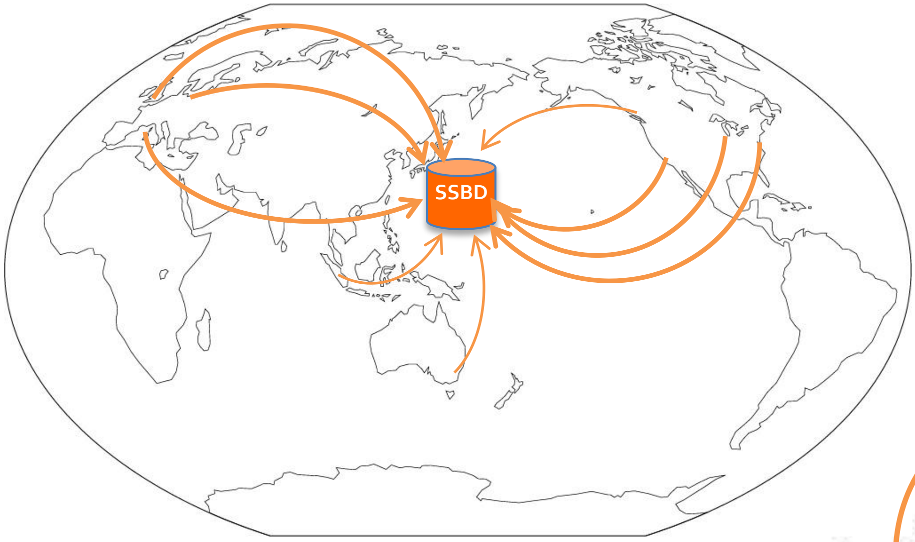
2a. データベース統合化の中長期的な体制の構築
2b. 生命動態の定量データを記述するフォーマットの大規模データへの対応と国際連携
2c. データベース登録作業等の簡素化・効率化

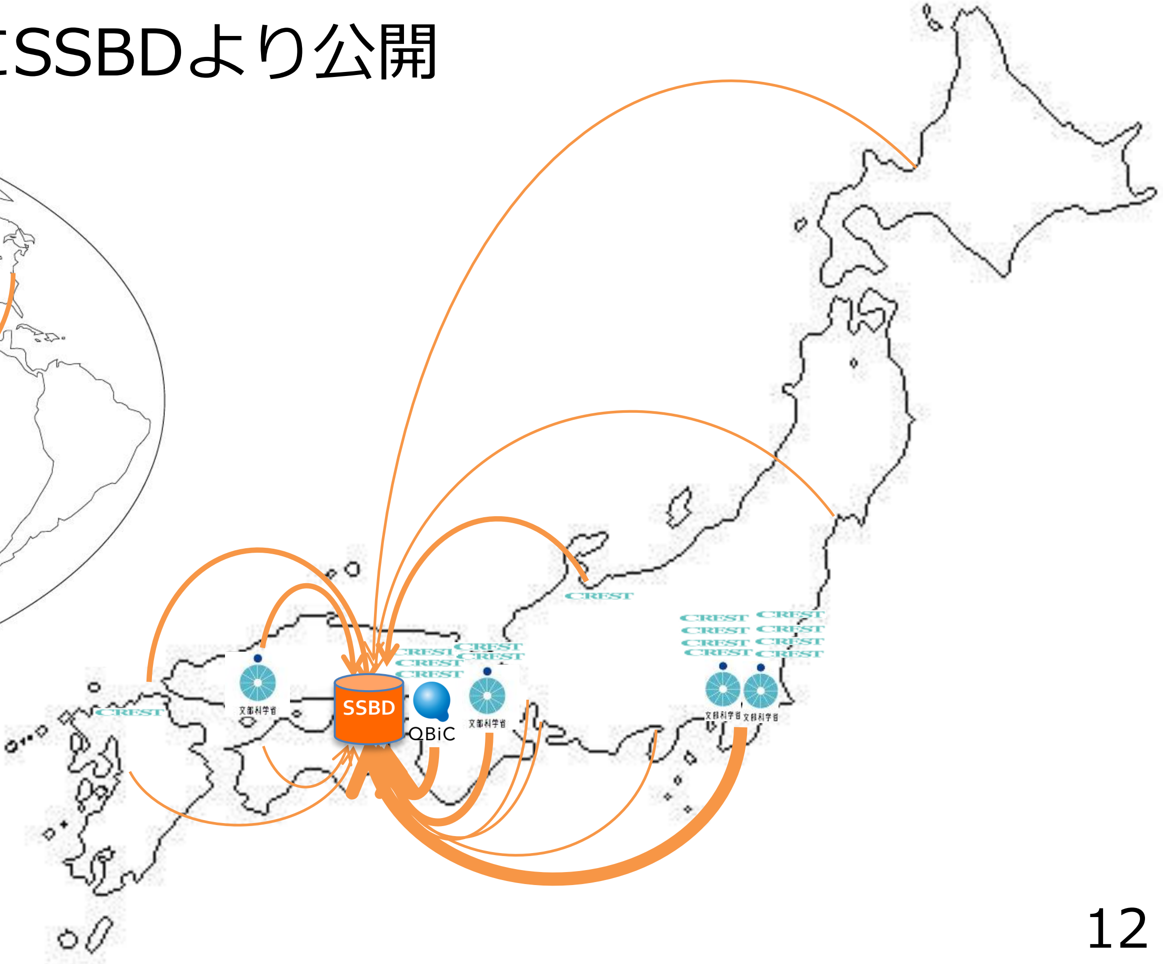
（３） 我が国の細胞生物学および発生生物学の画像データベースと生命動態システム科学のデータベースとの統合

3a. 細胞生物学および発生生物学の画像データベースの統合
3b. 画像処理ソフトなどに対応したアプリケーション・プラグインの開発

1a. 生命動態システム科学の最新データの継続的な統合

- 分野を代表する国内外のデータの統合
  - 理研QBiC、CREST生命動態、生命動態推進拠点からの成果発表の増加
- 論文発表後速やかにSSBDより公開

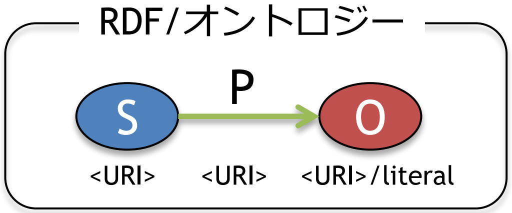




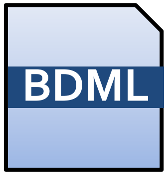
理研QBiC: H23～
CREST生命動態: H24～
生命動態推進拠点: H24～

1b. RDF/オントロジーを利用したデータベース連携の実現

- 時空間定量データのメタ情報をRDF/オントロジーを使って記述
  - BDMLファイルに記述
  - RDFトリプルストアを公開

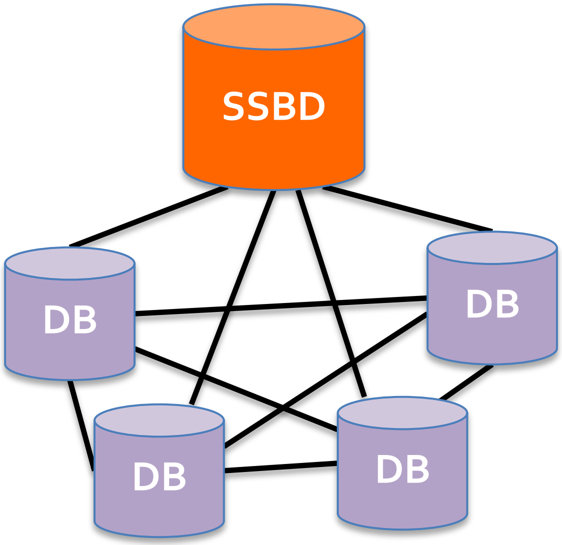


BDMLにRDF記述を追加

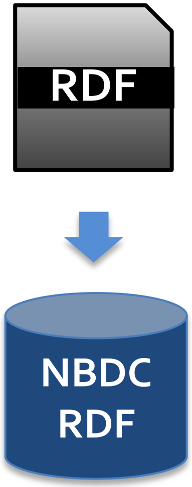



```
<rdf:Description rdf:about="xxx">
  <bqbiol:isVersionOf>
    <rdf:Bag>
      <rdf:li rdf:resource="xxx">
    </rdf:Bag>
  </bqbiol:isVersionOf>
  <bqbiol:hasTaxon>
    <rdf:Bag>
      <rdf:li rdf:resource="xxx">
    </rdf:Bag>
  </bqbiol:hasTaxon>
...
```

RDFトリプルストアを公開



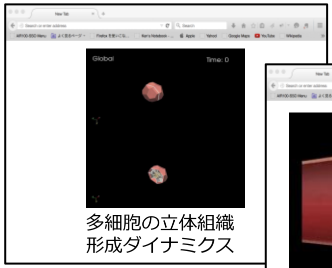
RDFデータファイルの提供



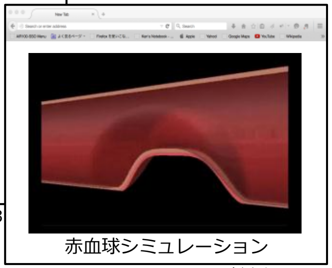
1c. データベースツールおよび解析ソフトウェアの充実

- オンラインツール
  - 可視化 etc.
- BDMLファイル対応ツール
  - 可視化、解析 etc.
- APIを充実

オンライン可視化ツール

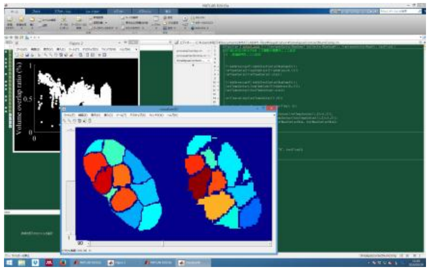


多細胞の立体組織形成ダイナミクス
Okuda et al., 2013

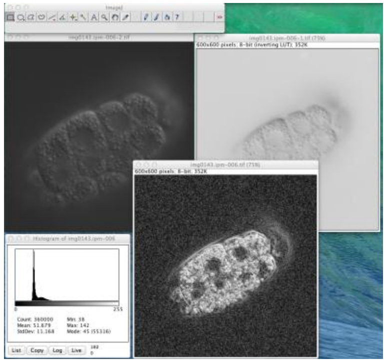


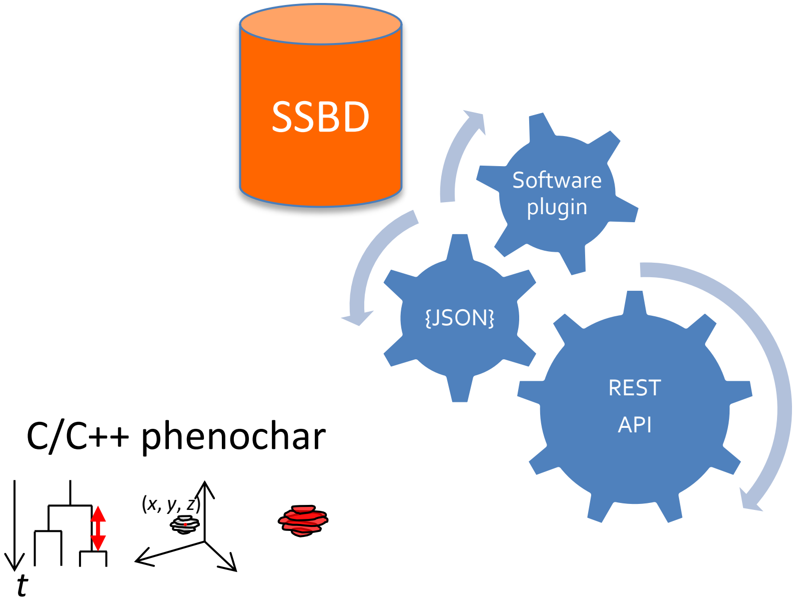
赤血球シミュレーション
杉山ら, 2013

R / Matlab plugin

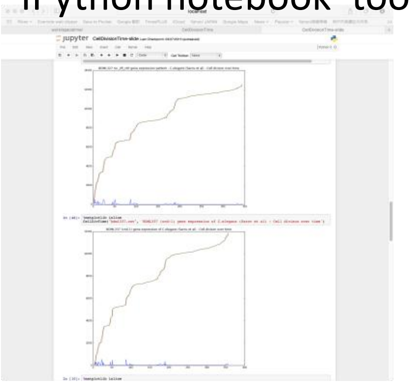


ImageJ / Java plugin



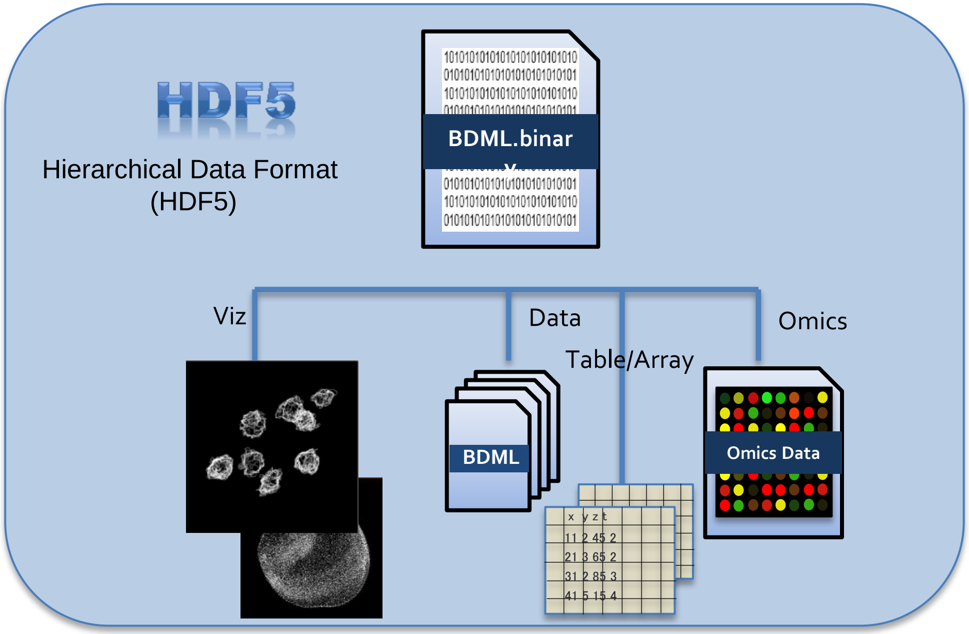


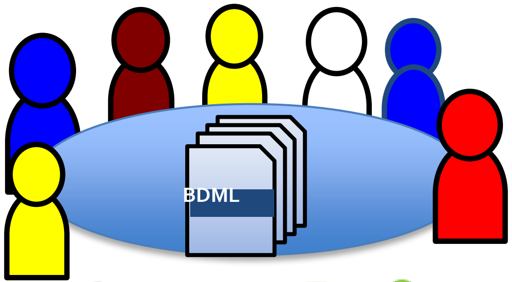
IPython notebook tools



2b. 生命動態の定量データを記述するフォーマットの大規模データへの対応と国際連携

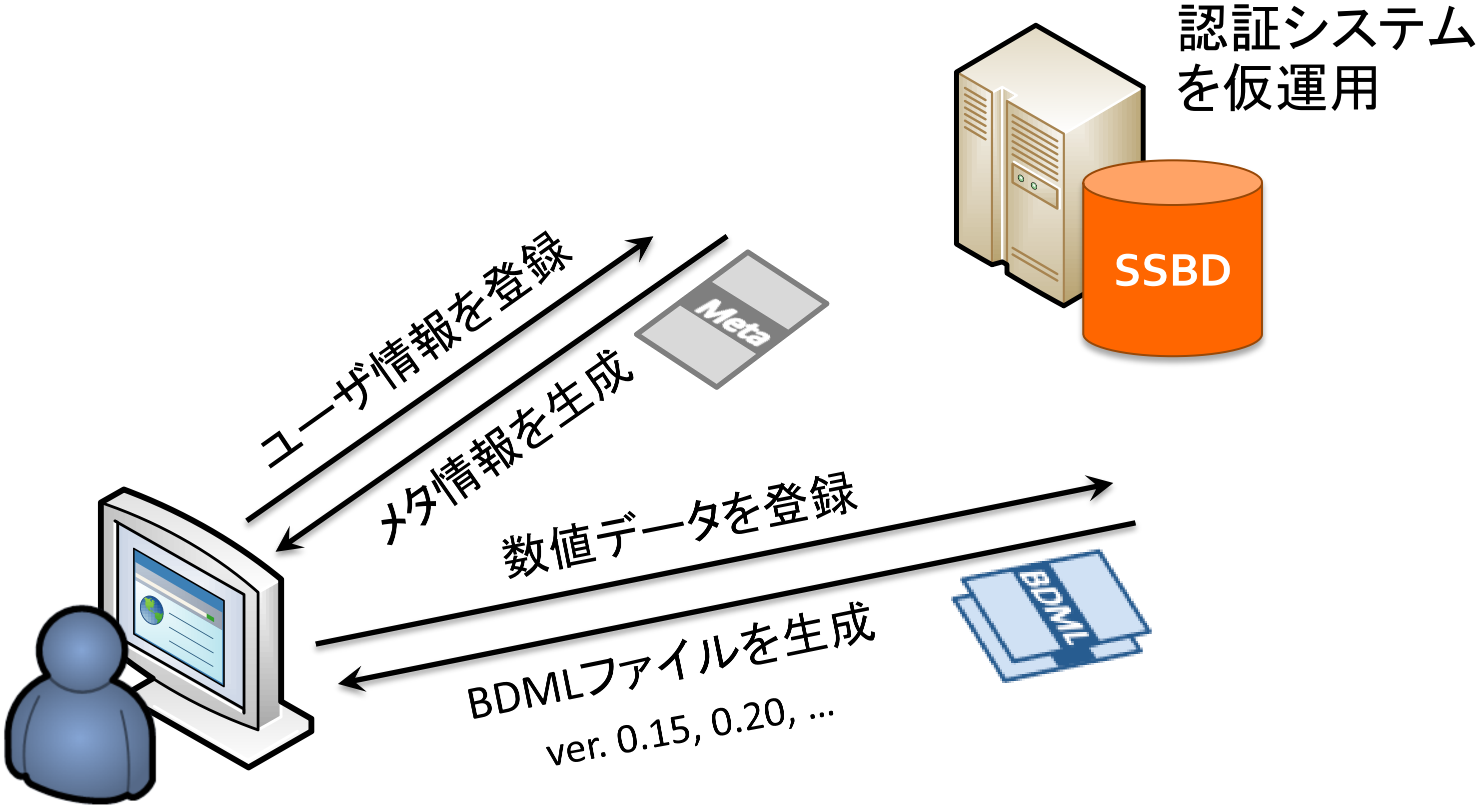
- BDMLのバイナリーフォーマット（HDF5）への対応
- 時空間オミクスデータ用フォーマット（Omics BDML）開発の推進
- 標準フォーマット開発の国際連携グループ（combineプロジェクト）との連携





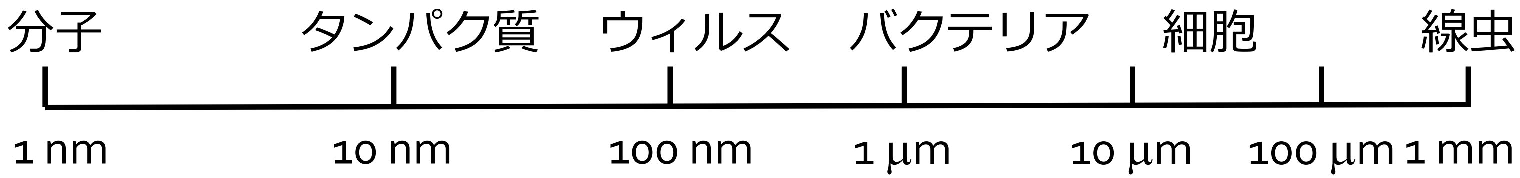
2c. データベース登録作業の簡素化・効率化

- データファイルを自動生成するシステムを構築
  - BDMLバージョンアップ時の作業の簡素化
- ユーザー認証のプロトタイプシステムの構築
  - データ登録作業の効率化



3a. 細胞生物学および発生生物学の画像データベースとの統合に向けた取り組み

- 分野を代表する国内のデータをメタ情報を含む形で統合
  - 定量化が期待されるデータ
  - 理研QBiC、CREST生命動態、生命動態推進拠点、「全国大学等バイオイメージング連携体制の今後のあり方を考える会」メンバーの代表的なデータ



SEM　STED　confocal　SPIM

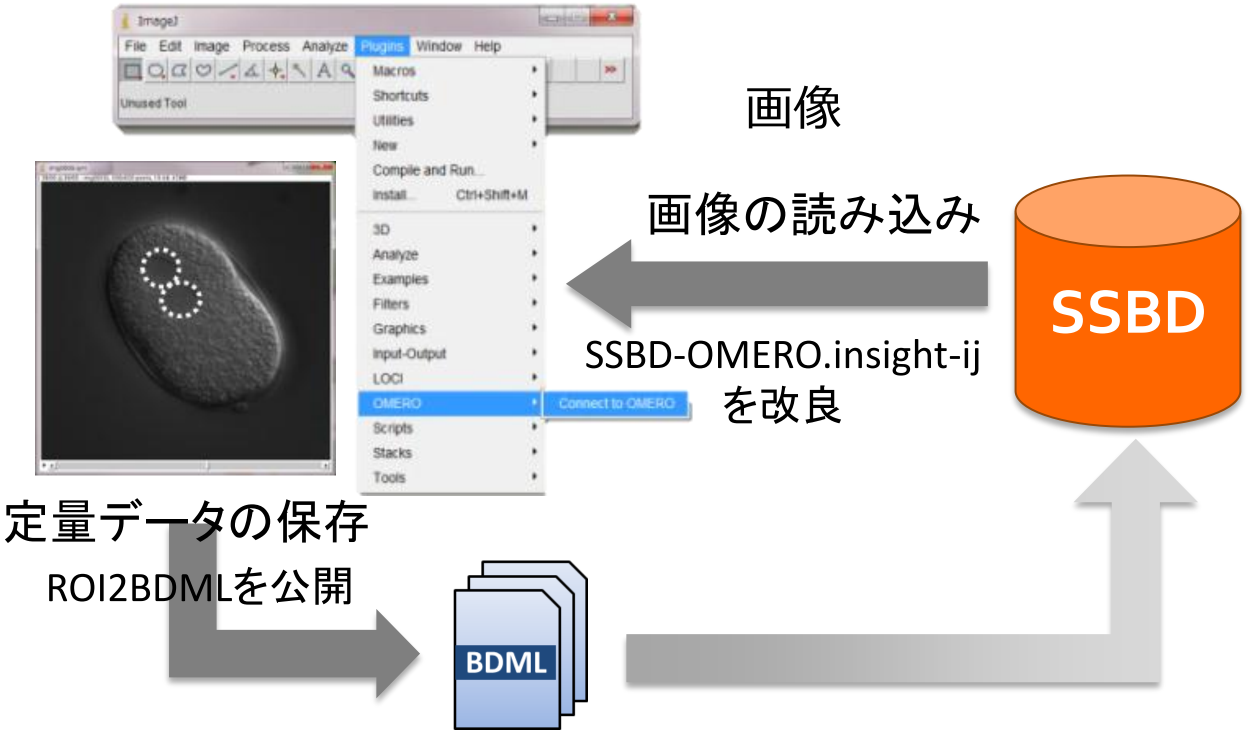
TEM　PALM　two-photon　DSLM

電子顕微鏡　超解像度顕微鏡　ライトシート顕微鏡

3b. 画像処理ソフトなどに対応したアプリケーション・プラグインの開発

- 主要な画像処理ソフトウェアに対してアプリケーション・プラグインを開発
  - SSBDに登録された画像の利用を促進

ImageJに対応したアプリケーション・プラグイン



研究開発のスケジュール

研究項目	H27年度	H28年度	H29年度
1a. 生命動態システム科学の最新データの継続的な統合	←	―	→
1b. RDF/オントロジーを利用したデータベース連携の実現	←	→- - -	- - -→
1c. データベースツールおよび解析ソフトウェアの充実	←	―	→
2a. データベース統合化の中長期的な体制の構築		・←	→
2b. 生命動態の定量データを記述するフォーマットの大規模データへの対応と国際連携	←	→- - -	- - -→
2c. データベース登録作業等の簡素化・効率化		・←	→
3a. 細胞生物学および発生生物学の画像データベースの統合	←	―	→
3b. 画像処理ソフトなどに対応したアプリケーション・プラグインの開発	←	―	→

将来展望

- 全世界のデータを統合した、生命動態研究の国際統合データベースとして継続的に発展させる
  - 塩基配列データにおけるGenBank/EMBL/DDBJに相当
  - 世界各地域の拠点と連携して運営
- 生命動態研究を推進する生命科学分野全体の情報インフラとして継続的に発展させる
  - イメージング
  - バイオイメージ・インフォマティクス
  - 数理モデル化
  - 計算シミュレーション
  - データ駆動型研究
- 医療や製薬、バイオ産業への生命動態研究の応用を推進