

Development of a Web Based Data Storage and Project Management System for Biological Electron Cryo-microscopy

Laurie Nason, Wah Chiu and Steven J. Ludtke

* National Center for Macromolecular Imaging, Baylor College of Medicine, Houston TX 77030

Issues facing Microscopists

As electron cryomicroscopy has increased in popularity and capability over recent years, the volume of data and number of projects underway has risen dramatically. Two problems now face this field, first, information sharing and coordination among geographically distant collaborators with diverse backgrounds, and second, maintaining information in a form suitable for processing and data mining. The data involved in collaboration is quite diverse. For instance, biochemists may provide purification notes with scanned images of gels, the microscopist needs to store image data, freezing conditions, microscope conditions, etc., and the data processor needs to store image assessments, particle locations, reconstructions, etc. In addition, these parameters are not strictly fixed with time. New microscopes often require storage of new parameters, and techniques continue to evolve over time. However, as new parameters may appear with time, it is also critical that existing parameters retain a consistent definition. Not only are the available data diverse, but as resolution of the reconstruction improves, the quantity of data is rising dramatically. A single reconstruction in some cases now involves hundreds of gigabytes of raw data. Storage of this quantity of data in a broadly accessible way is nontrivial even with rapid increase in disk and network capacity. While the molecular biology electron microscopy community is beginning to establish standards and associated databases for deposition of 3D reconstructions, such efforts are primarily for published projects at their conclusion. There is a dramatic need for a tool for use while a project is in progress. In addition, it would be useful if assessments of image quality from this project could be correlated to other projects going on at any given moment.

Requirements and Selection of Framework

To establish such a database, several requirements must be met. First, database access must be widely accessible and cross-platform compatible. Second, the database structure must be easily extensible without requiring extensive modification of existing data. Third, the metadata and data storage, as well as the access interface should have some level of redundancy. The first criteria can be met easily, through use of web-based technologies. The second requirement is less straightforward. While a standard relational database could be used, they lack the flexibility to easily store application-specific data within records in an existing database structure. Modern, object-oriented databases are better, but most are commercial products with high implementation costs. The final criteria can be met in a variety of ways, depending on the specific server used.

After examining a variety of tools, we eventually selected a free, open source package called Zope (<http://www.zope.org>). Zope is a Python-based object oriented framework for constructing web applications. The primary difference between Zope and a traditional web-server is that every page served by Zope is a dynamic document generated when requested by the web browser. Zope treats static documents and dynamic objects, such as database entries, in the same way. In addition, Zope objects are accessible through modern remote procedure call mechanisms such as XMLRPC, making

it easy to interface with existing software. The Zope server is available for most UNIX variants including Linux, Windows and MacOS. Its interface, including administration and programming is entirely web-based, and it is compatible with almost all web browsers. Data elements are object-oriented, and easily extensible without requiring modification of existing records of the same type. Zope can also interface with existing relational databases for easy interactions with other established databases.

Implementation

Each stage in the microscopy process has been broken down into components and analyzed. After extensive discussions with scientists involved at each stage of the project, an initial database structure was established. The database was designed with expandability in mind, so new fields and entirely new record types can be inserted through by following a simple procedure. Metadata is stored within the Zope database itself, while large files (such as scanned micrographs and CCD frames) are stored externally through Zope's external file mechanism. This allows the metadata (which remains relatively small) to be archived independently from the image data.

Currently implemented record types include: projects, subprojects, purifications, aliquots, freezing, microscopy sessions, micrographs, CCD camera frames, scanned micrographs, reconstructions, publications and generic lab-notebook entries (figures 1 & 2). It has already been partially integrated with our Zeiss scanner so scanned images are deposited easily into the database. The system has been in use at the National Center for Macromolecular Imaging (<http://ncmi.bcm.tmc.edu>) for over a year, and development is continuing. Future enhancements will include microscope integration for automatic deposition of CCD frames, stronger integration with analysis software such as EMAN and SAVR and expansion of record types to include more diverse experimental possibilities.

[1] This research is supported by NCRR RR02250, NSF through NPACI, and the Agouron Institute.

The screenshot displays the National Center for Macromolecular Imaging web interface. The top navigation bar includes the center's logo and the user's name, Laurie Nason, logged in on Thursday 07 February 2002. The main content area is divided into two panels. The left panel shows a tree view of the database structure, including Microscopy Sessions, Purifications, and Lab Notebook Entries. The right panel displays a Lab Notebook entry for a Cryo-EM reconstruction. The entry is titled "Lab Notebook" and is dated "Steve Ludtke on 02/14/02". The description reads: "11.5 A reconstruction of GroEL using ~5000 particles collected on the JEOL 4000 microscope." Below the description is a table with two columns: "Associated Files" and "Description". The "Associated Files" column lists "thread.final.EM.mrc(3.82 MB)" and includes a 3D reconstruction image of GroEL. The "Description" column contains technical parameters: "refine=8 proc=32 hard=25 ctfov=groel.sm sym=d7 ang=2 mask=42 pad=128 classkeep=.5 classifier=8 xfiles=2,7,700,9 amask=10, 7,3 3dit=2 3dit2=1". The image is labeled "(15.6 kB)". The interface also includes a navigation menu on the left, a project details sidebar, and a footer with the date and user information.

Figure 2 – Sample Lab Notebook entry