

G-Node: an integrated tool-sharing platform to support cellular and systems neurophysiology in the age of global neuroinformatics

Andreas V.M. Herz, Ralph Meier, Martin P. Nawrot, Willi Schiegel, Tiziano Zito

Abstract The global scale of neuroinformatics offers unprecedented opportunities for scientific collaborations between and among experimental and theoretical neuroscientists. To fully harvest these possibilities, a set of coordinated activities is required that will improve three key ingredients of neuroscientific research: data access, data storage, and data analysis, together with supporting activities for teaching and training. Focusing on the development of tools aiming at neurophysiological data, the planned German Neuroinformatics Node (G-Node) aims at addressing these aspects as part of the International Neuroinformatics Coordination Facility (INCF). Based on its technical and scientific scope, the Node could play a substantial role for cellular and systems neurophysiology as well as for the neuroscience community at large.

Motivation

The human brain is one of the most complex biological systems. It contains more than 10^{11} nerve cells and 10^{15} synaptic connections, and its functional elements extend over more than ten orders of magnitude in space – from molecular pathways in individual synapses and neurons to the entire brain. Similarly astounding is the fact that dynamical processes that last for less than a millisecond underlie life-long memories – a span of more than ten orders of magnitude in time. Moreover, the brain utilizes multiple nonlinearities and nested feedback-loops that severely limit any purely intuitive approach. Analyzing the dynamics and function of the human brain therefore continues to be a formidable challenge.

In the last decades, novel experimental methods such as patch-clamp recording and imaging techniques have changed basic as well as clinical neuroscience in a most dramatic way. Suddenly, single ion channels could be studied and functional magnetic resonance imaging is nowadays used in a routinely manner. It is most likely that all

these developments do not mark the end of a success story but rather represent some intermediate step. In addition, as shown by the recent discovery of adult neurogenesis (Erikson et al., 1998, Gould et al., 1999), neuroscience's firm dogmas of today may already be overthrown tomorrow.

Despite the great success of neuroscience over the last decades, there is thus no doubt that we are far from understanding the human brain. This concerns all levels of neural processes – from the regulation of molecular pathways, the dynamics of single synapses, and the information processing of small neural networks to the orchestrated function of the entire brain. As new studies are initiated on the basis of current interpretations of available data, long-term progress in the neurosciences will crucially depend on the broad availability of high-quality data (including pre-processed data as well as the underlying raw data) and high-quality data-analysis tools (e.g. Teeters et al. 2008). However, many of today's commercial recording tools are based on highly individual and proprietary data formats and come only with limited and typically closed-source software tools for data mining and analysis. This shortcoming severely complicates the access, storage, analysis and sharing of neuroscientific data and thus slows down the future development of brain research.

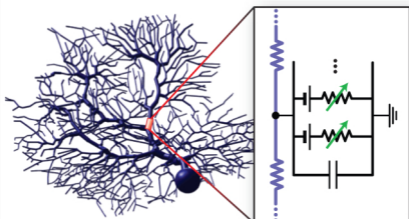
As a central element of the German Neuroinformatics Node (G-Node) within the International Neuroinformatics Coordination Facility (INCF) a novel software and hardware infrastructure will therefore be developed that eases the acquisition, storage and further processing of experimental data in the spirit of the overall INCF goals (e.g. Bjaalie and Grillner 2007; van Horn & van Pelt 2008). We will focus on cellular and systems neurophysiology for a number of reasons. First, the lack of common data standards is rather severe in this field; as a consequence, successful standardizations could have an enormous impact. Second, the complexity of an integrated and internationally coordinated approach to advanced data acquisition, storage and analysis goes beyond the capability of a single lab – one of the reasons for the slow progress seen so far – but can be tackled through a concerted effort. Third, without a thorough quantitative understanding of cellular and systems neurophysiology, there is no solid foundation for computational neuroscience and brain theory. In addition, the methodology and tools developed within the project could later also be used in other neuroscience areas. Especially early in the project, however, it will be helpful to focus on a specific task and research community to quickly reach critical mass.

The importance of high-quality physiological data and improved data-analysis tools for computational neuroscience is nicely illustrated by the task of deriving reliable single-neuron model parameters (see also the Figure). The standard strategy to deal with the key problem of cell-to-cell variability, i.e., population averaging, can be severely misleading because the dynamical characteristics of single-cell models are, in general, not a monotone function of their parameters. As a consequence, the mean behaviour within a class of models may strongly differ from that of a model with mean parameter values (Golowasch et al., 2002) and nearly identical dynamical characteristics may be implemented by rather different parameter combinations (Goldman et al., 2001). In addition, with increasing model complexity, the number of parameters to be estimated increases dramatically so that they have to be taken from different cells or even different preparations, further lowering the model's trustworthiness. Last not least, models are often calibrated using *in vitro* data, yet designed to capture the neural dynamics and computations of behaving animals. All these examples underscore the need for widely available high-quality experimental data as well as improved data-analysis tools – the latter being a key goal of the G-Node project.

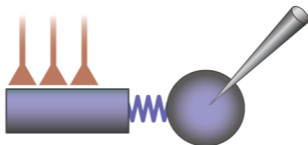
By directly addressing this need of a broad range of experimental and theoretical neuroscientists, the G-Node will thus support ongoing and future experiments in cellular and systems neurophysiology, encourage the standardization of data formats as well as analysis tools and thus directly facilitate the cooperation within and between different labs (Baxter et al., 2006). It is our hope that the G-Node, including its services as general neuroscience community site and its data archive functionality, will also attract a large group of neuroscientists that do not yet have close links to computational approaches or neuroinformatics. To foster international cooperation, the project will be carried out in close cooperation with the INCF secretariat and interested INCF nodes. To achieve its various goals, a larger group of neurobiologists and computer scientists will work on the G-Node project. At its host site (Ludwig-Maximilians-Universität Munich), there will be a staff of five full-time scientists/software developers. In addition, a group of five to ten collaborators from Germany and beyond will join the G-Node based on part-time contracts. Finally, there are also funds to support younger scientists that want to prepare their data and tools so that they can be integrated into the G-Node. Following the open-source concept of the forerunner Neuroinformatics Portal all tools developed within the G-Node will be made freely available to the national and international community. Together, these measures will incorporate the technological opportunities of global neuroinformatics into everyday

Herz1.eps

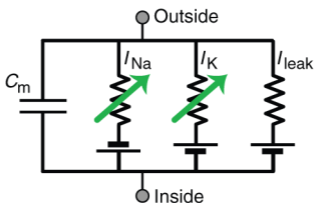
Level I: Detailed compartmental models



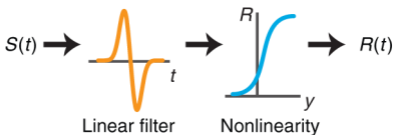
Level II: Reduced compartmental models



Level III: Single-compartment models



Level IV: Cascade models



Level V: Black-box models



neurophysiological routine and thus help establish a new scientific culture.

In the following sections, the four components of the project are summarized. Each component will be motivated by a short “Problem” discussion, followed by a description of the proposed “Solution and Specific Measures”. As there are substantial overlaps between the four project components, the solutions will overlap as well.

Data Access

Problem

There exists a plethora of physiological recording devices, many of which come with proprietary data formats. This makes it difficult and sometimes even impossible to share data with colleagues. In addition, the lack of standard data formats complicates the analysis of raw data by a person different from the one who originally recorded the data. Collaborations between different experimental labs as well as co-operations between theoreticians and experimentalists, which have been one key ingredient for the advancement of the neurosciences in recent years, are severely inhibited.

After recording, physiological data can often only be analyzed with the software shipped with the specific recording hardware. Such programs offer limited functionality. To carry on with advanced data analysis, or to test new analysis methods, the experimenter or data analyst is forced to bypass the software constraints through hand-made jugglery. This results in undesired fragility and unnecessary complexity of the experimental workflow, a day-to-day experience in many labs.

Solution and Specific Measures

- The G-Node will provide an open-source tool to allow direct access to a broad range of proprietary data formats. The tool will be initially based on NeuroShare (see <http://neuroshare.sourceforge.net/>) and its extensions within the FIND Matlab toolbox (Meier et al., 2008). Note, however, that the scope of the NeuroShare API is not sufficient in this respect because (a) it is platform specific – Windows only – and based on a proprietary format; (b) it describes how data are to be retrieved and delivered, but does not foster any standard regarding how data are stored; (c) it does not offer a reference implementation; (d) the libraries offered by hardware manufacturer are closed-source. In order to use the libraries, users need to buy the corresponding hardware and/or commercial software. Bug reports and feature requests are usually not possible. The FIND toolbox inherits these NeuroShare – specific shortcomings with respect to its data import and requires significant

expenses for the commercial Matlab software licenses.

- The G-Node will offer a platform-independent, open-source “data import” tool that will be made available in common programming languages.
- To assure long-term usability and distribution of the data, the establishment of a widely used unified data format is necessary. A unified data format has proved to be the crucial means towards standardization for many research areas (see for example the European Data Format EDF+ for EEG data analysis, Kemp and Olivan, 2003, or also <http://www.edfplus.info/>). Various efforts have been devoted recently in this direction for neural data (see for example the BrainML initiative by the NIH, Gardner et al., 2003, or <http://brainml.org/>), but no drop-in solution has been produced. Within the German Neuroinformatics Node, we intend to design such a format in collaboration with experimentalists from participating labs and in close interaction with the INCF, to tailor the format to the community needs.
- The user-oriented development and readily available implementation of the Node project will assure a widespread use of the unified data format, which will in turn provide an incentive for manufacturers of recording hardware and software to endorse this format as a standard (on a smaller scale, this happens already with NeuroShare). Note that the planned unified data format is not incompatible with NeuroShare, but complements it by specifying in detail how data should be stored, and thus makes it possible to work with existing hardware and software without inhibiting the use of newly developed tools.
- Unified export functions will be provided to basic formats (CSV, XLS, XML, binary, Matlab arrays, etc.), to databases (SQL, Oracle, ROOT, HDF5, etc.), and to the unified data format developed at the G-Node. The database back-end can be provided by the individual user. The G-Node will also offer standard database templates and “HowTo”s to allow maximal flexibility for advanced users and maximal ease of use for the novice. The design of the database template will be realized in strict collaboration with the participating labs.
- The G-Node will develop and offer an electronic lab book template (see, for example, the ELB article at Wikipedia, <http://de.wikipedia.org/wiki/Laborjournal/> or a similar project at NeuroScholar.org). The template will be designed such that it is easily customized according to local and experiment-specific needs. It will be possible to automatically integrate the electronic lab book in the database and to store raw data together with

metadata. The lab book will also feature an optional security layer, so that it can act as a full replacement or at least a helpful complement for the “paper” lab book required by law. Filling in a paper lab book tends to be cumbersome. The electronic lab book could help in this respect, for example by offering precompiled fields for static data (e.g. author, date, room temperature, etc) and, when integrated with the recording hardware, even for setup-specific variables (pressure, voltages, amplification levels, etc). The electronic lab book will be searchable and will maintain links to the experimental raw data, even when those are stored remotely. An important advantage of a standardized lab book will be that it facilitates efficient information exchange, both internally and among different labs.

- All software tools will remain open source and freely available.

Data Storage

Problem

Modern recording techniques generate rapidly growing amounts of data. In many labs, the problem of archiving and long-term storage is tackled locally with suboptimal means such as DVDs or external hard-disks. As a result, old data are often practically unavailable for further analysis and are not secured against hardware failures. Not only is this in stark contrast with basic rules of good scientific practice, see, for example, the suggestions by the German Science Association, which enforces the storage of scientific data used for publications and for projects funded by DFG for ten years (<http://www.dfg.de/antragstellung/gwp/index.html/>); sloppy data storage also hinders the spreading of scientific knowledge and damages the public reputation of neuroscience.

Solution and Specific Measures

Funded through a grant from the German Federal Ministry of Education and Research, the G-Node will offer data storage and management services, initially restricted to projects and international collaborations that involve German institutions. Support will be both, remote and local, and include hardware consulting, design and realizations, as well as maintenance and backup services. Upon request, the technological concepts and software solutions will be made available on an international level.

- Remotely, the G-Node will offer a central Data Storage System, which will handle large volumes of data and will be designed such that it can also meet increasing future

demands. The data upload and download for the end user will be transparent, i.e., as easy as moving files to and from an internal drive. Data can be searched and (optionally) integrated in the database developed in the “Data Access” part of the Node project. Data will be secured against hardware failures and stored for a time period chosen by the user – the storage of lab books will be indefinite. The access to the stored data will be granted by an authorization module fully configurable by the user. If different labs want to share data, all they have to do is to set up permissions accordingly. The end user does not need to perform any backups or to care about the scalability of local storage devices.

- Connected with the Data-Storage System, a data mining service will be offered for groups with limited local computation capabilities. Results can be offered for further analysis in different formats depending on the user needs.
- At a later stage of the project, services could also be offered on a pay-basis to a wider group of users. This could be an interesting option not only for individual researchers or labs but also for publishers, e.g., of conference proceedings, who do not have the facilities to store and make available supplementary material.

Data Analysis

Problem

The complexity of modern physiological data often requires a careful data analysis that goes far beyond standard statistical tests (see, for example, Brown et al., 2004; Kass et al., 2005; Grün 2008). The underlying techniques are, however, not routinely available to main-stream neuroscientists. For example, a thorough understanding of the responses of a neuron to dynamic stimuli may only be possible using a theoretical model described by a set of nonlinear differential equations. At this level of sophistication, experimentalists may lack the appropriate mathematical and computational background (and often also the time) to carry out a model-based analysis of their data – and resort to traditional methods. These methods, however, do not allow the researchers to make best use of their data. As a consequence, data are often only used to provide certain evidence for some model whereas, with proper statistical scrutiny, they could have been used to refute an alternative model and thus be far more valuable scientifically. Similarly, theoreticians frequently propose new models and analysis algorithms without having tested them in depth with real data. Both aspects lead to an almost grotesque situation:

some complex but potentially highly valuable data are not studied with state-of-the-art analysis tools, and some potentially excellent tools are not brought to the attention of a general neuroscience audience but only published in theory-oriented journals. On both, the experimental as well as theoretical side, substantial resources are thus wasted.

Solution and Specific Measures

- The G-Node will offer a public and well-annotated code repository for the analysis and modeling of neurophysiological data. The repository will encourage theoreticians to submit their model implementations and experimentalists to test novel analysis methods. To ease broad access, no specific programming language will be requested for code contributions. To assure the quality of available material, standard practices of open-source development will be enforced, such as a) quality management and transparent code review: a researcher who submits code suggests an external reviewer, and both are listed to demonstrate their responsibility; b) proper credit assignment: every contributor of a tool offered by the Node will be explicitly mentioned, together with relevant original publications of the respective method. In addition, contributors will be encouraged through the Node to publish their tools in method-oriented journals. This will underscore the usefulness of the tools and further increase the standing and publicity of the project; c) extensive documentation and official guidelines; d) documented examples and test data.
- To make the algorithms and their usage as attractive and transparent as possible for the end user, this repository will be supplemented by a web portal that contains discussion forums, “Wiki”s, mailing lists, tutorials, “HowTo”s, FAQ, etc.
- Within the G-Node, a toolbox will be designed and implemented that offers a general framework for data analysis. This toolbox extends the repository – essentially an unstructured “big bag” of peer-reviewed tools for data analysis – to a coherent and tightly cross-linked data-analysis package. The term “Toolbox” is used here in its widest possible meaning - i.e. it comprises a collection of quite different routines. All these routines will interact with a central interface to access the data. Individual users can then focus on their core scientific tasks while the toolbox takes care of the implementation, performance, and data handling details. Users will be able to submit bug fixes, patches, feature requests, and new algorithms. The long-term goal is to implement toolboxes in different programming languages, possibly based on existing software (see, e.g., <http://mdp-toolkit.sourceforge.net/> and <http://chronux.org>) but with common API. To allow us to incorporate existing analysis code we will therefore provide interfaces to different programming languages / interpreters.

The toolboxes will be based on the export/import functions and unified formats specified in "Data Access". We believe that in the long run Python will become the most commonly used programming language for analysis of neural data in the future. Currently, Matlab plays this role and will thus stay in use with a large community for the near future. An excellent working example of seamless integration of different programming languages are the Boost Libraries (<http://www.boost.org/>), highly optimized C++ routines, which can be efficiently called from Python.

- The toolbox and the code repository will remain open source and freely available.
- The success of the toolbox will depend not only on its scientific quality but also on the ease with which it can be used by a general neuroscience audience (Hidetoshi et al., 2007). Here, a crucial aspect is the "look and feel" of the G-Node Web Portal. Accordingly, a prerequisite for the portal's success is an attractive graphical layout. More importantly, the portal must also provide all information in an optimally user-friendly way and allow both experts and traditional scientists to interact as easily as possible with the tools and services provided by the Node. Last not least, the portal should also be designed such that it facilitates the communication between interested scientists and the Node team as much as possible.

Teaching and Training

Problem

During regular courses, neuroscience students are rarely exposed to neuroinformatics, computer science or advanced data analysis; some senior neuroscientists may not even be aware of recent developments in these fields. It is therefore mandatory to complement the infrastructure of the G-Node with parallel teaching and training activities. To give just one example: The introduction of "good practices" for experimental and computational studies is likely to fail if it is not coupled to intensive teaching. For their own benefit, neuroscience students and senior researchers alike should thus learn to use electronic lab books such as that offered through the G-Node, to securely store the results in an organized and rational way using modern database structures, and to apply the appropriate analysis tools.

Solution and Specific Measures

- The G-Node will offer courses in best practice from day-to-day work at the bench to successful long-term data management. The courses will be primarily directed towards

master and PhD students and be based on recommendations obtained within the "Data Access", "Data Storage", and "Data Analysis" sections of the Node project. Courses and workshops on software tools, toolboxes for data-analysis, and electronic lab books could be integrated in the teaching programs of the Bernstein Centers for Computational Neuroscience and other interested institutions.

- Together with the Bernstein Coordination Secretariat (BCOS, Freiburg, Germany), the G-Node will make tutorials and course notes freely available on its portal site. This could be done within a "Bernstein Lecture Notes" series and include the publication of refined curricula and course material, such as sample data and numerical exercises of data analysis as well as video lectures. Freely available and citable high-quality lecture notes will also be helpful to attract the best students around the world as these lecture notes will demonstrate the scientific and teaching standards of the Bernstein Centers. The structure and organization of this part of the G-Node Portal could be modeled after that of the Toolbox, including specific rules and quality standards for submissions from the community.

- To provide an incentive for young investigators to join the G-Node, short-term contracts will be provided to students and postdocs for preparing and annotating data, analysis tools and computational models from their own projects such that they can be made available via the G-Node. This measure will have two benefits: first, many valuable resources will be added to the freely available repositories of the G-Node and increase its attractiveness for other researchers, especially in its early phase. Second, a significant group of young neuroscientists will be exposed to the idea of data-sharing. As all contributions will be reviewed, there will also be a strong quality incentive that will influence the scientific standards of the involved young scientists and help create a new scientific spirit along the INCF goals.

- Public activities and cooperation with international neuroscience institutions (e.g. Riken, Sloan-Swartz-Centers, etc.) will be coordinated with BCOS and INCF.

Outlook

The success of the G-Node project depends critically on whether the tools to be created for data access, data storage, and data analysis will be used by a broad neuroscience audience. We believe that our emphasis on the following objectives will help to achieve a wide acceptance:

First, user friendliness will be a core guiding principle for the design of all G-Node tools.

In fact, the project will involve close collaborations with neuroscientists that are interested to apply recent techniques of computer science, neuroinformatics and computational neuroscience.

Second, all tools will be developed within the “open source” philosophy. This will allow higher flexibility for advanced users that may add their own analysis tools and test their performance and stability within the toolbox framework. Those experienced users may thus become highly-motivated contributors to the project.

Third, we intend to use already existing computing resources within the user community. This strategy strongly reduces the initial costs of this project and, even more importantly, avoids bottlenecks in data transfer and sharing of computer power. Moreover, it avoids conflicts with the research and investment plans of participating labs.

Forth, although the G-Node will offer extensive data storage, its use will not be mandatory. This aspect will lessen widespread psychological barriers towards external (and thus not fully controlled) data storage and possible sharing (Teeters et al. 2008).

Finally, tools provided by other initiatives within and beyond INCF will be integrated wherever useful. This concerns also ontology packages and other support measures, as developed, for example, within the Japanese Node (<http://www.neuroinf.jp/>) of INCF (see also Yamaji et al., 2007). Similarly, the Chronux open-source software package for the analysis of neural data (<http://chronux.org/>) will be adapted and/or incorporated. We will also interact with other initiatives aiming at standardization and integration of neuroscience data, such as the Blue Brain Project (<http://bluebrain.epfl.ch/>). Clearly, close interactions with any of these initiatives will be based on strict reciprocity.

We strongly believe that cellular and systems neurophysiology will gain tremendously from global neuroinformatics and hope that the G-Node will help foster this exchange. With improved data quality and data availability, the scientific value of physiological data increases as well in that they can be used to clearly discriminate between different existing theories instead of only providing vague evidence for yet another hypothesis. This strategy served physics extremely well. Step by step, and based on testable model predictions, this will lead to a true brain theory from which, last but not least, clinical neuroscience as well as the biotechnical application domain will profit.

Acknowledgements: We would like to thank Jan Benda, Michael Bendels, Michael Brecht, Frank Endler, Martin Göpfert, Anja Gundlfinger, Jan Idzikowski, Harald Luksch, Matthias Munk, Björn Nadrowski, Andreas Neef, Florian Pieper, Raphael Ritz, Christoph Schmidt-Hieber, Shiro Usui, and Thomas Voigt for most helpful discussions about the scope of the G-Node. Part of this discussion took place during a two-day workshop in July 2007 that was supported by BMBF through the BCCN Berlin. All authors acknowledge BMBF support through BCCN Berlin (AH, MN, WS and TZ) and BCCN Freiburg (RM, MN).

References:

- Baxter, S.M., Day, S.W., Fetrow, J.S., Reisinger, S.J. (2006). Scientific Software Development Is Not an Oxymoron. *PLoS Computational Biology* **2**, e87.
- Bjaalie, J.G. and Grillner, S. (2007). Global Neuroinformatics: The International Neuroinformatics Coordinating Facility. *J. Neurosci.* **27**, 3613-3615.
- Brown, E.N., Kass, R.E., Mitra, P.P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience* **7**, 456-461.
- Eriksson, P.S., Perfilieva, E., Bjork-Eriksson, T., Alborn, A.M., Nordborg, C., Peterson, D.A., Gage, F.H. (1998). Neurogenesis in the adult human hippocampus. *Nature Medicine* **4**, 1313-1317.
- Gardner, D., Toga, A.W., Ascoli, G.A., Beatty, J.T., Brinkley, J.F., Dale, A.M., Fox, P.T., Gardner, E.P., George, J.S., Goddard, N., Harris, K.M., Herskovits, E.H., Hines, M.L., Jacobs, G.A., Jacobs, R.E., Jones, E.G., Kennedy, D.N., Kimberg, D.Y., Mazziotta, J.C., Miller, P.L., Mori, S., Mountain, D.C., Reiss, A.L., Rosen, G.D., Rottenberg, D.A., Shepherd, G.M., Smalheiser, N.R., Smith, K.P., Strachan, T., Van Essen, D.C., Williams, R.W., Wong, S.T. (2003). Towards effective and rewarding data sharing. *Neuroinformatics* **1**, 289-295.
- Golowasch, J., Goldman, M.S., Abbott, L.F., Marder, E. (2002). Failure of Averaging in the Construction of a Conductance-Based Neuron Model. *J. Neurophysiol.* **87**, 1129-1131.
- Goldman, M.S., Golowash, J., Marder, E. Abbott, L.F. (2001). Global Structure, Robustness, and Modulation of Neuronal Models. *J. Neurosci.* **21**, 5229-5238.
- Gould, E., Reeves, A.J., Fallah, M., Tanapat, P., Gross, C.G., Fuchs, E. (1999). Hippocampal neurogenesis in adult Old World primates. *Proc. Natl. Acad. Sci. USA* **96**, 5263-5267.
- Grün S (2008) Data-driven significance estimation for precise spike correlation. *J Neurophysiol* (submitted)
- Herz, A.V.M., Gollisch, T., Machens, C.K., & Jaeger, D. (2006). Modeling single-neuron dynamics and computations: a balance of detail and abstraction. *Science* **314**, 80-85.
- Hidetoshi, I., Takuto, N., Ryohei, K., Yoichi, S. (2007). Development and application of CMS-based databasemodules for neuroinformatics. *Neurocomputing* **70**, 2122-2128.
- Horn, van J and Pelt, van J (2008) 1st INCF Workshop on Sustainability of Neuroscience Databases, in preparation
- Kass RE, Ventura V and Brown E (2005) Statistical Issues in the Analysis of Neuronal Data. *J Neurophysiol* **94**: 8-25

Kemp, R., Olivan, J. (2003) European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data. *Clinical Neurophysiology* **114**, 1755-1761.

Meier, R., Egert, U., Aertsen, A., Nawrot, M.P. (2008). FIND - a unified framework for neural data analysis, to appear in: *Neural Network*, Special Issue on Neuroinformatics.

Teeters JL, Harris K, Millman JM, Olshausen BA, Sommer F (2008) Data Sharing for Computational Neuroscience. *Neuroinform.* doi 10.1007/s12021-008-9009-y

Yamaji, K., Sakai, H., Okumura, Y., Usui, S. (2007). Customizable neuroinformatics database system: XooNlps and its application to the pupil platform. *Comput Biol Med.* **37**, 1036-1041.

Figure Caption:

Computational models of single-neuron dynamics and computations – a core component of modern cellular and systems neurophysiology - require high-quality physiological data on every level of abstraction, as demonstrated by five representative levels of single-cell modeling. **I**: Detailed compartmental model of a Purkinje cell. The dendritic tree is segmented into many electrically coupled Hodgkin-Huxley-type compartments (Level III). **II**: Two-compartment model. The dendrite receives synaptic inputs and is coupled to the soma where the neuron's response is generated. **III**: Hodgkin-Huxley model, the prototype of single-compartment models. The cell's in- and outside are separated by a capacitance C_m and ionic conductances in series with batteries describing ionic reversal potentials. Sodium and potassium conductances (g_{Na} , g_K) depend on voltage; the leak g_{leak} is fixed. **IV**: Linear-nonlinear cascade. Stimuli $S(t)$ are convolved with a filter and then fed through a nonlinearity to generate responses $R(t)$, typically time-dependent firing rates. **V**: Black-box model. Neglecting biophysical mechanisms, conditional probabilities $p(R|S)$ describe responses R for given stimuli S . With increasing model complexity (from bottom to top) more and more data are required to constrain the rapidly growing number of model parameters – many thousand for neurons with complex dendritic trees; but even on the most abstract level of black-box models, faithful experimental estimates for the relevant probability distributions require advanced experimental paradigms and data analysis. This is a huge challenge for cellular and systems neurophysiology and a major motivation for our own initiative. Global neuroinformatics offers the tools to address this challenge – within the INCF, the G-Node project aims at building an integrated tool-sharing platform to support ongoing and future international efforts (Figure redrawn from Herz et al. 2006).