

CNS*2000 Computational Neuroscience Meeting Brugge, Belgium, July 16-20, 2000

A COMMON DATA MODEL FOR RECORDING AND SIMULATION DATASETS AND DESCRIPTIVE METADATA

Daniel Gardner, Robert DeBellis, Michael Abato, Kevin H. Knuth, and Steven M. Erde. Department of Physiology, Cornell University Medical College, New York, NY USA.

We have developed two databases to archive experimental neurophysiology data, targeting identified invertebrate neuron records and somatosensory cortical spike trains. The original schema underlying these databases made provision for Genesis, SNNAP, and NEURON parameter files, but the need to serve the modeling community as well as additional communities of experimental neuroscientists has led us to expand the schema into a Common Data Model for neurophysiology, one that explicitly includes data, quasi-data, simulations, protocols, and bibliographic information. A goal is to ease the interchange of data and descriptive metadata between neural simulations and actual physiological recordings.

Each first-class CDM element derives from one of five superclasses, each compatible with a wide range of neuroscience descriptors. Data elements include datasets from either recordings or simulations. For neurophysiological preparations these include time series encompassing voltage-clamp and patch clamp records and extracellular records of spike trains and local field potentials, as well as histograms describing channel amplitude and duration, and spike timing. The equivalent data for neural models may be time-series representations of solutions to systems of differential or integral equations, abstractions or canonical representations. Experimental or model-based driving functions including stimulation protocols such as voltage-clamp commands can also be data elements. Method elements can represent experimental protocols, manipulations, and modeling engines or schemas. Model elements include both parameter sets describing individual neuronal models as well as hypotheses that are related to either experiments or simulations. Model elements should be compatible with biophysical to phenomenological explorations of single neuron or system behaviors, including cellular or kinetic models, neural network representations, and conductance or cable based equivalents. Site elements are broadened to include structures ranging from subneuronal compartments to larger-scale networks or regions at any anatomical level. Reference elements encode bibliographic information.

Extensive controlled-vocabulary metadata represented in a hierarchical Attribute-Value schema enable queries within and among conformant data resources. This representation is useful for cross-disciplinary investigations. The emerging standard of eXtensible Markup Language (XML) wraps datasets and queries and is used to describe the data models that underlie neurophysiological and model databases, extensible as well to data and data models characterizing other domains of neuroscience. The CDM and its XML representations are platform and architecture-independent, designed for compatibility between existing and future data recording, analysis, and simulation methodologies.

Data and metadata that are used to specify a target experiment or protocol are inherently different in describing either neurophysiology or neural models. However, the proposed schema exists at a level of abstraction that can encompass both datatypes concisely and consistently. Experimentally-recorded cortical neurons are subtypes of site element. Here major descriptive identifying metadata include the neuron's receptive field, location, and firing patterns. Equally dominant descriptive characteristics of neural simulations or models of neuronal structures are more variable, reflecting the large number of approaches that modelers use to represent systems ranging from neural networks to population interactions. Metadata for such models are contained not only in site elements (level or scale of system being modeled, i.e. single unit or network), but have equally important information in method and model elements (equations, hypotheses and underlying theories, and simulation parameters).

We do not attempt to duplicate modeling tools or environments such as NEURON, Genesis, SNNAP,

or Matlab, nor supplant existing data structures that exist for other physiological databases such as those for ion channels and receptors. We seek instead to offer a framework and a high-level methodology for the description of data that implicitly contain complex interdependencies and whose understanding requires compilation of results from multiple disciplines within the neurosciences. The proposed standard allows as many models as possible to be expressed and therefore be available for interchange between both model and recorded data and information describing related sites and methods. The Common Data Model and its XML representation are designed to provide a common schema for:

1. representing electrode and other sensor data as well as simulation results, thus facilitating comparisons,
2. describing metadata and features such as neuronal properties characterizing both simulations and electrophysiological data, and
3. allowing developers using Genesis, NEURON, and similar tools to specify their underlying data models in CDM, enabling more sophisticated and detailed exchange of simulations linked to both simulation- and electrode-derived data. Toward this end, we coordinate efforts with NEURON developers including M. Hines, as well as several members of the Genesis group.

The Common Data Model strives to unify the goals of experimental neuroscience, while providing a self-defining and consistent means to describe data. The extension of the CDM to encompass neural models complements its original goal of describing living neurons. A major purpose of the presentation is to encourage modelers, in the cooperative spirit of the Open Source movement, to suggest enhancements to the Common Data Model and to propose and test specific instantiations of these data structures. In providing such a schema, we hope to foster and enhance the level of interplay between experiment and theory.

Supported by NIH-NIMH and Human Brain Project via MH57153 and NIH-NINDS via NS36043.