A CELL-CENTERED DATABASE FOR MULTI-SCALE MICROSCOPY DATA MANAGEMENT

Shenglan Zhang[*], Xufei Qian[**], Amarnath Gupta[**], Diana Price[*], Mona Wong-Barnum[*], Mark H. Ellisman[*,**], and Maryann E. Martone[*,**]

[*] National Center for Microscopy and Imaging Research, Center for Research in Biological Structure and Dept. of Neurosciences, University of California, San Diego, La Jolla, CA 92093-0608, USA

** San Diego Supercomputer Center, University of California, San Diego, CA 92093-0505, USA

Running title: multi-scale microscopy data management

Correspondence to:

Maryann Martone, Ph.D.

National Center for Microscopy and Imaging Research

Department of Neurosciences

University of California, San Diego

San Diego, CA 92093-0608, USA

Tel: (858) 822-0745

Fax: (858) 822-0828

Email: mmartone@ucsd.edu

Abstract: We have developed an on-line database system (http://ncmir.ucsd.edu/CCDB) for multi-scale 3D structural and protein localization imaging data from light and electron microscopy, including correlated microscopy. The Cell-Centered Database (CCDB) takes advantage of the newly created Biomedical Informatics Research Network (BIRN) Infrastructure (http://www.nbirn.net) to foster microscopy image data federation and distributed transparent retrieval of microscopy multimedia documents. The CCDB was developed using Oracle and models the entire process of 3D reconstruction, from experiment to segmentation, while providing analytical query capabilities. The purpose of the CCDB is to provide a resource for biologists and computational modelers interested in the analysis of distribution of macromolecules in the context of subcellular, cellular and tissue structure. This paper describes the design, implementation and application of the state-of-the-art relational image database management system.

Key words: neuroinformatics, immunocytochemistry, electron tomography, dendritic spines, light microscopy, electron microscopy, multi-scale

INTRODUCTION

Databases are widely utilized today by the molecular biology community for storing gene and protein structural information (Baxevanis, 2002). Data on genes and gene products must be integrated with data derived from other disciplines to aid biologists in unraveling key biological processes ranging from macromolecular structure, to sub-cellular, cellular and tissue organization. Microscopic imaging at both the light and the electron microscopic level provides unique information on protein localization and interactions that extends and enriches that obtained from molecular and biochemical techniques. Efficient storage, fast retrieval and secure sharing of microscopy images are an ongoing challenging. Our goal is to utilize existing information technology and infrastructure to make management of cell level microscopy images easy, fast and comprehensive while providing the means to query information contained in images in meaningful ways. This paper will provide an example of how information technology can be applied for the management of biological microscopy images.

Microscopy image management systems are still in their infancy and most biologists still maintain their images without any metadata structure or quantitative description. Some on line biology image databases (Andrews et al., 2002; Carazo & Stelzer, 1999; Gonzalez-Couto et al., 2001; Flickner et al., 1995; Niblack et al., 1993; Wang et al., 2000) have tried to facilitate image exchange and management, e.g., the QBIC (Flickner et al., 1995), BioImage (Carazo & Stelzer, 1999), OME which focus on biological fluorescence imaging (Swedlow, et al., 2003) and PSLID systems (Wang et al., 2000). The image research community, however, does not have a complete scientific microscopy image management system to maintain and manage rich image files and image metadata acquired by light and electron microscopy techniques. Current data management approaches, including structured flat files or XML-based methods, relational Database Management Systems, object-relational DBMSs and object-oriented DBMSs (Achard et al., 2001) can easily handle the relatively simple requirements of molecular biology data but not the more varied and complex needs of microscopy imaging data. The design and implementation of such image management systems faces several technological and sociological challenges including: 1) Lack of standardization; 2) The large size of image files; 3) The complexity of the data; 4) The reluctance of many biologists to make original data available before they have published on their findings or finished mining it.

We have entered into an era where scientists need to come to a consensus on data sharing (Chicurel, 2000). In the fall of 2001, the Biomedical Informatics Research Network (BIRN) was initiated by the National Institute of Health (NIH) (Marx, 2002). BIRN is aimed at creating a test bed to address biomedical researchers' need to access and analyze data at a variety of levels of aggregation located at diverse sites throughout the country. In its initial phase, the BIRN network involves a consortium of approximately 12 universities and provides the infrastructure for a federation of databases that house brain imaging data, including structural and functional MRI scans of human and mouse subjects, and multi-scale microscopic analysis of mouse models of human disease. Groups are working on large scale, cross-institutional imaging studies on Alzheimer's disease, depression and schizophrenia and animal models relevant to the study of multiple sclerosis, attention deficit disorder, brain cancer and Parkinson's disease. These studies are being used to drive the creation of a biological data grid and persistent archive of biological data.

Using the BIRN infrastructure, our group has integrated the Storage Resource Broker (SRB), a distributed file management system (Rahasejar et al., 2002) and Oracle9i to create a system for storing and retrieving microscopy image files and microscopy image metadata. More importantly, we have designed and implemented a unique data model for microscopy image data for a publicly accessible database for 3D cellular level structural and protein localization data, the Cell-Centered Database or CCDB (Martone et al., 2002a, 2002b, 2003). The CCDB data model addresses complexity, security, and standardization issues of multi-resolution 3D experimental image data. It is supported by Oracle9i and is able to scale, extend and link to other bioinformatics databases through the BIRN infrastructure. The federation of the CCDB microscopy image data management system with other databases will allow the biologists to navigate through many levels of biological complexity and come closer to the goal of understanding biological systems across scales and functionalities. The goals and overall description of the CCDB project have been described elsewhere (Martone et al., 2002a, 2003b, 2003). In this paper, we provide technical details on the CCDB architecture, data model, and approaches to implement query interfaces. Some of technical detail has been addressed elsewhere in Zhang et al. (2003).

MATERIALS AND METHODS

CCDB System Architecture

The CCDB system has been implemented with a three-tier design (Gallaugher & Ramanathan, 1996), which includes a web-based application front-end, a business and transaction logic middle-tier and a data access layer (as illustrated in Figure 1).

Application Front-End: The application front-end provides user access using a standard web browser. The application presents a series of dynamically generated HTML screens that are driven by the application logic layer.

Application Logic: The middle tier contains the application logic and the web server. The applications are written using JavaBeans, Java Servlets, Java Server Pages (JSP), and Java utility classes. The web server delivers dynamically generated web pages to the client and parses client requests. In addition, this layer contains the appropriate drivers to connect to the data access layer.

Data Access layer: The data access layer is comprised of a relational database management system (Oracle) and a SRB server (Rahasejar et al., 2002). The SRB was chosen to manage different formats of heterogeneous CCDB data files distributed on different types of storage devices over the network. The descriptive data is managed on Oracle 9.0.1.4.0 platform (Figure 1), which is currently the fastest, most reliable and most scalable relational database management system (RDMS) on the market (Elmasri & Navathe, 2000). The database is accessed from the application level through a Java Database Connectivity (JDBC) driver. There are three main advantages to store CCDB image files in the separate SRB file storage system. First, the SRB is a grid-based, federated server system provides location transparency and improves both the reliability and availability of image data by replicating it in different storage systems on different hosts (Rahasejar et al., 2002). Second, because most of the image files are binary data and are not standardized across relational database systems, the separation of the file storage from the database will maintain easy accessibility by other tools that need to manipulate the data. Finally, some of the image data will consume gigabytes of space, and thus disentangling the metadata from the data files results in much more efficient I/O access.

CCDB Data Sources

The CCDB was designed to store 2D and 3D light and electron microscopy images and image descriptors in Oracle and SRB (Figure 1). Images that are handled by CCDB can be raw or post-processed such as 3D reconstructions or segmented images. Raw (unprocessed) images, and related post- processed images are defined as one record with the same base name stored in the CCDB. In most current molecular databases, a single entry represents a single protein structure or gene sequence with a unique identifier. The CCDB is distinguished from current molecular databases in that a single record comprises a related set of images and a record is associated with other related sets of records.

Microscopy data resources contain heterogeneous multimedia information. The potential types of files in the CCDB include: 1) Images: micrographs that are encoded in standard formats (e.g. JPEG); 2) Mixed multimedia data: Compressed image files and parameter files that are bundled together as one 3D volume tar file; 3) Animations: A sequence of images (e.g. AVI or MPEG) at different tilt angle or different time to illustrate 3D cell structure; 4) Graphics: Drawings or illustrations that are encoded using some descriptive standards (e.g. PICT); 5) Spread sheet: Formatted files containing analytical information derived from images (e.g. ASCII, Excel).

Although the core of the CCDB is formed by relevant multi-dimensional, multi-resolution, and correlated microscopy images, a description of each microscopic image object also is essential in order to provide identification of the images and information about their content. A comprehensive description of images will make interdisciplinary usage, image retrieval and image federation possible. This image descriptive information is referred to as metadata. Every image object in the CCDB is accompanied by descriptive, structural, and administrative metadata. The descriptive metadata includes experimental descriptions, file parameters, image annotations, image and experimental evaluations, and segmented object measurement and analysis. The experimental description includes subject properties (e.g. animal group, animal treatment, animal age), sample preparation (e.g. tissue thickness, tissue storage location), instrument details and imaging parameters (e.g. instrument, resolution, voltage), and data processing method (e.g. reconstruction algorithm, reconstruction programs), etc. File (image file or non-image file) information includes location, name, file size, file format, file comments, etc. Structural metadata provides an anatomical characterization of the object. This information is important for the federation of images in CCDB with internal correlated microscopy images or images external to the CCDB. Administrative metadata represents information necessary for management of the image object, such as the date it was created, the security level of the image, and the ownership, etc.

CCDB Data Model

The CCDB data model was designed to implement the entire process of 3D reconstruction and integrate different categories of cell-level image data and derived image data products (Figure 2). We use an object relational data model to handle the complexity of the experimental data in the CCDB. The current data dictionary comprises more than 65 tables, as briefly illustrated in the Entity-Relational (ER) diagram in Figure 3 and defined in detail at http://donor.ucsd.edu/CCDB/IMG/ER_092403.jpg. Additional description of the purpose of the CCDB project and the types of data it was designed around are described in Martone et al. (2002a, 2003). This section will provide a detailed description of the data model and an update on new and enhanced features.

Hierarchical Structure: The data model of the CCDB was designed to be extensible. As new microscopy technologies emerge, and additional types of tissue preparation, e.g., cell culture, subcellular fractions, or protocols are developed, these details can easily be added within the database structure as needed. A major effort in developing the CCDB data model was directed towards providing a structured representation of essential collection information, key entity attributes and relationships. Federation, interpretation and comparison of image datasets are heavily dependent on the how the images are obtained. In the CCDB data model, each microscopy dataset must be associated with project, experiment, subject group, subject, animal, tissue, and processing step information. An illustration of the attributes and description of these seven entities can be examined at http://donor.ucsd.edu/CCDB/table_descr.shtml. These seven entities are required and essential components that form the backbone of the CCDB. Overall, one to one/many relationships from project to experiment, from experiment to subject group, from subject group to subject, from subject to animal, from animal to tissue, from tissue to processing step, from processing step to microscopy product are enforced within the CCDB. For example, every experiment must belong to a project, and every project must have at least one experiment associated with it.

Abstract Entity: A primary focus of the CCDB is 3D reconstruction. One of the challenges of designing a microscopy image data model is the variety of microscopy techniques that can be used to obtain primary microscopy data and to perform subsequent 3D reconstruction, segmentation and quantification of data derived from the original images. To account for this instrumentation-related aspect of complexity, we created an entity called the “microscopy product”. A single microscopy product consists of a series of images related in some way, e.g. a set of 2D images to be used for a 3D reconstruction are related in some systematic way, e.g., time series. In order to allow varied reconstruction specification, and to avoid field and entity redundancy in the CCDB model, we designed several abstraction entities to describe microscopy images (Figure 4). At the top is the microscopy product, which specifies the type of microscope used and general image information such as pixel resolution and size. The top layer of abstraction divides products into light or electron microscopy. The common attributes for light or electron microscopy images are identified and included in the entity. The light microscopy image has two child entities, which are transmitted light and fluorescent images. Fluorescent images are also an abstraction entity that has several child entities: confocal, multiphoton and epifluorescent. Both light and electron microscopy images may belong to an image set that are related through one or more variables. For example, a tilt image series refers to a set of images collected at certain angular increments over a specific range of angles. Mosaic images are sets of images taken at specified X-Y positions that will be used to form a larger montaged image. Optical section series, through focus series and serial sections refer to images related through z position. Time series data are related through time increments. Sets of images that are not related in a systematic way through time or position but may belong to a single session can also be grouped together under survey images. Because any combination of microscopy techniques can be used in one microscopy product and produce a reconstruction, e.g., a microscopy product may be both a mosaic and an optical section series, the above entities are all related to a single microscopy product.

The CCDB not only stores raw images, but also post-processed images such as derived images, reconstructed images, or segmented images. Thus, a single dataset in the CCDB will have multiple representations: 2D images, 3D reconstructions, 3D segmentations, correlated data from multiple microscopies. For correlating multiple derived data products from a single set of microscopy images, three entities including 2D image, reconstruction images and segmented images are designed into the CCDB data model. Because segmentation comes from a reconstruction image, the one to zero or many dependency relationship is enforced in the CCDB from reconstruction to segmentation steps.

Self-defined Object: Microscopy images may be very complex. A fluorescent image may represent a sample with one or more fluorophores, with each fluorophore having several attributes, such as color, excitation wavelength, and emission wavelength. We took advantage of Oracle9i object-relational feature in the CCDB data model design to accommodate this complexity. For example, we implemented a varray (variable number of elements in an array) of fluorophore color objects (Figure 5) in the fluorescent images entity to allow a researcher to define the number of labels represented in an image along with each of the attributes such as color, excitation wavelength, and emission wavelength etc. Similar objects have been created for other fields where the number of objects is not fixed. Another example of CCDB objects is the protocol for specimen preparation. The detailed protocol is stored as a plain text file in CLOB field. Key details of the specimen preparation regime that would form the basis of rational queries or are required for image interpretation are stored in extra tables without breaking the backbone of the CCDB data model. Tables are currently implemented for fixation, embedding, protein localization (probe, protein, antibody, probe detection), and stain (http://donor.ucsd.edu/CCDB/table_descr.shtml). The protein localization tables describe which proteins (ex. Neurotransmitter, receptor, or other protein) are localized in a tissue and the method used to identify them. Proteins referenced in the CCDB can be linked to protein structures contained in the PDB through the PDB id. Because at this early stage of the project, much of the data being added to the CCDB represents legacy data obtained over the past 10 years, many experimental details may not be available for a given data set. Thus, the above information on tissue processing is not required currently for these legacy datasets, however, this extra experimental information is necessary for any researcher who desires to reanalyze or utilize CCDB data, and thus these fields are required for newly acquired datasets in the future. The fields were chosen not only to provide a complete description of the particular protocol used for a specimen, but also provide the ability to query key details, e.g., antibody or fixative to allow researchers to investigate possible sources of variability between experiments and experimenters.

<u>Metadata View and Security Levels</u>: A comprehensive view that includes correlated attributes of the image metadata is constructed based on the user's major interest and the security information. This feature is designed to control access to the CCDB images and metadata. Based on different reconstruction techniques, we have designed single tilt reconstruction view, optical section series view and through focus view to group the metadata associated with different reconstruction techniques. There are currently three security levels to control data access in CCDB: 1) private; 2) group (e.g. NCMIR, NIH); 3) public. The private data is only accessible by the owner. Group data is accessible by a self-defined group of users with access permission. Public data is accessible to any user who has a web browser and is registered through the CCDB site. Security information is associated with each record by tagging with ownership, security level and security id in the CCDB. Only the user (or group) who has ownership or a permission of a particular data set or image is able to retrieve the original image. All data stored in the CCDB, however, is potentially available to other users by contacting the owner through e-mail, who controls access to the data via permissions. To promote the spirit of collaboration and explore the significance of scientific images, a brief description, thumbnails, and animation of

each image stored in CCDB are available to all public users. Many data sets in the CCDB require no additional permissions to download the original data.

Ontology Link: An ontology link is implemented in the CCDB data model to support semantic queries and data sources federation. This major ontology source for the CCDB is the Unified Medical Language System (UMLS) (Humphreys et al., 1998). The UMLS is a large vocabulary and metathesaurus designed to aid the development of systems that help health professionals and researchers retrieve and integrate electronic biomedical information from a variety of sources and to make it easy for users to link disparate information systems, including computer-based patient records, bibliographic databases, factual databases, and expert systems. Utilizing of the UMLS for mapping databases can help to overcome retrieval problems caused by differences in terminology and the scattering of relevant information across many databases. The UMLS will meet our needs to federate the CCDB data sources with other biological database resources. We use two types of UMLS knowledge resources, the Metathesaurus and Semantic Network. The Metathesaurus provides a uniform, integrated distribution format from over 100 biomedical vocabularies and classifications (the majority in English and some in multiple languages) and links many different names for the same concepts. The Semantic Network contains information about the types or categories (e.g., "Disease or Syndrome," "Virus") to which all Metathesaurus concepts have been assigned and the permissible relationships among these types (e.g., "Virus" causes "Disease or Syndrome"). All the data records in CCDB are labeled with concept ID and semantic ID for future data semantic mediation (Gupta et al., 2002).

Atlas Link: All the images in CCDB are registered within atlas coordinate systems and associated atlas identifiers defined by different scientists. The atlas link concept was applied in neuroimaging by using anatomical templates positioned in a standard coordinate system.

Experimental data are aligned into this coordinate system and consequently inherit labels for numerous neuroanatomical compartments. In addition to establishing homology at the level of anatomical structures, coordinates can be used directly to address equivalent points across images. The atlas link for the images in CCDB enable the spatial query of the images using a specially designed GIS-based brain atlas tool called the Smart Atlas (Martone et al., 2002b). Atlas coordinates (x, y & z coordinates) can come from different coordinate systems. This atlas link provides a powerful imaging framework to house and correlate localization/expression of molecular constituents (e.g. proteins, receptors, etc) with anatomical information drawn from microscopy imaging technology as well as other technologies.

CCDB Query Tools

The purpose of the query tools is to help CCDB users to search and retrieve the images in a more meaningful way without the need to understand the underlying system implementation. After a user registers with CCDB and obtains an account, he/she may log into the CCDB and issue either simple or customized advanced queries using the query tools integrated into the CCDB website. A forms-based query mode in CCDB provides comprehensive search capabilities through a menu of options, without the need to know SQL syntax or the internal structure of the database. Several key query types were identified for our forms-base query mode.

The simple query permits the user to query the image dataset based on project keyword, project id, person’s name, reconstruction technique, cell type, structure, organ, protein and species etc. All query fields are joined by a logical “AND”. The query form provides the available choices for some query items to avoid zero hits.

The fields in the simple query form were chosen because they are likely to be useful to the majority of biologists seeking to retrieve specific data sets from the CCDB. The simple query form does not, however, provide the capability to query every image attribute in the CCDB. In order to meet different CCDB user's query preference, we have implemented a customizable advanced query tool that permits the user to build a query according to the reconstruction technique and more detailed project, experimental, subject, analytical and anatomical details contained in the CCDB. An example of a customized query form is shown in Figure 6. In this example, the user selected project and neuron summary information (obtained from a Neurolucida segmented object) as the criteria to search the CCDB images. Neurolucida (Microbrightfield, Vermont) is an image analysis tool for performing quantification of neuronal structures. Parameters that can be analyzed by Neurolucida include dendritic and axonal lengths, diameters, numbers of spines etc. Using the advanced query form, the user can construct a more complex query such as:

CQ1: Find records for neurons that have been segmented using Neurolucida and has projectID as P0000 and where the cell body is less than 9 μm in diameter and the total number of spines are less than 382 and the volume is less than 7838.33 μm^3, etc.

All results are returned from either a simple or advanced query in the form shown in Figure 7. The results representation is formatted into user-friendly web pages. Each CCDB data record is one "microscopy product" which includes multiple images. A brief description of the dataset is presented along with thumbnails of all imaging and quantitative data available for that dataset. Examples and explanation of correlated and brief results pages are shown in Figure 7.

Recently, we have implemented an evaluation function for each data set where users can enter an evaluation of the quality of the experiment, imaging, reconstruction and segmentation results. Microscopy images will be used extensively in many aspects of biological research in different ways. We envision that the CCDB will not only serve as an image collection space but also a knowledge exploration and new discovery environment. Scientists may have different interpretations and/or additional annotations for the same dataset. The CCDB will provide a forum for the scientist to post their comments on the images by providing additional analysis or annotation or by providing an evaluation of the image data.

CCDB Data Submission

Currently, all the data that has been validated as flat data files are entered into CCDB through Oracle SQL loader utility. This SQL loader utility is very useful and efficient for large data set loading. As scientific data submission environment and a friendly user interface, however, are essential since CCDB data submitters are mostly neuroscientists, and microscopists. We therefore designed the data submission interface for the CCDB as an electronic lab notebook with straight forward uploading and downloading capabilities. After an experimenter enters the secure user interface, the experimenter can manage data in CCDB from three different administrative levels: 1) User level; 2) Project Level; 3) Image Level. From the user level, the user can grant other users permission to edit or view his/her data or forbid any access of his/her private data. The experimenter chooses to create a new project or add data to an already existing project. All the data records on the user level are presented and organized according to the ownership of the records. For the project level, the data records are presented and organized according to the project domain. Any user that is assigned to that project team should be able to access the project and edit the data. The image administrative level is more related with the image annotation or description prospects of the images. The experimenter who creates the images can insert the image annotation or description. Any data manipulation in the CCDB should be logged at the database level for future trace of any data errors.

Straight forward uploading and downloading capabilities of data entry through the Telescience Portal are currently under development at NCMIR (Peltier et al., 2003). It is intended that we will support both xml data file submission and on-line form submission. Any data submission to the CCDB will first be stored in the staging area before validation and security level tagging. XML data file submission will be evaluated as a preferred submission strategy because offline uploading will avoid network traffic and database constrain violation and provide a flexible user-interface for the users.

Upon submission, a data set is immediately available for retrieval. However, the new submission is noted with a conditional tag to alert users that the data set has yet been reviewed by the administrator. Once approval, the conditional tag is removed and the data set is available unconditionally.

CONCLUSIONS

We have successfully designed and implemented a web-based database system for storing, organizing, retrieving, and federation of microscopy imaging data. Currently the CCDB is populated with 3D light and electron microscopy datasets and supports a wide range of users and institutional applications. The CCDB uses a flexible and robust framework to meet the needs of project oriented experimental imaging data management by using the BIRN infrastructure and security features to protect privileged information**.** The BIRN infrastructure represents the first attempt to develop a "protocol" for collaborative research among neuroscientists and medical scientists through database federation. The CCDB as a component of BIRN will be one of the test nodes of doing large-scale medical science. BIRN has developed and continues to refine useful ways of sharing data rapidly over the Internet, with appropriate controls over data. In order to support the BIRN project, CCDB will continue to work with multidisciplinary scientists to populate the CCDB with a variety of cell related imaging data.

The CCDB's major contribution to the microscopy community is to provide a robust and hybridized architecture and an extensible, secure, and comprehensive data model with advanced features to address the management of complex scientific and multidisciplinary imaging data. Our efforts complement other ongoing efforts such as the Open Microscopy Environment to bring to bear modern information technology in the management and mining of microscopic imaging data. The ontology and atlas link in the data model will enable the federation of CCDB microscopy imaging data with other related biological data resources and allow the performance of semantic and spatial queries on imaging data. In addition, quantitative results from the imaging segmentation can be plugged into the data mining tools for further analytical query capability in the CCDB. As the number of data records in the CCDB increases, future research topics will include new indexing technology and new tool kit development. The specific focus will be on imaging data mining. New data types and functionality will be added to model and statistically analyze image objects. In addition, new query capabilities will be implemented to support the collaboration among multiple institutions.

It is our ultimate goal to streamline the whole process from data collection, data validation, data storage, data retrieval and data analysis for the microscopy imaging data. We hope that the CCDB will contribute to the scientific community by providing: 1) an imaging resource to fill in the resolution gaps from tissue down to the subcellular level; 2) an open resource for biologist, computational neuroscientists, and computer scientist for sharing 3D imaging data of cells and tissues; 3) an imaging management system for tracking, analyzing imaging data and; 4) an interoperable resource for large scale data federation under the BIRN project.

ACKNOWLEDGEMENTS

This project is supported by NIH grants from NCRR RR04050, RR08605 and the
Human Brain Project grant DA16602 supported by the National Institute on Drug Abuse, the National Institute of Biomedical Imaging and Biotechnology, and NSF grants supporting the National Partnership for Advanced Computational Infrastructure NSF-ASC 97-5249 and MCB-9728338. The Bioinformatics Research Network is supported by NIH grants RR08605-08S1 (BIRN CC) and RR043050-S2 (Mouse BIRN). We thank the San Diego Supercomputer Center for their technical support.

REFERENCES

ACHARD, F., VAYSSEIX, G. & BARILLORT, E. (2001). XML, bioinformatics and data integration. *Bioinformatics* **17**, 115-125.

ANDREWS, P., HARPER, I. & SWEDLOW, J. (2002). To 5D and beyond: quantitative fluorescence microscopy in the postgenomic era. *Traffic* **3**, 29-36.

BAXEVANIS, A.D. (2002). The Molecular Biology Database Collection: 2002 update. *Nucleic Acids Research* **30**, 1-12.

CARAZO, J.M. & STELZER, E.H.K. (1999). The BioImage database project: organizing multidimensional biological images in an object-relational database. *J Structural Biology* **126**, 97-102.

CHICUREL, M. (2000). Databasing the brain. *Nature* **406**, 822-825.

ELMASRI, R. & NAVATHE, S.B. (2002). *Fundamentals of Database Systems*. Reading, Massachusetts: Addison-Wesley.

FLICKNER, M., SAWHNEY, H., NIBLACK, W., ASHLEY, J., QIAN, H., DOM, B., GORKANI, M., HAFNER, J., LEE, D., PETKOVIC, D., STEELE, D. & YANKER, P. (1995). Query by image and video content: the QBIC system. *Computer* **28**, 23-32.

GALLAUGHER J. & RAMANATHAN S. (1996). Choosing a client/server architecture. A comparison of two-tier and three-tier systems. *Information Systems Management Magazine* **2** (13), 7-13.

GONZALEZ-COUTO, E., HAYES, B. & DANCKAERT, A. (2001). The life sciences Global Image Database (GID). *Nucleic Acids Research* **29**, 336-339.

GUPTA, A., LUDĂSCHER, B. & MARTONE, M.E. (2002). Registering scientific information sources for semantic Mediation. In *21st International Conference on conceptual modeling (ER),* Spaccapietra, S., March, S.T. & Kambayashi, Y. (Eds.), pp.182-198. Berlin, New York: Springer.

HUMPHREYS, B.L., LINDBERG, D.A.B., SCHOOLMAN, H.M. & BARNETT, G.O. (1998). The Unified Medical Language System an Informatics Research Collaboration. *J Am Med Inform Assoc*. **5**, 1-11.

MARTONE, M.E., GUPTA, A., WONG, M., QIAN, X., SOSINSKY, G., LUDAESCHER, B. & ELLISMAN, M.H. (2002a). A cell-centered database for electron tomographic data. *J. Structural Biology* **138**, 145-155.

MARTONE, M.E., ZHANG, S., GUPTA, A., QIAN, X., PRICE, D.L., WONG, M., SANTINI, S. & ELLISMAN, M.H. (2003). The cell-centered database: a database for multiscale structural and protein localization data from light and electron microscopy. *Neuroinformatics* **1**, 379-395.

MARTONE, M.E., GUPTA, A., LUDASCHER, B., ZASLAVSKY, I. & ELLISMAN, M.H. (2002b). Federation of brain data through knowledge-guided mediation. In *Neuroscience Databases: A Practical Guide*, Kotter R. (Ed.), pp. 275-292. Boston: Kluwer Academic Publishers.

MARX, V. (2002). Beautiful Bioimages for the Eyes of Many Beholders. *Science* **297**, 39-40.

NIBLACK, W., BARBER, R., EQUITZ, W., FLICKNER, M., GLASMAN, E., PETKOVIC, D., YANKER, P., FALOUTSOS, C. & TAUBIN, G. (1993). The QBIC project: Querying images by content using color, texture, and shape. In *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases*, Niblack, W. (Ed.), pp.173-187. Bellingham, Wash., USA: SPIE.

PELTIER, S., ABEL W.L., LEE, D., MOCK, S., LAMONT, S., MOLINA, T., WONG, M., DAI, L., MARTONE, M.E. & ELLISMAN, M.H. (2003). The Telescience Portal for advanced tomography applications. *J. Parallel and Distributed Computing* **63**, 539-550.

RAHASEJAR, A., WAN, M. & MOORE, R. (2002). MySRB & SRB – Components of a Data Grid. In *The 11th International Symposium on High Performance Distributed Computing*, pp.301-310. Los Alamitos, Calif.: IEEE Computer Society.

SWEDLOW, J.R., GOLDBERG, I., BRAUNER, E. & SORGER, P.K. (2003). Informatics and quantitative analysis in biological imaging. *Science* **300**, 100-102.

WANG, K., LIN, J., GAJNAK, J. & MURPHY, F. (2002). Image content-based retrieval and automated interpretation of fluorescence microscope images via the protein subcellular location image database. *IEEE*, 325-328.

ZHANG S., QIAN, X., GUPTA, A. & MARTONE, M.E. (2003). A practical approach for Microscopy Imaging Data Management (MIDM) in neuroscience. In *Proceedings of the 15th International Conference on Scientific and Statistical Database Management*, Nittel, S. & Gunopulos, D. (Eds.), pp. 237-240. Los Alamitos, California: IEEE Computer Society.

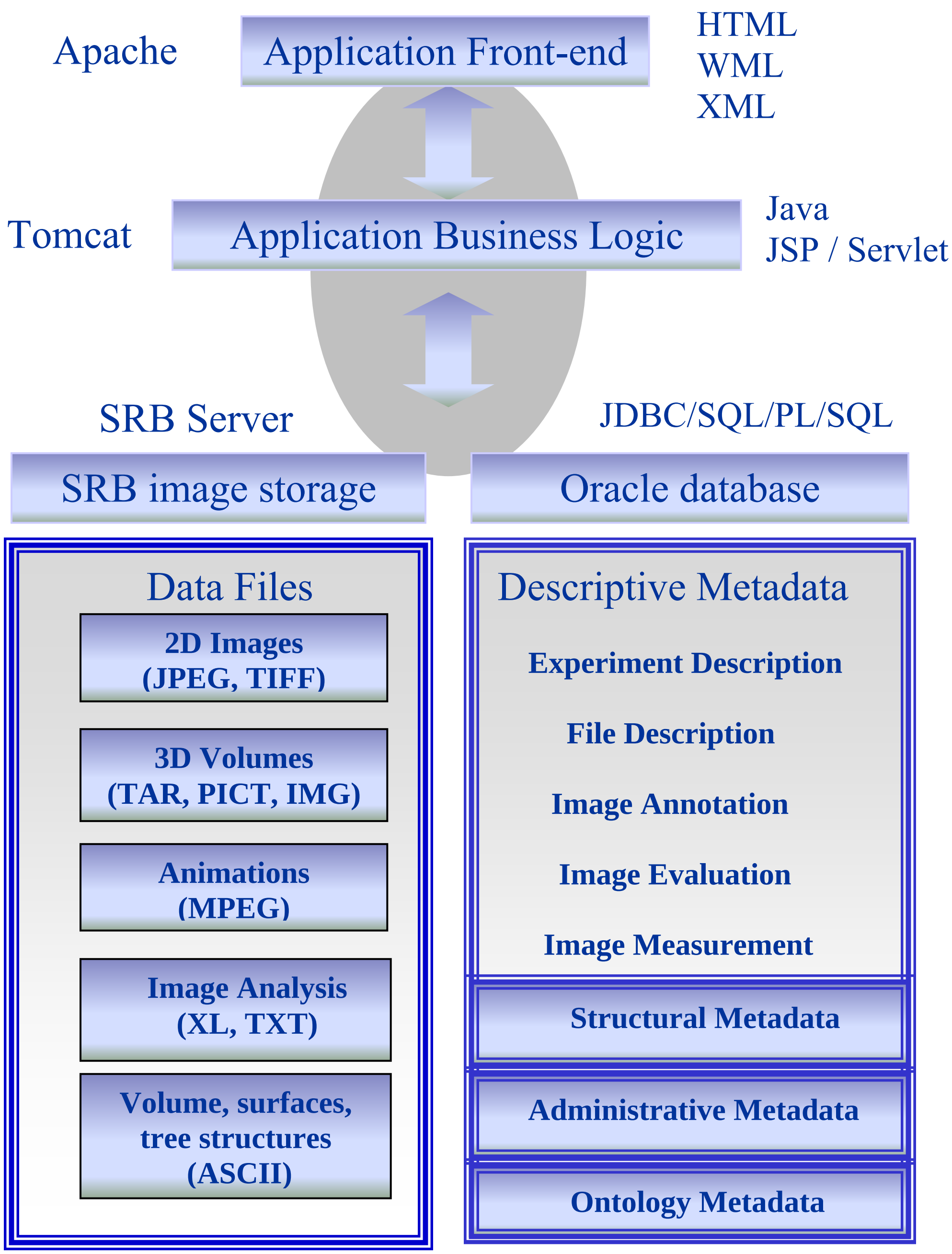

Apache
Application Front-end
HTML
WML
XML
Tomcat
Application Business Logic
Java
JSP / Servlet
SRB Server
JDBC/SQL/PL/SQL
SRB image storage
Oracle database
Data Files
2D Images (JPEG, TIFF)
3D Volumes (TAR, PICT, IMG)
Animations (MPEG)
Image Analysis (XL, TXT)
Volume, surfaces, tree structures (ASCII)
Descriptive Metadata
Experiment Description
File Description
Image Annotation
Image Evaluation
Image Measurement
Structural Metadata
Administrative Metadata
Ontology Metadata

Figure 1. The design of CCDB system architecture. The image data files and descriptive metadata are stored in SRB and CCDB database. The distributed data files with different formats such as 2D image, 3D volumes, animations, image analysis, and surfaces, tree structures are managed by SRB server. Descriptive, structural, and administrative metadata is managed by Oracle9i. The data files in SRB are registered in Oracle 9i.

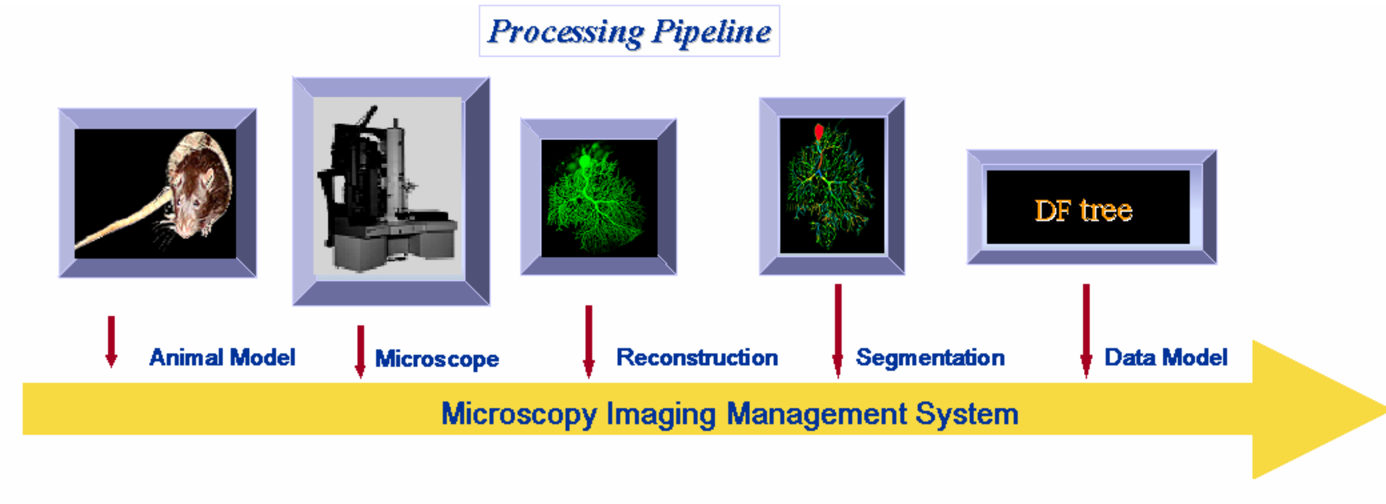

Processing Pipeline
DF tree
Animal Model
Microscope
Reconstruction
Segmentation
Data Model
Microscopy Imaging Management System

Figure 2. The pipeline for CCDB data retrieval. All the data that are related with animal model, microscope parameters, reconstruction images and segmentation measurement, mathematical data models are stored in the microscopy image management system.

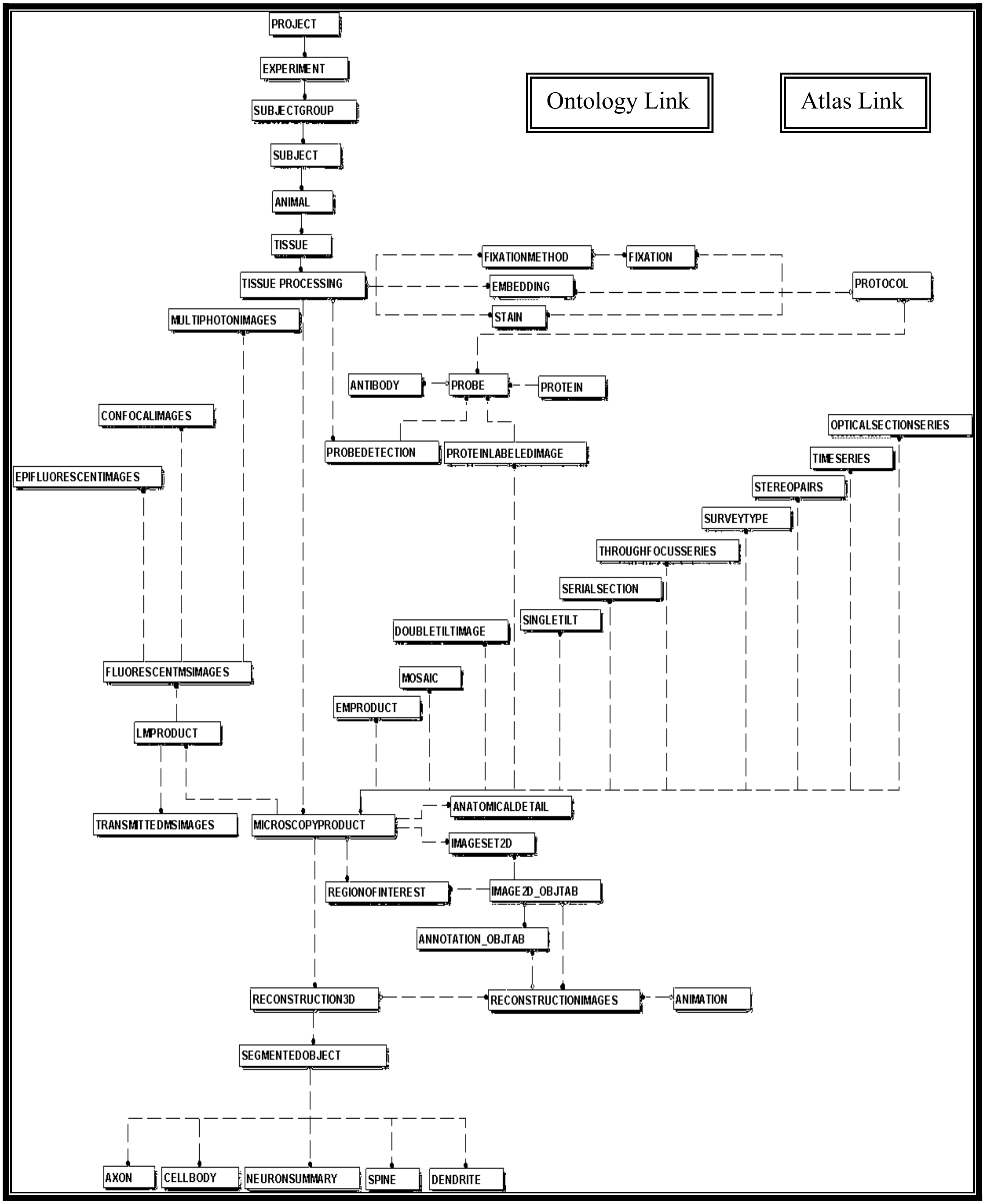

PROJECT
EXPERIMENT
SUBJECTGROUP
SUBJECT
ANIMAL
TISSUE
TISSUE PROCESSING
Ontology Link
Atlas Link
FIXATIONMETHOD
FIXATION
EMBEDDING
STAIN
PROTOCOL
MULTIPHOTONIMAGES
ANTIBODY
PROBE
PROTEIN
CONFOCALIMAGES
PROBEDETECTION
PROTEINLABELEDIMAGE
OPTICALSECTIONSERIES
TIMESERIES
STEREOPAIRS
EPIFLUORESCENTIMAGES
SURVEYTYPE
THROUGHFOCUSSERIES
SERIALSECTION
SINGLETILT
DOUBLETILTIMAGE
FLUORESCENTMSIMAGES
MOSAIC
EMPRODUCT
LMPRODUCT
ANATOMICALDETAIL
TRANSMITTEDMSIMAGES
MICROSCOPYPRODUCT
IMAGESET2D
REGIONOFINTEREST
IMAGE2D_OBJTAB
ANNOTATION_OBJTAB
RECONSTRUCTION3D
RECONSTRUCTIONIMAGES
ANIMATION
SEGMENTEDOBJECT
AXON
CELLBODY
NEURONSUMMARY
SPINE
DENDRITE

Figure 3. Main component of CCDB ER diagram implemented on Oracle9i. This schematic does not show all the tables in the CCDB, but gives a summary of the relationship between major tables. The detail diagram is available at http://donor.ucsd.edu/CCDB/IMG/ER_092403.jpg

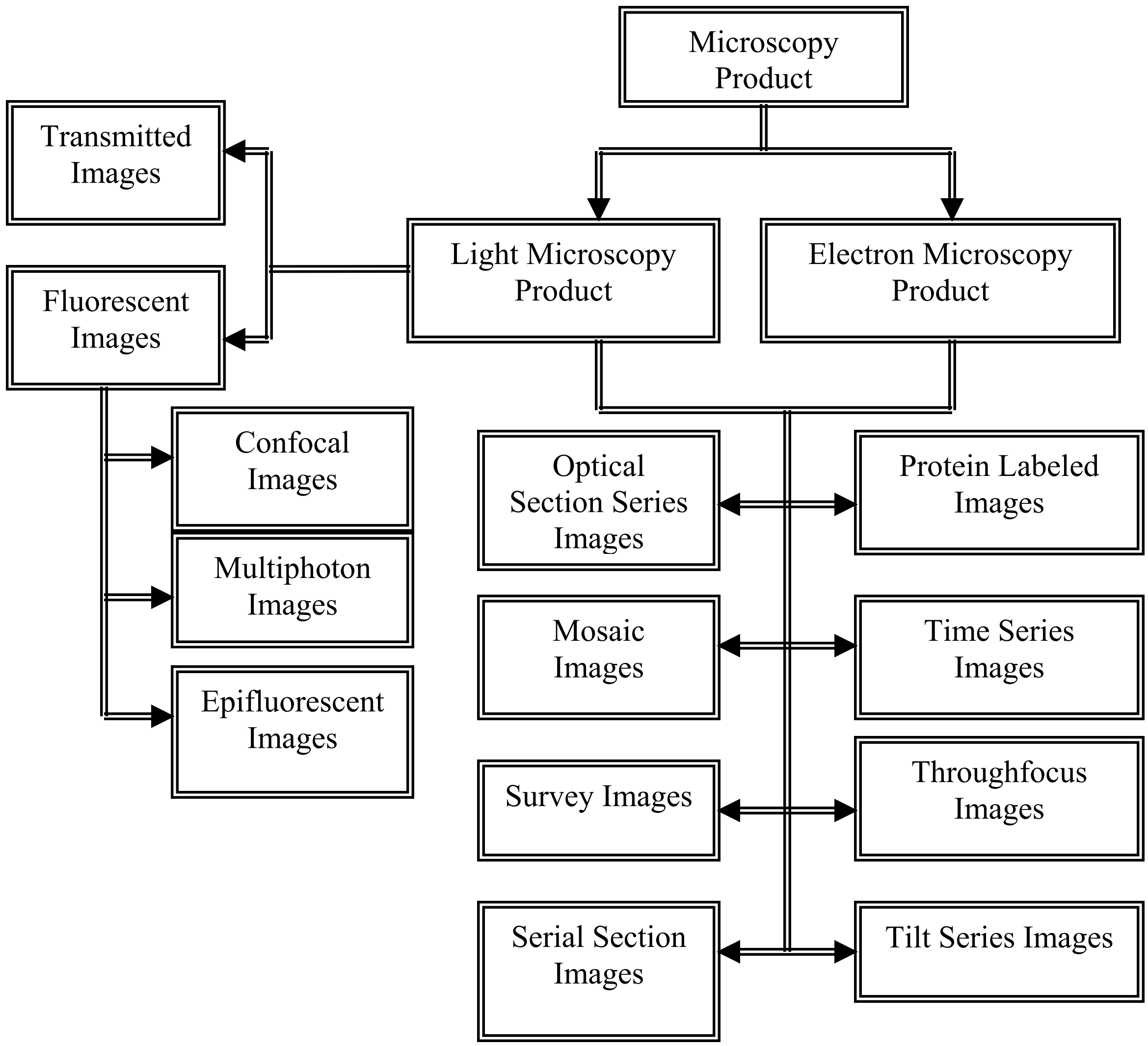
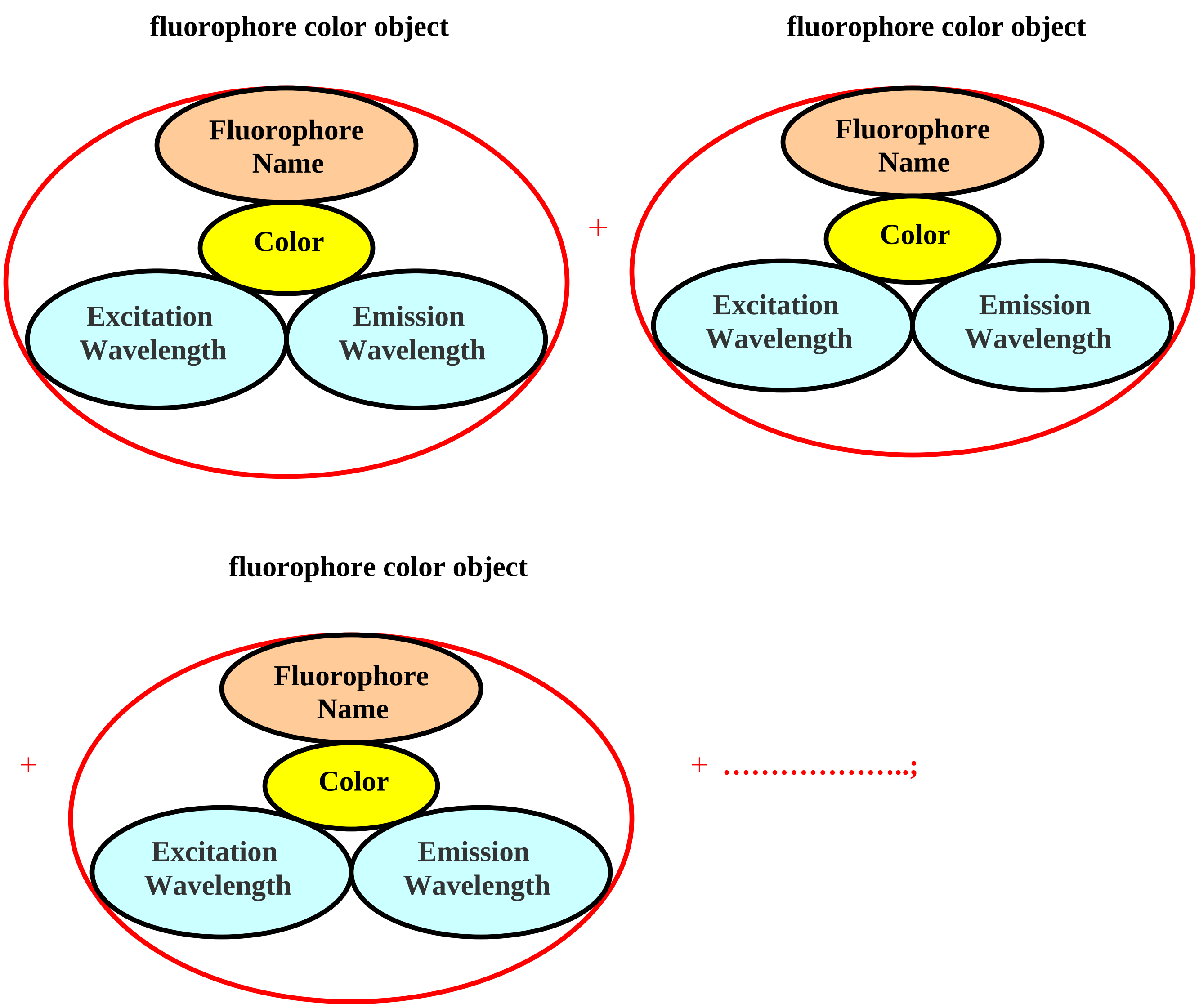

Microscopy Product
Transmitted Images
Fluorescent Images
Light Microscopy Product
Electron Microscopy Product
Confocal Images
Multiphoton Images
Epifluorescent Images
Optical Section Series Images
Protein Labeled Images
Mosaic Images
Time Series Images
Survey Images
Throughfocus Images
Serial Section Images
Tilt Series Images

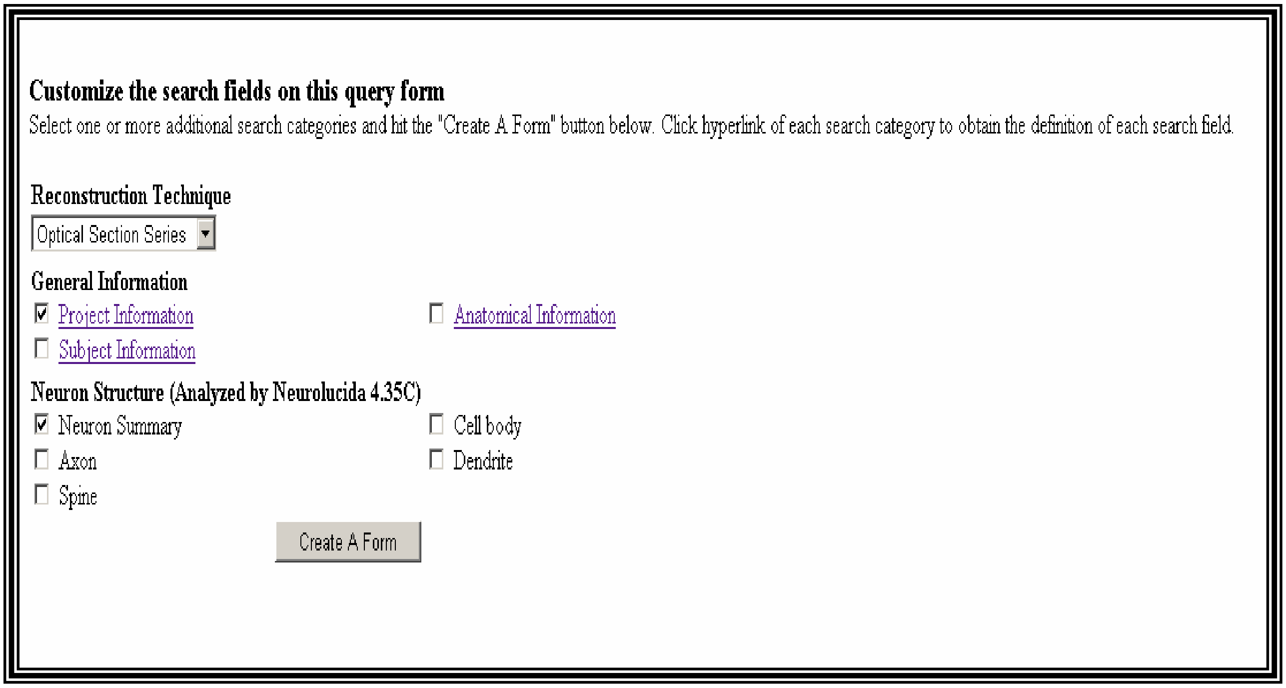
Figure 4. The hierarchical structure and abstraction of microscopy product entities. The top layer of abstraction divides microscopy image products into light or electron microscopy. The light and electron microscopy products are the abstraction of optical section Series, tilt series, through focus, time series, or protein labeling products. The common attributes for light or electron microscopy images are identified and included in the entity. The light microscopy image has two child entities, which are transmitted light and fluorescent images. Fluorescent images are also an abstraction entity that has several child entities: confocal, multiphoton and epifluorescent.



= fluorophore colors varray

(a collection of variable number of fluorophore color objects)

Figure 5. A cartoon representation of the fluorophore color object and fluorophore color varray. Oracle objects (SQL objects) are composite data structures that group related data items, such as fluorophore color attributes (fluorophore name, color, excitation wavelength, emission wavelength) into a single data unit. An object type is functionally similar to a Java class--you can populate and use any number of individual objects of a given object type, just as you can instantiate and use individual objects of a Java class.



Customize the search fields on this query form

Select one or more additional search categories and hit the "Create A Form" button below. Click hyperlink of each search category to obtain the definition of each search field.

Reconstruction Technique

Optical Section Series

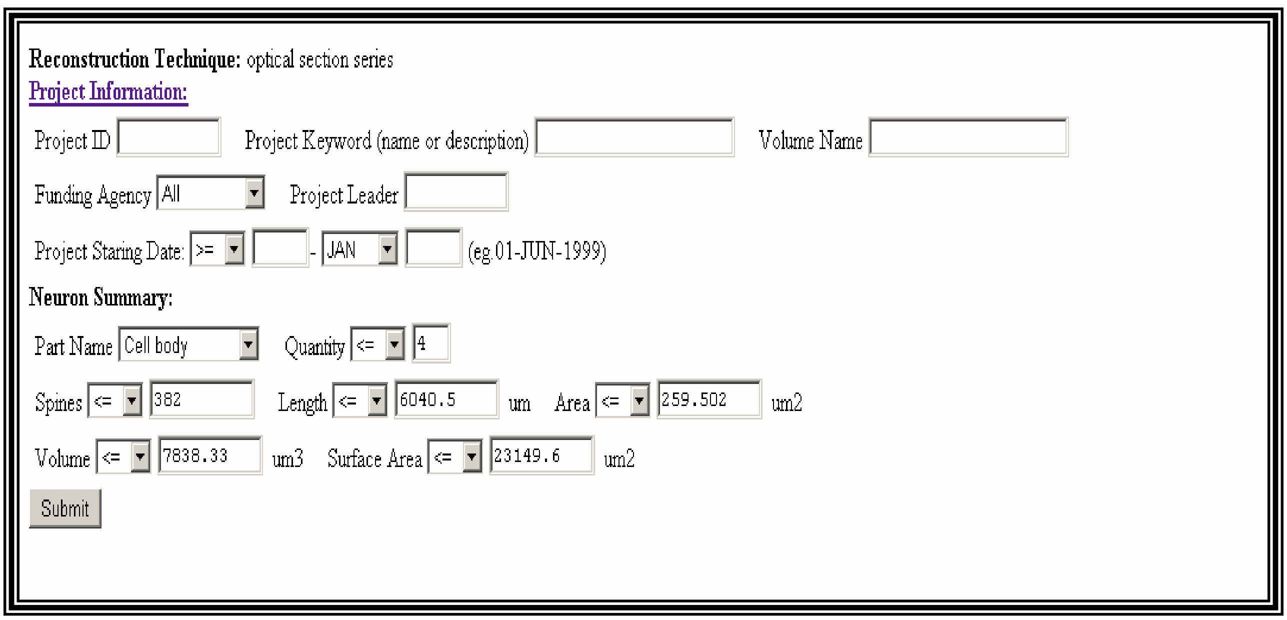
General Information

- [x] Project Information
- [ ] Anatomical Information
- [ ] Subject Information

Neuron Structure (Analyzed by Neurolucida 4.35C)

- [x] Neuron Summary
- [ ] Cell body
- [ ] Axon
- [ ] Dendrite
- [ ] Spine

Create A Form



Reconstruction Technique: optical section series

Project Information:

Project ID | Project Keyword (name or description) | Volume Name

Funding Agency: All | Project Leader

Project Staring Date: >= - JAN (eg.01-JUN-1999)

Neuron Summary:

Part Name: Cell body | Quantity <= 4

Spines <= 382 | Length <= 6040.5 um | Area <= 259.502 um2

Volume <= 7838.33 um3 | Surface Area <= 23149.6 um2

Submit

Figure 6. A screen copy showing customized query forms. The top panel illustrates the form initially presented to the user used to create the customized query form. The bottom panel shows the resulting customized query form generated by selecting the project information category under General Information and the neuron summary category under the Neuron Structure are chosen. The quantitative values represent the maximum value for each field calculated from the database.

Query Result

1 record was found!

(1)

Record : 1
View Details (2)
View Protocol (3)

Project: P1119

Description: Measurements of spine parameters using light microscopy and electron tomography

Species: rat

System: CNS

Organ: brain

Region: cerebellum

Cell type: Purkinje Neuron

Structure: spiny dendrite

Volume ID: pccor10/pccor10_vol.tar

(4)
Click on View Correlated Data link to see a correlated through focus series reconstruction image set

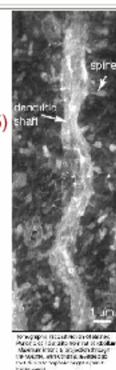
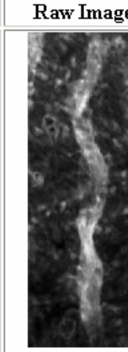
Volume	Segmented Object	Raw Image
(5)  (6) Downloads (19M/TAR)! (7) Image Description (8) Submit Evaluation (9) List Evaluation	 (10) (11) List & Measurement View 3D image by Jviewer! (12) Downloads (xls) (13) Downloads	 click here to view the animation (14) Downloads (TAR) (15)

Figure 7. A screen copy showing part of the results from simple or advanced query in CCDB. 1) Display the results of cell type in the SenseLab database at http://senselab.med.yale.edu/senselab; 2) Display full detail information on the project, experimental, subject group, subject, animation, tissue information, processing step, microscopy product, reconstruction and segmentation etc. information of the record; 3) Display a detail protocol for specimen preparation if available; 4) Retrieves correlated volume, e.g., correlated light and electron microscopic volumes of the same specimen. The correlative light and electron microscopy is critical for understanding the functionality and large-scale structure of tissues and organs; 5) Click the thumbnail image to obtain the full resolution of the 2D volume image; 6) Download the original image, provide the links to obtain the publication abstract in the PubMed database; 7) Display brief description of thumbnail image; 8) Display a form for the user to enter the reconstruction quality evaluation; 9) List all evaluation related with this dataset submitted by different CCDB users; 10) List any measurements stored with segmented objects in reconstruction; 11) View and manipulate 3D segmented reconstruction using a web based viewer; 12) Download the Excel spreadsheet containing measurements; 13) Download segmented reconstruction files; 14) Display movie of 2D images used to create 3D reconstruction; 15) Download set of raw images used to create reconstruction.