


Searching and Coordinating Databases: OWLdb

C. Forbes Dewey, Jr.
Massachusetts Institute of Technology

CBI Safety Surveillance Workshop, March 13, 2008

© cfdewey 2007



My Drug Development Colleagues

| | |
|-----------------------|-------------------------|
| <i>Peter Hunter</i> | <i>Paul Matsudaira</i> |
| <i>Richard Kitney</i> | <i>Harvey Lodish</i> |
| <i>Mark Wilkinson</i> | <i>Steve Tannenbaum</i> |
| <i>Judith Blake</i> | <i>John Essigman</i> |

My CytoSolve, OWLdb and ExperiBase Collaborators


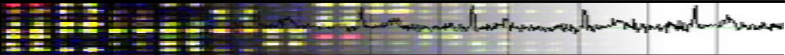
Jason Zhang
Shiva Ayyadurai
Ceryen Tan
Kurt Stihl
Khanh Dang
Kaori Hagihara

My MIT-Singapore Collaborators

Paul Matsudaira
Sourav Bhowmick
Harry Lu

Support: NIH, Singapore-MIT Alliance, NIH, PNNL


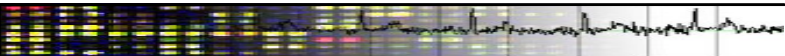
© cfdewey 2007



Why am I here?

- ❖ Background in experimental biology and computers
- ❖ A glutton for punishment (dreary semantics issues)
 - ◆ Treating each experimental method with a common view: *ExperiBase*
 - ◆ Database federation: Scalable queries across multiple data sources: *OWLdb*
- ❖ A sucker for unsolved systems problems
 - ◆ Scalable solution methods for massively parallel biological pathways: *Cytosolve*
- ❖ Energized buy the challenge of drug development

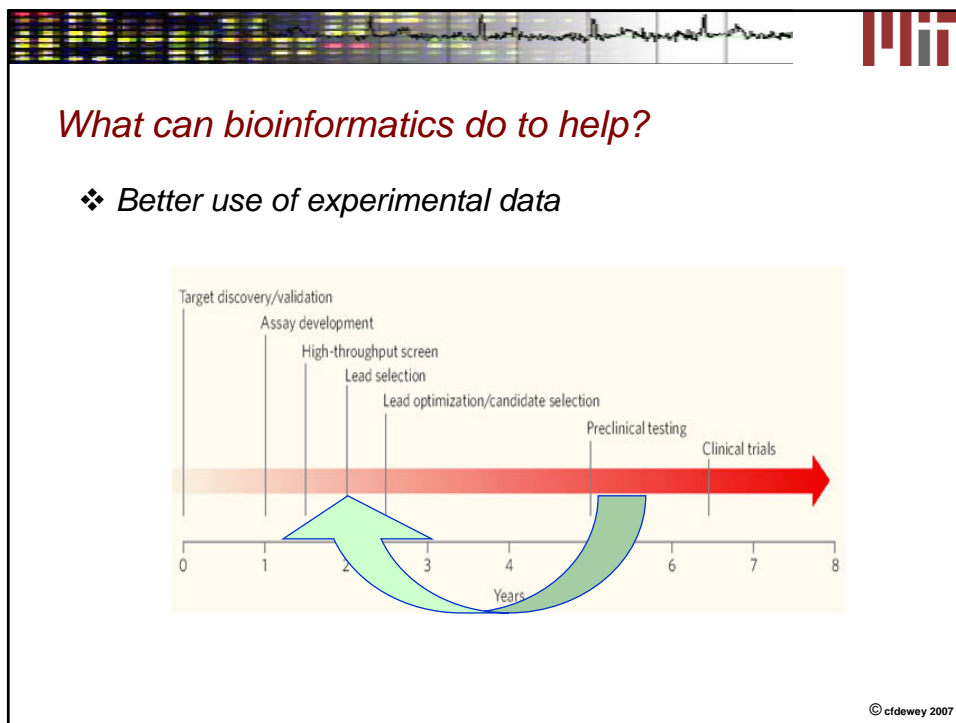
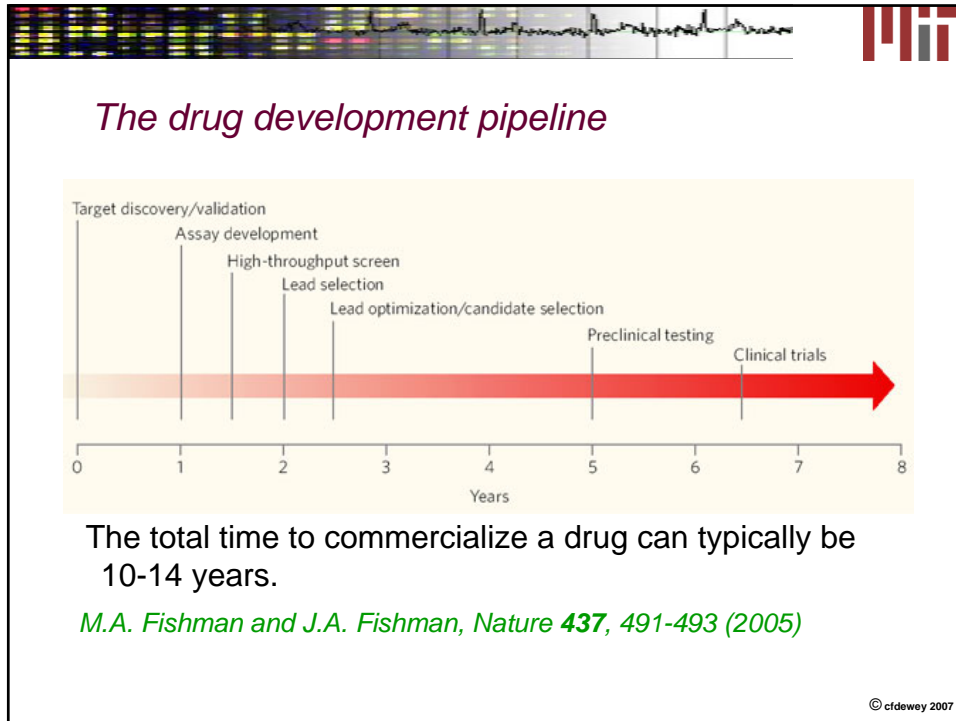
© cfdevey 2007

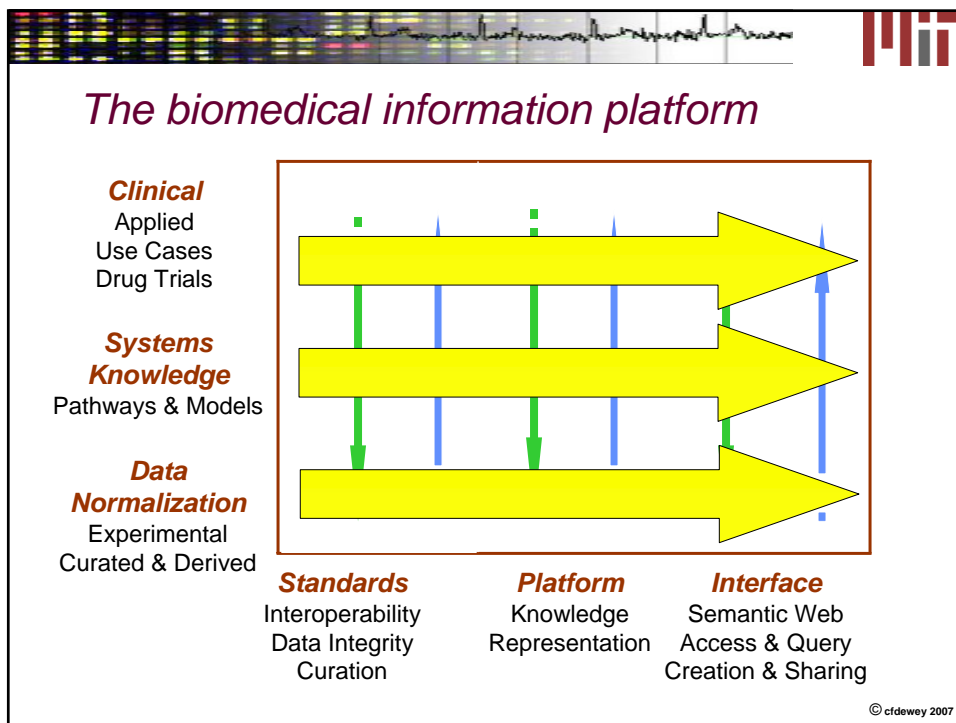
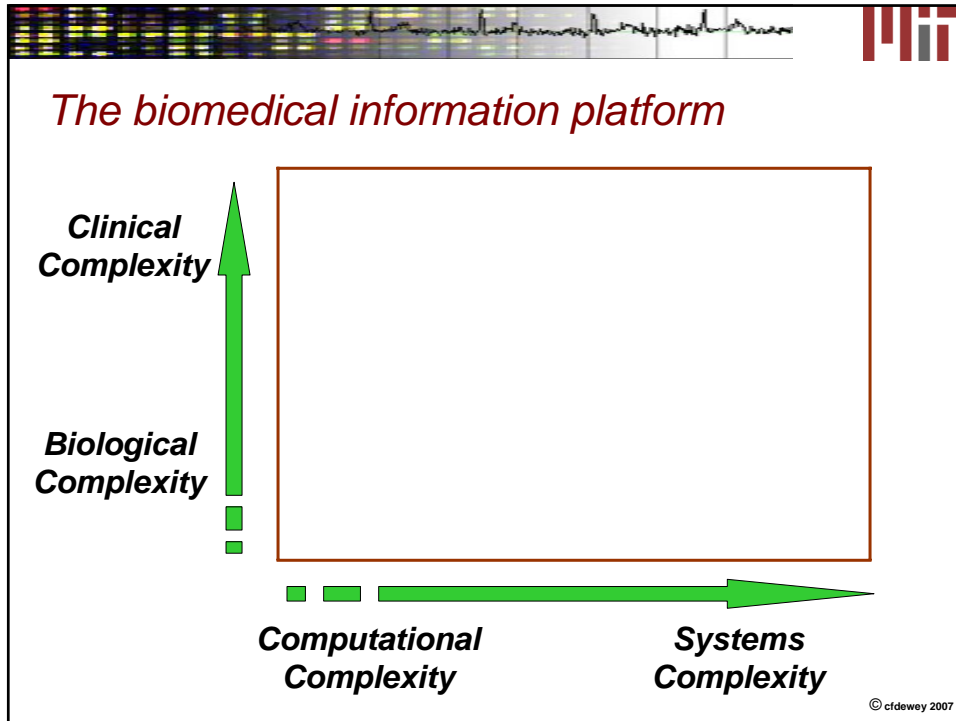



Here's today's agenda . . .

- ❖ Introductory remarks about drug development
- ❖ Systems biology issues for drug development
- ❖ Some solutions for experimental biology
 - *ExperiBase*: Treating each experimental method with a common view
 - *Cytosolve*: Scalable computations for multiple pathways
- ❖ Extracting information definition to ontologies
 - *OWLdb*: Rendering databases in ontological form
 - *Unimodels*: Biological pathways in ontological form
- ❖ Final philosophical comments

© cfdevey 2007








*Intelligent experimental data archives . . .
Why do we need them???*

- ❖ Archival data storage and retrieval (e.g. FDA)
- ❖ Collaboration between different biological experiments
 - Multiple data types: images, microarrays, mass spec, FACS
- ❖ Facilitate analysis involving many data sources
- ❖ Create systems that can provide complex information and lead to unique discoveries and solve serious problems like surveillance

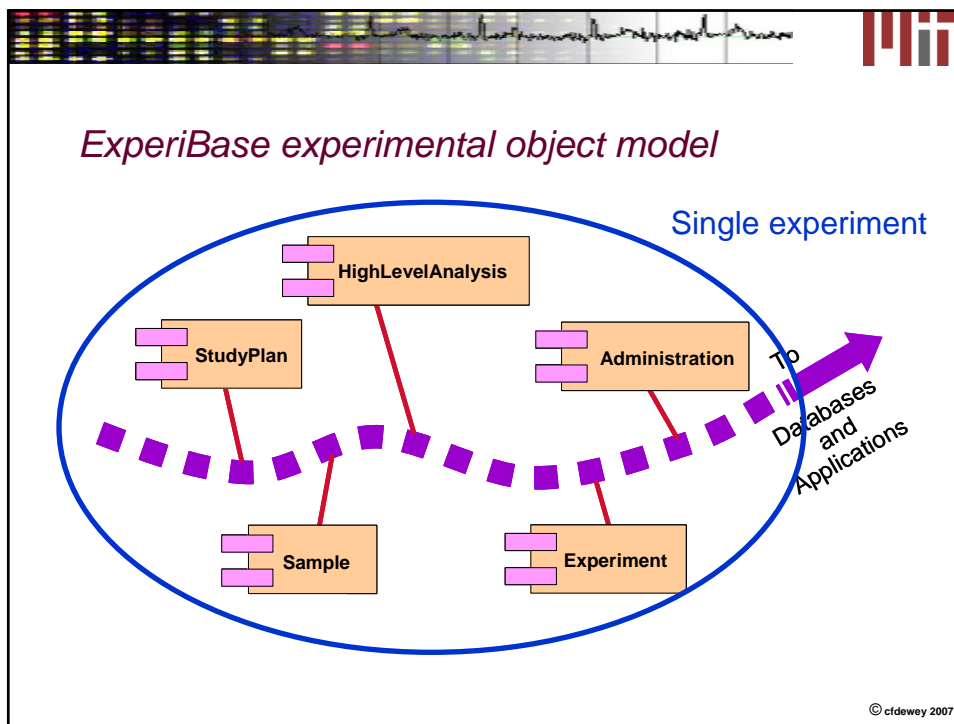
© cfdewey 2007




ExperiBase

- ❖ Based on ontology standards
- ❖ Conceptual consistency between different experimental methods
- ❖ Reuse of concepts between different experimental methods
- ❖ Portable platform independent of OS

© cfdewey 2007



-
- Current support by ExperiBase*
- ❖ Gel Electrophoresis
 - Western Blot
 - 1D Gel (Converted from HUPo)
 - 2D Gel
 - ❖ Flow Cytometry / FACS
 - ❖ Microarray Experiments
 - ❖ Mass Spectrometry
 - ❖ Microscope Images
- Complete — In progress — Preliminary
- © cfdewey 2007




Large-scale biological pathways . . .

The next grand challenge

- ❖ Biology producing protein-protein associations
- ❖ The thrust is to identify “systems”
- ❖ The current art is connection drawings
- ❖ The goal is *quantitative* biological models

. . . . at the cell level and beyond

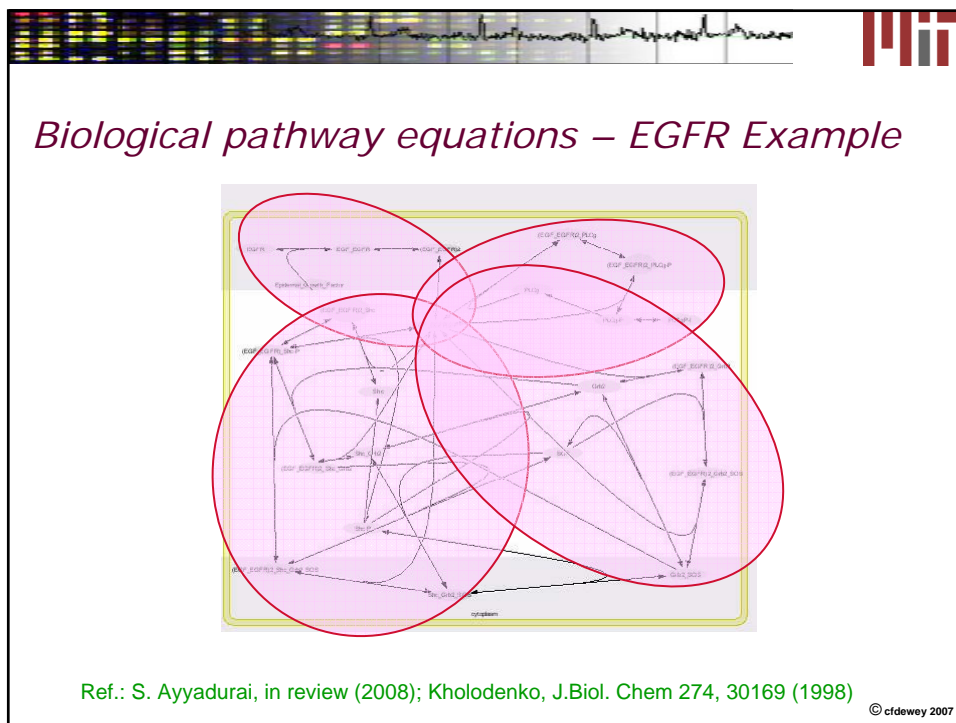
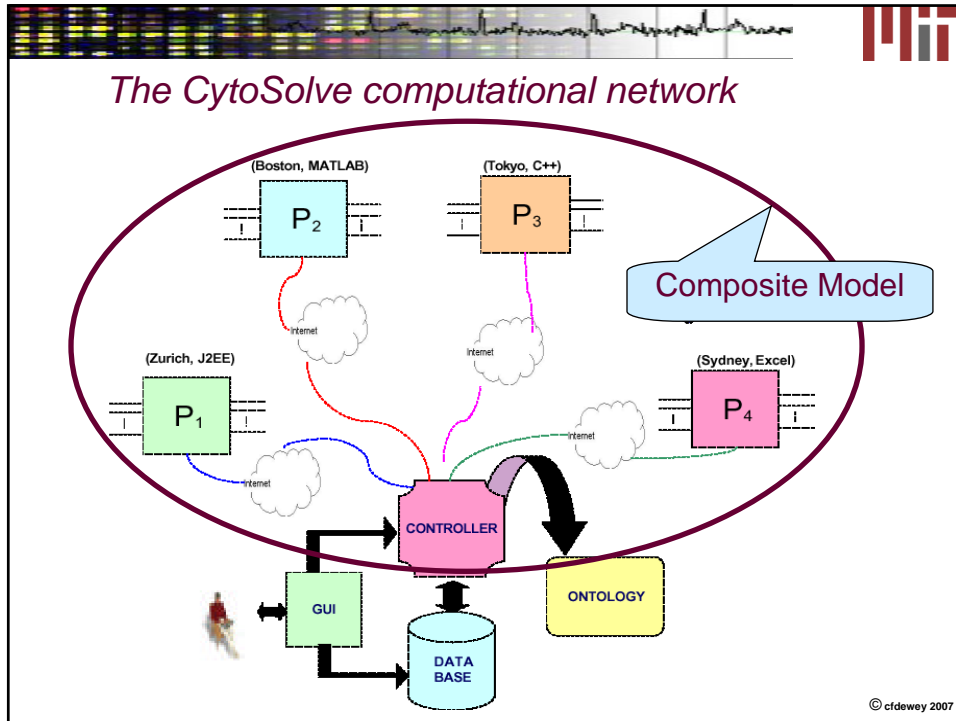
© cfdewey 2007



CytoSolve

- ❖ A globally-distributed pathway architecture
- ❖ Connects independent pathways into one system
- ❖ Scalable to large problems ($>10^4$ equations)
- ❖ Based on existing computational semantics
 - [SBML](#), [CellML](#), [MML](#) mathematical descriptions
- ❖ Tolerant of other computational environments:
 - [C](#), [C++](#), [Matlab](#), other with appropriate interface
- ❖ Reuse of semantics between different solution methods
- ❖ Portable platform independent of OS

© cfdewey 2007



What are the common problems for ExperiBase and CytoSolve?

- ❖ Semantics
 - Need unique definitions
 - Need easy way to compare semantics
- ❖ Curation
 - Need to update at a higher level and subsequently update all derivative documents
- ❖ Functionality
 - Need a way of tagging related information and associating it with the original data/solution

© cfdewey 2007

OWLdb: A new paradigm for creating databases

```

    graph LR
        subgraph User_Function [User Function]
            CDB[Create Database]
            CDO[Change Database Organization]
            GD[Get Data]
        end
        subgraph Graphical_Ontology_Editor [Graphical Ontology Editor]
            BO[Build OWL Ontology]
            MO[Modify OWL Ontology]
            RFO[Read Fields From Ontology]
        end
        subgraph Conversion_Software [Conversion Software]
            SC[SQL Converter]
            SM[SQL Merge]
            SR[SQL Request]
        end
        subgraph Database [Database]
            SQLDB[SQL Database]
        end
        CDB --> BO
        CDO --> MO
        GD --> RFO
        BO --> SC
        MO --> SM
        RFO --> SR
        SC --> SQLDB
        SM --> SQLDB
        SR --> SQLDB
    
```

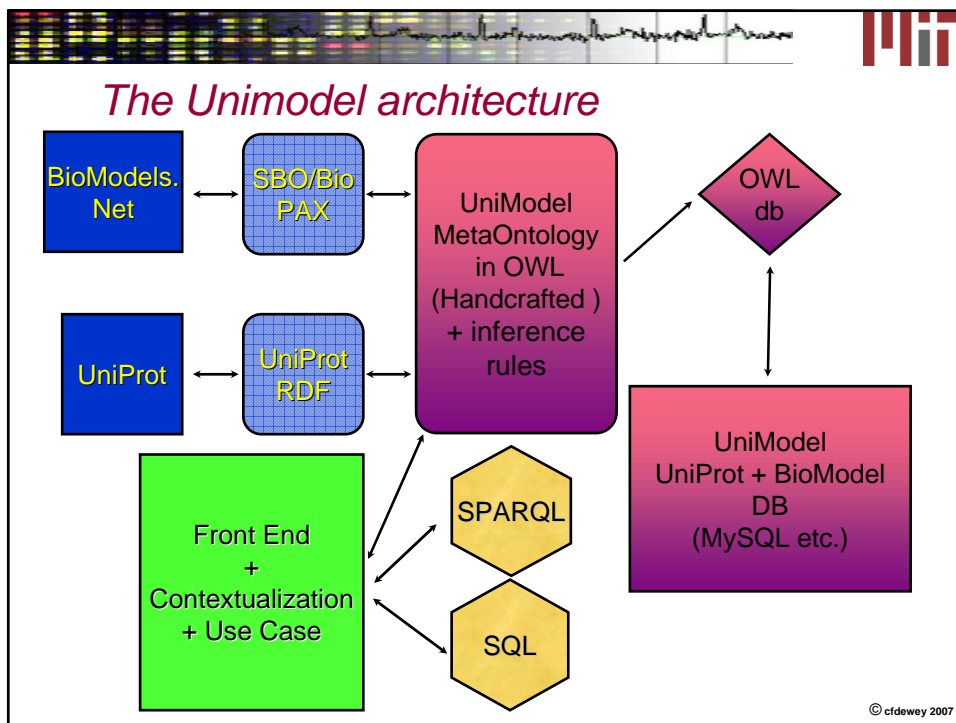
... . Courtesy of Kurt Stihl

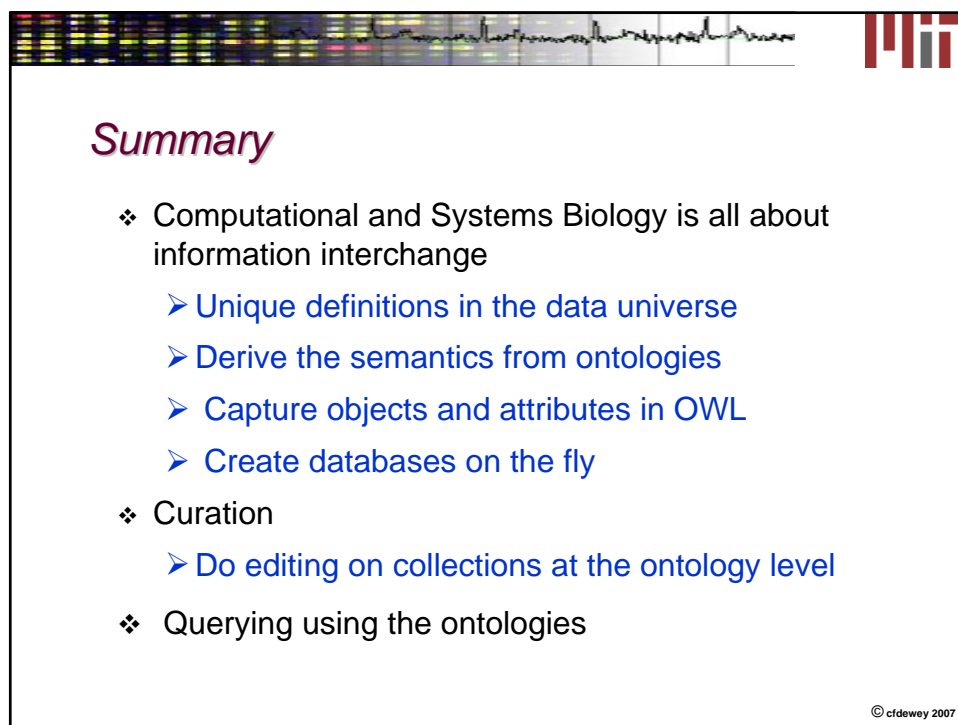
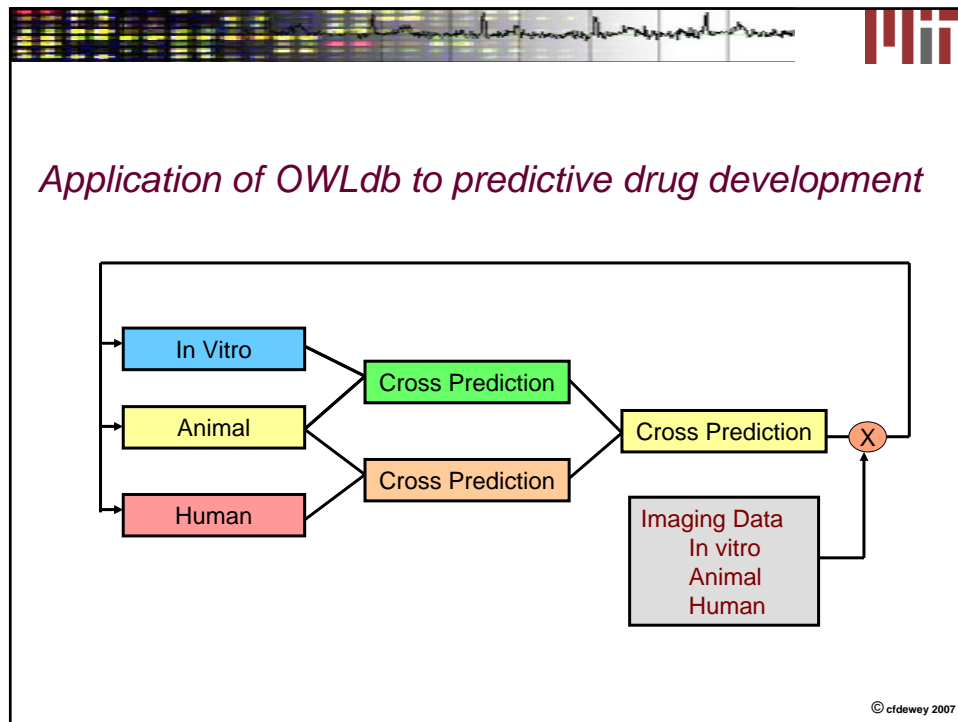
© cfdewey 2007


Key Components

| | <i>UniModels</i> |
|--------------|--------------------------------------|
| Registry | LSID |
| Interface | UniModels Ontology |
| Nomenclature | UniModels / SBO / UniPROT RDF |
| Editor | Protege/SBMLEditor |
| Parser | Jena |
| Reasoner | Jena/OWLdb |
| Storage | Oracle / MySQL / PostgreSQL / Sesame |
| Query | SPARQL / SQL / D2RQ / SeRQL |

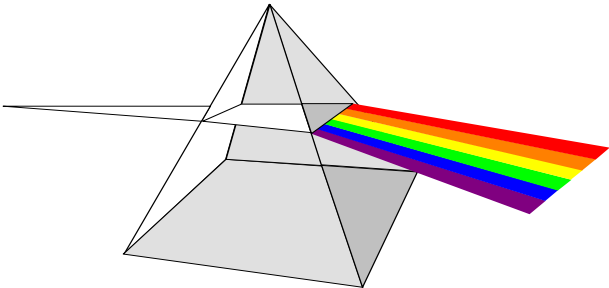
© cfdewey 2007







Engineering for Improved Health Care



© cfdewey 2007