datasharing

INCF Program on Standards for Data Sharing

New perspectives on workflows and data management for the analysis of electrophysiological data

December 4-5, 2013
Jülich, Germany

incf

Table of Contents

Introduction 3

Workflows in electrophysiology 3

Existing approaches discussed in the workshop 3

Data structures and software libraries

Documentation and provenance

Workflow management systems

Supercomputing

INCF activities in this domain

Discussion 6

Table 1 8

Table 2 8

Table 3 9

Recommendations 9

References 10

Introduction

Breakthroughs in recording technologies and analysis approaches have led to unprecedented levels of complexity in electrophysiological experiments. Most notably, the ability to perform massively parallel recordings from hundreds of neurons in the brain paired with a strong interest in complex, natural stimulation and behavioral paradigms, leads to a surge of intricately interwoven incoming data that should be analyzed with sophisticated methods to uncover the network dynamics by exploiting the parallel aspect of the data.
The combination of these factors leads to highly complex projects that pose a challenge for researchers who, over the course of years, are confronted with planning of the analysis, organization of workflows in larger teams, programming of software, and bookkeeping of the results obtained by constantly evolving analysis methods.

The complexity of electrophysiological research has reached a level where a professionalization of the data analysis workflow, both conceptually and in terms of supporting software infrastructures, has become a necessity. Thus, the INCF Program on Standards for Data Sharing convened a 2-day workshop to provide a discussion platform for researchers and developers who work on components that are instrumental in such workflows. The presentations and discussions of the workshop focused on three major issues related to workflows in electrophysiology: (i) the development of data structures and software libraries that enable interfacing of data from various sources and integration of methods for data manipulation and analysis, (ii) documentation and provenance tracking solutions to support reproducible analysis processes able to cope with the required levels of flexibility and data size, and (iii) workflow management systems that allow automatic and extensible processing of complex analysis workflows, in particular on high performance computers.

During the course of the workshop, the participants discussed the evolution of workflows in electrophysiology as well as existing approaches for data structures/software libraries, documentation/provenance, workflow management systems, and supercomputing to identify the challenges to workflows and data management in electrophysiology. At the conclusion of the workshop, the participants devised a set of recommendations for what INCF should do to improve the workflow/data management situation based on the key challenges identified and INCF's current activities within this domain.

Workflows in electrophysiology

In the typical electrophysiological workflow of the past, data was recorded or simulated, stored, pre-processed, and analyzed. Once the analysis was complete, the results of the experiment were published in a paper. Today, breakthroughs in recording technologies and approaches have increased the complexity of typical electrophysiological workflows. Modern electrophysiological experiments are large in size (ranging from hundreds of electrodes in vivo up to thousands in vitro, e.g., Maccione et al (2014)), contain various data types, complex stimuli/behavior (e.g. Riehle et al (2013), Hatsopoulos and Suminski (2011)), complex preprocessing of the data, like spike sorting (e.g., Quiroga et al (2004), Lewicki (1998), Harris et al (2000)) , and even more complex analysis paradigms (e.g. Denker et al (2011), Kim et al (2011), Torre et al (2013)), and, unlike the typical electrophysiological experiments of the past, involve collaborative effort. With this increase in complexity and collaboration, the need for data/results management, increased computational load, linking analysis, and reproducibility, as well as code verification, methods validation, and workflows/collaboration has become more apparent (Denker et al (2014)). Fortunately, there are potential solutions to the challenges faced in modern electrophysiology. Standards for data and metadata, as well as ontologies, databases, and I/Os should reduce the challenges of increased experimental complexity by providing standards and methods to manage data and results. Code sharing and provenance tracking techniques should aid code verification and reproducibility, respectively, while workflow systems can be used to validate methods. Improved interoperability between different software and data types should help to better link analyses, and the use of parallelization techniques should reduce the barriers of computational load.

Existing approaches discussed in the workshop

Data structures and software libraries

odML (open metaData Markup Language, g-node.org/projects/odml) is a format for collecting and exchanging metadata in an automated, machine-readable format (Grewe et al 2011). Metadata is stored as extended key-value pairs (properties) in a hierarchical structure, which can be grouped into sections and subsections. In odML, only the format of metadata is defined,

not the content; so it is extensible and can be easily adapted to the specific needs of the researcher (Zehl et al (in prep)). To offer the possibility to use common standard key terms, an odML terminology has been defined and can be extended (g-node.org/projects/odml/terminologies). This terminology contains a small number of key concepts in electrophysiology, including terms defined by the CARMEN initiative (http://www.carmen.org.uk/). Mapping customized terminologies to the standard terminologies allows users to name and organize metadata as preferred without impacting the validity of the format. However, the current format of the terminology presents some limitations. To address these limitations, the Ontology of Experimental Neurophysiology has been created based on the odML terminology (Bruha et al., 2013; Le Franc et al., 2014).
The G-Node data sharing platform (g-node.org/gndata) integrates data and metadata by combining the Neo data model for electrophysiological data with the odML approach for metadata organization (Sobolev et al 2014).

A **mobile app for odML** is being developed to facilitate metadata capture in electrophysiological experiments. Users have the ability to create entry forms, annotate data, and export the information in odML format for easy integration with metadata acquired from other sources.

Spyke Viewer is an open source GUI application used to browse, visualize, and analyze data from electrophysiological experiments and neural simulations (Pröpper et al 2014). It is built on top of the Neo library, a Python implementation of an object model for electrophysiological data (Garcia et al 2014). Like Neo, Spyke Viewer focuses primarily on local field potential and spike data (there is also basic support for EEG data). With Spyke Viewer, users can load any file format supported by Neo and browse its contents. Analysis plug-ins provide data visualization and analysis capabilities. Data stored in non-standard formats can also be analyzed in Spyke Viewer through I/O plug-ins. Users can quickly implement custom analysis and IO plug-ins using Python. The data analyzed in Spyke Viewer can easily be shared with users of other tools since it exports files in multiple standard formats using Neo.

CRCNS.org (Collaborative Research in Computational Neuroscience, crcns.org) is a website that was established to promote data sharing in computational neuroscience (Teeters et al 2008). It provides a marketplace and discussion forum for sharing tools and neuroscience data (physiological recordings from sensory and memory systems, as well as eye movement data). There are forums for each dataset specifically, as well as a general discussion forum. For each dataset, there is a summary of the data (what it is, what procedures were performed), information about the data format, and links to both download the data and the data's discussion forum. The website also contains community resources such as information about courses, hosted projects, and a list of publications that used the data available via crcns.org.

SEEK4Science (seek4science.org). The SEEK platform is a web-based resource for sharing heterogeneous scientific research datasets, models or simulations, processes and research outcomes. It preserves associations between them, along with information about the people and organisations involved. Underpinning SEEK is the ISA infrastructure, a standard format for describing how individual experiments are aggregated into wider studies and investigations. Within SEEK, ISA has been extended and is configurable to allow the structure to be used outside of Biology.
SEEK is incorporating semantic technology allowing sophisticated queries over the data, yet without getting in the way of users. SEEK is the platform underpinning two major ERANets, and the Virtual Liver Network. It is now funded for 5 more years under ISBE and ERASysAPP.

Documentation and provenance

Provenance W3C standard (w3.org/TR/prov-overview) was conceived as a method to enable the wide publication and interchange of provenance on the Web and other information systems. The model enables the representation and interchange of provenance information using widely available formats such as RDF (Resource Description Framework, a family of W3C, World Wide Web Consortium, specifications originally designed as a metadata data model), and XML (Extensible Markup Language, a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable) and provides definitions for accessing provenance information, validating it, and mapping to Dublin Core Metadata Initiative vocabulary. The core of the standard is a conceptual data model that defines a common vocabulary for describing provenance which is instantiated by various serializations that are used by implementations to interchange provenance.

Sumatra (neuralensemble.org/Sumatra) is a tool developed to provide automated capture of all information required to reproduce computational experiments (simulation/analysis) for arbitrary command-line launched programs/scripts or GUI (Graphical User Interface)-based tools that incorporate the Sumatra library. Sumatra functions like an electronic laboratory notebook in that all the information about the simulation/analysis is stored in a record store that can be browsed, queried, and annotated using either a command-line or web-based interface. In addition, Sumatra permits multiple users and projects to use a common record store. Sumatra also has a web-service record store, thus allowing any client with web-access to use the record store.

LabLog (lablog.sourceforge.net) is an open-source tool used to document neuroscience projects and to store and manage metadata in the odML framework. It is intended to assist scientists in keeping track of recorded data and storing project related information.

Other provenance collection tools are now available: see ProvWeek. In particular in PROV format.

Research Objects (researchobject.org) is a framework and set of standards for structuring and exchanging reproducible results. Pioneered in the wf4ever EU project for the reproducibility of workflows, it is now extended to packaging interlinked components of research. Work is ongoing to integrate nipypes with research objects. ROs use community standards for dependencies, provenance, checklists, versioning, aggregation, annotation and identity.

The open source ISA metadata tracking tools (isa-tools.org) help to manage an increasingly diverse set of life science, environmental and biomedical experiments that employing one or a combination of technologies. Built around the 'Investigation' (the project context), 'Study' (a unit of research) and 'Assay' (analytical measurement) general-purpose Tabular format, the ISA tools helps you to provide rich description of the experimental metadata (i.e. sample characteristics, technology and measurement types, sample-to-data relationships) so that the resulting data and discoveries are reproducible and reusable. ISA is the heart of the NPG Scientific Data initiative.

Workflow management systems

CARMEN (Code Analysis, Repository, and Modelling for E-Neuroscience, carmen.org.uk) is an e-science project funded by the UK Engineering and Physical Sciences Research Council (EP/E002331/1) and the Biotechnology and Biological Sciences Research Council (BB/I000984/1). It provides a virtual laboratory for electrophysiology that enables sharing and collaborative exploration of data, analysis code, and expertise with a primary focus on neural activity recordings (signals and image series). The virtual laboratory is a federation of server nodes that allows distributed data to be stored local to acquisition where analysis code can be uploaded and executed, thus eliminating the need to transport derived datasets over low bandwidth connections. Data and analysis codes are described by structured metadata, thus providing an index for search and audit of workflows. In addition to the virtual laboratory, CARMEN has also produced several standards: (i) MINI (Minimum Information about a Neuroscience Investigation): minimum reporting guidelines for the annotation of data and other resources and (ii) NDTF (Neurophysiology Data Translation Format): provides a vendor-independent mechanism for translating between raw and processed electrophysiological data in the form of time and image series. The CARMEN website also provides users with links to electrophysiology and data sharing resources, software, and an events calendar.

Nipype (Neuroimaging in Python Pipelines and Interfaces, nipy.org/nipype) is an open source Python initiative that provides a uniform interface to existing neuroimaging software and facilitates interaction between these packages (e.g. SPM , FSL , FreeSurfer , Camino , MRtrix , MNE , AFNI , and Slicer) within a single workflow. Nipype enables users to both interact with tools and combine processing steps from different software packages. It also allows users to develop new workflows and process data faster, as well as enable users to share their workflows with the community.

Taverna Workflow Management System (taverna.org.uk) is an open source software tool for designing, executing, and sharing workflows that was developed by the myGrid Consortium that allows users to integrate many different software components, including SOAP or REST web services, Grid services (including UNICORE), scripting platforms, command tools as well as offer users the ability to import new service descriptions and new platform/infrastructure plug-ins. A Suite of tools include: a Taverna Workbench and a Taverna Online web based tool for authoring workflows; Taverna Server and command line that allows Taverna workflows to be run on other machines, computational grids, clouds, web pages, and portals, as well as online workflow designer and enact. So Taverna workflows do not need to be executed within the Taverna Workbench. The Taverna Player is an embeddable widget to run workflows in situ of applications such as iPython Notebook and web pages.

Taverna Workbench includes the ability to monitor workflows and examine provenance in PROV format. Taverna is used in a wide variety of disciplines, such as bioinformatics, cheminformatics, biodiversity, virtual physiology, digital preservation and social science or. Taverna is used as the platform of the VPH-SHARE and VPH-DARE projects.

Taverna workflows are deposited, along with 21 other workflow systems, in a cloud repository for workflow exchange – myExperiment.org. This platform is being reworked to be RO compliant.

The Human Brain Project (HBP) Platform Solution was developed to solve the problems of installation, data sharing, and traceability (a sub-set of the reproducibility problem). The web-based, HBP platform has been designed to enable analysis sharing, data sharing, rich viewers (all web-based), some interactive capabilities, and provenance tracking with a focus on computational and biological neuroscience. The platform's task format is a Python wrapped software component that has a default GUI presented in the user interface. Tasks are registered in a task repository with all software dependencies. The task format has declared inputs/outputs and is executed as a job using a REST API, and provenance is automatically tracked for each job. Tasks are registered with an automatically generated GUI or the programmatic service API client. The platform's workflow uses a simple API that is computationally and source code friendly. Workflow provenance information is provided by the task format, and analysis metadata is included in the file.

Supercomputing

PyCOMPS is a programming model for running parallel computations in Python that aims to simplify the development of workflows and execution in distributed computing environments. It is composed of 2 components: (i) a programming mode that is implicitly parallel and infrastructure agnostic and (ii) a Java-based runtime system that manages the parallelization on the level of the existing hardware and software infrastructure.

UNICORE (unicore.eu) is a general purpose, open source, European Grid Technology that follows the latest standards from Grid Technologies and Web Services. UNICORE provides users with an intuitive graphical interface and command line client to access high-performance computing resources and execute user-defined workflows. It provides application integration mechanisms on the client services and resource level, and tightly integrated workflows which can be modified for different, domain-specific workflow languages, as well as interoperability through common open standards. UNICORE (Uniform Interface to Computing Resources) offers a ready-to-run Grid system including client and server software. UNICORE makes distributed computing and data resources available in a seamless and secure way in intranets and the internet.

INCF activities in this domain

- A standard format for storing electrophysiology data and related metadata that aims to enable the efficient sharing of this type of data is being developed. It is based on HDF5 (Hierarchical Data Format, a set of file formats and libraries designed to store and organize large amounts of numerical data) and specifies what needs to be stored for commonly used electrophysiology data types. In addition to the physiological raw measurements, the standard includes metadata required for interpretation of the raw data. A full document specifying the requirements for the standard has been formulated and is currently being prepared for broad dissemination.
- Ontology for Experimental Neurophysiology (OEN) is intended to make the description of electrophysiology data coherent and searchable. OEN is developed in collaboration with researchers from the German INCF Node, NIF , neuroelectro.org, and the task force. The development of the OEN focuses on two main ontology branches: neurophysiological concepts and concepts for devices and methods. The neurophysiological terms describe experimental data features (spike properties, etc.). These terms are included in NeuroLex and currently they are derived mostly from neuroelectro.org (neurolex.org/wiki/Category:Action_potential_characteristic). The device and method concepts describe experimental setups and methods, and are based on different metadata terminologies (odML (g-node.org/odml), MINI (carmen.org.uk/standards), Helmholtz (doi: 10.3389/conf.fninf.2013.09.00025). Since the OEN device branch has some overlap with the Ontology for Biomedical Investigation (OBI), it is developed as an extension of the OBI ontology in tight collaboration with the OBI developer community. The device/method branch is available as an OWL file, which contains the full OBI ontology (github.com/G-Node/OEN/tree/gh-pages). Version 1.0 of OEN will be released at the end of 2014.
- Neuroimaging Informatics Data Model (NI-DM), developed by the task force and the BIRN Derived Data Working Group, is a domain specific extension of the W3C PROV data model for representing provenance information (w3.org/TR/prov-overview). It is a framework for the generation, storage, and query of metadata. Its base layer adopts standard Semantic Web Technologies (i.e. RDF and SPARQL). NI-DM object models represent different data structures in brain imaging (i.e. results of SPM, FSL, DICOM tags, FreeSurfer Directory, and XNAT). Examples of implementation of the NIDASH Data Model (NI-DM) for software can be found on github at github.com/ni-/ni-dm.
- Web page on resources for data sharing (datasharing.incf.org/ep/Resources) contains an inventory of useful resources to support finding information on formats, tools, and services to facilitate data sharing.

Discussion

The aim of the discussion session was to assess the current state of workflows in electrophysiology as well as existing approaches for data structures/software libraries, documentation/provenance, workflow management systems, and supercomputing to identify what is needed to improve workflows and data management to promote reproducibility of electrophysiological data. During the discussion, the participants identified what they considered to be the major challenges to reproducible research in electrophysiology (see Table 1), as well as what they considered is required to solve these challenges (see Table 2). The participants also devised potential solutions to some of the identified challenges using the approaches/tools presented during the workshop. The discussion concluded with the participants developing a list of general recommendations for the field (see Table 3).

The challenges were identified in 5 areas: metadata, provenance, tools/documentation, supercomputing, and community information exchange. Briefly, in terms of metadata, the participants remarked that data and its related metadata are often stored in separate domains and that there are currently no accepted standards for file formats, semantics, or metadata. In consequence, participants identified significant challenges to reproducibility to be a lack of interoperability between existing tools and an incomplete provenance trail. Also, they believed that the community lacked appropriate forums where experimenters could exchange information and that community access to parallelization and high performance computing is needed. The participants agreed that although the challenges pose impediments to reproducibility, they are not insurmountable. Integration of data and its associated metadata and the development of standards for data files, semantics, and metadata, as well as the development of workflows to fully track provenance, were considered relatively straight-forward fixes for improving the situation; however, the problem is that there are no approaches/tools currently available with all of the required capabilities. The participants strongly believed that organizations such as INCF should take an active role in lobbying vendors of electrophysiological recording systems to push for common formats for data and metadata in their tools/applications. Also, since most electrophysiology laboratories use proprietary tools, the participants believed that there was also a need to promote interoperability between formats.

In terms of potential solutions based on existing approaches, one option to enable users to query data by metadata is to integrate the Neo data model with odML metadata (Sobolev et al 2014). Sumatra and the optimization plugin in Taverna Workbench represent two options for tracking provenance, while the odML Mobile App could be used to improve metadata collection by experimenters.

Table 1. Challenges to reproducible research in electrophysiology

Metadata:

- Data and related metadata are stored independently making it hard to link the two
- Getting metadata from the experiment to data storage
- No standard file formats, semantics, or metadata formats currently
- Standards for post-processing required
- Partial metadata information
- Many researchers do not know that metadata is as important as data. Also, metadata should be recorded for both, successful and failed experiments, which will help to tune experiments or for queries across data sets for particular conditions.

Tools and documentation:

- Many tools target people with some programming experience, not people who are not able to write code.
- Not enough automation
- Lack of electrophysiology modules within existing workflow tools
- Need for interfaces for available tools to local data analysis environment
- Lack of interoperability between proprietary tools

Provenance:

- Incomplete provenance information for all steps of human and machine workflows
- Need for workflow and provenance solutions that track information for long periods of time

Supercomputing:

- Difficulty of parallelization
- Lack of access to high performance computing

Community information exchange:

- No way to evaluate which analyses are appropriate for a specific type of scientific question
- No way to evaluate which tools are best for your analysis
- Lack of feedback on what parameters should be recorded for analysis by others later

Table 2. Requirements for reproducible research in electrophysiology

Metadata:

- Integration of data and metadata-- stronger links between data and metadata
- User friendly tools for easy metadata capture and management—need to determine what metadata are important (may be difficult since you will never be able to satisfy everyone's needs for reusing your data)
- Develop format standards for metadata, file formats, and semantic information
- Develop standards for post-processing

Community exchange where researchers can share:

- Practical information about experiments
- What type of metadata is required for certain experiments
- That also includes a forum for feedback in which parameters they should record for other people to reuse their data
- No way to evaluate which tools are appropriate for your type of evaluation—a user forum where other researchers can explain what they used the tool for and how it worked
- Tutorials, examples for different tools must be accessible and discoverable—should be a part of a community site

Tools and documentation:

- Tools should be easy to learn, use, and well documented:
- Electrophysiology-specific modules should be included in existing workflow engines
- Interface available workflow tools to local data analysis environment
- Most tools that are available are not convenient for an experimenter to change to. There is a need to focus on the needs of the experimenter during the process to develop tools that they do not need to learn. It would be beneficial if the tools could collect metadata as you go presumably with the data
- Machine readable vendor information
- Develop best practices for software developers for interoperability

Provenance:

- Develop best practices for software developers to handle provenance tracks
- Include provenance information at every stage of data analysis, including interactive manual processing and exploratory data analysis:
- Want to be able to track information for long periods of time whether the experiment works or not (exploratory data analysis)--Workflow/provenance solutions that cater to this
- Data from machines and human data should be captured in a standard format

Supercomputing:

- The use of parallelized software should be made easy and seamless (in particular, available to experimenters)
- Access to high performance computing should be made easier

Table 3. General recommendations for the electrophysiology community:

- Define the different levels of the workflows and develop apps that perform the tasks
- Lobby funding agencies and journals to require reproducible research
- Lobby funding agencies to improve the chance of funding granted to laboratories with a data management system
- Establish a repository to publish data (get a doi so researchers can track who is using the data)
- Lobby funding agencies to fund data management resources.
- Vendors should be encouraged to adopt community standards and focus on interoperability
- Presentation of the importance of metadata collection and provenance tracking at courses to younger researchers.
- INCF should fund a course on versioning, software, and relevant concepts for neuroscientists
- INCF could have an initiative for communities to contribute tools for others to know what is being done

Recommendations to INCF

At the conclusion of the workshop, the participants refined the general recommendation to a set of specific recommendations that they believed INCF should pursue to improve reproducibility in electrophysiological research. Specifically, INCF should:

- Establish best practice guidelines for developers working on provenance and interoperability.
- Develop standards for metadata. INCF's work on metadata should focus on capturing metadata, integrating it with data, and easy entry/management, not on determining which metadata are important since this will vary with user needs. This work should be coordinated with efforts of existing INCF task forces working on metadata, provenance, and workflows.
- Create a website for community exchange. The participants believed the current webpage with resources for data sharing in electrophysiology should be expanded to a website for community exchange that contains practical information about experiments, the type of metadata required for certain experiments, and a forum feedback on experimental design (cf., e.g., SEEK4Science). The site should also contain tool tutorials.
- Engage with vendors to coordinate standardization. Specifically, INCF should encourage vendors to adopt common formats for data/metadata, include annotation tools that include standard terms, have automatic metadata capture at all steps, and automated save function. Also, the participants thought that INCF should also encourage vendors to develop easy, user-friendly tools.
- Support training activities on versioning, software, and relevant concepts. This recommendation was motivated by the fact that many of the tools used today target people with some programming experience, not people who are unable to write code. INCF should hold courses to teach users basic coding, as well as offer courses to help researchers develop easy to learn, easy to use, and well-documented tools (see also Software Carpentry, http://software-carpentry.org/) and promote standards for data management (e.g., ELIXIR, https://www.elixir-europe.org/). In addition, the INCF should encourage organizers of data analysis courses to base their teaching on open standards, formats and tools to promote their use to the young generation of scientists.

References

Denker M, Roux S, Lindén H, Diesmann M, Riehle A, Grün S (2011) The local field potential reflects surplus spike synchrony Cereb. Cortex 21 (12): 2681-2695. doi: 10.1093/cercor/bhr040

Denker M, Einevoll ET, Franke F, Grün S, Hagen E, Kerr J, Nawrot M, Ness TB, Ritz R, Smith L, Wachtler T, Wojcik D (2014) Report from 1st INCF workshop on validation of analysis methods. Tech Report International Neuroinformatics Coordination Facility (INCF). https://incf.org/documents/workshop-reports/report-of-the-1st-incf-workshop-on-validation-of-analysis-methods

Garcia S, Guarino D, Jaillet F, Jennings TR, Pröpper R, Rautenberg PL, Rodgers C, Sobolev A, Wachtler T, Yger P and Davison AP (2014). Neo: an object model for handling electrophysiology data in multiple formats. Front. Neuroinform. 8:10. doi: 10.3389/fninf.2014.00010

Grewe J, Wachtler T, Benda J (2011). A bottom-up approach to data annotation in neurophysiology. Front. Neuroinform. 5:16. doi: 10.3389/fninf.2011.00016

Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsaki G (2000) Accuracy of Tetrode Spike Separation as Determined by Simultaneous Intracellular and Extracellular Measurements. J Neurophysiol 84:401-414.

Hatsopoulos NG and, Suminski AJ (2011) Sensing with the motor cortex. Neuron. 72(3):477-87. doi: 10.1016/j.neuron.2011.10.020

Kim S, Takahashi K, Hatsopoulos NG, Coleman TP (2011) Information transfer between neurons in the motor cortex triggered by visual cue. Conf Proc IEEE Eng Med Biol Soc. 2011:7278-81. doi: 10.1109/IEMBS.2011.6091697;

Lewicki M (1998) A review of methods for spike sorting: the detection and classification of neural action potentials. Network: Computation in Neural Systems 9:R53-R78.

Maccione A, Hennig MH, Gandolfo M, Muthmann O, van Coppenhagen J, Eglen SJ, Berdondini L, Sernagor E. (2014) Following the ontogeny of retinal waves: pan-retinal recordings of population dynamics in the neonatal mouse. J Physiol. 1(592):1545-63.

Pröpper R and Obermayer K (2013) Spyke Viewer: a flexible and extensible platform for electrophysiological data analysis. Front. Neuroinform. 7:26. doi: 10.3389/fninf.2013.00026

Quian Quiroga R, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. Neural Comp 16:1661-1687.

Riehle A, Wirtssohn S, Grün S and Brochier T (2013) Mapping the spatio-temporal structure of motor cortical LFP and spiking activities during reach-to-grasp movements Front. Neural Circuits 7: 48 doi: 10.3389/fncir.2013.00048

Sobolev A, Stoewer A, Leonhardt AP, Rautenberg PL, Kellner CJ, Garbers C, Wachtler T (2014) Integrated platform and API for electrophysiological data. Front. Neuroinform. 8:32. doi: 10.3389/fninf.2014.00032

Teeters JL, Harris KD, Millman KJ, Olshausen BA, Sommer FT (2008) Data Sharing for Computational Neuroscience. Neuroinformatics 6(1): 47-55.

Torre, E, Picado-Muiño, D, Denker, M, Borgelt, C and Grün, S (2013) Statistical evaluation of synchronous spike patterns extracted by Frequent Item Set Mining. Front Comput Neurosci doi: 10.3389/fncom.2013.00132

Zehl L, Denker M, Wachtler T, Grewe J, Stoewer A, Sobolev A, Jaillet F, Brochier T, Riehle A, Grün S. (in prep) A Practical Approach to Implementing odML in a Neurophysiology Laboratory.

datasharing