FACETS

FP6-2004-IST-FETPI 15879

Fast Analog Computing with Emergent Transient States

Report on Comparison of Existing Software Structures and Data Models in Single Neuron Databasing

Report Version: 1.0

Report Preparation: Asif Jan, Felix Schürmann

Classification: Pub

Contract Start Date: 01/09/2005 Duration: 4 Years

Project Coordinator: Karlheinz Meier (Heidelberg)

Partners: U Bordeaux, CNRS (Gif-sur-Yvette, Marseille), U Debrecen, TU Dresden, U Freiburg, TU Graz, U Heidelberg, EPFL Lausanne, Funetics S.a.r.l., U London, U Plymouth, INRIA, KTH Stockholm

Project funded by the European Community under the "Information Society Technologies" Programme

DELIVERABLES TABLE

Project Number: ***FP6-2004-IST-FETPI 15879***

1)Project Acronym: FACETS

Title: ***Fast Analog Computing with Emergent Transient States***

Del. No.	Revision	Title	Type[1]	Classification[2]	Due Date	Issue Date
12	**1.0**	**Report on Comparison of Existing Software Structures and Data Models in Single Neuron Databasing**	**R**	**Pub**	**28/02/06**	**31/05/06**

[1] *R: Report; D: Demonstrator; S: Software; W: Workshop; O: Other – Specify in footnote*

[2] *Int.: Internal circulation within project (and Commission Project Officer + reviewers if requested)*
Rest.: Restricted circulation list (specify in footnote) and Commission SO + reviewers only
IST: Circulation within IST Programme participants
FP5: Circulation within Framework Programme participants
Pub.: Public document

DELIVERABLE SUMMARY SHEET

Project Number: FP6-2004-IST-FETPI 15879

Project Acronym: FACETS

Title: Fast Analog Computing with Emergent Transient States

Deliverable N°: 12

Due date: 28/02/06

Delivery Date: 31/05/06

Short description:

This report surveys various single cell databases and their underlying principles and from there derives a database architecture for storing single cell data for the FACETS project. Experimental data on neurons is comprised of manifold information such as electrophysiological response to stimulus, the morphology of the cell, the gene expression data of ion channels, immunohistochemical data on ion channel distributions. Furthermore, it is important to conserve information about the environment of the neuron, ie. the microcircuit they are part of. It is thus desirable to store in addition properties of the microcircuit such as locations and types of synapses, voltage clamp data revealing the ion currents in neurons and pharmacological data on the different pre and postsynaptic receptors used in a microcircuit, numbers and frequencies of occurrence of specific neurons etc. Yet, the context of the FACETS project requires not only the integration the data of several experimental labs but also the availability of this data for the modeling workflow as well as the organization of the individual cell's model data itself and potentially model data of microcircuits. Thus it is imperative for a single cell database to be able to store this multidimensional data and provide means for integrating and correlating the data at various levels of details.

Partners owning:

Laboratory of Neural Microcircuitry – EPFL (8b)

Partners contributed: Laboratory of Neural Microcircuitry – EPFL (8b)

Made available to:

II.INTRODUCTION

The amount of data on single cells accumulated over more than a hundred years in Neuroscience is immense and growing steadily. Unlike the Human Genome Project [1] which has been the unifying movement in genomics, primary data in the field of neuroscience – regardless the efforts of the Human Brain Project [2] – rarely leaves the individual lab. Theoretical neuroscience on the other hand requires easy access to the source data and besides individual collaborations; a standardized and unified exchange has not been established yet.

A consortium like FACETS - in which several experimental neuroscience labs as well as modeling groups are participating - thus may be an ideal test scenario for the establishment of collaboration wide databasing system for individual cell data, from the anatomy and electrophysiology to the functional complex models as well as simplified models that in the context of FACETS ultimately are implemented in hardware. Specifically, the goal of modeling V1 requires the consolidation of *in-vivo* and *in-vitro* data, yet even goes beyond the individual cells: The database should be able to hold as much of the context of a given cell (e.g. connections) to ultimately allow the reconstruction of the complete microcircuitry. In order to provide an optimal database platform supporting the needs of theoretical as well as experimental neuroscientists, a field wide survey of existing data management standards and implementation was carried out. This report contains the results from the fore mentioned survey.

This document is organized as following; Section II provides a survey of existing projects dealing with the issues of Neuroscience data storage, description and sharing, Section III proposes an architecture for the single cell database incorporating concepts and guidelines from Section II, and Section IV contains the conclusions.

III.Comparison of Existing Software Structures and Data Models

Recent years have seen significant efforts in designing the data management, modeling, analysis, simulation, and visualization tools for Neuroscience. This has resulted in multitude of informatics projects, supported mainly by the Human Brain Project, in the areas of data management, analysis, visualization, modeling and simulations. Due to the nature of the Neuroscience research each of these projects covers a subset of the knowledge acquired as part of the experiments i.e. Cell Centered Database (CCDB) [3] for collecting electron tomography data, Neurodatabase.org [4] for storing neurophysiology data, and SenseLab [5] for collecting multidimensional data of olfactory pathways. In addition there are number of efforts underway for designing common

vocabularies and ontologies; these efforts include but are not limited to BrainML [6] for describing and exchanging neurophysiology data, NeuroML [7] for supporting unified structure for modeling and simulations, and MorphML [8] for describing and exchanging morphology data. Furthermore there are tools for modeling and simulating the behavior of single neuron as well as that of networks of neurons e.g. NEURON [9], GENESIS [10], NeuroConstruct [11] etc. Recently there have also been significant efforts for designing mediator systems providing access to multiple heterogeneous data sources examples include Knowledge Integration of Neuroscience Data (KIND) [12] and Query Integrator System (QIS) [13].

In spite of all above-mentioned projects there is not a single database project providing facilities for storing single neuron data covering various neuron properties e.g. morphology, electrophysiology, computational models etc. The following paragraphs describe some of the important projects addressing data management issues for Neuroscience experiments in general. These projects can be classified into two categories i.e. guidelines for standardized data sharing and specific database projects.

A.Towards A Standardized Description And Sharing of Neuroscience Data

OECD Neuroinformatics Working Group Report: This report [14], published in June 2002, examined the technical as well as social issues hindering the globally open collaborations in the field of Neuroscience. The report recognized the fact that there are hundreds of thousands of research groups working in the field of Neuroscience, collecting data on various aspects of brain as diverse as chemical, biophysical, structural, morphological, physiological aspects. The data is collected at different levels i.e. single molecule to whole brain level, and often with different time scales i.e. ranging from microseconds to several days or even weeks. Due to difference in the methods, purposes and scales of data collection, it is often difficult for researchers to share this data for the benefit of research community as a whole. The report then established few basic guidelines aimed at streamlining the future work in the field of Neuroinformatics. These guidelines include integration of existing projects via a globally accessible web portal. This portal serves as starting point for locating various database/informatics efforts thus facilitating collaboration at a global scale. Another important guideline deals with using standards for data and method sharing, and development of common ontology. Both of these objectives are addressed by projects like CDM, and various other efforts such as NeuroML, MorphML, NeuronNames etc (as described in subsequent paragraphs). Lastly, the report encouraged

individual scientists to share their primary data.

Common Data Model (CDM) for the Neuroscience: The "Common Data Model for Neuroscience (CDM)" [15] defines core entities for Neuroscience data. These abstract elements provide the essential foundations for designing interoperable data models. They can be extended by the designer of new data models in order to suit site-specific requirements. The CDM is organized around a set of five basic entities, i.e. *data*, *site*, *reference*, *model* and *method* - each encapsulating a distinct aspect of the data model. All other entities in the data model are to be derived from these basic entity types (or super-classes as referred to by the authors). The model has extensions for representing different neurophysiologic entities such as neurons, connections, axons, dendrites, experimenter information, references to scientific publications relating to the data, and references to the raw data as part of the scientific experiment. As indicated by its name, CDM does not restrict itself to storing or describing single cell data but rather aims to provide vocabularies and schemas for sharing Neuroscience data in general. The focus of the project is on data sharing and the higher level schemas are independent of underlying databasing schemes, i.e. the only constraint is to provide a mediator layer intercepting queries from CDM system and returning response to CDM system.

Modeling Languages and Ontologies: In addition to these major database initiatives, there are numerous efforts for designing common ontologies, vocabularies and modeling languages for representing the neuroscience data from different perspectives. These languages include but are not limited to: "NeuroML" [7] for modeling computational neuron models, "MorphML" [8] for representing neuronal morphological data, "BrainML" [16] for providing meta format for exchanging neuroscience data, and NeuroNames [17] for brain structure names.

Apart from the efforts in promoting common vocabularies, ontologies and data models, there are a number of database projects dealing with the issues and challenges of storing Neuroscience data. As highlighted in the OECD report and stressed further in the CDM project; each of these databases represents needs of a specific research group and does not provide a universal platform for storing, managing and sharing data related to various aspects of Neuroscience research. Furthermore most of these projects have grown from simple prototypes whose sole objectives were to make data accessible via a web page. Consequently these projects did not include many guidelines and standards as advocated by OECD working group or CDM etc. Nevertheless there is a growing realization for using standard-based data description and sharing leading to

significant remodeling of existing projects and/or adding mediator layers publishing propriety data formats to a standardized (mostly XML based) format.

B. Neuroscience Databases

Cell Centered Database (CCDB): The Cell Centered Database (CCDB) is designed to store and manage morphological and protein localization data at cellular and sub-cellular level derived mainly from electron tomography. The CCDB manages and records the whole process of image reconstruction from experiment preparation, to tissue processing, reconstruction and analysis. CCDB is being developed in the context of data produced at the National Center for Microscopy and Imaging Research (NCMIR) (http://ncmir.ucsd.edu). It is implemented using an Oracle database for managing the descriptive metadata and Storage Resource Broker (SRB) server for storing the images and related data files. The oracle database contains information about experimental setup, imaging and reconstruction details, and links to the physical locations of the data. The example datasets stored as part of CCDB include low resolution tomograms of neuronal spiny dendrites, tomograms of multi-component structures like Node of Ranvier, protein localization data derived from immunocytochemistry, enzyme histochemistry, protein specific dyes etc. CCDB deals mainly with morphology and tomography data and is not intended to be used for storing other properties of neurons such as electrophysiology, computational models etc. An important component of the CCDB project is a knowledge based data mediator called KIND (Knowledge Integration of Neuroscience Data) [12]. Using the conceptual model of the underlying data sources the KIND mediator forms semantic networks of terms and relationships, thus allowing interoperability amongst different data sources. Similar to CDM, any underlying database system can be made conformant to the KIND mediator.

SenseLab: The SenseLab project [5] aims to integrate multidisciplinary models of neurons and neuronal system in order to facilitate advanced analysis tools and techniques in Neuroscience research. The project contains seven different database systems arranged in three categories namely “neuronal databases”, “olfactory databases” and “disease databases”. Of interest to this report are the neuronal databases that include “CellPropDB” for storing data regarding membrane channels, receptor and neurotransmitters; “NeuronDB” containing three types of neuronal properties i.e. voltage gated conductances, neurotransmitter receptors, and neurotransmitter substances; and “ModelDB” for storing computational models for neurons. The databases are designed as per the EAV/CR (entity-attribute-value with classes-relationships) representation.

The EAV-CR representation allows an efficient design of database schemas that are frequently updated. Neuronal databases provide a nice framework for storing key properties of neurons and linking them to various neuron models; but the databases are limited in a sense that they do not provide any means to store raw data (resulting from electrophysiology recordings etc.) or data about different dimensions of single cell experiments i.e. morphology, gene expression etc. Additionally, the databases are designed mainly for displaying the data as web pages which makes it difficult for different application programs to interact with the database.

Allen Brain Atlas Project: The goal of the Allen Brain Atlas project [18] is to produce a genome scale collection of gene expression profiles throughout the brain of the mouse. It will contain cellular resolution insitu hybridization (ISH) data across the entire brain for every gene. The gene selection and probe design has been achieved using data contained in Refsq, TIGR, Celera and Riken FANTOM3 databases. So far, approximately 20,000 genes have been assayed in the sagittal plane, and 3500 genes have been processed in the coronal plane. Over 600,000 mouse brain section have been generated and processed as part of the project. The project is concerned with the detailed accumulation of gene expression data only and not with data at single cell level and/or for data belonging to morphology, electrophysiology or computational models.

NeuroSys: The NeuroSys project [19], another data management effort supported by the Human Brain project, aims to provide an easy-to-use semi-structured data management system for the individual scientists and research laboratories. The project tries to separate the complexity of designing the database schema and data entry as well as query interfaces from the database usage. The focus of the project is to use the XML languages for describing and storing the data. The database contents are then exported via dynamically generated GUI components. NeuroSys can be considered to be a lab inventory system rather than a system capable of storing single neuron data.

As can be seen from above section, most of the database projects are focused around a very specific aspect of Neuroscience data thus fulfilling the needs of the local research group or a research group working for specific scientific objective. Presently there is not a single project that provides a framework for storing single cell data at multiple dimensions - thus making it imperative to design and subsequently implement a database scheme capable of storing single neuron morphology, electrophysiology, gene expression, and computational models data. At the same time, such an effort should provide methods for integrating this data and exposing a unified

interface for users and application programs to browse, analyze and download the data. Furthermore, the database platform shall be designed in a manner which facilitates integration with existing informatics projects, thus compiling with OECD guidelines in general.

The following section describes the proposed architecture for single neuron database including an illustration of the data model used for the project highlighting security and data sharing issues.

IV. Proposed Architecture for Single Neuron Database

The current database project is being undertaken in order to allow archiving and searching of data and the easy exchange thereof between locally separated scientists. It should enable active utilization of the data for informatics research as well as support automatic retrieval by large amounts of data for third party analysis applications or by other databases and search engines conforming to standards as advocated by CDM and OECD. Following are some of the important design considerations for the single cell database architecture:

1. **The Architecture Should Reflect the Underlying Data Hierarchy**

   Microcircuit data has many multidimensional hierarchical levels. The structure of a neuron is analogous to a tree with a large number of "roots" (axons) and "branches" (dendrites). Each neuron has a dendritic tree that receives and an axonal tree that forms a specified number of synaptic connections. Specific parts of the axons (the presynaptic innervation pattern) touch the dendrites of a certain fraction of neighboring neurons in a very particular manner (the postsynaptic innervation pattern). The result is that each neuron is packed with 1-10,000 synapses according to specific rules. When several hundred or thousands of neurons are connected together then multiple constraints must be taken into consideration to allow all the "pieces" to fit together. In addition, each synapse has potentially unique physiological properties. We therefore propose to store not only the attributes of each cell, but also the information that links the attributes thus making the data structure inherently relational. The data model must therefore have the capability to store the hierarchical data with minimum overhead. This will also be essential for mining the single cell data and for building custom versions of the microcircuit.

2. **The Architecture Should be Extensible and Support Schema Evolution**

   As the scientific data may expand and even change form (such as introducing ion current, pharmacology, neuromodulation, and presynaptic and postsynaptic receptor data), it

introduces the problem of schema evolution i.e. where the schema of the existing database is not appropriate for storing the new data. This normally results in upgrading or redesigning the existing database, which is an expensive operation. Therefore, any architecture for storing the neuroscience data must provide a mechanism to store newly discovered facts without having to redesign the existing database. Therefore, the database structure should be composed of various hierarchies that can be integrated together in order to provide a unified view to single cell data.

3. The Architecture Should Preserve Data Privacy While Encouraging Data Submission

Populating neuroscience databases by many different research groups is a major challenge because of a) lack of interest by some groups, b) privacy concerns, and c) lack of agreement on the structure of data storage and the standards implemented in the database. The architecture should therefore provide a privately useful tool for different research groups, allow for adaptation of the forms of data submission and enable easy data submission. A researcher shall be able to configure the access rights to the underlying data in terms of who can access the data and for what purposes. Furthermore, a focus on proper description of the data will facilitate subsequent efforts for improving data interoperability.

4. The Architecture Should Explicitly Address the Issues of Interoperability

The architecture of the data management solution must explicitly address the issue of interoperability with existing Neuroinformatics resources. It must have the capability to publish its data into existing Neuroscience databases; at the same it shall also provide mechanisms for searching and correlating the data with other databases. The architecture, thus, must conform to the design guidelines as presented by the latest research in the field of information integration across Neuroscience databases. The architecture shall provide provisions of interoperability at the schema level, i.e. by exporting the database structure as a self describing XML document, as well as at the knowledge level i.e. by implementing and adhering to the existing ontology and taxonomies.

5. The Architecture Should Facilitate Advance Data Mining and Analysis

The current focus in the Neuroscience databases is on presenting a web based system for displaying the database contents for the interested users [20]. Useful as it may be, this has resulted in minimal utilization of the Neuroscience data. The database architecture should be

optimized in terms of its use by advanced data mining tools with considerable knowledge regarding the database semantics and analysis environments. Therefore the architecture must provide provisions for correlating individual neuron and synapse profiles at multiple levels of details, and facilitate offline data mining and analysis.

The following paragraph describes the data model for single neuron database designed utilizing the above mentioned considerations.

A.Data Model for Single Neuron Database

The data model for the single neuron database is based on two key entities i.e. neurons and synapses. The key properties of neurons and synapses are numerically categorized as "profiles" [21]. Neurons may be characterized in terms of their morphological, physiological and gene expression profiles and synaptic connections in terms of their morphological and physiological profiles. Besides improving the structure of the data, the categorization also allows for implementation of an extensible data model for neurons and synapses. For example, a laboratory involved in collecting ion channel data about neurons may wish to include this information in the database. Now since properties of neurons are arranged as part of "profiles", a new profile reflecting the information regarding neuron ion channels may be created and associated with a neuron. Furthermore, the categorization of neuron and synapse properties as distinct profiles also facilitates the correlation of key properties across various neurons and synapses profiles. In addition to these profiles there is data regarding classification of neurons i.e. morpho-electrophysiological classes as well. The synaptic connections are characterized in terms of their anatomy (numbers and distributions of connections, total numbers of boutons/synapses, structural rules of connectivity, functional connection statistics, etc) and physiology (synaptic biophysics, synaptic dynamics, unitary responses, pharmacological properties etc). These data together allow the systematic reconstruction of a microcircuit with as much biological realism as possible, and provides basis for building advance informatics applications for simulating, visualizing and mining functional and structural aspects of neocortical microcircuits. Figure 1 depicts elementary building blocks of information needed for reconstructing neocortical microcircuit.

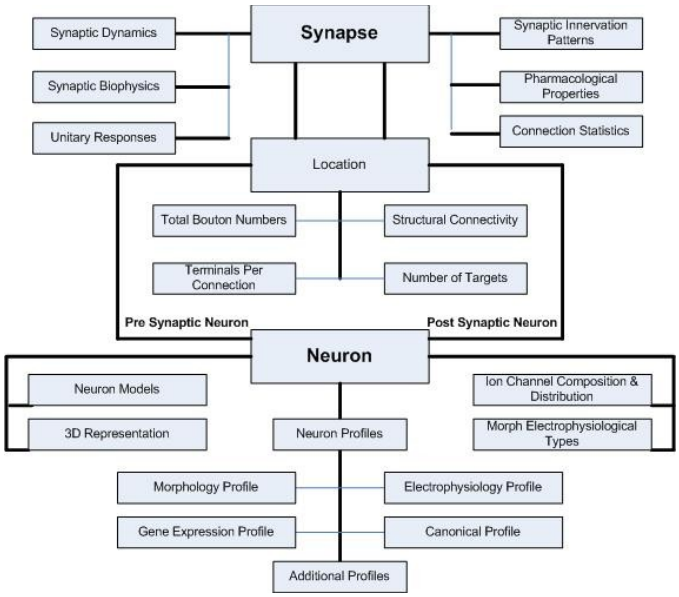

Figure 1: Elementary building blocks of information for neocortical microcircuit

The profiles and classification refer to the numerical characterization of key properties of neurons and synapses in the microcircuit. These numerical values are extracted from the raw data resulting from performing various experiments. This raw data is referred to as the properties of neurons in the current document and includes electrophysiology recordings, morphological reconstructions, image stacks etc. Figure 2 depicts the logical relationship between a neuron and its classification, profiles and properties. Taken collectively profiles, classification and properties allow us to look at various functional as well as structural aspects of neurons and look at the neuron's data from different levels of details. For example, if we want to look at the pyramidal cells from layer 5, then we use classification information to get a list of all pyramidal cells for layer 5. This list will then be filtered based on the presence and absence of key numerical parameters stored as part of neuron's electrophysiology and morphology profiles. Now if we are interested in looking at detailed electrophysiology recordings or morphological image stacks we can browse the properties for the neuron. Thus, taking an analogy from world wide web (WWW), a neuron acts like an index page containing links to its classification information, key numerical characterization and corresponding raw data. At a different granularity, individual neuron's data is linked together to provide a detailed representation for neocortical microcircuit.

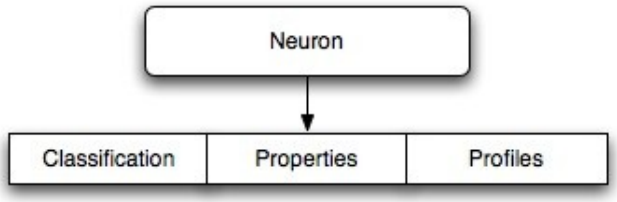

Figure 2: Categorizing neuron's data

B.Design And Implementation Details

As mentioned in Section III.A and depicted in Figure 2, there are three distinct logical levels at which data is managed in the current system i.e. providing an indexing structure, managing the raw data, and managing key numerical properties. These three levels in turn are composed of various sub levels each addressing specific objectives. In practice these levels correspond to the manner in which data is collected from experiments, archived as part of central data repository, analyzed in order to extract key numerical parameters, and searched for the presence and absence of various criteria. The process consists of various steps such as constraining users to use appropriate metadata at the data archival time, constructing a central index for storing the metadata information, providing structure for storing raw data corresponding to various properties of neurons, and providing means to store the key numerical parameters. Figure 3 depicts the design of the data management environment, which is explained in subsequent paragraphs.

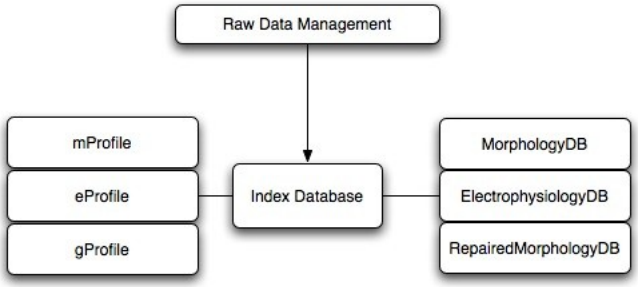

Figure 3: Various components of the data management platform

Managing the raw data

Experiments are performed by individuals or groups of individual associated with a research organization. The experiments are defined in the context of the organization's scientific

objectives, which in turn are organized as projects. Larger projects are decomposed into smaller projects, each focused around a specific aspect of the overall research, thus resulting into a multi-level project hierarchy. Experiments conducted as part of these projects yield many datasets representing the biological phenomenon in question. These datasets, which can be morphology image stacks, electrophysiological recordings etc, constitute the raw data for the experiments. This raw data is then further processed in order to extract key parameters useful for higher level analysis. These key parameters from a number of related experiments form a basis for scientific publication reporting newly discovered knowledge. Figure 4 depicts the conceptual structure of how experiments are grouped in various projects; subsequent paragraphs provide further details for the above-mentioned concept.

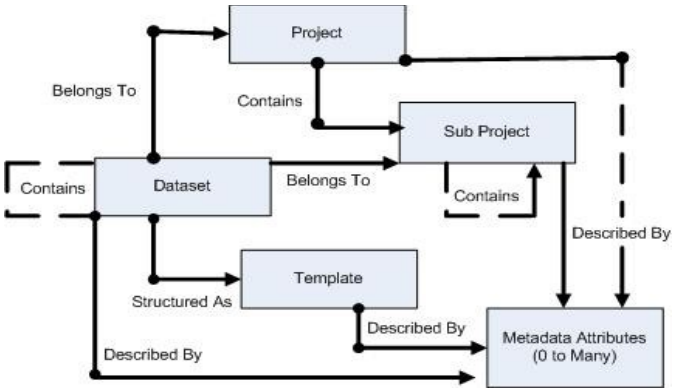

Figure 4: Managing the raw data for experiments

As shown in Figure 4, there are various structural constructs that are used for facilitating proper description of the experimental data. As mentioned above, an experiment is conducted for achieving scientific objectives of a specific project (or subproject), and results into multiple datasets. Thus it can be seen that each experiment will have a specific structure. In addition to the structure of the underlying datasets, each experiment also contains various metadata attributes that describe the experimental conditions and objectives. The structure and the metadata attributes for the experiment are abstracted into a structural template. That provides a blue print for managing raw data resulting from an experiment and also for annotating this raw data with useful metadata attributes. The system provides notions of "System Templates" that mimic standardized structure of experiments belonging to a specific project and used by all individuals at the site; and "User Templates" representing customized structure for individual users.

Managing Data Index

The Index database is the central inventory that contains metadata information relating to various neurons recorded in the laboratory. After collecting the raw data for the experiments, experimenters make an entry to the index database indicating number of neurons recorded for a particular experiment, listing connections for these neurons, and signal as to which properties have been obtained. Also the experimenter indicates information about classification of the given neurons. However, the index database is implemented in a manner that allows experimenters to fill in as many details as possible at the index creation time, and come back later to add new attributes and/or update values of existing attributes. Separating indexing mechanism from the actual data was a design choice taken in favor of easily adding new properties and profiles for neurons. For example, adding a new property for storing ion channel models for a neuron will result into implementing an IonChannelDB and linking it to central index database. Now as models become available for various neurons, this information can be updated in the index database. However, with this separation comes the cost of adding data consistency logic at the application layer ensuring that indexing structure and corresponding profiles and properties are consistent with each other.

Managing Single Cell Properties and Profile Data

Different labs as part of the FACETS project are working at different aspects of single cell experiments i.e. studying electrophysiological behavior *in vitro* and *in vivo*. Thus it is imperative that the database system provides a rather flexible scheme for storing this data. Keeping this objective in mind, the database allows for storing various cell properties (also referred to as raw data) such as morphological reconstruction data, electrophysiological recordings, computational models etc. This data is then linked to the index database in order to expose these properties to the users and application programs. As mentioned above each of these properties database is maintained separately thus allowing us to add new properties without having to change the database schema. Single cell profiles, on the other hand, refer to the key numerical parameters that are extracted from the properties data using custom made analysis programs. At present, only morphometric analysis data (or mProfile) is available via the database. In future data regarding other profiles such as electrophysiology, gene expression etc will also be made available.

Implementing Single Cell Database – Technologies and Web Based Access

Storage Resource Broker (SRB) [22] is used for managing the raw data, and SRB metadata

catalog was extended to provide indexing information. SRB is a client-server middleware providing a scalable and secure wide-area communication infrastructure and is widely used in a number of data grid projects worldwide. SRB also provides a metadata catalog for storing the metadata attributes for the data. SRB server can be accessed using various application programming interfaces i.e. C/C++, Java, Python etc. The current prototype uses the Java based SRB API for implementing the logical structure of the data management. Neuron properties are stored in dedicated spaces in on the SRB server and are linked to the index database. Figure 5 shows the implementation details for the data management platform.

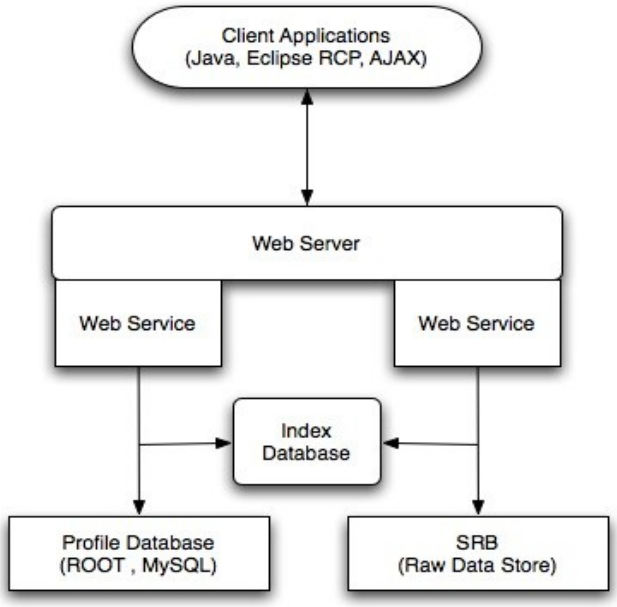

Figure 5: Database Environment

Neuron profiles are implemented using a ROOT [24] based database and a MySQL [23] database management system. One of the aims to have the dual implementation for profile databases was to evaluate the usefulness of ROOT for storing the data as against using a relational database (and SRB based) backend supported by a ROOT based analysis layer. One of the important design choices for implementing server side logic is the ubiquity of client access. The server side logic is built on top of a web services layer thus allowing various clients, including java swing clients, Eclipse RCP [25] client and AJAX [26] bases clients to access the platform and use the data. As an alternate SRB wire protocol based access is also implemented. And it is left to the capabilities of the client applications to choose as to use the SRB protocol or web services protocol. For example, AJAX based clients may always use web services layer where as clients implemented using Java Swing and Eclipse RCP may use both approaches. However, the profile databases are always accessed via a web service interface. As seen from Figure 5, the main access to the system

is via a web server. The web server hosts a number of web services responsible for fulfilling a specific clients requests. One web service connects to SRB based properties database in order to browse, upload and download raw data for the experiments. The other web service connects to the profile database allowing users to look at the key numerical data extracted from the properties database. Both of the web services are connected to the index database thus allowing users to search for specific cells etc. In future, the web services for profile and properties database will be integrated allowing the automatic execution of analysis programs extracting profile information from raw data.

A prototype realizing the described design goals has been developed and was presented to the consortium at month 6 of the project. Its features and capabilities are described in detail in a report accompanying D6.

V. Conclusions

This report highlighted various research projects dealing with data management in the field of Neuroscience and highlighted the challenges for providing a fully functional database platform capable of serving experimental as well theoretical neuroscientists. One of the important issues hindering the global sharing of the neuroscience data in general and Single Cell data in particular is the difference in the data collection methods, objectives and levels – a challenge which for the FACETS consortium is tackled at many places of FACETS but especially with WP8. Meanwhile, for the advancement of the knowledge in the field it is imperative that the data be managed at multiple levels of details and be correlated across different cell properties in order to support advance research in the field. Standardizing the data description and multilevel organization of the underlying database structure can facilitate achieving some of the above mentioned objectives. The report described a prototype database platform demonstrating these principles and provides a basis for discussions and further development in the context of the FACETS project.

References

1. Human Genome Project at http://ornl.gov/sci/techresources/Human_Genome/project/about.shtml
2. Shepherd, G.M. et al. The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. Trends Neurosci. 21, 460–468 (1998).
3. Martone, M.E. et al. A cell-centered database for electron tomographic data. J. Struct.

Biol. 138, 145–155 (2002).

4. Portal for Neuroscience Database at www.neurodatabase.org
5. Miller, P.L. et al. Integration of multidisciplinary sensory data: a pilot model of the human brain project approach. J. Am. Med. Inform. Assoc. 8, 34–48 (2001).
6. BrainML at www.brainml.org
7. Goddard NH, Hucka M, Howell F, Cornelis H, Shankar K, Beeman D. Towards NeuroML: model description methods for collaborative modelling in neuroscience. Philos Trans R Soc Lond B Biol Sci. 2001 Aug 29;356(1412):1209-28. Review.
8. MorphML at www.morphml.org
9. NEURON Simulation Environment, http://www.neuron.yale.edu/neuron/
10. Bower J, Beeman D and Hucka M, The GENESIS Simulation System, The Handbook of Brain Theory and Neural Networks, Second edition, (M.A. Arbib, Ed.), Cambridge, MA: The MIT Press, 2002
11. NeuroConstruct program at www.neuroconstruct.org
12. Gupta, A., Ludaescher, B. & Martone, M.E. Knowledge-based integration of neuroscience data sources. in Proc. 12th Int. Conf. Scientific Statist. Database Management IEEE Comput. Soc. (2000).
13. Marenco L, Wang TY, Shepherd G, Miller PL, Nadkarni P. QIS: A framework for biomedical database federation. J Am Med Inform Assoc. 2004 Nov-Dec;11(6):523-34. Epub 2004 Aug 6.
14. OECD NEUROINFORMATICS working group report June 2002
15. Gardner et al, Common Data Model for Neuroscience Data and Data Model Exchange, J Am Med Inform Assoc. 2001 Jan-Feb;8(1):17-33
16. BrainML at www.brainml.org
17. Bowden, D.M. & Dubach, M.F. NeuroNames 2002. Neuroinformatics 1, 43–59 (2002)
18. Allen Brain Atlas at http://www.brainatlas.org/aba/
19. Pittendrigh S, Jacobs G., NeuroSys: a semistructured laboratory database. Neuroinformatics. 2003;1(2):167-76
20. Nadkarni P, Marenco L, Chen R, Skoufos E, Shepherd G, Miller P.Organization of Heterogeneous Scientific Data Using the EAV/CR Representation. J Am Med Informatics Assoc 1999; 6:478-493)
21. Markram H, Xiaozhong L, Silberberg G, Toledo-Rodriguez M, and Gupta A.(2003) The Neocortical Microcircuit Database (NMDB). Databasing the Brain:From Data to

Knowledge; Koslow SH and Subramaniam S (Ed) Wiley Press

22. Baru C, Moore R, Rajasekar A, Wan M, The SDSC Storage Resource Broker , Proc. CASCON'98 Conference , Nov.30-Dec.3, 1998, Toronto, Canada
23. MySQL Database Management System at www.mysql.org
24. Brun R and Rademakers F, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also http://root.cern.ch/
25. Eclipse Rich Client Platform at www.eclipe.org (technical article)
26. Crane D. and Pascarello E, AJAX In Action, Manning Publishing 2006