Common Data Model Extensions for Exchanging Simulations and Comparing Recorded to Computed Datasets.

Robert DeBellis, Michael Abato, Steven M. Erde, Kevin H. Knuth & Daniel Gardner. Laboratory of Neuroinformatics, Dept. of Physiology, Weill Medical Coll. of Cornell Univ., NY, NY. 748.16

Extending the Common Data Model To Link Simulations With Data

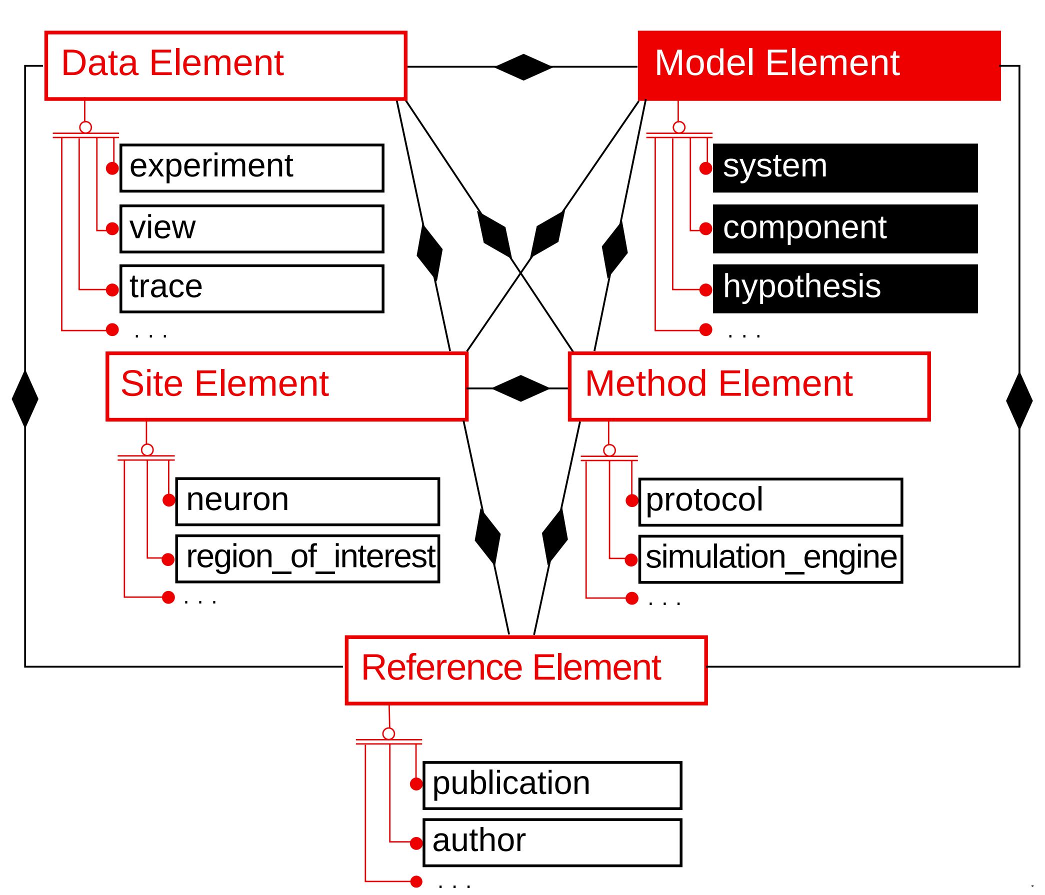
We designed a Common Data Model to underlie neurophysiology databases and to provide interoperability among disparate neuroscience data resources. Our XML-based Biophysical Description Markup Language (BDML) serves as the interface to mediate data exchange and data model coordination.

We here specify Common Data Model extensions designed to aid description and exchange of simulations, and comparison of sensor- and model-derived data. These supplement modeling tools with a schema enabling searchable databases of models and thier outputs.

To affirm the utility of our standards, we ask viewers to examine how well these extensions serve to describe and specify their models, model data or model archives, and to propose tests or suggest enhancements to accommodate additional classes of models.

Unlike the schema for our cortical neurodatabase (see adjacent poster) canonical attributes and controlled vocabulary defining Model Elements will depend upon this ongoing testing against a spectrum of models and are at present a work in progress.

Common Data Model Superclasses



1. Model Element Extends the Hierarchical Descriptive Schema of the Common Data Model

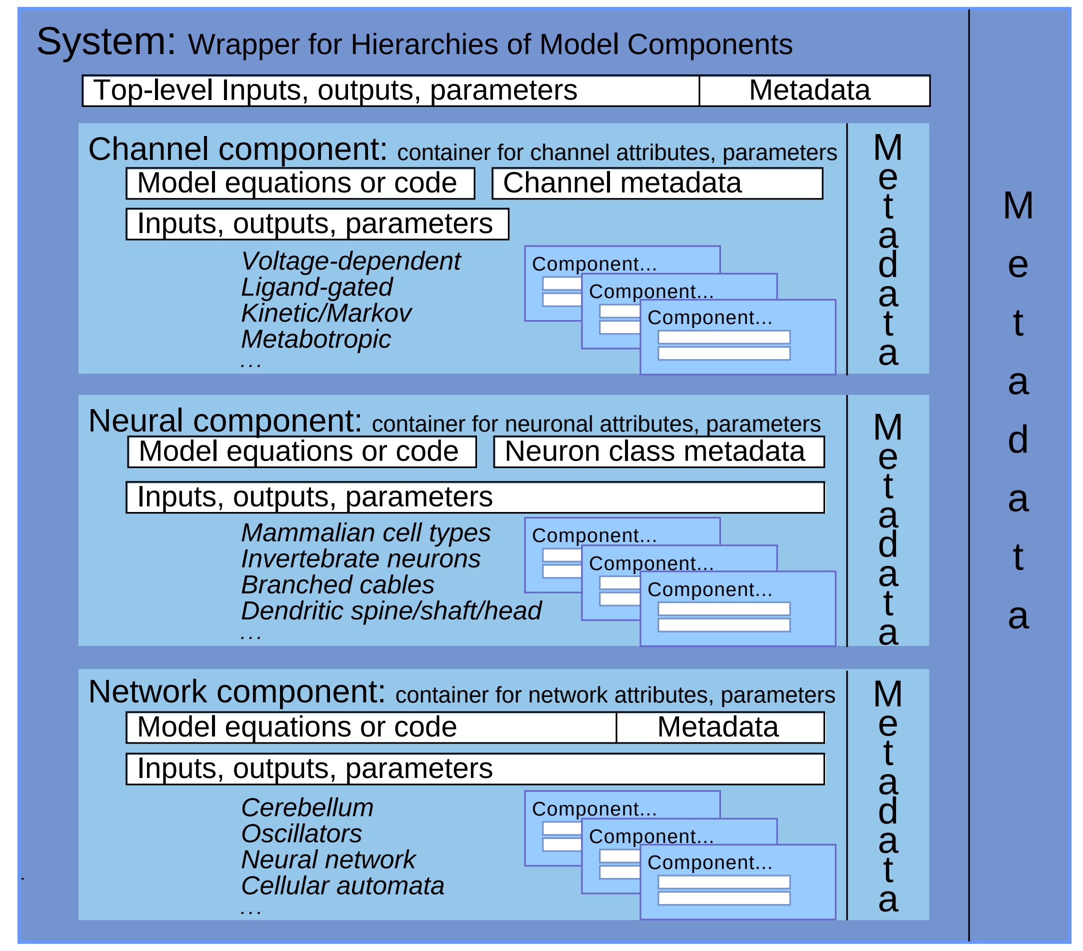
The Model Element (highlighted) is one of five superclasses of the Common Data Model initially designed to describe neurophysiology data. In conjunction with the other superclasses, it can describe phenomenological or physical abstractions as well as hypotheses that underlie or test experimental findings. Because the data model is hierarchical, users may characterize data or models, and specify database searches, at user-defined scope or granularity.

Figs. 2, 3, 4, and 5 detail Model Element and Data Element designs and metadata for describing simulations and linking them to sensor data and model quasi-data.

URLs:

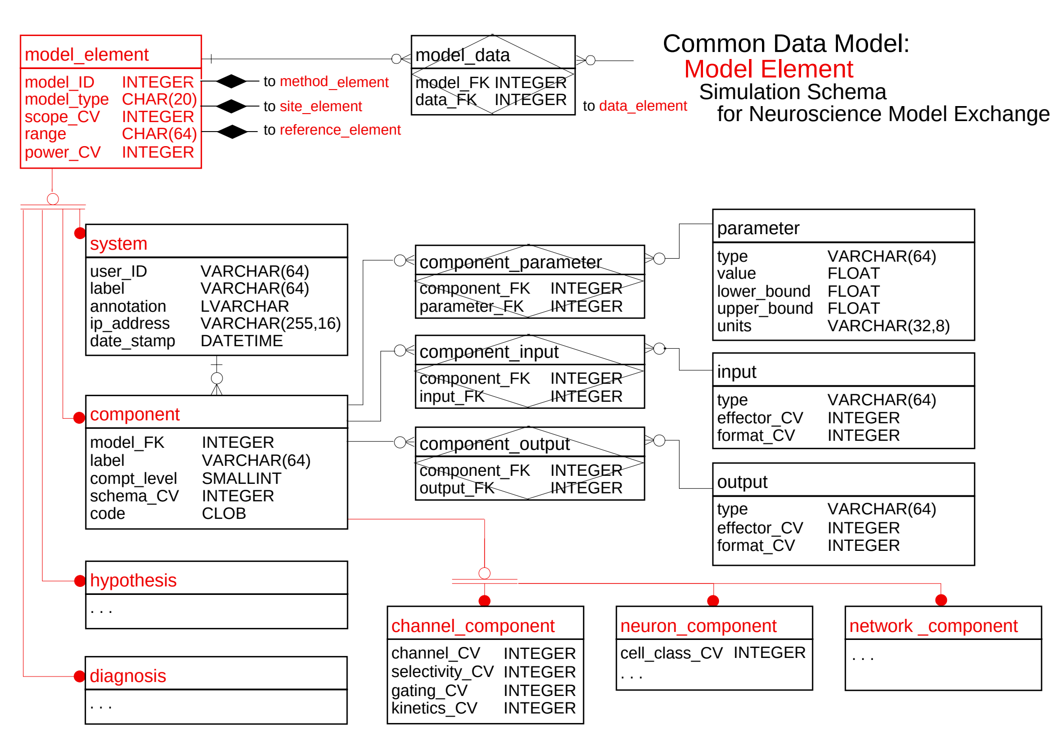
- Cortical database: cortex.med.cornell.edu
- Aplysia database: mollusc.med.cornell.edu

Schema, Attributes, and Vocabulary to Specify Neural Models and Data



2. The Complexity of Current Neural Models Favors a Hierarchical Schema to Characterize Them

Contemporary models vary in their scope, underlying schema, and characteristic inputs, outputs, and parameters. Models often incorporate multiple lower-level components. Accommodating as broad a set of existing models as possible, we propose an intuitive hierarchy using **system** as a wrapper for individual **components** (three types shown). Descriptive searchable metadata attributes characterize each level.



3. Model Element Abstracts Simulations, Hypotheses, Diagnoses, and Curve Fits

All Model Elements include attributes for scope (phenomenon), range (applicability) and power (verifiability). For simulations, the system wrapper specifies submitter identity and timestamp and encompasses one or more components, characterized by schema, parameter sets, inputs, and outputs. Parameters include values or ranges; inputs/outputs include effector or energy (common to Method Element) and format (static, signal, symbolic, list, graphic). Attribute names ending in _CV have controlled vocabulary values.

Other subtypes characterizing hypotheses, diagnoses, curve fits, and coordinate-based models for MRI analysis not shown here.

Cortical neuron functional area:
somatosensory
- SI
- SII
- insular/retroinsular
- postauditory/Ig/Id
- posterior parietal

motor
- MI/F1
- SMA/preSMA/F3/F6
- cingulate
- MEF
- PMd/F2/f7
- PMv/F4/F5
- FEF

preF/preM/orbitoF
visual/multi
- V1
- V2
- V3
- V3a
- V4
- VP
- MT
- temporal
- parietal
  - MST
  - LIP
  - VIP
  - AIP

auditory
- AI
- . . .

. . .

Recording technique:
extracellular
- single electrode
- multielectrode array

intracellular
- voltage-clamp

patch clamp
- cell-attached
- o-o
- i-o
- whole-cell

macroelectrode
nerve cuff
- field/surface
  - scalp
  - cortical

optical
- neuro-intrinsic
- neuro-via-dye
- light intensity

MEG
- radial
- planar

mechanical
- pressure
- force
- displacement
  - translation
  - rotation
  - radial

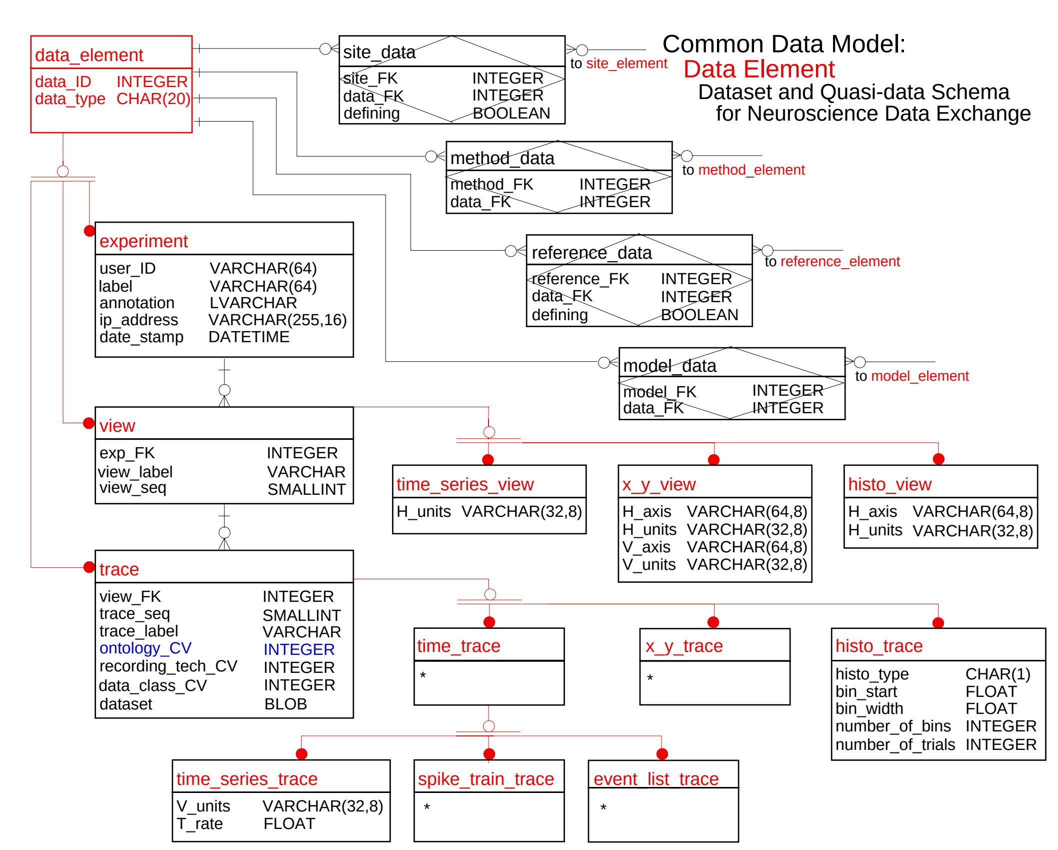
sound intensity
thermal
. . .

Model power:
de novo
based on single
based on multiple
verifiable single
verifiable multiple
predictive single
predictive multiple
. . .

Model I/O format:
static
signal
symbolic
list
graphic
. . .

4. Sparse Hierarchical Controlled Vocabularies Enable Specificity and Interoperability

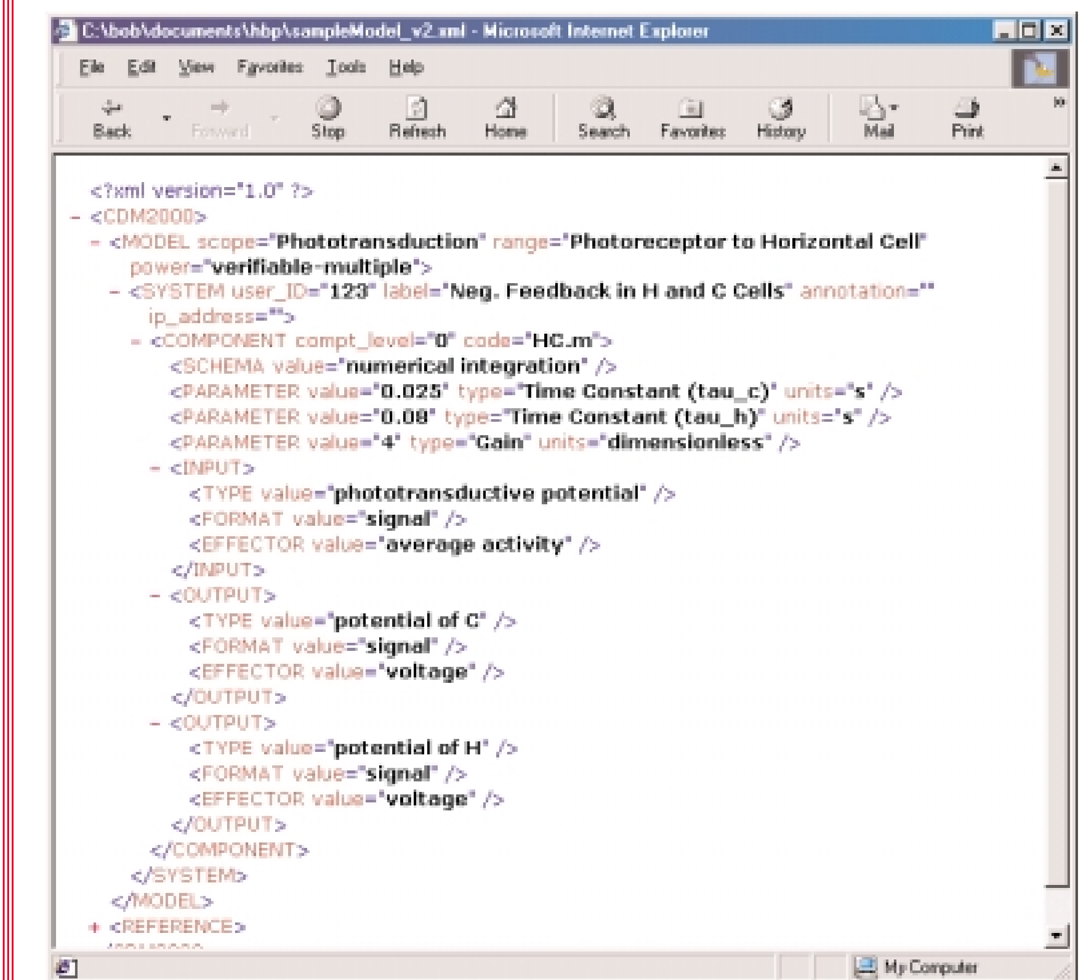
Four sample controlled-vocabulary trees describe neuronal location and recording technique for datasets, and power and input/output format for model elements. Such sparse vocabularies, many hierarchical, enable interoperability and selectable specificity by allowing searches combining broad with focused terms, a strategy analogous to natural-language processing.



5. Data Element Represents Datatypes and Wrappers for Sensor Data and Simulation Quasi-data

Data Element implements an intuitive Experiment>View>Trace hierarchy for neurophysiology datasets. In order to link quasi-data such as simulation results with sensor data, the Data Element data model has been enhanced by substituting an ontology attribute in the trace datatype, expanding the functionality of the stim_resp flag previously characterizing time_trace.

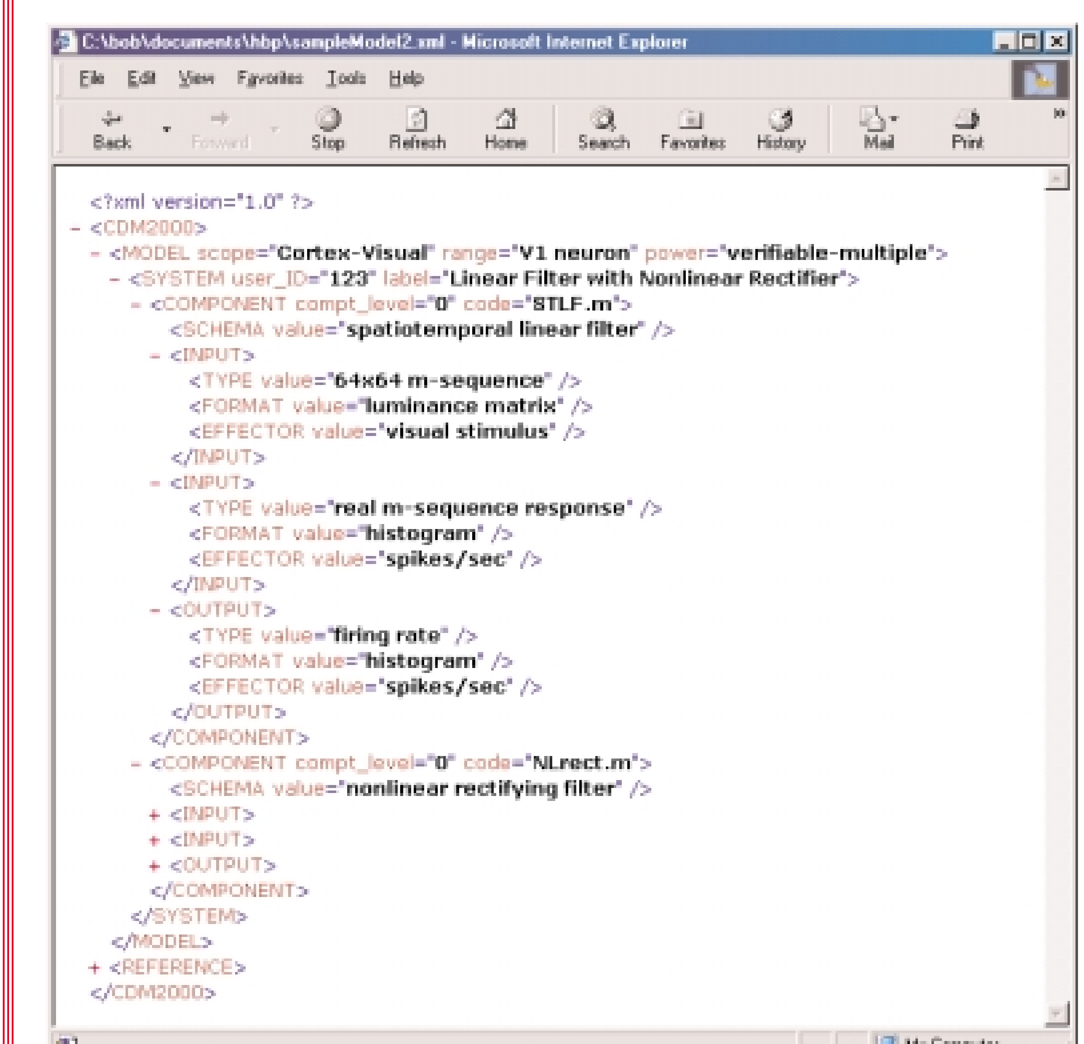
XML-Based BDML Interfaces Describe and Link Simulations


C:\bob\documents\hbp\sampleModel_v2.xml - Microsoft Internet Explorer

```
<?xml version="1.0" ?>
- <CDM2000>
 - <MODEL scope="Phototransduction" range="Photoreceptor to Horizontal Cell"
    power="verifiable-multiple">
  - <SYSTEM user_ID="123" label="Neg. Feedback in H and C Cells" annotation=""
     ip_address="">
   - <COMPONENT compt_level="0" code="HC.m">
      <SCHEMA value="numerical integration" />
      <PARAMETER value="0.025" type="Time Constant (tau_c)" units="s" />
      <PARAMETER value="0.08" type="Time Constant (tau_h)" units="s" />
      <PARAMETER value="4" type="Gain" units="dimensionless" />
    - <INPUT>
       <TYPE value="phototransductive potential" />
       <FORMAT value="signal" />
       <EFFECTOR value="average activity" />
      </INPUT>
    - <OUTPUT>
       <TYPE value="potential of C" />
       <FORMAT value="signal" />
       <EFFECTOR value="voltage" />
      </OUTPUT>
    - <OUTPUT>
       <TYPE value="potential of H" />
       <FORMAT value="signal" />
       <EFFECTOR value="voltage" />
      </OUTPUT>
     </COMPONENT>
    </SYSTEM>
   </MODEL>
 + <REFERENCE>
```

6. Prototype BDML Neuron Model Description

Sample BDML definition of a two-cell-class photoreceptor model from Schnapf *et al.* Prototype attributes, elements and controlled-vocabulary values illustrate metadata describing the model and therefore usable for searching a database of models.

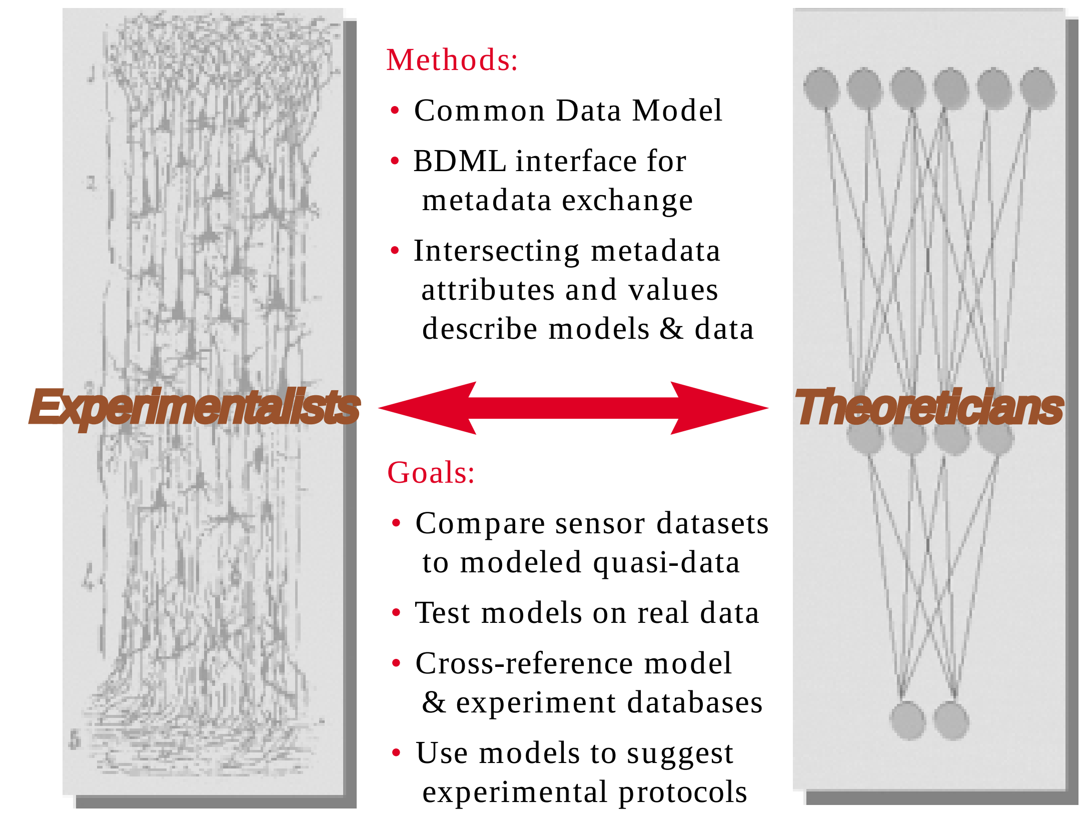

C:\bob\documents\hbp\sampleModel2.xml - Microsoft Internet Explorer

```
<?xml version="1.0" ?>
- <CDM2000>
 - <MODEL scope="Cortex-Visual" range="V1 neuron" power="verifiable-multiple">
  - <SYSTEM user_ID="123" label="Linear Filter with Nonlinear Rectifier">
   - <COMPONENT compt_level="0" code="STLF.m">
      <SCHEMA value="spatiotemporal linear filter" />
    - <INPUT>
       <TYPE value="64x64 m-sequence" />
       <FORMAT value="luminance matrix" />
       <EFFECTOR value="visual stimulus" />
      </INPUT>
    - <INPUT>
       <TYPE value="real m-sequence response" />
       <FORMAT value="histogram" />
       <EFFECTOR value="spikes/sec" />
      </INPUT>
    - <OUTPUT>
       <TYPE value="firing rate" />
       <FORMAT value="histogram" />
       <EFFECTOR value="spikes/sec" />
      </OUTPUT>
     </COMPONENT>
   - <COMPONENT compt_level="0" code="NLrect.m">
      <SCHEMA value="nonlinear rectifying filter" />
    + <INPUT>
    + <INPUT>
    + <OUTPUT>
     </COMPONENT>
    </SYSTEM>
   </MODEL>
 + <REFERENCE>
  </CDM2000>
```

7. Prototype BDML Network Model Description

A higher-level network responding to visual stimuli with a cortical firing pattern is also describable using BDML. As above, descriptive metadata controlled-vocabulary values of attributes enable searches. Example courtesy of D. S. Reich.

Towards Data–Model Interoperability



8. The Extended Common Data Model Will Enable Interoperability Between Experimenters and Theorists

Our new methods include the extended Common Data Model, BDML interfaces, and intersecting metadata for data and models.

Goals include developing a data model that could concisely represent neurophysiological models in a hierarchical descriptive metalanguage to facilitate rapid searching of neural models in data bases, to easily draw links between neurophysiological experimental datasets and data generated by models, and provide a concise and consistent descriptive language to be used to define crucial elements of neural models.

Not Conclusions, But Questions:

- Would your research benefit from methods for linking simulations with data?
- Are our schema and attribute set adequate to describe the models you create or use?
- If you design or maintain neurodatabases or simulation resources, will our data- and model-descriptor metalanguage BDML aid description, archive, and exchange of your data or models?

Support:

This Human Brain Project/Neuroinformatics research is funded by NIMH and NINDS via MH57153.

Thanks to M. Hucka, M. Hines, D. Baxter, T. White for discussions.

LNI Laboratory of Neuroinformatics
Weill Medical College of Cornell University