The German Neuroinformatics Node -
Database tools for data management and collaboration
in neurophysiology

Thomas Wachtler, Philipp Rautenberg, Andrey Sobolev, Andreas Herz

German Neuroinformatics Node
Ludwig-Maximilians-Universität München
Science, Feb 11, 2011
International Neuroinformatics Coordinating Facility (INCF)

OECD Global Science Forum

Recommendations for international development of neuroinformatics

→ International Neuroinformatics Coordination Facility (INCF)
  - established 2005
  - INCF Secretariat in Stockholm
  - National Nodes

www.incf.org
INCF Activities: Programs

INCF Programs address important scientific issues in specific areas of neuroinformatics. In these programs, INCF initiates and coordinates international groups of scientists to develop solutions in the form of standards, guidelines, products or services.

Current Programs:

- Digital Brain Atlasing
- Ontologies of Neural Structures
- Multi-Scale Modeling
- Standards for Data Sharing
INCF Programs (I)

Program on Digital Brain Atlasing

Coordinate atlasing projects (rodent)

Standardization: Waxholm Space

Establish Digital Atlasing Infrastructure
INCF Programs (II)

Program on Ontologies of Neural Structures

Establish platform for translation and clarification of terminologies
Integration of other ontology projects
Linked to Digital Atlassing program

Developments:

- **Structural Lexicon**
  Consistent definition of structural terms

- Neuronal Naming
  **Neuron registry** and convention for naming
INCF Programs (III)

Program on Multi-Scale Modeling

Promote interoperability and sharing of model development and simulation tools

- Standards and Guidelines for Large-Scale Modeling:
  Language standard for spiking neuron models (nineML)

- MUSIC Multi-Simulation Coordinator
  Standardized interface for communication between neural network simulators

- Blue Gene/L Supercomputer Access
INCF Programs (IV)

Program on Standards for Data Sharing

Development of methods for automated metadata acquisition and standards for interoperability between data management systems

Focuses on
- imaging data
- electrophysiology
INCF National Nodes

Belgium
Czech Republic
Finland
France
Germany
India
Italy
Japan
Netherlands
Norway
Poland
South Korea
Sweden
Switzerland
United Kingdom
United States

http://www.incf.org/about/organization/nodes
J-Node: Japanese Node of the INCF

Each platform has been developed to organize neuroinformatics databases, by which research results can be shared with the public.

**Visiome Platform**
- [http://visiome.neuroinf.jp/](http://visiome.neuroinf.jp/)
Visiome Platform (VP) integrates research papers, experimental data, mathematical models and tools, and other information associated with research on vision (i.e. the retina, psychophysics, and higher-level visual functions) and makes this data publicly available.

**Brain Machine Interface Platform**
- [http://bmi.neuroinf.jp/](http://bmi.neuroinf.jp/)
Brain Machine Interface Platform (BMI PF) accumulates BMI (Brain Machine Interfaces)-related experimental data, mathematical models, and tools generated in neuroscience, computational theory, and robotics.

**Invertebrate Brain Platform**
- [http://invbrain.neuroinf.jp/](http://invbrain.neuroinf.jp/)
Invertebrate Brain Platform (IVB PF) integrates experimental data, mathematical models, and research tools relevant to the study of invertebrate brains, neurons, and behavior.

**Integrative Brain Research**
- [http://www.toge.neu.nips.ac.jp/](http://www.toge.neu.nips.ac.jp/)
Integrative Brain Research (IBR) Project is a grant group of neuroscientists funded by the Ministry of Education, Culture, Sports,

J-Node: Japanese Node of the INCF

The structure and function of the J-Node

*Some platforms are under development and are not released yet.
G-Node: German Node of the INCF

www.g-node.org

- Integral part of the National Network Computational Neuroscience
- Focus on cellular and systems neurophysiology
- Development of tools and infrastructure for data management and analysis, to facilitate collaboration and scientific progress
- Teaching & Training in neuroinformatics
Teaching & Training

Training in neuroinformatics for students and postdocs

• Courses on advanced data analysis methods
  next:
    3rd G-Node Winter Course on Neural Data Analysis
    March 7-11, 2011, Munich, Germany

• Courses on scientific programming techniques
  next:
    Advanced Scientific Programming in Python
    Sep 11-16, St Andrews, UK
    http://python.g-node.org

• Online teaching material
Services

Infrastructural and development support of neuroinformatics resources for cellular and systems neuroscience

Example: New version of CoCoMac database

(in collaboration with Rembrandt Bakker, Nijmegen and Markus Diesmann, Jülich)

- Web based access for automated queries by other applications
- Developing advanced user interface for community-driven development
Development of Tools for Efficient Data Management

Why do we care about data management?

- Progress in neuroscience increasingly depends on collaborative efforts, exchange of data, re-analysis of data
- Advances in technology and methodology dramatically increase volume and complexity of the data recorded
- Typically, each lab develops methods for managing and analyzing the data.
  - Duplicate work
  - Collaborative work or re-use of data is hampered by the effort it takes to understand how to access the data
Development of Tools for Efficient Data Management

*What is needed?*

- Tools for organizing and exchanging data that can be used by different laboratories
  - save time and resources
- Standardized methods to specify the metadata that are necessary to access and analyze data
  - ensure future access to data and facilitate collaborative work
- Integrated organization of data and metadata to facilitate data management and data analysis
1 - **Web-based Platform for Data Management and Data Sharing**

(Django Web Framework, Python, postgresQL Database)

Provides secure, web-based access to your data:

- **Data Storage and Retrieval**
- **Data Organization**
  - Projects - Experiments - Datasets
- **Data Sharing**
  - Private - Shared - Public

http://portal.g-node.org/data
2 - Method for Metadata Specification and Exchange

- **Format**: metadata properties are described in a hierarchical structure of key-value pairs, e.g., in XML

- **Content**: recommended terminologies

  http://www.g-node.org/odml

Grewe et al, submitted
odML: Representing The Structure of Metadata

- format is simple, flexible, inherently extensible
  ➔ can be adapted to the specifics of the lab or experiment

- can carry any metadata
  ➔ no information is lost

- enables automated collection of metadata starting in the laboratory

http://www.g-node.org/odml
odML: Terminologies for Neurophysiology

Terminologies (available at www.g-node.org/odml)

<table>
<thead>
<tr>
<th>name</th>
<th>type</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>analysis/generic</td>
<td>Descriptions of an analysis.</td>
</tr>
<tr>
<td>PSTH</td>
<td>analysis/psth</td>
<td>Properties to describe a peri stimulus time histogram.</td>
</tr>
<tr>
<td>PowerSpectrum</td>
<td>analysis/powerspectrum</td>
<td>Properties to describe a power spectrum.</td>
</tr>
</tbody>
</table>

Stimulus Terminology

name: Stimulus
type: stimulus/generic
baseURL: http://g-node.org/odml/terminologies/v1.0/
definition: Section to describe a generic stimulus. This section is basis of various related sections that specify more specific stimuli.

Properties:

<table>
<thead>
<tr>
<th>name</th>
<th>type</th>
<th>unit</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>string</td>
<td></td>
<td>The author of this stimulus.</td>
</tr>
<tr>
<td>Description</td>
<td>string</td>
<td></td>
<td>A textual description of the stimulus.</td>
</tr>
<tr>
<td>Duration</td>
<td>float</td>
<td>s</td>
<td>The duration of the stimulus in seconds.</td>
</tr>
<tr>
<td>StartTime</td>
<td>time</td>
<td></td>
<td>The time the stimulus started.</td>
</tr>
<tr>
<td>EndTime</td>
<td>time</td>
<td></td>
<td>The time the stimulus ended.</td>
</tr>
<tr>
<td>InterstimulusInterval</td>
<td>float</td>
<td>s</td>
<td>The time between successive stimulus presentations.</td>
</tr>
<tr>
<td>Repetitions</td>
<td>int</td>
<td></td>
<td>The number of stimulus presentations.</td>
</tr>
<tr>
<td>Intensity</td>
<td>float</td>
<td></td>
<td>The intensity of the stimulus.</td>
</tr>
<tr>
<td>Modality</td>
<td>string</td>
<td></td>
<td>The stimulus modality, e.g. visual, auditory, etc.</td>
</tr>
<tr>
<td>OutputChannel</td>
<td>string</td>
<td></td>
<td>The physical output device (e.g., an analog output channel, a monitor screen, a loudspeaker, etc.). This information can be used to explicitly express that several stimuli (that share the same OutputChannel) are jointly presented.</td>
</tr>
<tr>
<td>StimulusFile</td>
<td>binary</td>
<td></td>
<td>The stimulus file used. Including it will increase the size of the metadata file and we recommend to rather point to a file using an URL.</td>
</tr>
<tr>
<td>StimulusFileURL</td>
<td>URL</td>
<td></td>
<td>The URL of the applied stimulus file.</td>
</tr>
<tr>
<td>DC</td>
<td></td>
<td></td>
<td>MyStimulus - [stimulus/generic]</td>
</tr>
<tr>
<td>WhiteNoise</td>
<td></td>
<td></td>
<td>Duration = 2.5 s</td>
</tr>
<tr>
<td>Sinewave</td>
<td></td>
<td></td>
<td>Duration = 1.0 s</td>
</tr>
</tbody>
</table>

500 ms
odML: Automated Collection of Metadata

via libraries provided (Java, C++, Python), odML can be integrated into data acquisition, stimulation, and analysis tools, e.g.:

LabLogBook (Jan Grewe, Munich)  
http://lablog.sourceforge.net

Relacs (Jan Benda, Munich)  
http://www.relacs.net
3 - Development of Integrated Database System for Data Management

How to organize your data?

- The answer depends on how you want to control your data -
Development of Integrated Database System for Data Management

How to organize your data?

Organization of data and metadata should represent the structure of the experiment
Development of Integrated Database System for Data Management

Goal: Database schema for neurophysiology experiments

What is the benefit?
Development of Integrated Database System for Data Management

Don't search - ask!

"Give me all spike trains where

• $age=5$ AND
• stimulus color=$red$ ... AND ... 
• ... OR ..."

Integrated organization of data and metadata in database enables **efficient access and automated selection** of data for analysis.
Development of Integrated Database System for Data Management

Interfaces to various applications, languages and formats enable **automated selection of data** and **integration with data analysis** workflow.
Development of Integrated Database System for Data Management

Extension: Integration of morphological data and simulation tools

see Poster: Rautenberg et al
Summary

- Key to efficient data management is the unified access to data and metadata
- Using database technology, we develop methods for neurophysiology data management that facilitate data access and data exchange.
- Interfaces to common software tools and programming languages will enable automated data selection and integration in the data analysis work flow
Acknowledgments