Available online at www.sciencedirect.com
ScienceDirect
Journal of Structural Biology 161 (2008) 220–231
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

The Cell Centered Database project: An update on building community resources for managing and sharing 3D imaging data

Maryann E. Martone [a,b,*], Joshua Tran [a,b], Willy W. Wong [a,b], Joy Sargis [a,b], Lisa Fong [a,b], Stephen Larson [a,b], Stephan P. Lamont [a,b], Amarnath Gupta [b,c], Mark H. Ellisman [a,b]

[a] Department of Neurosciences, University of California at San Diego, San Diego, CA 92093-0608, USA
[b] Center for Research in Biological Systems, University of California, San Diego, CA 92093-0446, USA
[c] San Diego Supercomputer Center, University of California, San Diego, CA 92093-0515, USA

Received 2 July 2007; received in revised form 4 October 2007; accepted 5 October 2007
Available online 16 October 2007

Abstract

Databases have become integral parts of data management, dissemination, and mining in biology. At the Second Annual Conference on Electron Tomography, held in Amsterdam in 2001, we proposed that electron tomography data should be shared in a manner analogous to structural data at the protein and sequence scales. At that time, we outlined our progress in creating a database to bring together cell level imaging data across scales, The Cell Centered Database (CCDB). The CCDB was formally launched in 2002 as an on-line repository of high-resolution 3D light and electron microscopic reconstructions of cells and subcellular structures. It contains 2D, 3D, and 4D structural and protein distribution information from confocal, multiphoton, and electron microscopy, including correlated light and electron microscopy. Many of the data sets are derived from electron tomography of cells and tissues. In the 5 years since its debut, we have moved the CCDB from a prototype to a stable resource and expanded the scope of the project to include data management and knowledge engineering. Here, we provide an update on the CCDB and how it is used by the scientific community. We also describe our work in developing additional knowledge tools, e.g., ontologies, for annotation and query of electron microscopic data.

Keywords: Electron tomography; Bioinformatics; Ontology; 3D reconstruction

1. Introduction

The Cell Centered Database (CCDB) project was launched in 2002 as an on-line repository of high-resolution 3D light and electron microscopic reconstructions of cells and subcellular structures (Martone et al., 2002, 2003, 2007). The CCDB contains data covering the dimensional range known as the "mesoscale", roughly encompassing the structures that sit between gross morphology and molecular structure, e.g., cellular networks, cellular and subcellular microdomains along with their macromolecular constituents. The study of mesoscale structures, like dendritic spines, continues to present a challenge to experimentalists, because their dimensions fall squarely between the capabilities of current imaging technologies. Investigations of physiology, structural dynamics, coarse molecular distributions, and large scale distributions of dendritic spines are typically accomplished by optical microscopies. Appreciation of the fine structural detail on internal structure, cytoskeletal organization, localization of molecular constituents, location of synaptic contacts, and detailed views of the immediate microdomain such as pre-synaptic boutons and glial processes require 3D electron microscopic imaging. To build a comprehensive understanding of complex tissues in this dimensional range requires the ability to aggregate data obtained by multiple researchers across techniques and spatial scales.

* Corresponding author. Department of Neurosciences and Center for Research in Biological Systems, University of California, San Diego, San Diego, CA 92093-0446, USA. Fax: +1 858 822 3610.
E-mail address: mmartone@ucsd.edu (M.E. Martone).

1047-8477/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2007.10.003

Of current techniques, electron tomography is providing some of the most significant and spectacular information about mesoscale structures, with its ability to situate macromolecules in their 3D cellular contexts (Lucic et al., 2005; Marsh et al., 2004). One of the main motivations in the creation of the CCDB was to provide a forum for the very rich and valuable data sets produced by electron tomography to be made available to the public. The original CCDB was first proposed to the electron tomography community at the 2nd International Conference on Electron Tomography held in Amsterdam in 2001. At that time, the CCDB existed more as a concept than an actual product. By the time the special issue of *Journal of Structural Biology* arising from that conference was published in 2002, however, the first public version of the CCDB was on-line (Martone et al., 2002). The support for the CCDB was provided by a grant through the Human Brain Project (Wong and Koslow, 2001), a program designed to produce computational tools and databases for sharing scientific data with the broader scientific community. Over the past 5 years, we have continued to refine the architecture of the CCDB and have moved it from a prototype to a stable infrastructure. At the same time, we have had to refine our vision of the CCDB in response to community feedback, technological advances in knowledge engineering and our own experiences with sociological, technical and biological aspects of data sharing. In this paper, we present an overview of the current CCDB, our experiences in its creation, and plans for future development.

2. Materials and methods

2.1. Current architecture of the CCDB

The public CCDB is available at http://ccdb.ucsd.edu. The data model of the CCDB is illustrated in Fig. 1, which shows a highly simplified view of the schema. The CCDB was built using a combination of enterprise software components and cyberinfrastructure developed largely in an academic setting. The current CCDB utilizes Oracle 10g as the relational database management system with additional applications written in Java. Data entry forms for the CCDB were built using Gridsphere, an open source project for building secure java-based web portals (www.gridsphere.org). Because Gridsphere components, called portlets, are built to a common specification, the CCDB input forms may be easily incorporated into any Gridsphere-compliant portal.

The CCDB utilized the basic architecture developed by the Biomedical Informatics Research Network (BIRN; Grethe et al., 2005) and Telescience (Peltier et al., 2003) projects for distributed file storage and access. The BIRN project is an example of a so-called ''grid'' project, predicated on a model of distributed hardware and software. The basic idea behind most cyberinfrastructure projects like BIRN is that it should not matter where a resource is located physically or what hardware it is using. Programmatic access and security should be uniform across all of these resources. This uniformity is provided by software layers, ''middleware'', that sit between the physical resource and the programs required to access it. The CCDB utilizes both the distributed collections manager called the Storage Resource Broker (SRB; Grethe et al., 2005) and the authentication mechanisms for CCDB files (Peltier et al., 2003).

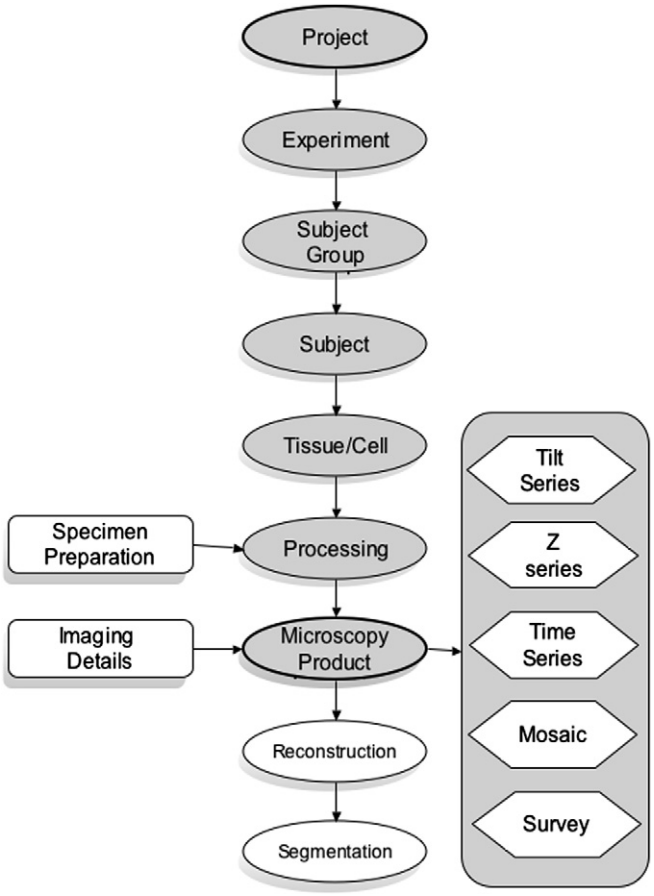

Fig. 1. Simplified view of the CCDB schema, showing the main classes of information contained in CCDB tables. The Microscopy Product provides the unique identifier for the CCDB database. Each oval represents a single table in the CCDB; other shapes are categories of information that are elaborated in multiple tables. The relationship between the core tables of the CCDB (ovals) is one to many, that is, one project can have many experiments and so on. All microscopy products must be registered within their experimental contexts so the first seven tables are required (gray ovals). The tissue and processing tables contain a minimal set of specimen preparation details, while the microscopy product contains a minimal set of imaging details. More detailed specimen preparation and imaging protocols are stored in additional tables. The general classes of microscopy products are illustrated in the gray box to the right.

2.2. Data model

The CCDB was designed around the process of reconstruction from 2D micrographs, capturing key steps in the process from experiment to analysis. The core tables shown in Fig. 1 represent the backbone of the CCDB, each