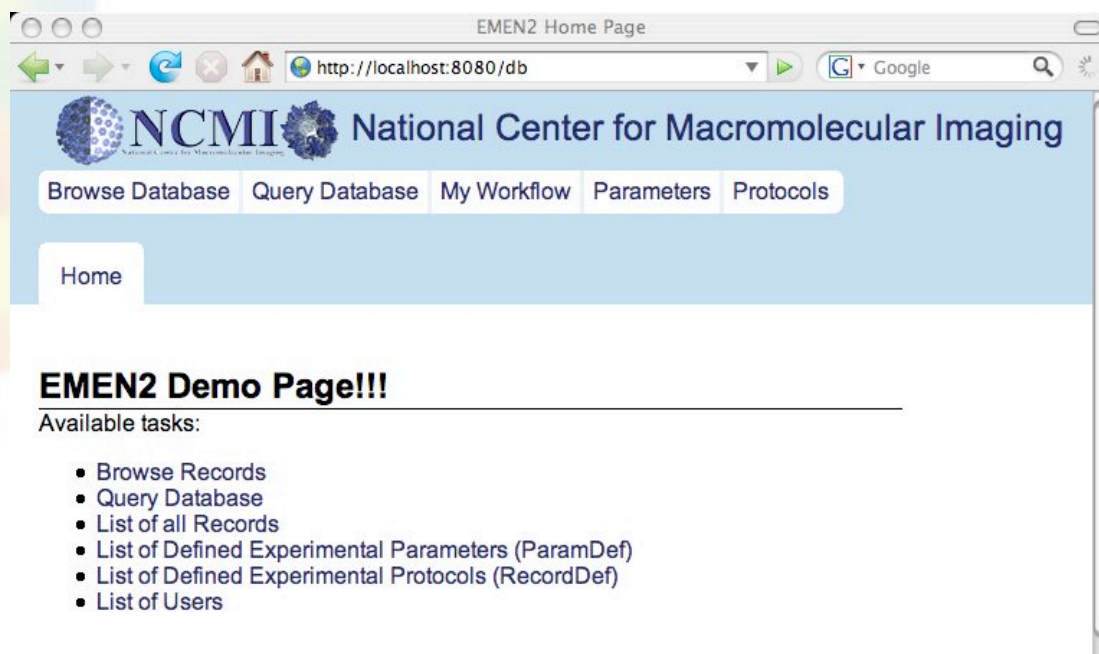# EMEN2: An Extensible, Collaborative, Electronic Lab Notebook

## Problem:

Traditional databases lack the flexibility needed for an electronic lab notebooks. They usually rely on fixed schemas that can be difficult to extend, foresight in preparing tables, and designs optimized for millions of similar records instead of the lower-volume but somewhat heterogeneous data of a labnotebook.

However, unstructured documents like traditional notebooks lack the consistency and searchability of databases.

Is there a way to provide an environment where neither flexibility nor consistency is compromised?

## EMEN2 Goals:

Data sharing internally and externally

Organization in a complex, changing environment

Standardization of user practices

User friendly and extensible by "normal" people

Flexibility and adaptability to a variety of lab types

Security, in-line viewing of record histories

Instrument integration

Structure through flexible hierarchies

Data-mining and analysis

## Overview:

Parameters:

Well defined, semantically correct variable names to store experimental parameters in a consistent way

Protocols:

Free-form, human readable description of an experiment, listing what should be done and what should be recorded

Records:

An instance of a protocol and collected values

# Parameters:

All data is associated with a Parameter type that describes what the value represents, what kind of data is stored, default units or possible value choices, and how the Parameter is related to other Parameters.

The use of well-defined Parameters helps ensure the goal of global consistency which is necessary for meaningful data mining. The use of semantically meaningful Parameter names, the Parameter hierarchy, and concise descriptions helps the user choose the most appropriate Parameter type to record a value.

# Protocols:

### Simple description of an experiment, similar to lab notebook:

A preprepared $grid_prep was placed in a pair of forceps and loaded into the plunger. $cryogen was preprepared below the plunger.$grid_volume of specimen was deposited on the front of the grid using a pipette. The grid was then blotted on $grid_blot_side using $filter_paper_type and the plunger was triggered after a $grid_plunge_delay to rapidly submerse the grid in the cryogen.

### Parameters are embedded into the description and filled-in with each new record created.
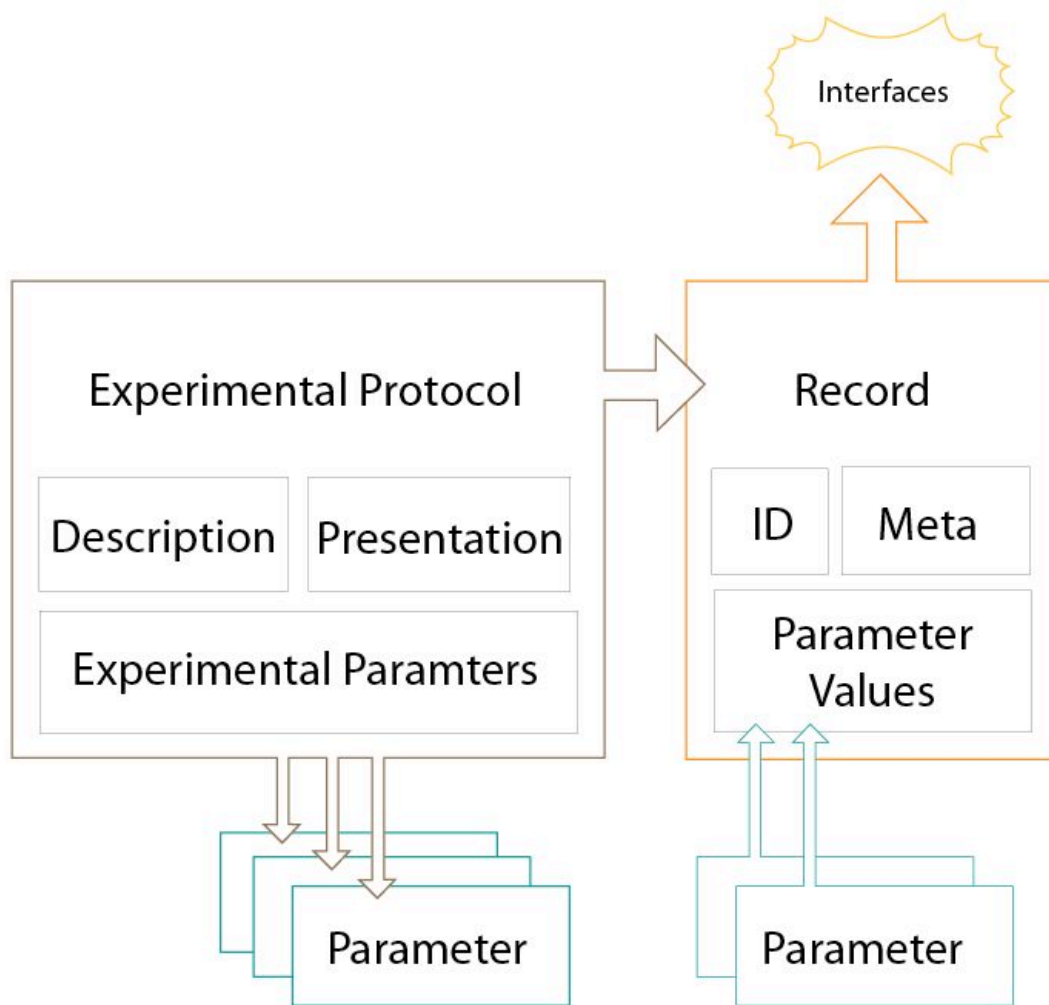
### Protocols also define their own presentation options in the user-interface using a similar syntax. A hierarchy is also used to structure protocols to enhance searching and organization.

## Records:

Each record is an instance of an experimental protocol, like a typical page in a lab notebook. The user (or instrument) enters all the data collected during the experiment as defined by the Protocol.

A key to EMEN2's flexibility is the ability to attach additional parameters and values at any time, beyond what is defined by the protocol. These are fully indexed and integral to the record.

Organization of data and projects is provided through the use of multiply-linked, user defined hierarchies.

Each record inherits base parameters from an Experimental Protocol. Additional parameters may be attached to the record at any time by the user. The Experimental Protocol also defines various layouts for the presentation of the information.
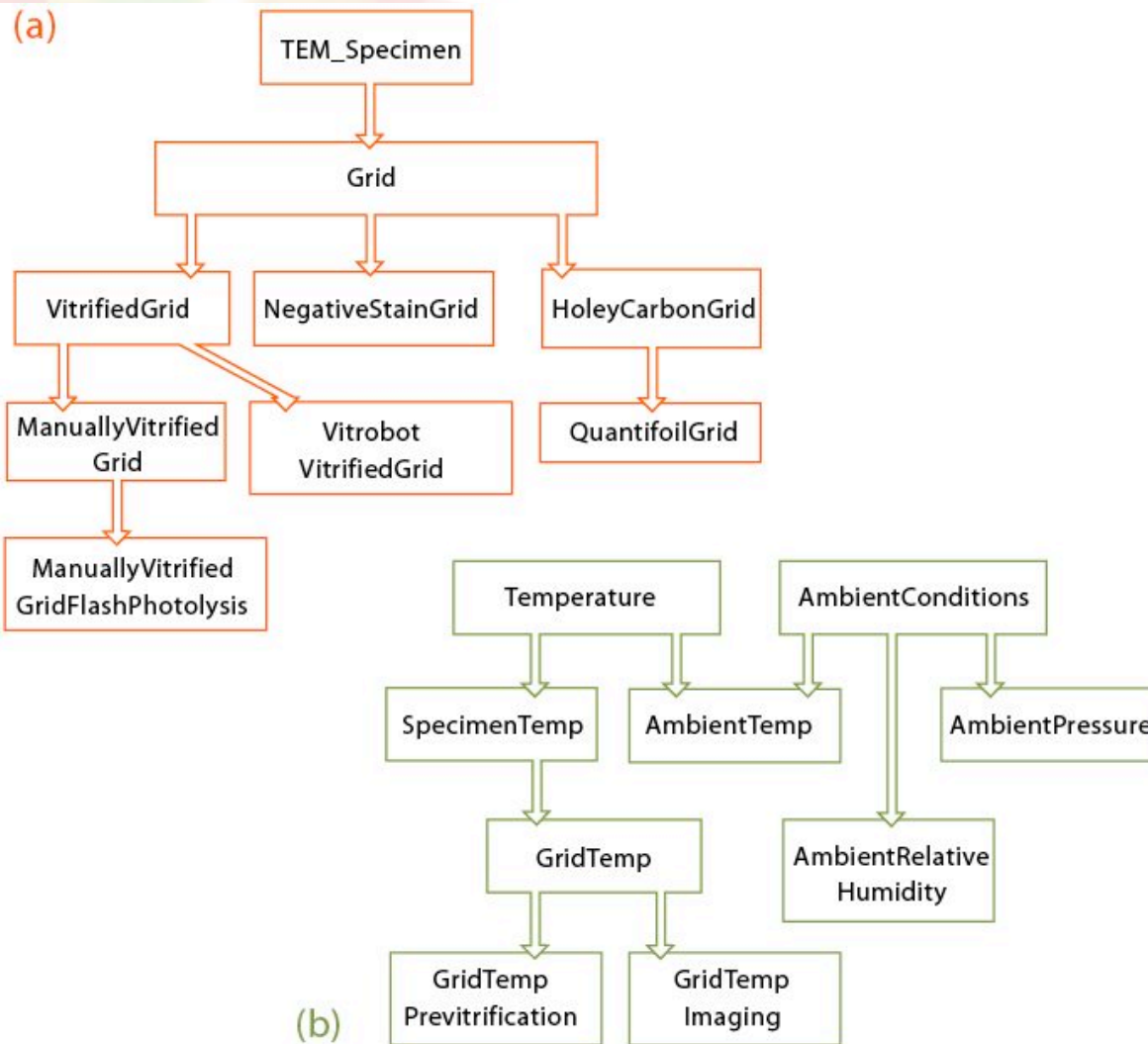
Examples of sections of Experimental Protocol and Parameter hierarchies. Well defined hierarchies are helpful in creating new parameters and protocols while preserving the maintainability and structure of the database.

Database Core:

> Berkeley DB
>
> BSDDB Python Library
>
> Stores Pickled Python objects
>
> Native Python data types
>
> Security checked before all operations
>
> Flexible indexing and storage options for large
>> binary files

Interface:

> Python
>
> Twisted.Web base library for handling HTTP
>
> AJAX used where appropriate to increase
>> responsiveness

# N C M I  |  National  Center  for  Macromolecular  Imaging

Friendly, efficient web-based interface with quick
access to common tools and information

## Hierarchical organization, useful table summaries



📁 EMEN2 → 📁 All Groups → 🗃 NCMI → 📖 Project ▆▆▆▆▆ → 📖 Project ▆▆▆▆▆ Wah Chiu

Project ▆▆▆▆▆ Wah Chiu » All Children » labnotebook (6)  microscopy (6)  purification (3)

# microscopy

(0-6 of 6)

| Date of Experiment | Link to microscope | Data entered by | Freezing session ID | Condenser used | Size of electron beam spot | Size of aperture | Magnification | Methods used | Ice thickness | m: childcount |
|---|---|---|---|---|---|---|---|---|---|---|
| 2004/10/20 | | | | 70 | 1.0 | 40 | 40.0 | FasTEM MDS | About right | 34 |
| Purpose of Experiment: collecting data | | | | | | | | | | |
| 2004/03/24 | | | | 70 | 1.0 | 40 | 50.0 | Yoshi box | About right | 102 |
| Purpose of Experiment: collecting data | | | | | | | | | | |
| 2004/10/28 | | | | 150 | 2.0 | 60 | 40.0 | MDS | About right | 281 |
| Purpose of Experiment: collecting data | | | | | | | | | | |
| 2004/03/31 | | | | 70 | 1.0 | | 50.0 | Yoshi box | too thin to about right | 136 |
| Purpose of Experiment: collecting data, Sample E2 box | | | | | | | | | | |
| 2004/04/05 | | | | 70 | 1.0 | 40 | 50.0 | Yoshi box | About right | 427 |
| Purpose of Experiment: collecting data | | | | | | | | | | |

# Integrated Viewing of Data:

Image Browser:

Google-Maps Style
interactive viewing
of image data with
fast zooming and
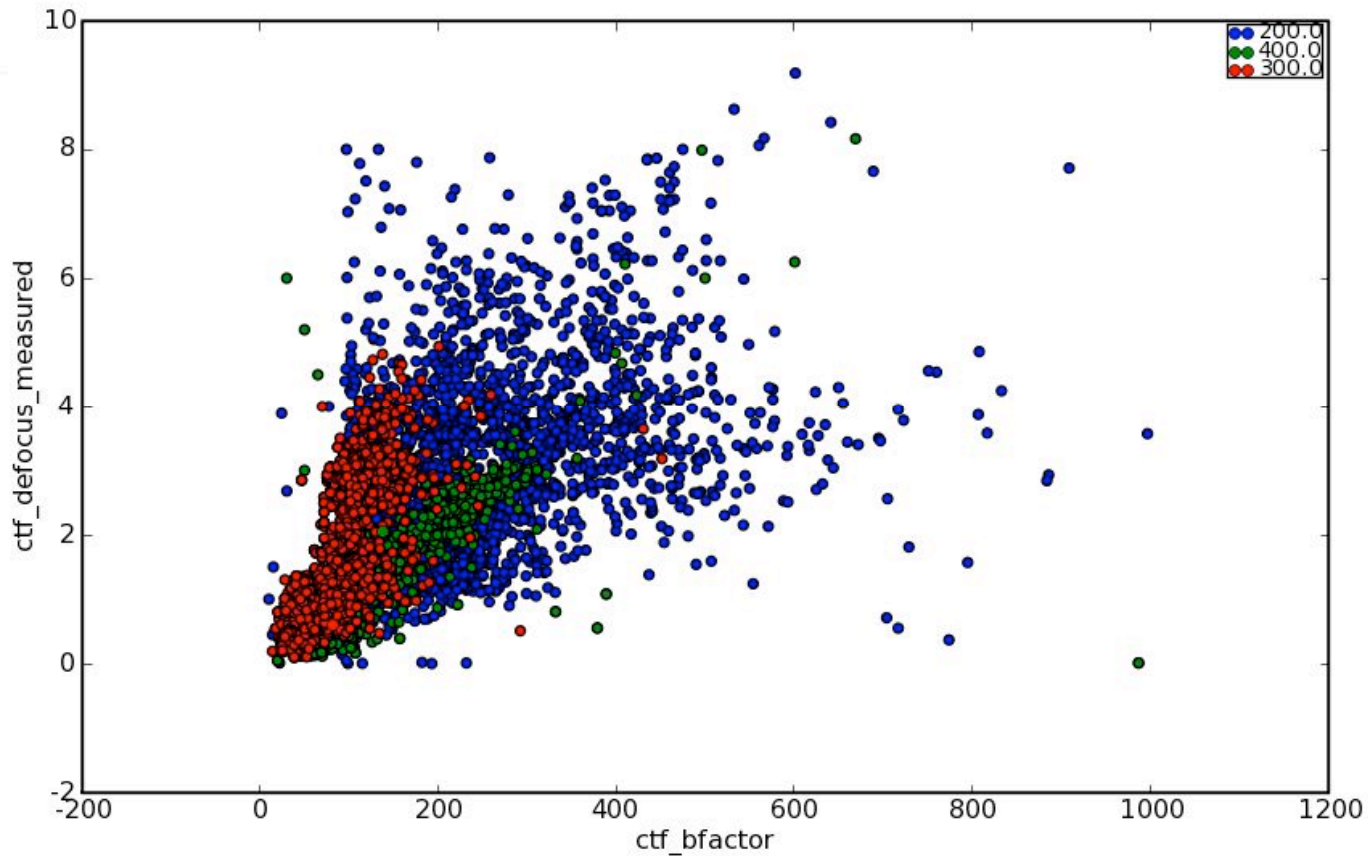panning

Quick access to
data quality
evaluation tools
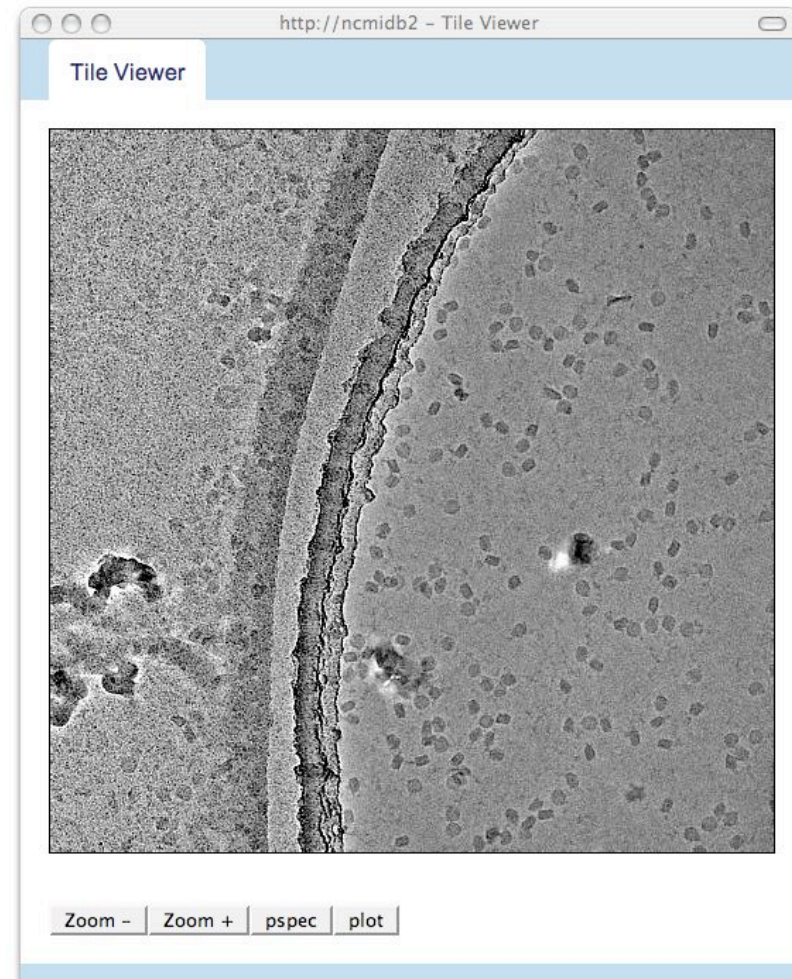
## Flexible query language with integrated graphing

# Acknowledgments:

### NCMI

Wah Chiu

Steven Ludtke

Chris Booth

Haili Tu

Hari Damodaran

Kurt Welgehausen

### Support:

National Center for Research Resources (NCRR)

Agouron Institute