

# A Dynamic Publication Model for Data-driven Neurodatabases Implements Interoperability for Neuroscience Information Resources. Daniel Gardner, Michael Abato, Kevin H. Knuth, Robert DeBellis & Steven M. Erde. Lab. of Neuroinformatics, Dept. of Physiology, Weill Med. Coll. of Cornell Univ., NY, NY. 748.15

## A DYNAMIC PUBLICATION MODEL FOCUSES NEUROPHYSIOLOGY DATABASES

In a series of presentations at prior SfN and Biophysics meetings, we have detailed the design and implementation of two neurophysiology databases, including their data model, database schema, midlayer processing, protocols, user tools, and controlled vocabularies.

In this poster, we show how a data-driven dynamic publication model underlies our design. A familiar journal article metaphor specifies the type, scope, organization, and quantity of data, informs design goals, and facilitates ease of use and interoperability.

In the adjacent poster, we extend both underlying model and tools to cover simulations and enable comparison of data records to simulation results.

## DATA-DRIVEN DATABASES FEATURE DATASETS, DATA VIEWS AND METADATA

- + Data are persistent and model-independent
- + Conforms to journal metaphor
- + Reanalyzable data are multiply useful
- + Data quality is largely observer-independent
- + In contrast, the knowledge base is complex and fluid
- + Need for database moderators or curators is reduced
- Delivery of data and views requires advanced methodology, although
- + Methodologies are applicable to quasi-data
- ~ Fits some domains of neuroscience better than others

### 1. Data, Not Inferences, Enable Databases

A data-driven schema provides graphical neurophysiology data from which interpretations may be derived, resembling Results sections of journal papers. Unlike inferences or statements based on data, which are conditional, complex and changeable, datasets are ideally persistent and model-independent. We thus design methods and standards for archiving, searching, browsing, displaying, and exchanging reanalyzable datasets—with related recording sites, methodology, and models—rather than attempting to encode, index, and deliver inferences or conclusions. The adjacent poster shows such methods are applicable as well to quasi-data from simulations.

Signs signify + advantages, - disadvantages, and ~ neutral characteristics of data-driven databases.

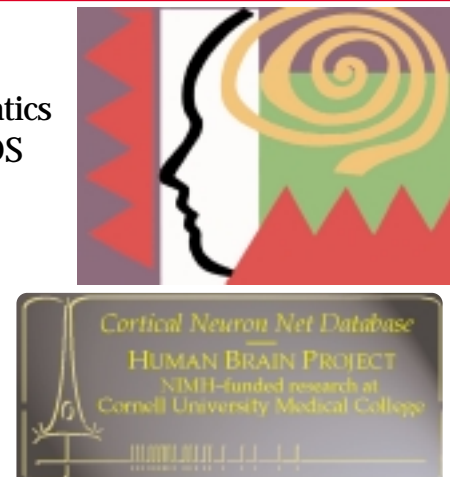
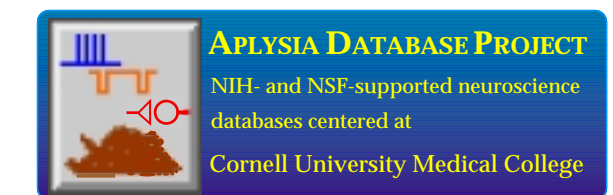
### URLS:

- Cortical database: [cortex.med.cornell.edu](http://cortex.med.cornell.edu)
- Aplysia database: [mollusc.med.cornell.edu](http://mollusc.med.cornell.edu)

### SUPPORT:

This Human Brain Project/Neuroinformatics research is funded by NIMH and NINDS via MH57153 and NS36043.

Data from the lab of Esther P. Gardner.



## THE DYNAMIC PUBLICATION MODEL ESTABLISHES SCOPE, TYPE, AND QUANTITY OF DATA

### 2a. Publication-Model Databases Complement and Transcend Papers

#### Similarities

• All entries are attributed, with submission analogous to authorship.

• An experiment wrapper organizes related data, as does a single publication.

• Like journal papers, reading or viewing is open but submission is restricted or reviewed.

• Multiple indexing terms enable emergent specificity for searches.

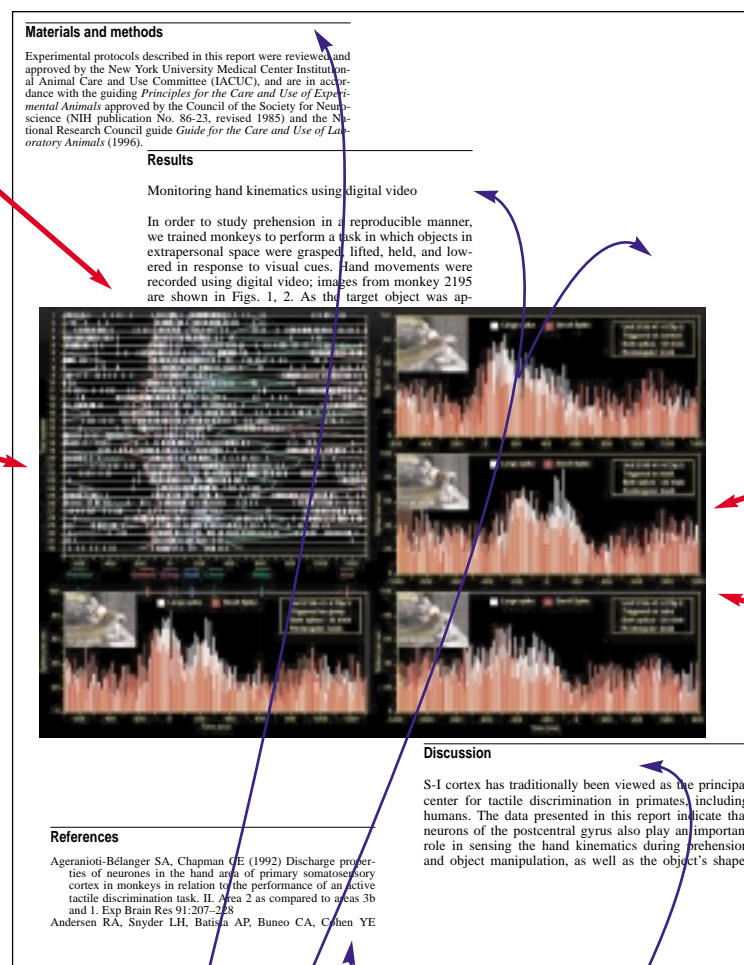
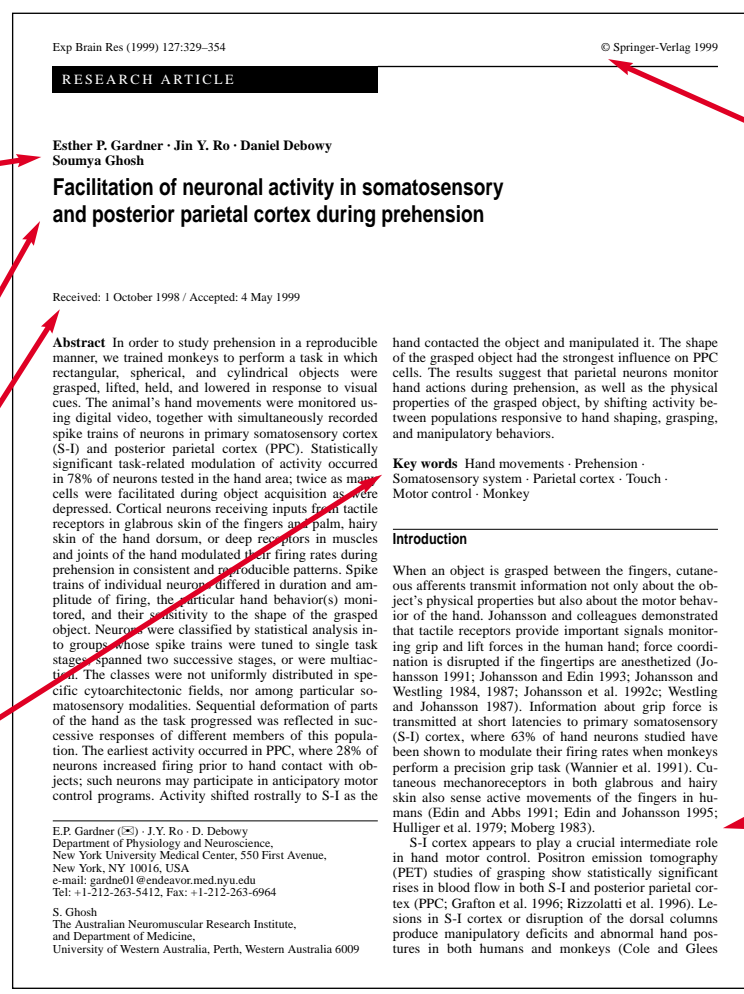
• A uniform but open framework organizes without restriction.

• Selected—characteristic or defining—data are shown, rather than a complete or exhaustive set.

• Data presented as graphs, parsed into figure-like views, comprised of one or more traces.

• Users favor submission coincident with acceptance of related papers.

• Work needed to prepare and submit data should be no more than needed to make comparable figure.



### 2c. The Common Data Model Abstracts a Data-Driven Schema Using Five Publication-Like Superclasses

- **Method Elements** use metadata to encode and describe protocol and preparation information.
- **Data Elements** include figures and metadata, analogous to a Results section.
- **Model Elements** include hypotheses, diagnoses, and other components of a Discussion section.
- **Reference Elements** formalize bibliographic links and attributions.
- **Site Elements** specify recording sites and anatomy found throughout journal articles. Implementing each of several anatomical schemes organizes without restricting choices.

#### Differences

• Copyright is retained by submitters.

• Descriptive metadata use controlled vocabulary to annotate datasets, recording sites, methodology, models and hypotheses.

• Metadata enable search by multiple attributes.

• The data-driven database model minimizes textual hypotheses, tests, interpretations, and commentary

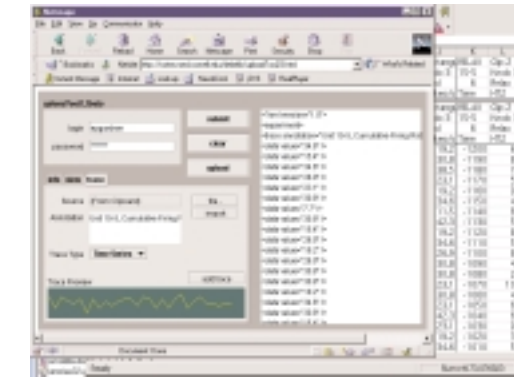
• Unlike paper-based or page-oriented PDF articles, neurophysiology data display and delivery—and controlled vocabulary hierarchies—require new methods for multiplatform delivery.

• The Virtual Oscilloscope permits dynamic control of display and extent.

• Dynamic data displays avoid infringement of journal copyrights.

• Access to the actual datasets permits reanalysis and comparison

## NEURODATABASE IMPLEMENTATION SHOULD EMULATE THE UTILITY, OPENNESS AND UNIVERSALITY OF PUBLICATIONS



### • DataEntry Tool

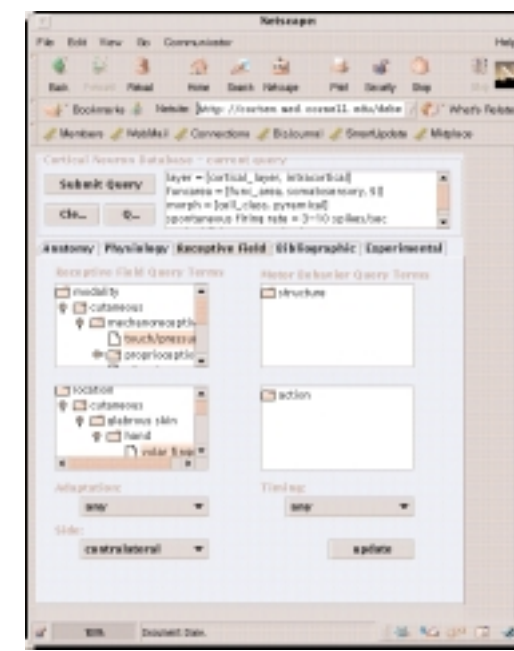
### 3a. Neurobiologists Upload Datasets and Select Descriptive Metadata

On any local computer, the DataEntry Tool allows submitters to organize their datasets as for figures, annotate by selecting descriptive controlled vocabulary metadata, and upload data and metadata to a compliant neurodatabase via XML, Java, and a midlayer DataServer that adds date and author tags.

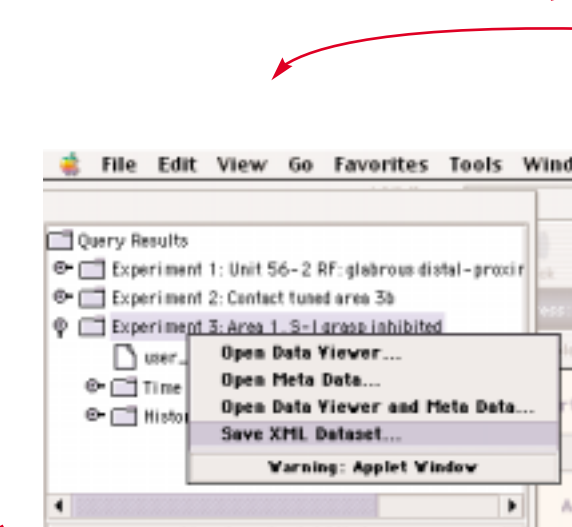
### 3b. Neurobiologists Select, View, and Acquire Datasets and Metadata

The QueryTool enables neurodatabase search by multiple controlled vocabulary descriptive attributes. The Query Results screen returns sets of experiments matching search criteria from the neurodatabase, offering a choice of datasets and also actions: viewing data using the Virtual Oscilloscope, examining metadata via the Metadata Viewer, or saving datasets in XML or other formats for further analysis.

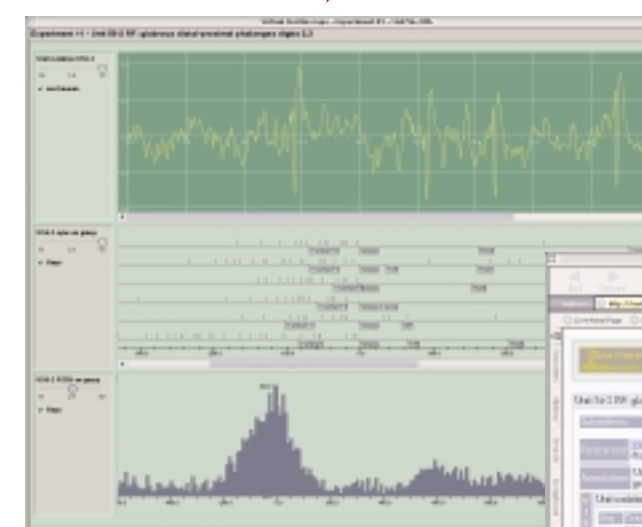
Data flow thus resembles the sequence database model, modified to conform to neurophysiology data.



### • QueryTool



### • Query Results selection and Action PopUp



### • Virtual Oscilloscope

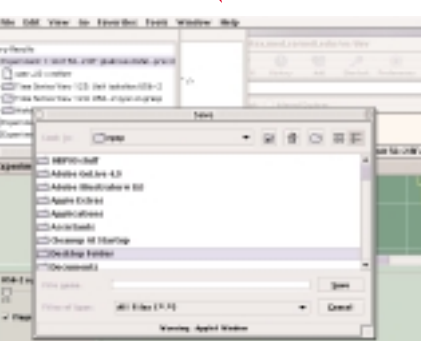
### • Metadata Viewer



### • Remote DataServer and Neurodatabase



### • Remote DataServer and Neurodatabase



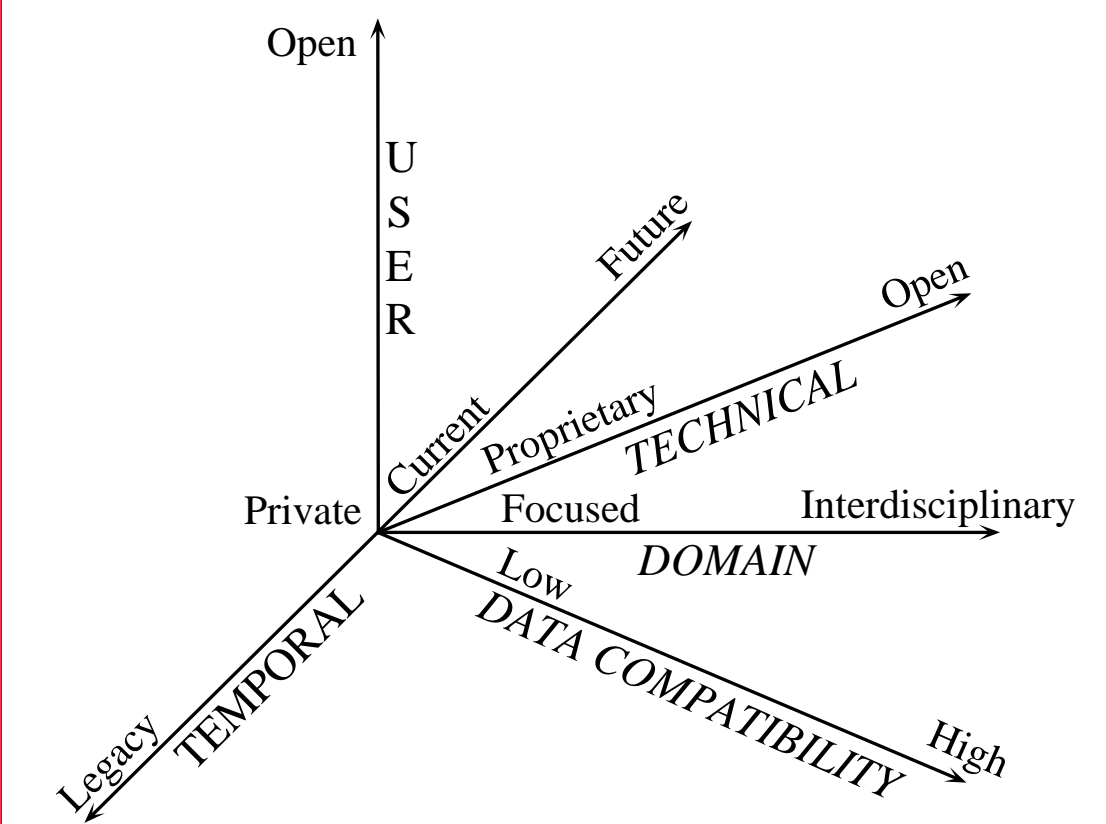
### • Save screen & XML

## IDEAL NEURODATABASES ARE AS USABLE AND INTEROPERABLE AS PAPER JOURNALS

- MacOS 8.1 to 9; Internet Explorer 5; MRJ
- Windows 98 or NT; Netscape; Java Plugin 1.1.3
- Five Unix versions:
  - SPARC Solaris >2.5.1; Netscape; Java Plugin 1.1.3
  - RedHat Linux >6.1; RedHat glibc Netscape 4.72rh
  - SGI IRIX >6.5; Netscape 4.72
  - HP-UX 10.20; Netscape 4.72
  - IBM AIX >4.2; Netscape 4.72
- Each enhanced by Sun's swingall.jar

### 4. Seven Reference Platforms Offer Users Full Database Access on Most Modern Desktops

Aiding database/journal complementarity, multiplatform user tools allow neurodatabases to be accessed almost as freely as print media.



### 5. Interoperability Has Multiple Dimensions

Interoperability of data-driven resources is schematized by escaping the origin in each of several dimensions. *User*: Community-wide databases and multiplatform software promote openness. *Technical*: Open architecture and nonproprietary standards and formats aid data exchange. *Domain* and *Data*: Reliable data and queries link databases covering domains of neuroscience with different data models, techniques, and practices. *Temporal*: Standards independent of current methodology, support permanence, including data persistence, evolution of technology, and inclusion of legacy data.

## CONCLUSIONS:

Data-driven publication-based models:

- Embody intuitive views and organization of data,
- Span neurophysiology datasets and metadata,
- Fit contemporary database & network methods,
- Focus on data useful for exchange, collaboration, comparison, and post-hoc analyses, and
- Promote interoperability