A Common Data Model for Recording and Simulation Datasets and Descriptive Metadata.

Daniel Gardner, Robert DeBellis, Michael Abato, Kevin H. Knuth & Steven M. Erde.

Laboratory of Neuroinformatics, Dept. of Physiology, Weill Medical College of Cornell Univ., NYC, NY, USA.

Can a Common Data Model Be Devised To Span Data and Simulations?

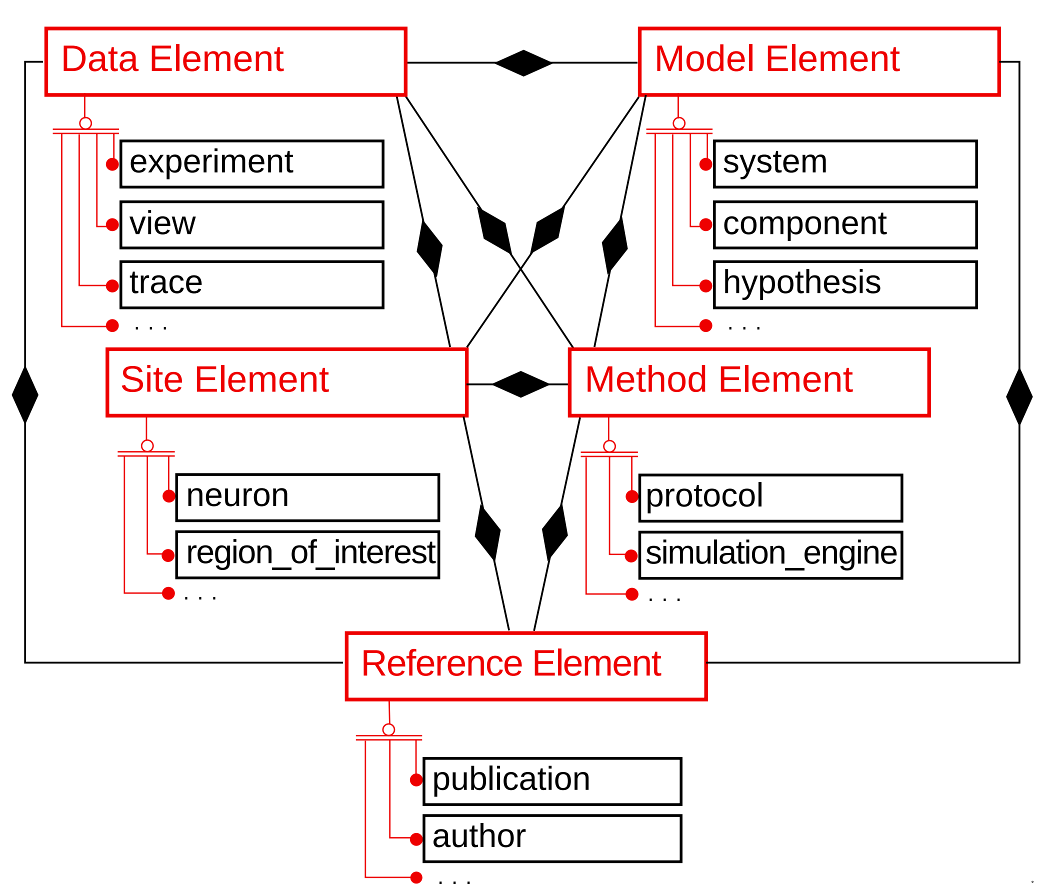
Why Does This Poster Offer Questions Rather Than Answers?

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 and enable interchange for modeling tools, rather than duplicating their function.

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 classes of models.

Common Data Model Superclasses



1. Top-Level Superclasses Span Neurophysiology

Components of the Common Data Model derive from one of five superclass Elements: Data (datasets, quasi-data), Model (hypotheses, simulations), Site (spanning compartments to neurons to systems), Method (protocols, modeling engines), and Reference (publications).

Figs. 4, 5, & 6 detail Model and Data Element designs for describing simulations and linking them to sensor data and model quasi-data.

Support

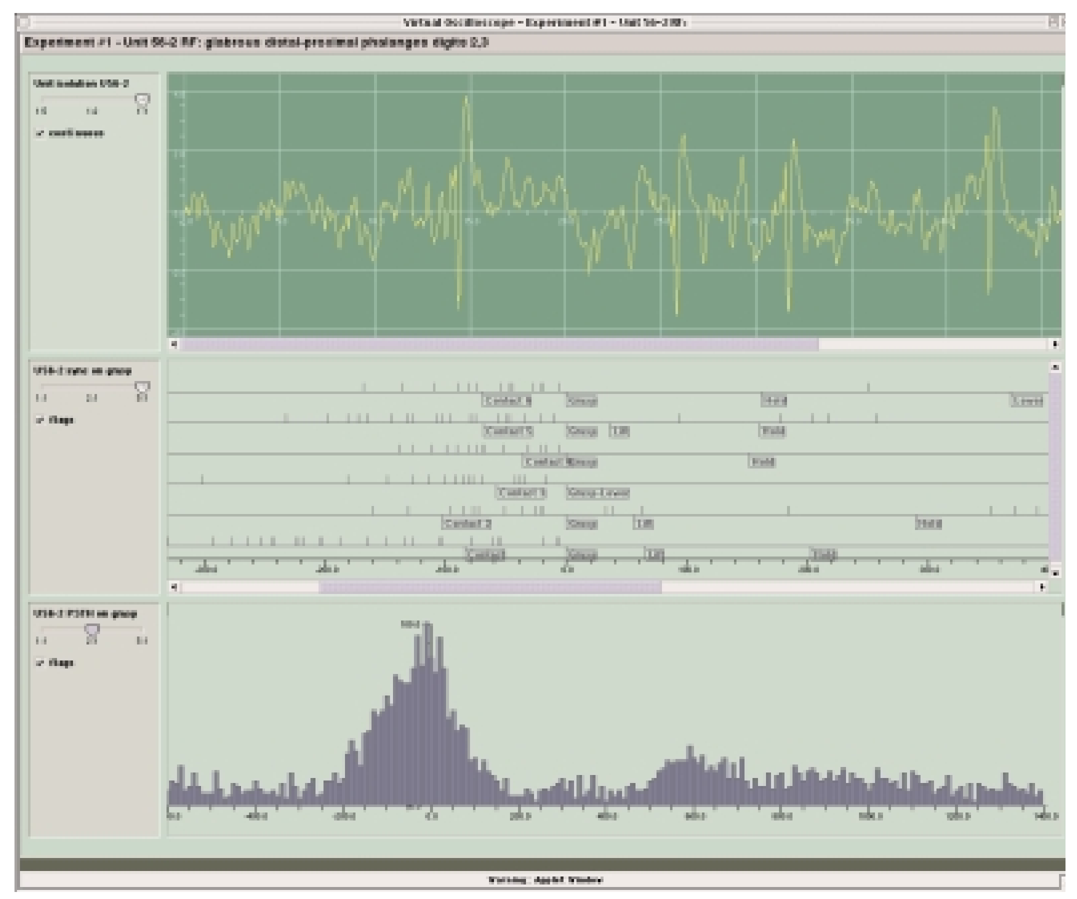
We acknowledge support of NIMH and the Human Brain Project via MH57153 and NINDS via NS36043

Thanks to M. Hucka, M. Hines, D. Baxter, T. White. Data shown are from the lab of Esther P. Gardner.

Aplysia Database Project
NIH- and NSF-supported neuroscience databases centered at
Cornell University Medical College

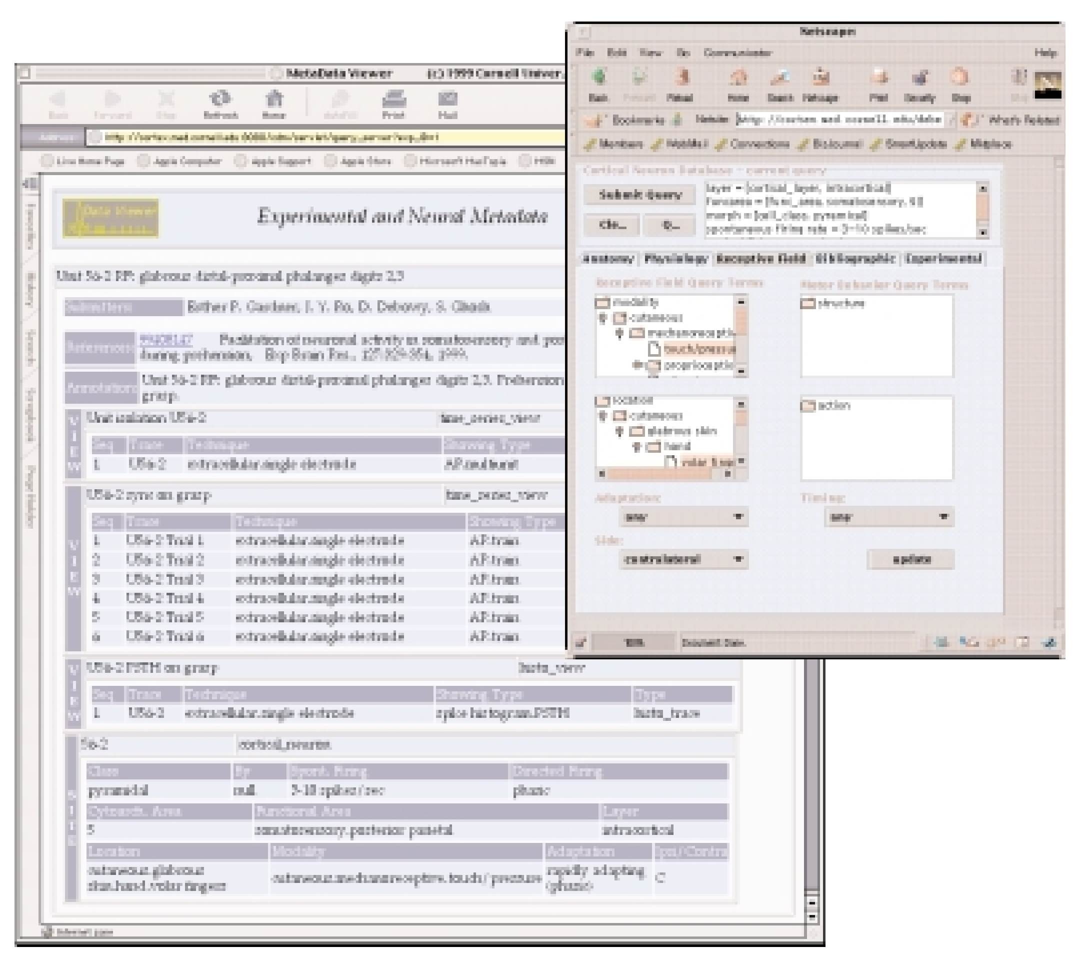
Cortical Neuron Net Database
Human Brain Project
NIMH-funded research at
Cornell University Medical College

Database Viewer and Query Tools Display the Common Data Model



2. VirtualOscilloscope Displays Database Data

Database datasets as well as Java viewers are downloaded to users, allowing dynamic control of sweep and other variables. Data are graphically parsed as view panes containing one or more traces. Tools are Java-based, enabling multiplatform data search and display.

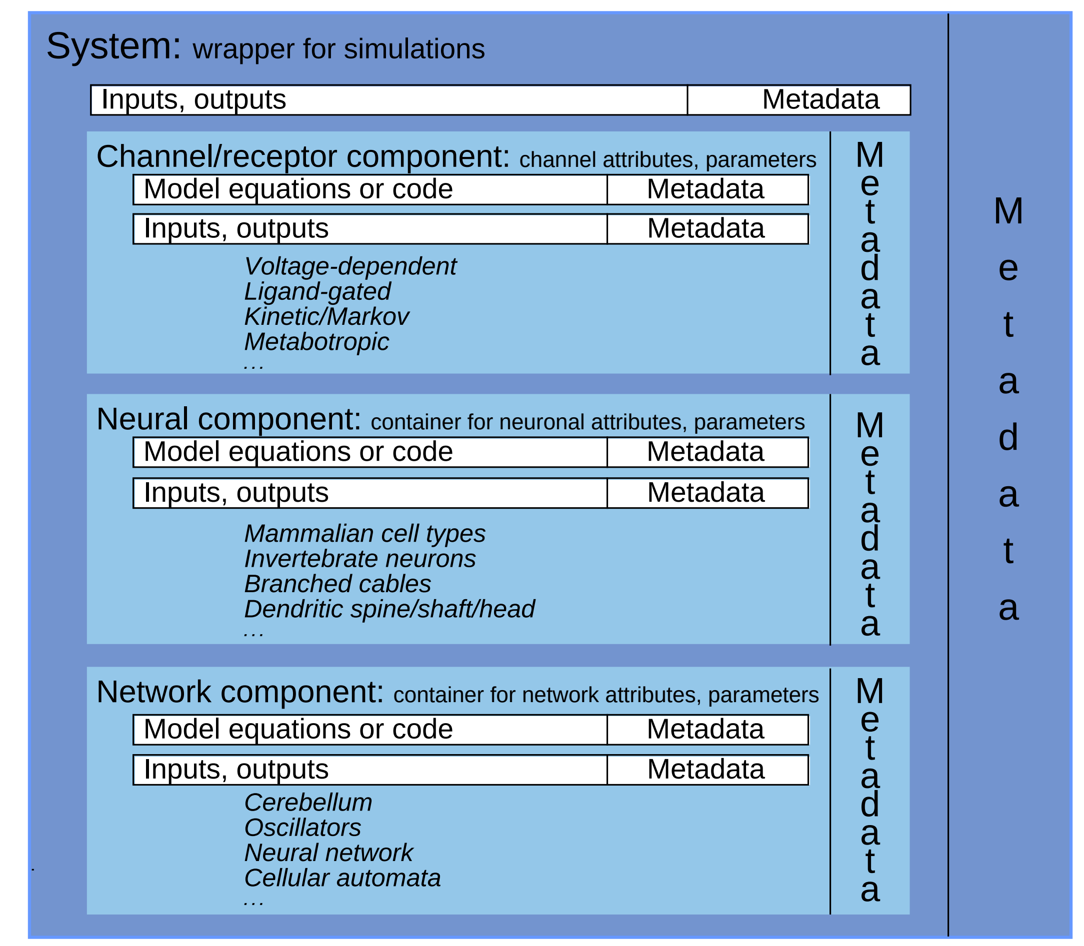


3. QueryTool and Metadata Viewer Use Controlled-Vocabulary Descriptive Attributes to Search and Describe Data, Sites, and References

QueryTool (above) provides database access via metadata searches. Values of most attributes are organized in a Hierarchical Attribute-Value controlled-vocabulary implementation. Metadata Viewer (below) uses the same controlled-vocabulary metadata to describe returned datasets and related neurons and protocols.

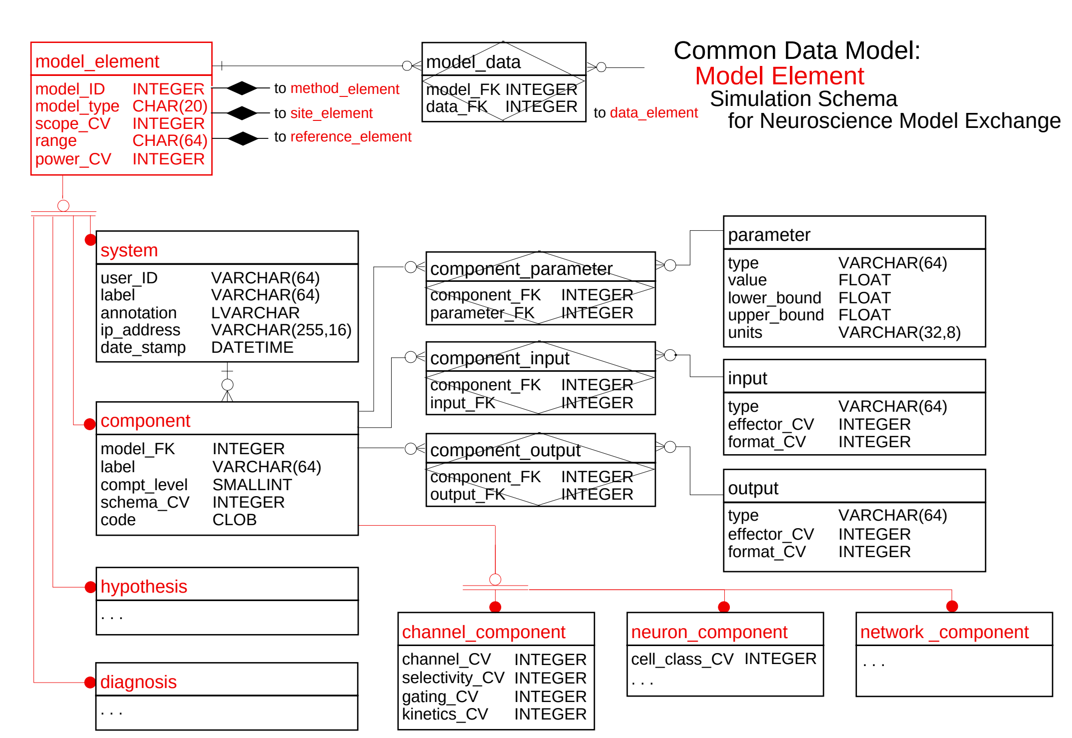
The reference is a live link, bringing up PubMed.

Which Descriptive Attributes Best Specify Neural Models and Data?



4. 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.



5. 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
  - . . .
- . . .

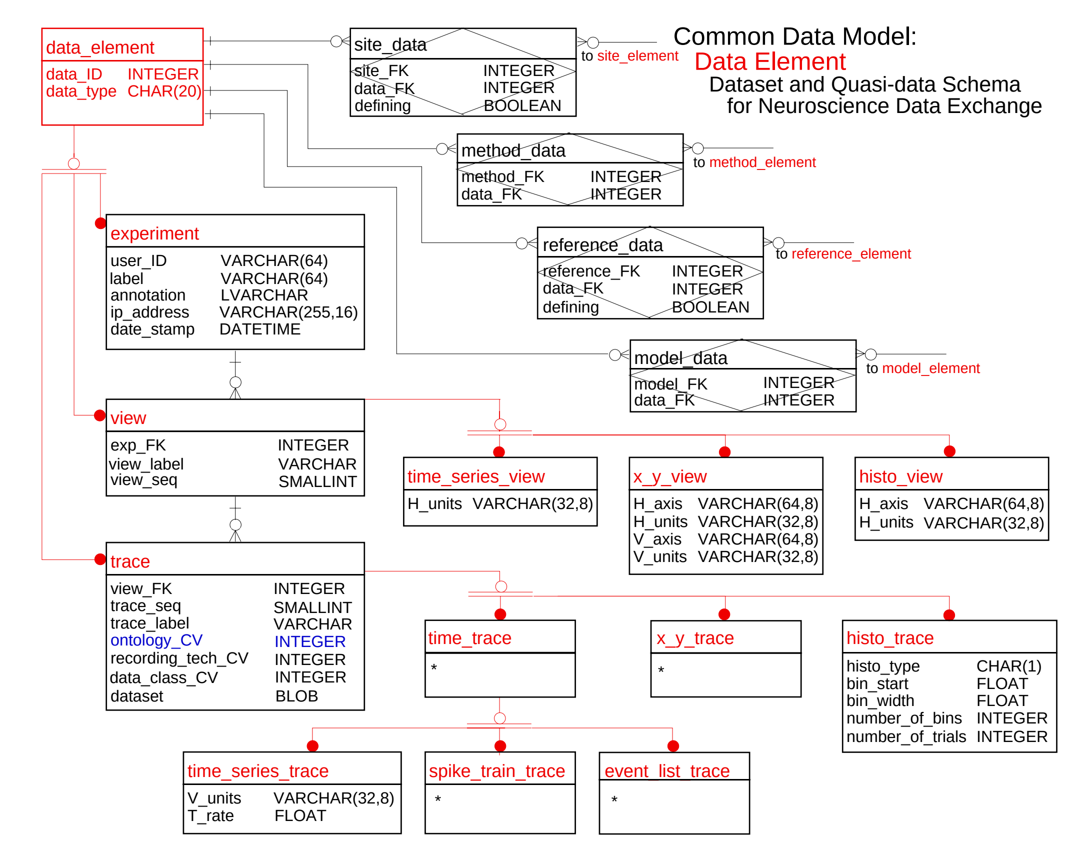
Recording technique:
- extracellular
  - single electrode
  - multielectrode array
- intracellular
  - voltage-clamp
- patch clamp
  - cell-attached
  - o-o
  - i-o
  - whole-cell
- macroelectrode
- nerve cuff
  - field/surface
    - scalp
    - cortical
- optical
  - neuro-intrinsic
  - neuro-via-dye
  - 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
- . . .

6. 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, allow selectable operational specificity.



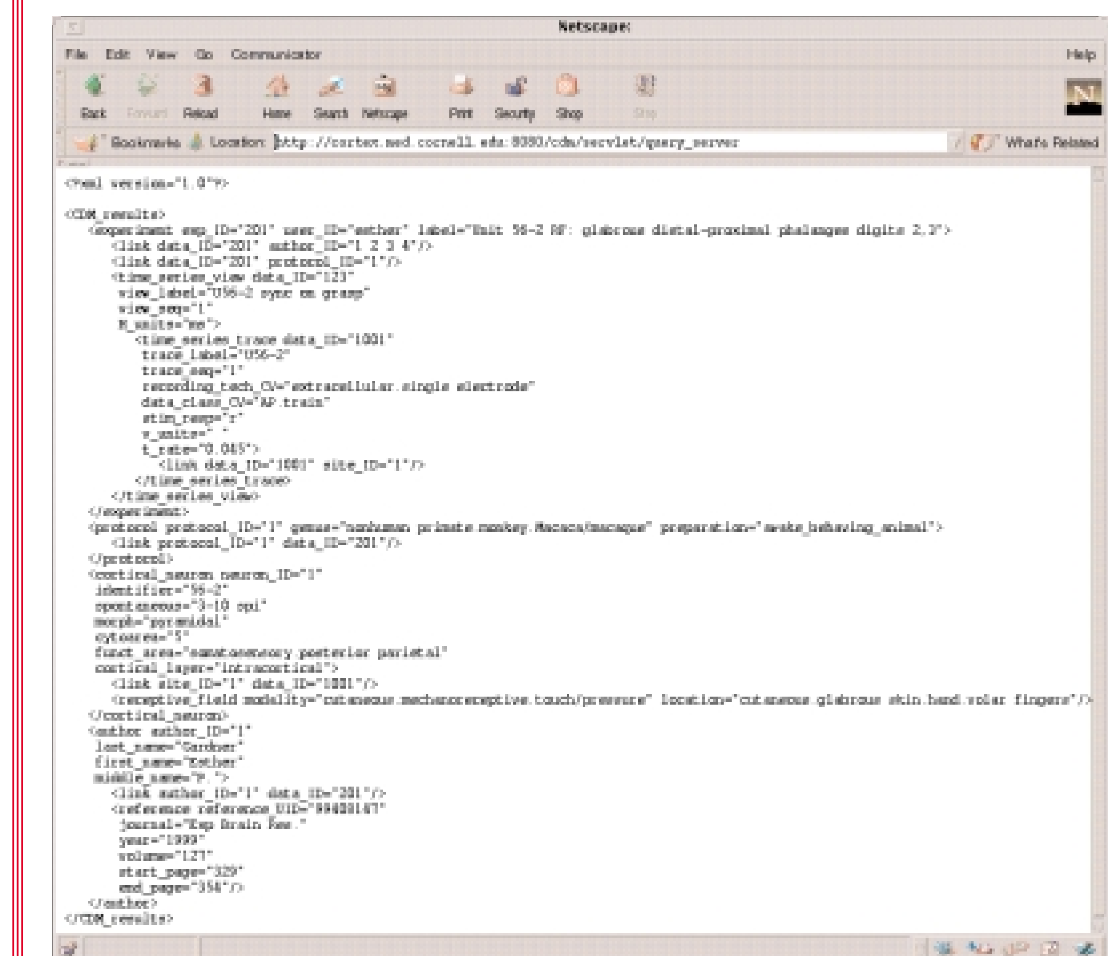
7. Data Element Represents Datatypes and Wrappers for Sensor Data and Simulation Quasi-data

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.

URLs:

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

Will XML Interfaces Link Disparate Databases and Simulator Archives?



8. BDML Enables Interoperable Data Exchange

Any data resource that posts an XML-based query can utilize these XML-wrapped responses from a BDML-enabled database.

```
<?xml version="1.0"?>
<CDM_Model>
<MODEL scope="Phototransduction" range="Photoreceptor to Horizontal Cell"
       power="verifiable-multiple">
   <SYSTEM user_ID="343" label="Neg. Feedback in H and C Cells" >
      <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>
   <AUTHOR>Schnapf et al</AUTHOR>
   . . .
</REFERENCE>
</CDM_Model>
```

9. Prototype Model Description Using BDML

Sample BDML definition of a photoreceptor model from Schnapf *et al.* Prototype attributes, elements and values for illustrative purposes.

Not Conclusions, But More Questions:

- Would you use 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?