BDML (Biological Dynamics Markup Language): an open format for representing quantitative biological dynamics data

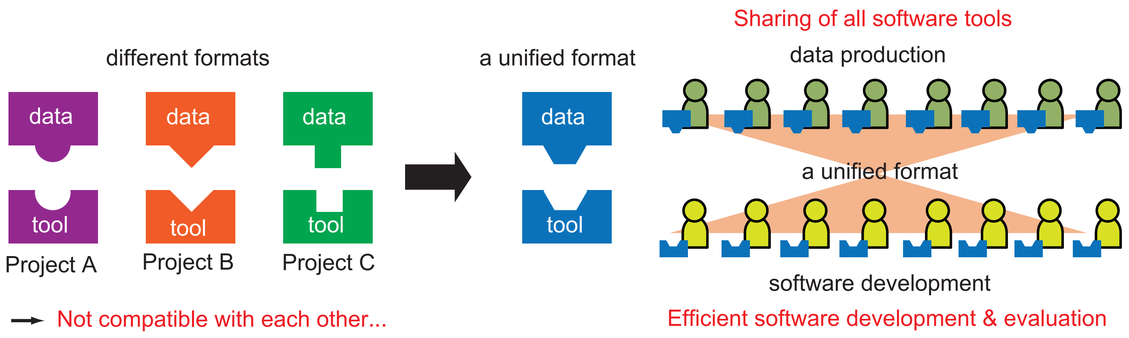
Koji Kyoda, Yukako Tohsato, Kenneth H.L. Ho, Shuichi Onami
Laboratory for Developmental Dynamics, RIKEN Quantitative Biology Center

Summary

We developed an open unified format, Biological Dynamics Markup Language (BDML), for representing quantitative biological dynamics data. BDML is based on eXtensible Markup Language (XML) whose advantages are machine/human-readability and extensibility. BDML will improve the efficiency of development and evaluation of software tools for data visualization and analysis.

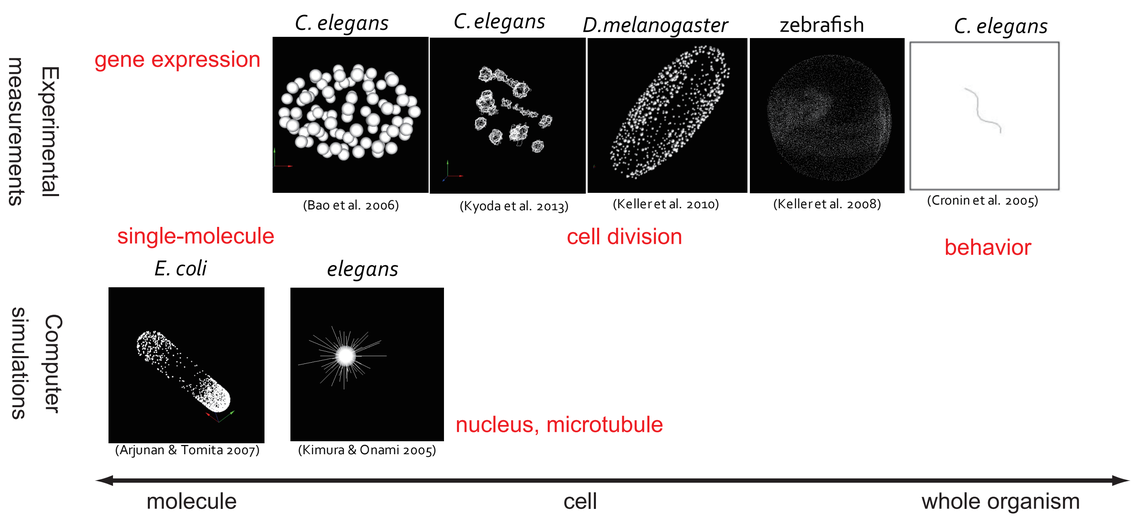
Objective

To create a unified format for representing quantitative biological dynamics data enabling efficient development of software tools and data exchange



Biological dynamics

Various types and scales of biological dynamics for different species can be supported.



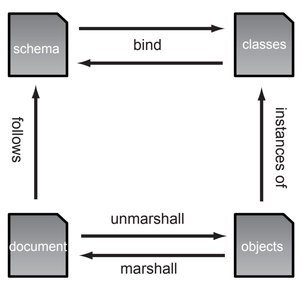
Methods

Biological Dynamics Markup Language (BDML) is based on XML.

Why XML?

Comparison XML with the other formats

	tab-delimited	XML	JSON	BSON	HDF5
hierarchical structure		○	○	○	○
readability	○	◎	○		
extensibility		◎	○	○	○
multi-language		○	○	○	○
web application		○	○	○	
server application		○	○	○	○
file size				○	○
numerical precision				○	○



machine-readable

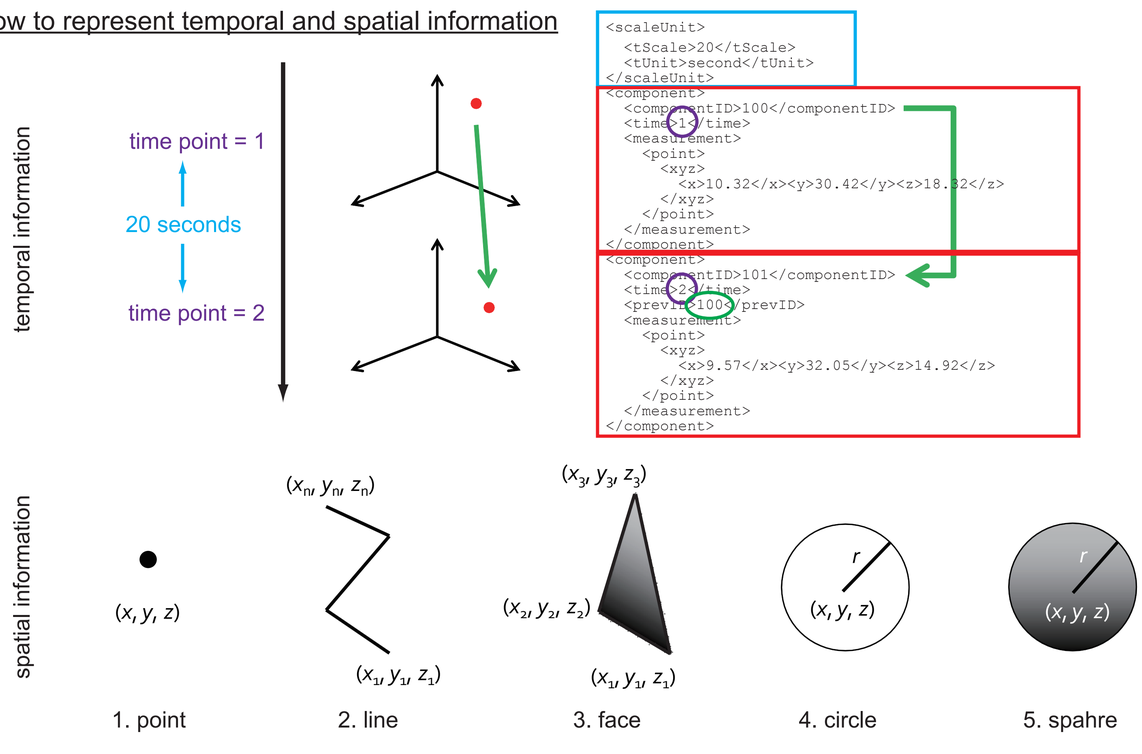
```
<?xml version="1.0" encoding="UTF-8"?>
<paper>
    <title>Biological Dynamics
                Markup Language</title>
    <author>
        <name>Koji Kyoda</name>
        <e-mail>kkyoda@riken.jp</e-mail>
    </author>
    ........
    <journal>
        <name>....</name>
        <volume>....</volume>
    </journal>
</paper>
```

human-readable

```
<?xml version="1.0" encoding="UTF-8"?>
<paper>
    <title>Biological Dynamics
                Markup Language</title>
    <author>
        <name>Koji Kyoda</name>
        <e-mail>kkyoda@riken.jp</e-mail>
        <address>Kobe 650-0047, Japan</address>
    </author>
    ........
    <journal>
        <name>....</name>
        <volume>....</volume>
        <page>...</page>
        <year>...</year>
    </journal>
</paper>
```

highly extensible

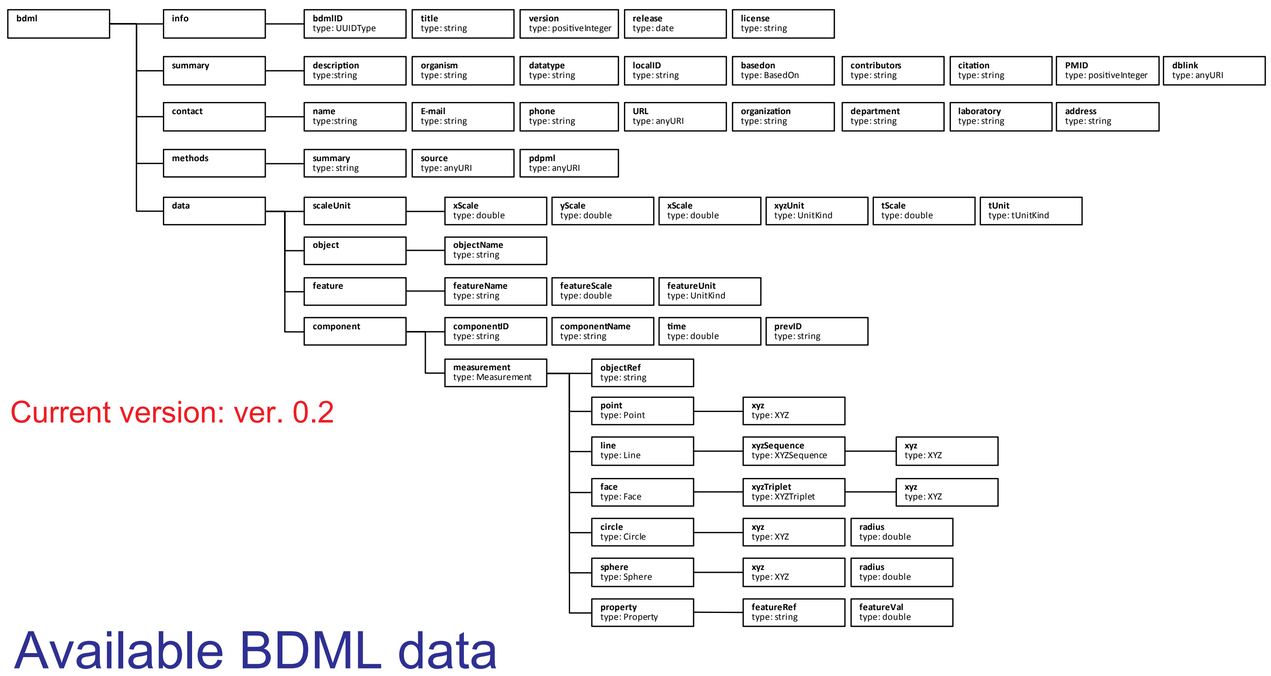
How to represent temporal and spatial information




```
<scaleUnit>
  <tScale>20</tScale>
  <tUnit>second</tUnit>
</scaleUnit>
<component>
  <componentID>100</componentID>
  <time>1</time>
  <measurement>
    <point>
      <xyz>
        <x>10.32</x><y>30.42</y><z>18.32</z>
      </xyz>
    </point>
  </measurement>
</component>
<component>
  <componentID>101</componentID>
  <time>2</time>
  <prevID>100</prevID>
  <measurement>
    <point>
      <xyz>
        <x>9.57</x><y>32.05</y><z>14.92</z>
      </xyz>
    </point>
  </measurement>
</component>
```

BDML structure

A BDML file can contain all the quantitative data and associate meta-information in a single file.

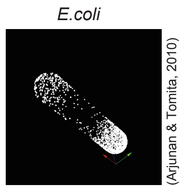


Current version: ver. 0.2

Available BDML data

Nearly 300 BDML data sets are currently available online at the SSBD database (http://ssbd.qbic.riken.jp).

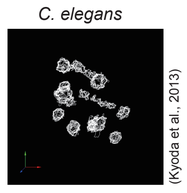
single-molecule
E.coli



(Arjunan & Tomita, 2010)

```
<measurement>
  <objectRef>molecule</objectRef>
  <point>
    <xyz>
      <x>5.06e-06</x>
      <y>5.44e-07</y>
      <z>1.04e-06</z>
    </xyz>
  </point>
</measurement>
```

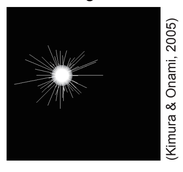
cell division
C. elegans



(Kyoda et al., 2013)

```
<measurement>
  <objectRef>nucleus</objectRef>
  <line>
    <xyzSequence>
      <xyz>
        <x>402</x>
        <y>436</y>
        <z>21</z>
      </xyz>
      <xyz>
        <x>403</x>
        <y>437</y>
        <z>21</z>
      </xyz>
      ...
    </xyzSequence>
  </line>
</measurement>
```

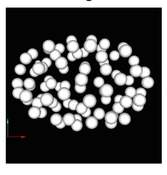
nuclear migration
C. elegans



(Kimura & Onami, 2005)

```
<component>
  <componentName>male</componentName>
  <measurement>
    <objectRef>pronucleus</objectRef>
    <sphere>
      <xyz>
        <x>1.97e-05</x>
        <y>3.94e-07</y>
        <z>-3.16e-07</z>
      </xyz>
      <radius>5e-06</radius>
    </sphere>
  </measurement>
  <measurement>
    <objectRef>microtubules</objectRef>
    <line>
      <xyzSequence>
        <xyz>
          <x>1.97e-05</x>
          <y>3.94e-07</y>
          <z>9.21e-06</z>
        </xyz>
        ...
      </xyzSequence>
    </line>
    ...
  </measurement>
<component>
```

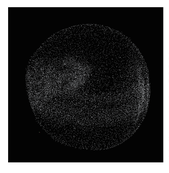
cell division & gene expression
C. elegans



(Bao et al., 2006)

```
<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>
```

cell division
zebrafish



(Keller et al., 2008)

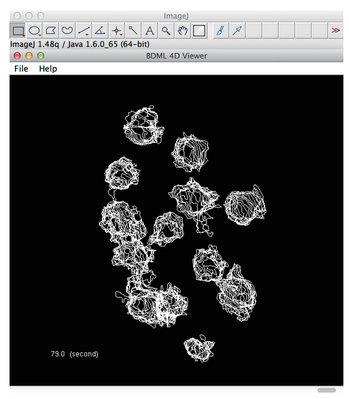
```
<measurement>
  <objectRef>nucleus</objectRef>
  <point>
    <xyz>
      <x>927.74</x>
      <y>624.693</y>
      <z>30.0409</z>
    </xyz>
  </point>
  <property>
    <featureRef>average ...</featureRef>
    <featureVal>925</featureVal>
    ....
  </property>
</measurement>
```

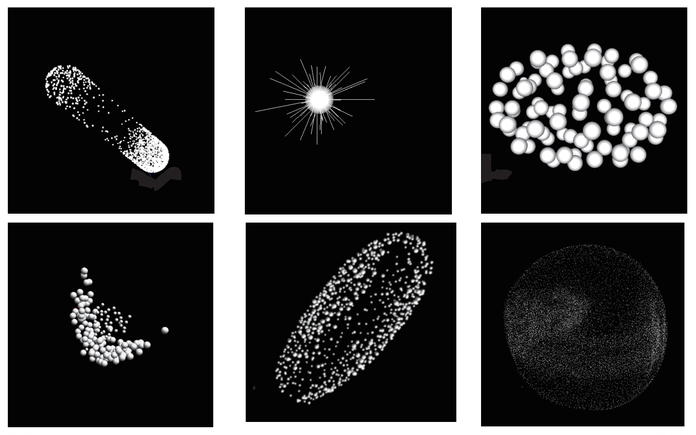
Development of software tools

Anyone can easily develop software tools for using the BDML format through open source libraries.

Visualization tool

BDML4DViewer is a visualization tool as a plugin for ImageJ.

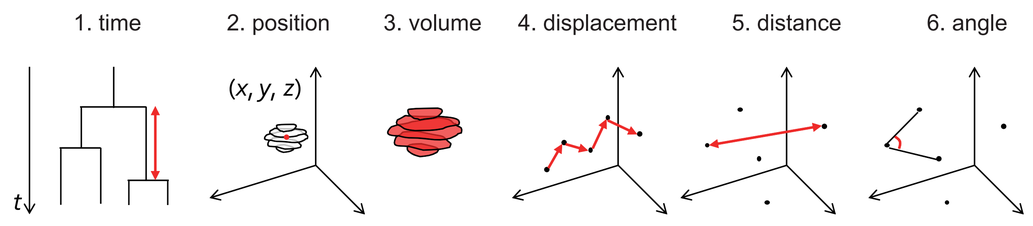


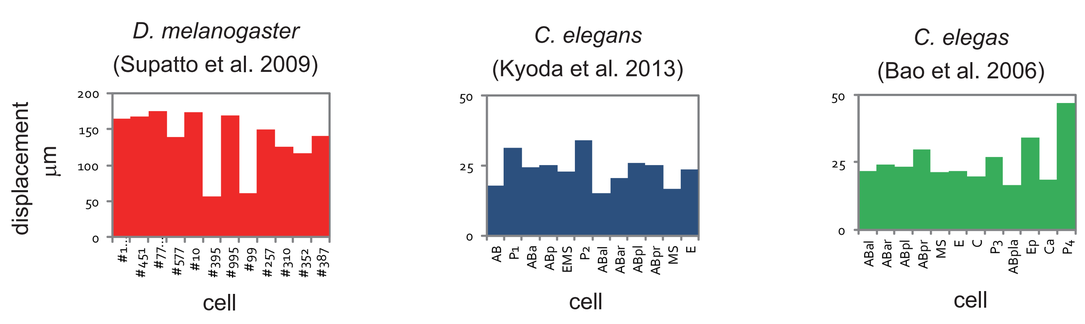


The *BDML4DViewer* can visualize any quantitative data described in BDML.

Analytical tool

phenochar is a analytical tool for extracting phenotypic characteristics from the BDML data.





The *phenochar* can be applicable to any quantitative data described in BDML.

Future plans

Extension of BDML format
* Supporting other basic types of spatial entities such as cube, cylinder
* Supporting binary formatted data (HDF5 data)

Supporting Omics data
* Extending BDML to represent spatiotemporal Omics data

Development of software tools for data analysis

Kyoda *et al*. (2015) Biological Dynamics Markup Language (BDML): an open format for representing quantitative biological dynamics data *Bioinformatics* **31**, 1044-1052.