1098 CD
DOI: 10.1017/S1431927606067663
Microsc Microanal 12(Supp 2), 2006

EMEN2: A Flexible & Mineable Data/Metadata Archival System for CryoEM

Haili Tu, Wah Chiu, Deepy Mann and Steven J. Ludtke

* National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX, 77030

Microscopy and related imaging disciplines face a common problem: how to archive massive quantities of image data and metadata in an efficient easily mineable fashion. Historically, there have been two options: an electronic notebook, offering excellent flexibility, but poor mineability, or a scientific database, offering good mineability, but requiring a great deal of management overhead and offering limited flexibility. In 2001, the NCMI developed EMEN (Electron Microscopy Electronic Notebook), a hybrid object oriented database and electronic notebook [1]. Based on the lessons learned during this successful project, we have now implemented EMEN2, designed from the ground up to offer a flexible and mineable solution suitable for virtually any sort of experiment.

EMEN2 not only archives the data and associated metadata for a particular experiment, but also incorporates archival of experimental protocols as a fundamental component of the database itself (Fig. 1). The definition of the metadata associated with an experiment is declared in the textual description of the experimental protocol or other object to be described. Unlike most databases, end-users are permitted to add new experimental protocols (classes) to the database. To add a new protocol, the user need only to describe the experiment as they would record it in a paper notebook, with embedded parameter names wherever a number or other parameter would be recorded (Fig. 2). As the experimental protocol evolves with time, 'subclasses' of existing protocols are created, requiring entry only of the changes between the new and original protocol. Definitions of experimental protocols and parameters are related through a hierarchical ontology, which describes the conceptual interrelationships between each of these items. Individual records in the database (instances of a specific experiment) contain the stored parameters along with arbitrary commentary and additional mineable parameters through a wiki/blogging mechanism (Fig. 1). Records are also stored and interrelated hierarchically, and can be interactively browsed similarly to browsing a computer’s filesystem.

Key to EMEN2's flexibility is its simple and intuitive query interface for data mining. The query interface currently supports searching, 2D plotting and histogramming. Fig. 3-4 show examples of some simple database queries. This query language is capable of navigating all three of the associated ontologies forming the database, to answer questions that could not be posed in a standard SQL database. The indexing mechanism is highly optimized, and most complex queries on the current 500,000 record database can be completed in less than 1 second.

EMEN2 is implemented using BerkeleyDB (www.sleepycat.com) for high speed indexing, Python (www.python.org) for internal database logic and Cheetah (www.cheetahtemplates.org) for HTML generation. It has both a customizable web interface as well as a programmatic interface through python (both with full security). Current test versions of EMEN2 are available upon request (SLudtke@bcm.edu), and upon release the completed database will be freely available with full source at http://ncmi.bcm.tmc.edu.

References
[1] Ludtke, S. J., Nason, L., Tu, H., Peng, L. and Chiu, W. Object oriented database and electronic notebook for transmission electron microscopy, *Microscopy and Microanalysis* **9** :556-565,2003
[2] Supported by P41RR02250, PN2EY016525 and the Agouron Institute.

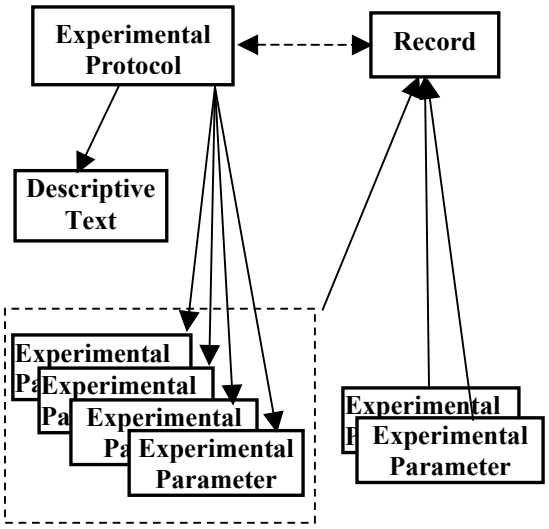

Figure 1 - Overview of EMEN2

Purpose of this microscopy session**:** *$$purpose*
Date: *$$event_date*
Warming-Up the Microscope & Cooling the ACD
The ACD (Anti-Contamination Device) was opened and filled with *$$acd_coolant*. It took ~45 minutes for the device to cool. Meanwhile the specimen position was set to (0,0,0), then the microscope target voltage was set to (*$$tem_voltage*);
and 'GO' was pressed to bring up the HT
Objective lens astigmatism was corrected using *$$alignment_specimen* at *$$magnification* magnification
Data will be recorded on *$$media_type*.
. . .

Figure 2 – A fragment of an experimental protocol

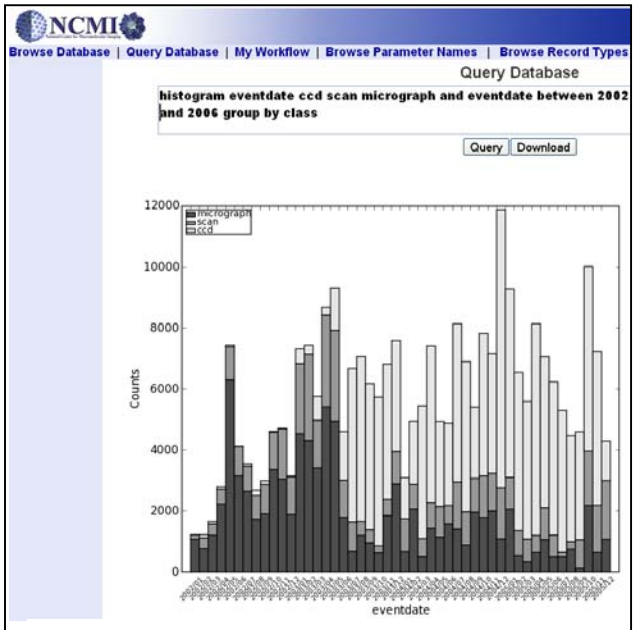

Figure 3 - Sample histogram query

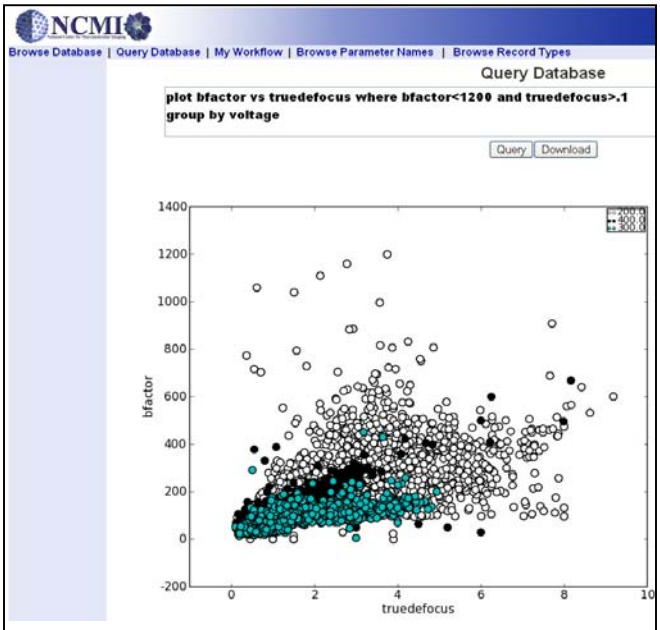

Figure 4 – Sample 2D plot query

https://doi.org/10.1017/S1431927606067663 Published online by Cambridge University Press