

第41回日本分子生物学会年会 MBSJ 2018
ワークショップ IPW2-09 生命科学のデータベース活用法2018

SSBD: 細胞・発生画像情報と 生命動態情報の統合データベース

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

Ho, K.H.L.¹, Tohsato, Y.^{1,2}, Kyoda, K.¹, Itoga, H.¹, Onami, S.¹

¹ Laboratory for Developmental Dynamics, RIKEN BDR

² Osaka Electro-Communication University

ホー・ケネス¹, 遠里 由佳子^{1,2}, 京田 耕司¹, 糸賀 裕弥¹, 大浪 修一¹

¹ 理化学研究所生命機能科学研究センター発生動態研究チーム

² 大阪電気通信大学 情報通信工学部 情報工学科



Need for public repository

- There is a need for public repositories for retaining important knowledge and reusing experimental data

The screenshot shows the Ensembl genome browser interface. At the top, it says 'Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog'. Below that, it says 'Human (GRCh38.p12)'. There is a search bar with 'Search all categories' and 'Search Human...'. Below the search bar, there are several sections: 'Genome assembly: GRCh38.p12 (GCA_000001455.2)', 'Gene annotation', 'Comparative genomics', and 'Variation'. Each section has a brief description and links to more information.

Ensembl

The screenshot shows the IGSR website. At the top, it says 'IGSR: The International Genome Sample Resource'. Below that, it says 'Providing ongoing support for the 1000 Genomes Project data'. There is a navigation menu with 'Home', 'About', 'Data', 'Portal', 'Analysis', 'Contact', 'Browser', and 'FAQ'. Below the navigation menu, there is a section titled 'Using data from IGSR' with a sub-section 'Data portal beta'. Below this, there is a table of samples:

Sample	Sex	Population	Example	Low coverage	High coverage	IBD phenotype	Complete
HG00513	Female	CHS		•	•	•	•
HG01112	Male	CLM		•	•	•	•
HG00759	Female	CDX		•	•	•	•
HG01500	Male	IBS		•	•	•	•

IGSR



Comment | OPEN | Published: 30 October 2018

A call for public archives for biological image data

Jan Ellenberg, Jason R. Swedlow, Mary Barlow, Charles E. Cook, Ugis Sarkans, Ardan Patwardhan, Alvis Brazma & Ewan Birney

Nature Methods 15, 849–854 (2018) | Download Citation ↓

The screenshot shows the KEGG MEDICUS website. At the top, there is a search bar with 'KEGG MEDICUS' and a search button. Below the search bar, there is a section titled 'KEGG MEDICUS' with a sub-section 'KEGG MEDICUS はヒトゲノム、疾病体ゲノム、様々なメタゲノムなどのシーケンス解析と有効利用を促進する統合リソースで、ゲノムの情報と疾患との関連、医薬品の作用・副作用との関連、薬剤的若や薬剤耐性との関連などが知識ベース化されています。また日本と米国すべての医薬品注記文書も統合されており、一般社会にとって有用なリソースです。このページもご覧ください。' Below this, there is a diagram showing the KEGG network with nodes for 'KEGG NETWORK', 'KEGG DISEASE', 'KEGG DRUG', and 'KEGG ENVIRON'. There are also links to 'KEGG MEDICUS 検索' and 'KEGG MEDICUS ヘルプ'.

KEGG/KEGG MEDICUS

The screenshot shows the PDBj website. At the top, there is a search bar with 'PDBj Protein Data Bank Japan'. Below the search bar, there is a section titled '初めての利用者のためのガイド' with a sub-section '必要なサービスを探す'. Below this, there is a table of services:

サービス	説明
全サービス	全サービスを表示するボタンを押すと、全サービスの概要が表示されます。
キーワード	キーワードを入力して、該当するサービスの一覧が表示されます。
キーワードボックス	キーワードを入力して、該当するサービスの一覧が表示されます。

PDBj

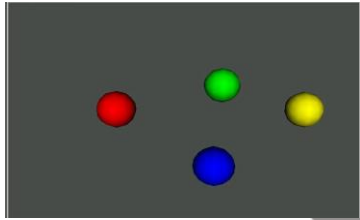
Public data archives are the backbone of modern biological research. Biomolecular archives are well established, but bioimaging resources lag behind them. The technology required for imaging archives is now available, thus enabling the creation of the first public bioimage datasets. We present the rationale for the construction of bioimage archives and their associated databases to underpin the next revolution in bioinformatics discovery.

(Ellenberg et al, 2018)

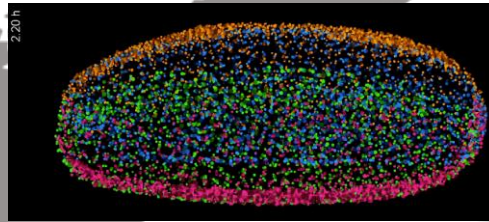
Problem: difficult to reuse

- Images and Datasets are scattered over the Internet
 - Difficult to locate and search
- Most quantitative data use different formats
 - Time consuming/difficult to reuse these datasets

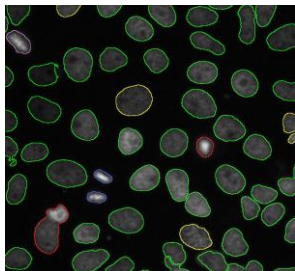
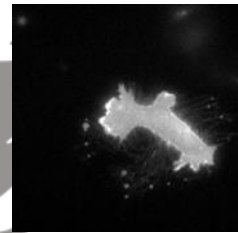
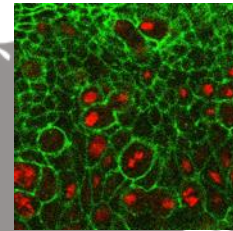
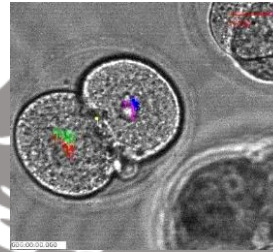
(Bao et al., 2006)



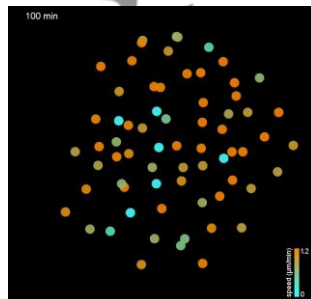
(Keller et al., 2010)



(Kurotaki et al., 2007) (Kondo & Hayashi 2013) (Kunida et al. 2012)



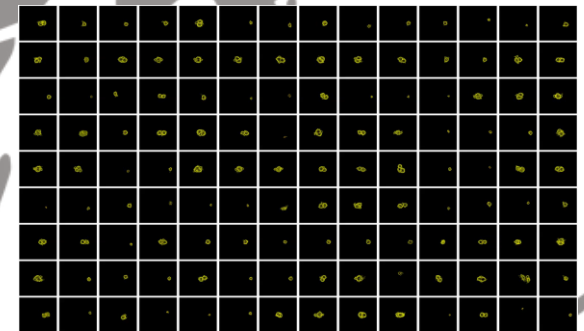
(Heldet et al., 2010)



(Keller et al., 2008)



(Yemini et al., 2013)



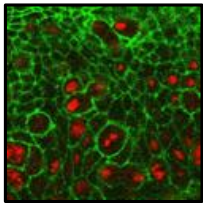
(Kyoda et al., 2013)

SSBD Database

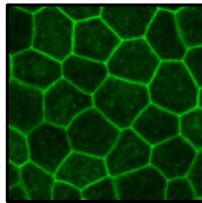
(Tohsato et al, Bioinformatics, 2016)

画像データ

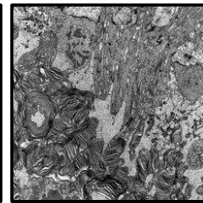
Image data



(Kondo & Hayashi 2013)



(Inomata et al. 2013)



(Shirai et al. 2016)



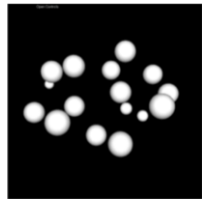
Original
format

定量データ

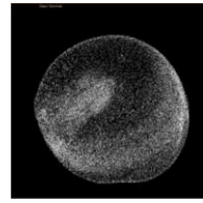
Quantitative data



(Arjunan & Tomita 2010)



(Bao et al. 2006)



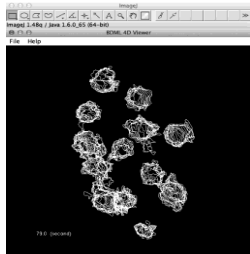
(Keller et al. 2008)



BDML/BD5
形式

ソフトウェアツール

Software tools



BDML/BD5
形式に対応

SSBD Database

Browse through categories:

Home Resources Manuals Publications News Software

Image Search

Data Search

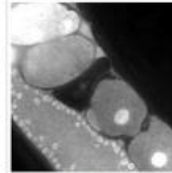
Archive Data

Introduction of SSBD

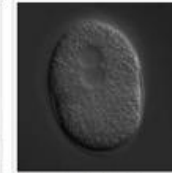
Systems Science of Biological Dynamics (SSBD) database provides a rich set of open resources for analyzing quantitative data and microscopy images of biological objects, such as single-molecule, cell, gene expression nuclei, etc. Quantitative biological data and microscopy image are collected from a variety of species, sources and methods. These include data obtained from both experiment and computational simulation.

Samples

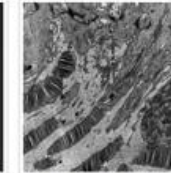
Microscopy images



Calcium response and shape changes in oocyte of *C. elegans*

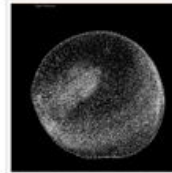


DIC image of nuclear division dynamics in *C. elegans* embryo

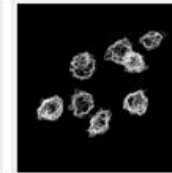


TEM image of retinal tissue from human embryonic stem cells

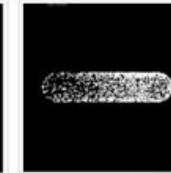
Quantitative data



Nuclear division dynamics in *D. rerio* wild-type embryo



Nuclear division dynamics in *C. elegans* wild-type embryo



Single molecule dynamics in *E. coli* wild-type

News and Events

OMERO Open Workshop in Kobe 2018

Thursday 25th October 2018 13:00-17:00

[Click here for more details](#)

Mar. 23, 2018: System maintenance - some disruption of service. We are migrating to a new server. Some of the services are not available during this period. We hope to provide full service asap.

Dec. 1, 2017: System maintenance notice (Date: Dec 4 JST) Due to system maintenance, SSBD database will be unavailable on Monday 4 December 2017.

[Older news ...](#)

Information

OMERO version 5.4.3: 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).

OMERO session ID: 8d220e77-f102-419e-b1e7-4c3b99f32087

Introducing SSBD Database

Introducing the SSBD Database
Systems Science of Biological Dynamics

Links

[NBDC RDF Portal](#) [RIKEN Meta Database](#) [WDDO](#)

[OME](#) [Ensembl](#) [WormBase](#)

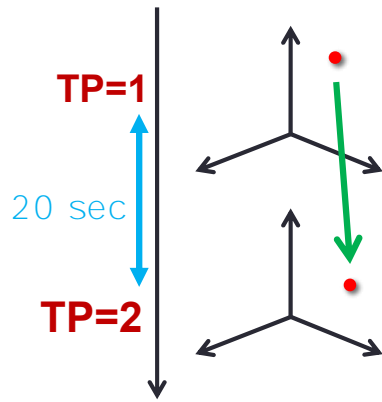


<http://ssbd.qbic.riken.jp>

BDML/BD5 – an open unified file format Biological Dynamics Markup Language

(Kyoda et al., Bioinformatics, 2015)

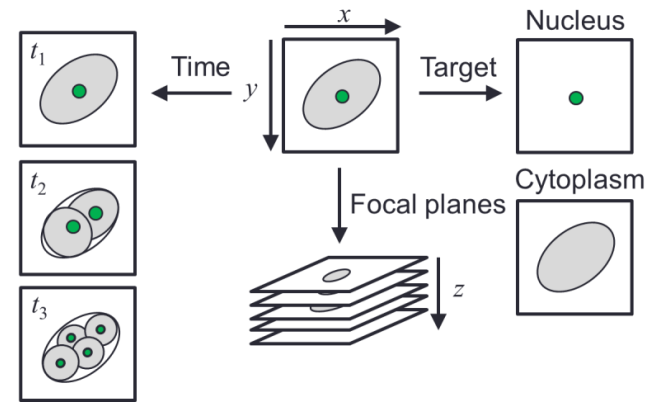
• Meta information in XML



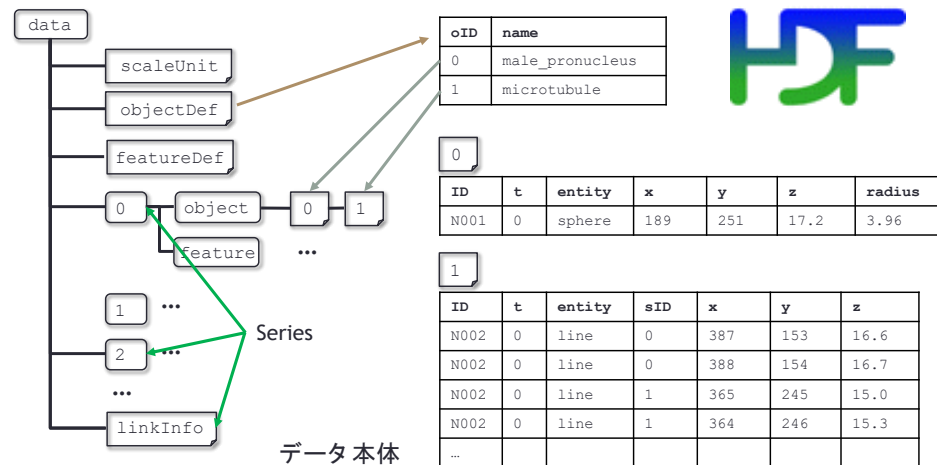
```

<scaleUnit>
  <tScale>20</tScale>
  <tUnit>second</tUnit>
</scaleUnit>
<component>
  <componentID>100</componentID>
  <time>1</time>
  <measurement>
    <point><xyz><x>1.3</x><y>3.4</y><z>1.2</z></xyz></point>
  </measurement>
</component>
<component>
  <componentID>101</componentID>
  <time>2</time>
  <prevID>100</prevID>
  <measurement>
    <point><xyz><x>0.9</x><y>3.2</y><z>1.1</z></xyz></point>
  </measurement>
</component>
  
```

Quantitative data in HDF



Modified from (Goldberg et al, Genome Biology 2005)



SSBD Big Data resource

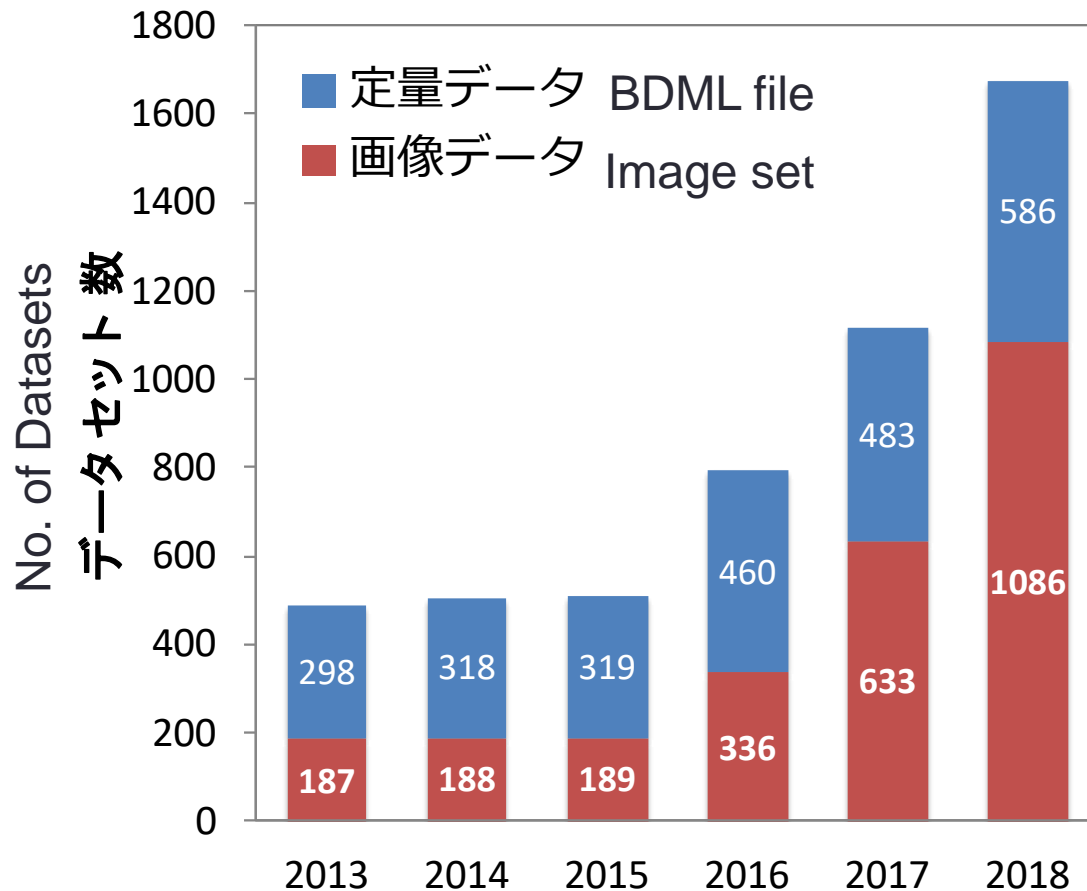
586 BDML-files, 64.7M entities
1086 image sets, 2.6M slices

Resources

As of Nov. 15, 2018 there are 586 BDML-files and 1086 image sets.

Organism	Type	Based on	Paper	Number of BDML-files	Number of images in SSBD	Released date	Update date						
<i>C. elegans</i>	Nucleus	Measurement	Bao et al. 2006	1	0	2013/10/03	2018/11/15						
<i>C. elegans</i>	Nucleus	Measurement	Kyoda et al. 2013	186	2,221,560	2013/09/02	2018/11/15						
<i>C. elegans</i>	Nucleus	Simulation	Kimura & Onami 2005	100	0	2013/10/03	2018/11/15						
<i>C. elegans</i>	Behavior	Measurement	Cronin et al. 2005	11	0	2014/10/03	2018/11/15						
<i>C. elegans</i>	Nucleus	Measurement	Toyoshima et al. 2016	14	41,380	2016/05/20	2018/11/15						
<i>C. elegans</i>	Cell	Measurement	Takayama & Onami 2016	119	77,306	2016/10/03	2018/11/15						
<i>C. elegans</i>	Cell	Measurement	Azuma & Onami 2016	1	144	2017/03/01	2018/11/15						
<i>C. elegans</i>	Molecule	Measurement	Arata et al. 2016	3	3,000	2017/10/03	2018/11/15						
<i>D. cf. damesi</i>	Individual	Measurement	Inoue & Kondo 2016	0	6,096	2017/10/03	2018/11/15						
<i>D. discoideum</i>	Molecule	Measurement	Komatsuzaki et al. 2015	1	5,400	2015/10/03	2018/11/15						
<i>D. discoideum</i>	Molecule	Measurement	Yasui et al. 2014	0	2,421	2017/10/03	2018/11/15						
<i>D. discoideum</i>	Cell	Measurement	Shibata et al. 2012	2	1,440	2017/10/03	2018/11/15						
<i>D. discoideum AX2</i>	Cell	Measurement	Kamimura et al. 2016	0	324	2018/11/14	-						
<i>D. discoideum</i> , Cultured cell (rat; PC-12), NA	Molecule	Measurement/Simulation	Watabe et al. 2015	4	61	2016/01/18	2018/11/15						
<i>D. melanogaster</i>	Nucleus	Measurement	Keller et al. 2010	2	0	2013/10/03	2018/11/15						
<i>D. melanogaster</i>	Cell	Measurement	Kondo & Hayashi 2013	0	85,928	2017/10/03	2018/11/15						
<i>D. melanogaster</i>	Cell	Measurement	Yasugi et al. 2017	0	12	2018/11/14	-						
<i>D. rerio</i>	Nucleus	Measurement	Keller et al. 2008	7	0	2013/10/03	2018/11/15						
<i>E. coli</i>	Molecule	Simulation	Arjunan & Tomita 2010	1	0	2013/09/02	2018/11/15						
<i>H. sapiens</i>	Cell	Measurement	Takagi et al. 2017	0	32	2018/11/14	-						
<i>H. sapiens</i> (ES)	Tissue	Measurement	Shirai et al. 2016	0	2	2017/10/03	2018/11/15						
<i>H. sapiens</i> (iPS)	Cell	Measurement	Kanemura et al. 2013	0	48	2017/10/03	2018/11/15						
<i>H. sapiens, M. musculus</i>	Cell	Measurement	Nojima et al. 2017	0	342	2018/11/14	-						
<i>M. musculus</i>	Nucleus	Measurement	Bashar et al. 2012	1	2,800	2014/10/03	2018/11/15						
<i>M. musculus</i>	Gene expression	Measurement	Masumoto et al. 2010	(OmicsBDML) 8	0	2014/10/03	2017/10/03						
<i>M. musculus</i>	Cell/Organ	Measurement	Ke et al. 2013	0	21,995	2017/10/03	2018/11/15						
<i>M. musculus</i>	Molecule/Cell	Measurement	Sakakibara et al. 2015	1	43,044	2017/10/03	2018/11/15						
<i>M. musculus</i>	Molecule	Measurement	Hirata et al. 2016	11	0	2017/10/03	2018/11/15						
<i>M. musculus</i>	Cell	Measurement	Noguchi et al. 2015	0	28,245	2018/11/14	-						
<i>M. musculus</i>	Organ	Measurement	Sasaki et al. 2014	0	12812	2017/10/03	2018/11/15						
<i>M. musculus</i>	Organ	Measurement	Sasaki et al. 2015	0	512	2017/10/03	2018/11/15						
<i>M. musculus</i>	Organ	Measurement	Tainaka et al. 2014	0	543	2017/10/03	2018/11/15						
<i>M. musculus</i>	Cell	Measurement	Herawati et al. 2016	0	5002	2018/11/14	-						
<i>M. musculus</i>	Cell	Measurement	Minegishi et al. 2017	0	12	2018/11/14	-						
<i>M. musculus</i>	Cell	Measurement	Ikedo et al. 2018	0	27	2018/11/14	-						
<i>M. musculus</i>	Cell	Measurement	Fumoto et al. 2017	0	201	2018/11/14	-						
<i>M. musculus</i>	Cell	Measurement	Kosodo et al. 2017	0	2738	2018/11/14	-						
<i>M. musculus</i>	Cell	Measurement	Koerberle et al. 2017	0	677	2018/11/14	-						
<i>M. musculus</i>	Cell	Measurement	Ishiki et al. 2014	0	5210	2018/11/14	-						
<i>M. musculus</i> (implanted with KPL-4; <i>H. sapiens</i>)	Cell	Measurement	Jin et al. 2016	0	24	2018/11/14	-						
<i>M. musculus</i> (implanted with KPL-4; <i>H. sapiens</i>)	Cell	Measurement	Tsuboi et al. 2017	0	39	2018/11/14	-						
<i>M. musculus</i> (iPS)	Organ	Measurement	Kinozaki et al. 2016	0	384	2017/10/03	2018/11/15						
<i>M. musculus</i> (iPS)	Cell	Measurement	Heasaki et al. 2016	0	39	2017/10/03	2018/11/15						
<i>M. musculus</i> (iPS)	Tissue	Measurement	Assawachananont et al. 2014	0	8	2017/10/03	2018/11/15						
<i>M. musculus</i> (iPS), <i>D. melanogaster</i> , Cultured cell (HEK293T; human)	Organ	Measurement	Ke et al. 2016	0	14,656	2017/10/03	2018/11/15						
<i>M. musculus</i> (iPS)	Cell/Particle	Measurement	Tanaka & Fujita 2015	2	14,573 (sf)	2016/10/03	2018/11/15						
<i>M. sieboldii</i>	Cell	Measurement	Tanaka et al. 2017	2	0	2018/11/15	-						
<i>X. laevis</i>	Molecule/Cell	Measurement	Inomata et al. 2013	0	9,000	2017/10/03	2018/11/15						
<i>X. laevis</i>	Cell	Measurement	Suzuki et al. 2016	0	1	2017/10/03	2018/11/15						
<i>X. laevis</i> , Cultured cell (HeLa; <i>H. sapiens</i>)	Cell	Measurement	Ishiwata et al. 2017	0	78	2018/11/14	-						
Cultured cell (CHO; hamster + MDCK; dog)	Molecule/Cell	Measurement	Matsuda et al. 2012	0	867	2017/10/03	2018/11/15						
Cultured cell (CHO; hamster)	Molecule/Cell	Measurement	Matsuda et al. 2015	0	764	2017/10/03	2018/11/15						
Cultured cell (DO11.10; mouse)	Cell	Measurement	Ichimura et al. 2016	0	48	2017/10/03	2018/11/15						
Cultured cell (Eph4; <i>M. musculus</i>)	Cell	Measurement	Yano et al. 2013	0	3	2018/11/14	-						
Cultured cell (ES; <i>M. musculus</i>)	Cell	Measurement	Maekawa et al. 2015	0	45	2018/11/14	-						
Cultured cell (HeLa; <i>H. sapiens</i>)	Cell	Measurement	Inomata et al. 2017	0	24	2018/11/14	-						

SSBD 2018 release datasets

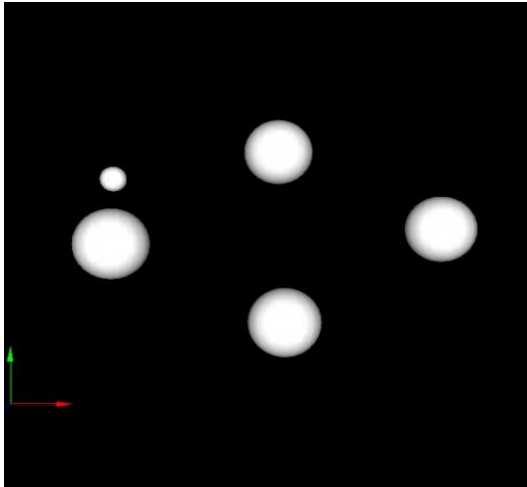


Increases in 2018

- BDML files: 103
 - 85,540 entities
- Image sets: 453
 - 58,919 slices

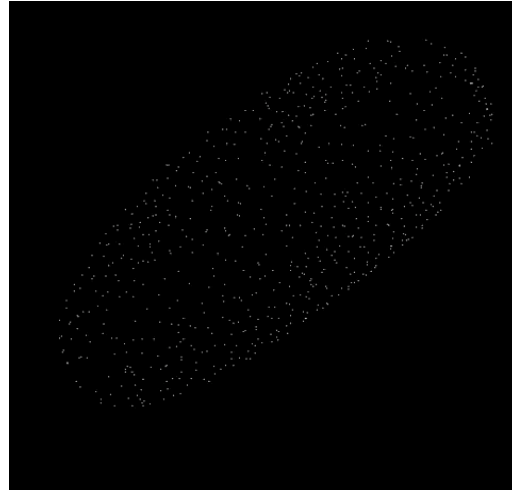
SSBD - examples of quantitative data

C. elegans embryogenesis



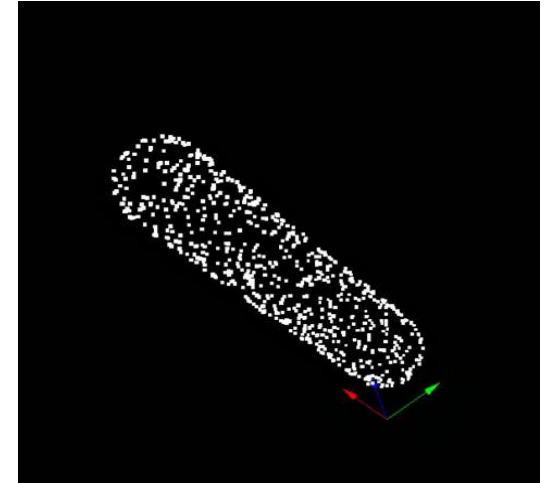
(Bao et al., 2006)

D. melanogaster embryogenesis



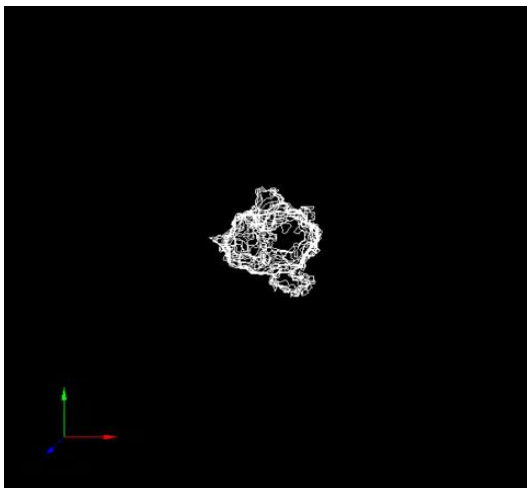
(Keller et al., 2010)

E. coli single-molecules



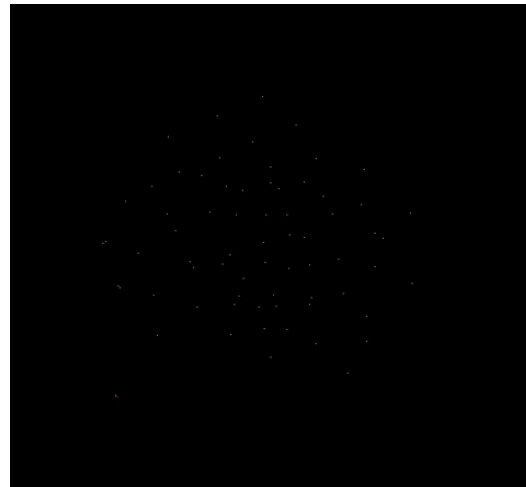
(Arjunan & Tomita, 2010)

C. elegans embryogenesis



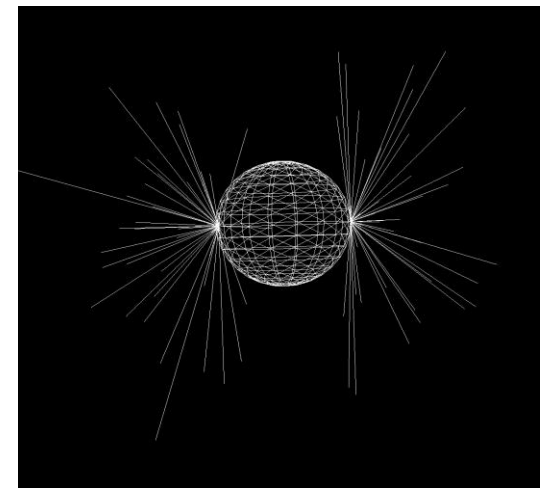
(Kyoda et al., 2013)

zebrafish embryogenesis



(Keller et al., 2008)

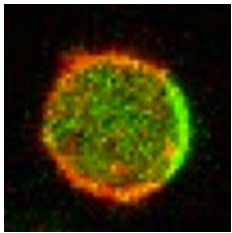
C. elegans pronuclear migration



(Kimura & Onami, 2005)

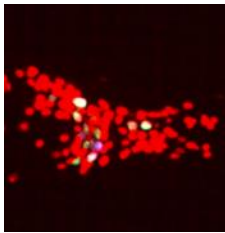
SSBD – examples of image data

Dictyostelium
cell



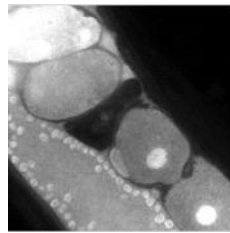
(Watabe et al. 2015)

C. elegans
neuronal cell



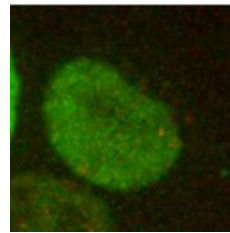
(Toyoshima et al. 2016)

C. elegans
oocyte



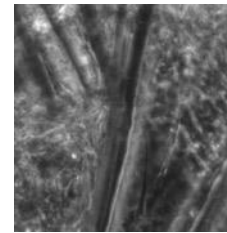
(Takayama & Onami 2016)

Mouse
ES cell



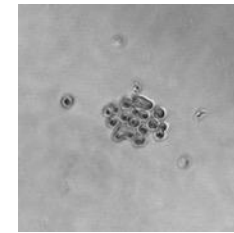
(Ochiai et al. 2015)

Mouse
iPS cell



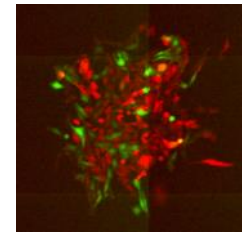
(Tanaka & Fujita 2015)

Mouse
cultured cell



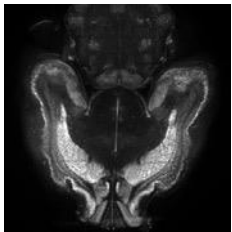
(Yonemura 2014)

Hamster
cultured cell



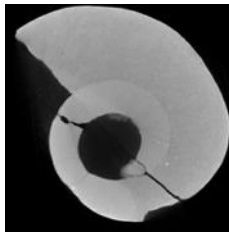
(Matsuda et al. 2015)

Mouse



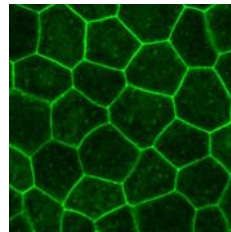
(Susaki et al. 2014)

D. cf. damesi



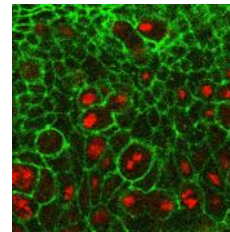
(Inoue & Kondo 2016)

X. laevis



(Inomata et al. 2013)

D. melanogaster



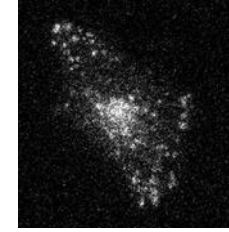
(Kondo & Hayashi 2013)

Rat



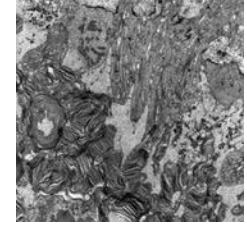
(Aoki et al. 2013)

Dog



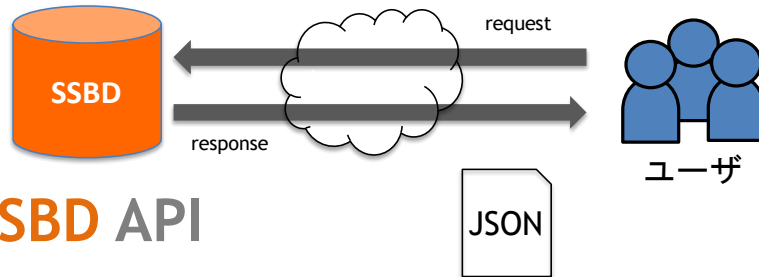
(Takai et al. 2015)

Human



(Shirai et al. 2016)

Analysis — REST API



Data can be accessed directly from SSBD REST API

SSBD API

Analysis - plotting the proliferation curve

```
In [24]: no_of_nucleus = []
timept = []
tp = 1
resultdata = retrieve_coord_tp('d15115', tp, display='off')
nn = resultdata['meta']['total_count']
while nn > 0:
    no_of_nucleus.append(nn)
    timept.append(tp)
    # print "tp="+str(tp)+" nn="+str(nn)
    tp=tp+1
    resultdata = retrieve_coord_tp('d15115', tp, display='off')
    nn = resultdata['meta']['total_count']
```

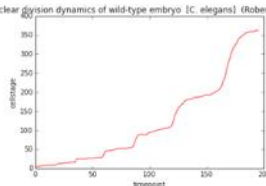
Get the title, organism and contact name of the dataset

```
In [25]: q.set_display('off')
resultmetadata = q.data(field='bdmIUUID', search='d15115')
for i in resultmetadata['objects']:
    title= i['meta_data']['title']
    name= i['meta_data']['name']
    pmid= i['meta_data']['PMID']
    organism = i['meta_data']['organism']
```

Plotting the curve

```
In [26]: fig=plt.figure()
plt.plot(timept, no_of_nucleus, 'r')
ax = fig.add_subplot(1,1,1)
ax.set_ylabel('cellstage')
ax.set_xlabel('timepoint')
plt.title = title+' [' +organism+' ] (' +name+' ) PMID: '+str(pmid)+' Cell division over time '
plt.title(plt.title);
```

BDML file for quantitative information about nuclear division dynamics of wild-type embryo [C. elegans] (Robert H. Waterston) PMID: 16477039 Cell division over time



7. Apply active contour model (snake)

```
In [257]: (x, y), radius = cv2.minEnclosingCircle(contours[max_i])
center = (int(x), int(y))
radius = int(radius) + 10 # make larger
s = np.linspace(0, 2*np.pi, 300)
x = center[0]+radius*np.cos(s)
y = center[1]+radius*np.sin(s)
init = np.array([x, y]).T

# alpha: length shape parameter. Higher values makes contract faster.
# beta: smoothness shape parameter. Higher values makes smoother.
# ganna: explicit time stepping parameter.
snake = active_contour(cv_img3, init, alpha=0.015, beta=200, gamma=0.001)

c = []
for s in snake:
    c.append([s[0], s[1]])
contour = np.array(c, dtype=np.int) # float -> int
#print ("snake.shape", snake.shape)
#print ("contour.shape", contour.shape)

cv_img4 = cv2.fillConvexPoly(np.zeros(size, dtype=np.uint8), contour, (255,255,255))

fig, (ax0, ax1, ax2) = plt.subplots(ncols=3, figsize=(10, 3))

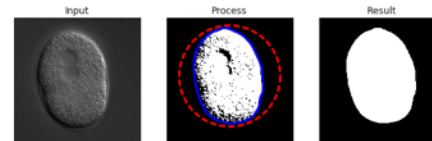
ax0.imshow(cv_img1, cmap=plt.cm.gray)
ax0.set_xticks([])
ax0.set_yticks([])
ax0.set_title("Input")
ax0.axis([0, cv_img1.shape[1], cv_img1.shape[0], 0])

ax1.imshow(cv_img3, cmap=plt.cm.gray)
ax1.plot(init[:, 0], init[:, 1], '-r', lw=3)
ax1.plot(snake[:, 0], snake[:, 1], '-b', lw=3)
ax1.set_xticks([])
ax1.set_yticks([])
ax1.set_title("Process")
ax1.axis([0, cv_img3.shape[1], cv_img3.shape[0], 0])

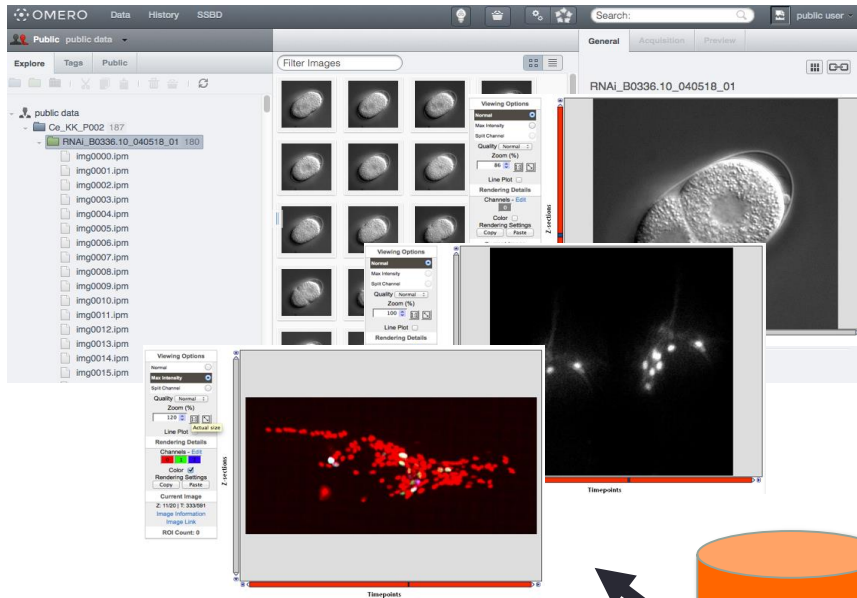
ax2.imshow(cv_img4, cmap=plt.cm.gray)
ax2.set_xticks([])
ax2.set_yticks([])
ax2.set_title("Result")
ax2.axis([0, cv_img4.shape[1], cv_img4.shape[0], 0])

plt.savefig('result3.tiff')

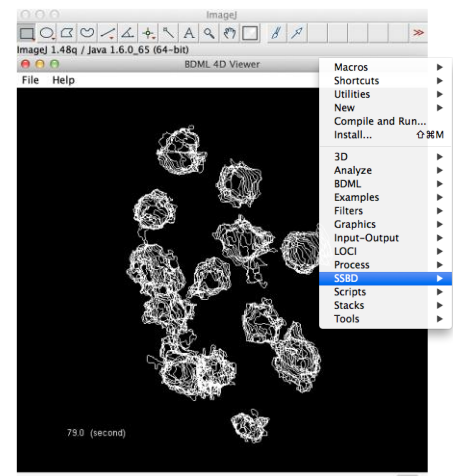
vec = np.array(contour, dtype=np.int).reshape((contour.shape[0], contour.shape[2]))
#print (vec)
np.savetxt('output.csv',vec,delimiter=',',fmt='%d',header="x,y")
```



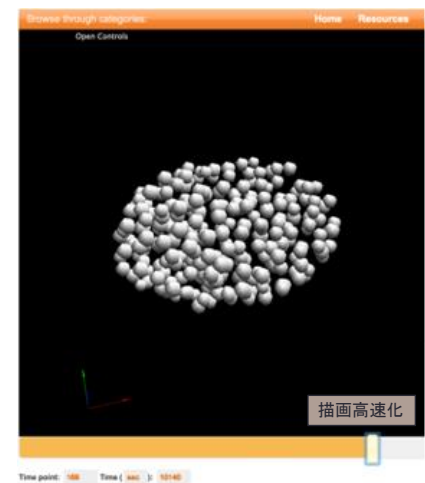
SSBD - Visualization



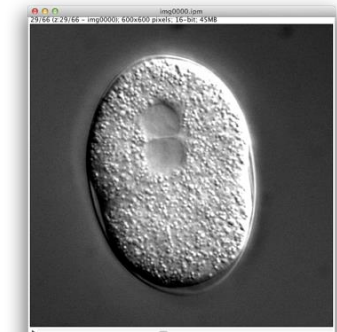
BDML対応ツール



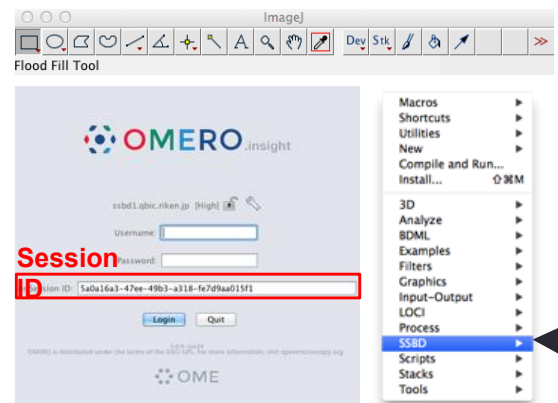
SSBD DBツール



ImageJプラグイン



JavaScriptアプリケーション



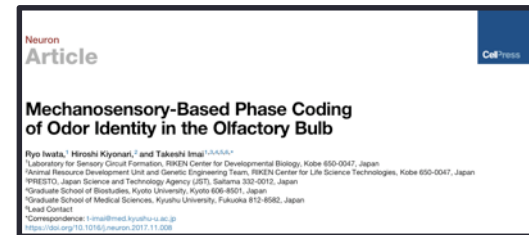
<http://ssbd.qbic.riken.jp/image/publicKey/>

SSBD repository service

- Submitting publications to journals with image and numerical data
- Reviewer has access to datasets before publications
- After publications, SSBD will include them in our release.



(Toyoshima et al. 2016)



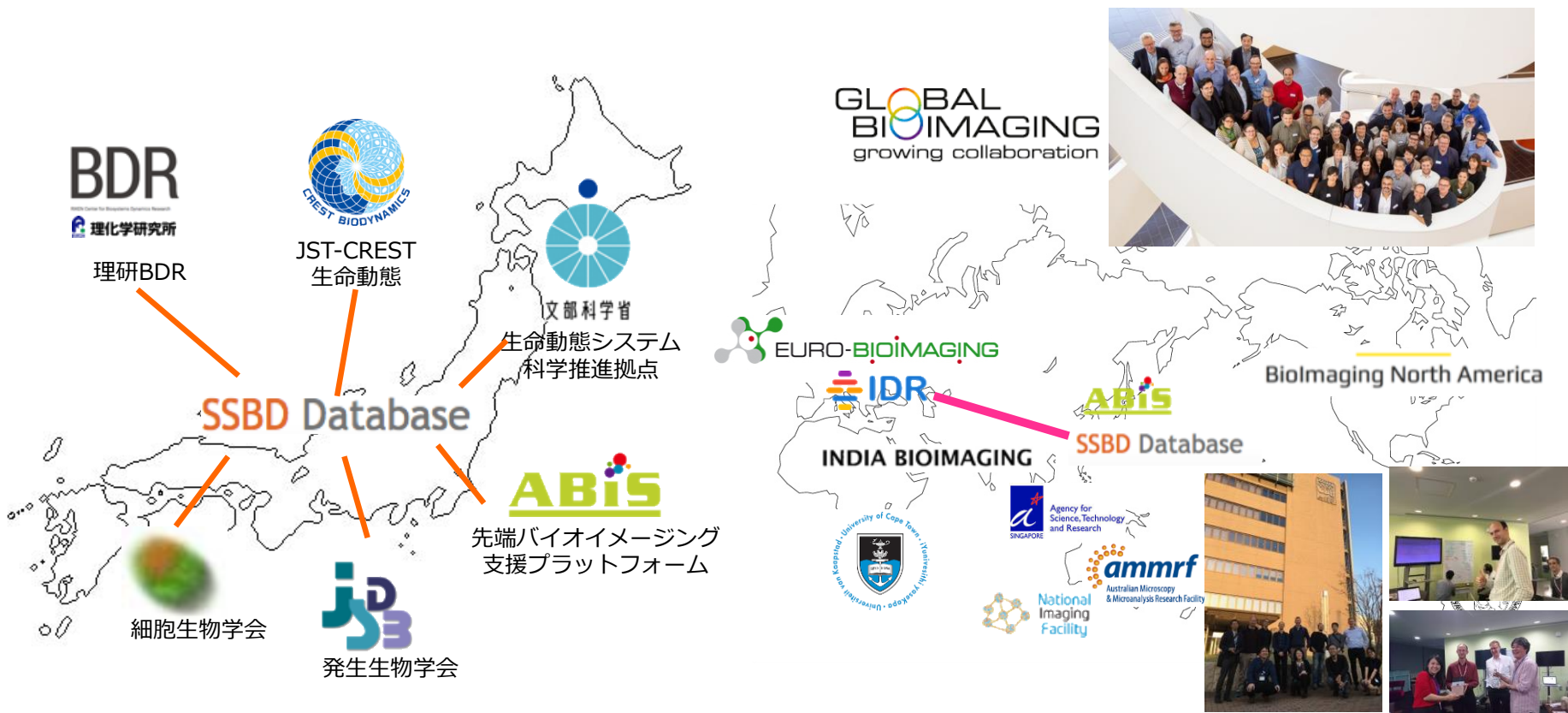
(Iwata et al. 2017)



(Azuma & Onami 2017)

Collaborations in Japan & Global partners

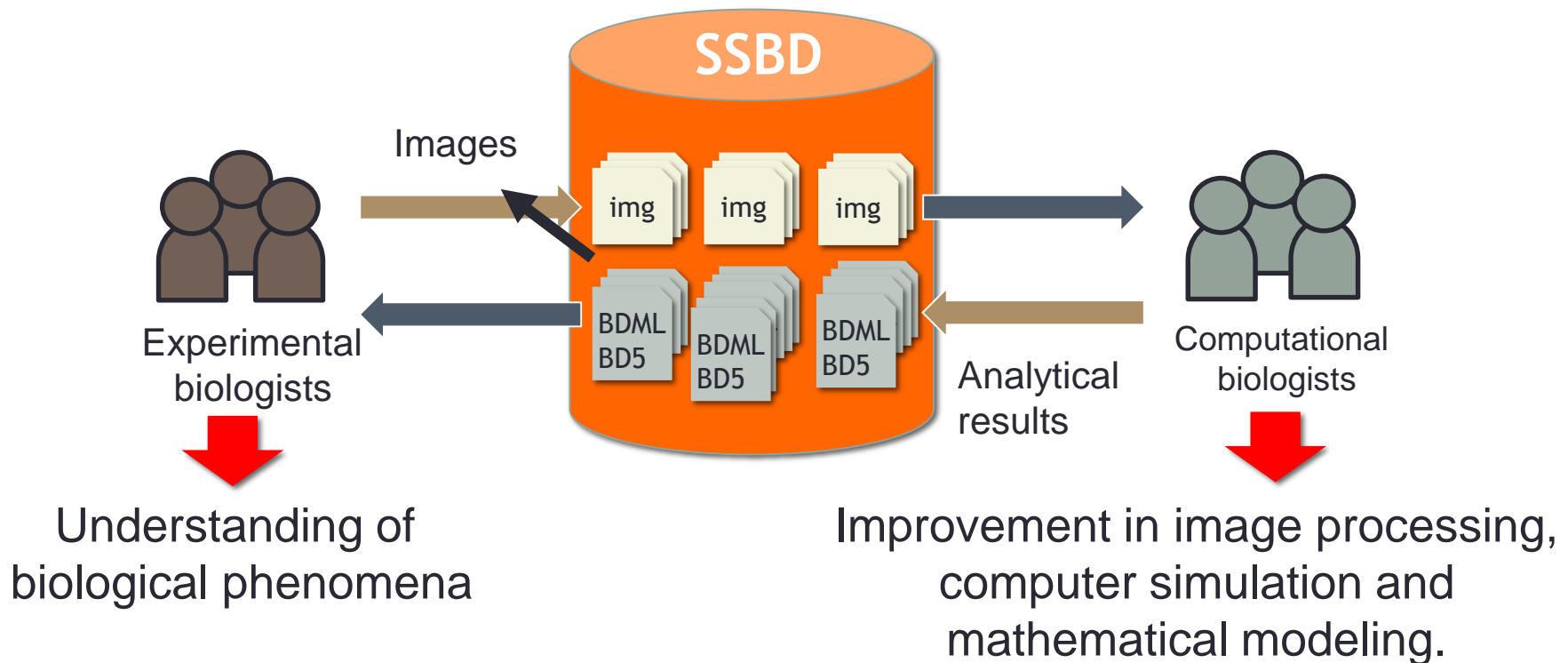
Working with various Japanese societies, institutes and research centers.



SSBD

Benefits of Open Science

- Public repository for sharing and reuse of data by biologists
- Open collaboration globally in developing computational tools/mathematical models



Acknowledgements

- **Project group**

Yuichi Iino (Univ. Tokyo)
Sigeo Ihara (Univ. Tokyo)
Hiroki Ueda (RIKEN)
Masahiro Ueda (RIKEN)
Ryoichiro Kageyama (Kyoto Univ.)
Kunihiko Kaneko (Univ. Tokyo)
Sinya Kuroda (Univ. Tokyo)
Minoru Ko (Keio Univ.)
Shigeru Kondo (Osaka Univ.)
Makoto Taiji (RIKEN)
Shinichi Tate (Hiroshima Univ.)
Michiyuki Matsuda (Kyoto Univ.)

- **Database development**

OMERO team
DBCLS
Hiroshi Masuya (RIKEN)
Norio Kobayashi (RIKEN)
Kai Lenz (RIKEN)

- **Onami Lab members (RIKEN)**

SSBD Database
Systems Science of Biological Dynamics



Email: ssbd@riken.jp

