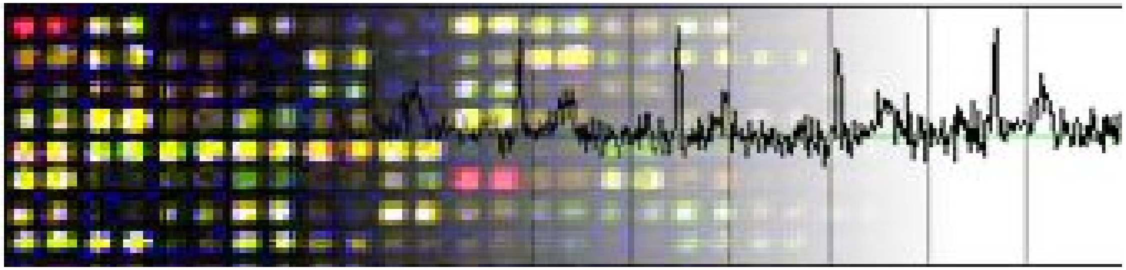
MIT OpenCourseWare
http://ocw.mit.edu

20.453J / 2.771J / HST.958J Biomedical Information Technology
Fall 2008

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.



Biomedical Information Technology

2.771J 20.453J HST.958J SMA5304 Spring 2008

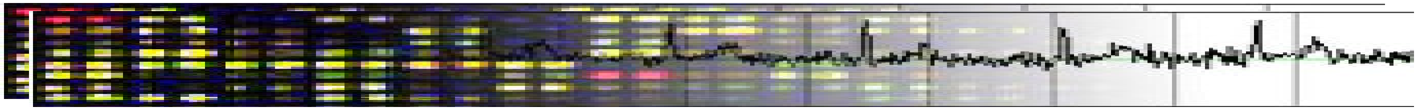
Lecture 4 September 2008

Introduction

Welcome to the class

- Information – the class web site
- Who to contact
- The schedule and homework issues
- Objectives and methodology
- The information-driven scientific method
- Ontologies and semantics for biomedical information
- Term paper instructions

© cfdewey 2008



Biomedical
Information
Technology

Meet the staff

Instructors:

Forbes Dewey	*MIT*
Sourav Bhowmick	*NTU*
Hanry Yu	*NUS*

Teaching Assistants:

Huey Eng Chua	*NTU*
Boon Soo Seah	*NTU*
Baracah Yankama	*MIT*

© cfdewey 2008

The Syllabus

Introduction (1 Week)

The course in outline

Term papers

Scope of Applications (1 Week)

Biological and medical data

Basic Technologies (4 Weeks)

Storing and querying biomedical data

Relational databases, querying methods

XML data

Ontologies (2 Weeks)

What they are and how they can be exploited

Biological Pathways(1.5 Week)

Quantitative descriptions of biology

Data Integration (1.5 Weeks)

Discussion of several large integration projects

Student Presentations and Summary (2 Weeks)

© cfdewey 2008

The Syllabus

Introduction (1 Week)

The course in outline

Term papers

Scope of Applications (1 Week)

Biological and medical data

Basic Technologies (4 Weeks)

Storing and querying biomedical data

Relational databases, querying methods

XML data

Ontologies (2 Weeks)

What they are and how they can be exploited

Biological Pathways(1.5 Week)

Quantitative descriptions of biology

Data Integration (1.5 Weeks)

Discussion of several large integration projects

Student Presentations and Summary (2 Weeks)

© cfdewey 2008

The Syllabus

Introduction (1 Week)

The course in outline

Term papers

Scope of Applications (1 Week)

Biological and medical data

Basic Technologies (4 Weeks)

Storing and querying biomedical data

Relational databases, querying methods

XML data

Ontologies (2 Weeks)

What they are and how they can be exploited

Biological Pathways(1.5 Week)

Quantitative descriptions of biology

Data Integration (1.5 Weeks)

Discussion of several large integration projects

Student Presentations and Summary (2 Weeks)

© cfdewey 2008

Reading and Homework

- Read the primary papers for each session *before* the session.
  - Use the secondary papers and any other material you can find to expand you understanding and answer homework.
- All homework is due at the beginning of the assigned class. “We take no prisoners.” Primarily electronic submission.
- Tutorials can be run from your own computers.
- Download Cell Designer:
  http://www.celldesigner.org/

© cfdewey 2008

Original work

There are three forms of student creativity that are recorded in this class. First, each student must show mastery of the materials in the homework problems. Second, we have tutorials in which each student should achieve a high level of competence. We do not test this formally except to observe the ability of each student to perform the stated exercise. Third, we have individual projects. You must make sure that your original work and creativity are apparent in the final presentation. Just reading from a review article is far short of our standard. In all three cases, If a student consults with others or uses published materials, this must be recorded and fairly described. Plagiarism will not be tolerated.

© cfdewey 2008

Term paper instructions

Each student in the course is required to present a term project that illustrates the use of the course material in a real information technology case in biology or medicine. The actual content of the case can vary depending upon the student's interests and existing skills. Projects can range from general studies of a class of problems and the recommendation of a solution to a detailed implementation in running software.

© cfdewey 2008

Medicine is an information science and a healing art

Our objective is to create information systems that serve the development of biology and medicine. In the end, we want to make this information available in human treatment of disease and the enhancement of life.

© cfdewey 2008

Active areas of modern biology

- *Biological discovery*
  - *How to knock out specific proteins*
  - *How to modify the genome*
  - *How to use stem cells appropriately*
- *Mechanisms of cells, tissues and organs*
  - *Predictive molecular dependencies*
  - *Designable living biological constructs*
  - *Direct intervention in disease states*
- *Genetically-aware personal healthcare*
  - *"What-if" scenarios for medical treatment*
  - *Personal genome for genetically-aware therapy*

© cfdewey 2008

Information technology in modern biology

- *Biological discovery*
  - *Queryable archives for high-throughput proteomics*
  - *Processing algorithms to understand data*
  - *Search for new relationships between known facts*
- *Mechanisms of cells, tissues and organs*
  - *Predictive models for interactions between proteins*
  - *Predictive models for molecular ensembles*
  - *Toxicity and dose predictions in drug development*
- *Genetically-aware personal healthcare*
  - *Predictive response to pharmaceuticals*
  - *Better control and prevention of chronic diseases (Alzheimer's disease, obesity, AIDS, and diabetes)*

© cfdewey 2008

Objectives

- Introduce the subject of bioinformatics
- Demonstrate the current state of the art
- Teach new methods to approach the field
- Provide design experience through a project
- Develop practical applications to biological and medical problems

© cfdewey 2008

Methodology

- Set the intellectual context
- Define the scientific and engineering challenge
- Design a solution
- Implement the solution
- Examine the consequences

The scientific method

© cfdewey 2008

Keys to biomedical computing standards

- Semantics

  Investigators can agree on meaning

  Ontologies for standardizing meaning

  Curation of ontologies – the *LSID* identifier

- Schema

  Share schema and concepts – *Ontologies*

- Scaleability

  The ability to scale to larger problems in the future

- Standard tools

  Common ontologies and schema for sharing data

  Reusable software!!!

© cfdewey 2008

Semantics and ontologies

Semantics:

The science of meanings . .
for communication and interchange of scientific information

Ontology:

A specification of a conceptualization

Tim Gruber, Stanford, circa 1993

© cfdewey 2008

The big impediment . . . sematics

- *What does the word "sample" mean?*

  (Quiz: "Football" = ?)

- *How can one establish meaning with certainty?*

  Use ontologies to define objects and concepts

- *How can meanings be compared and combined?*

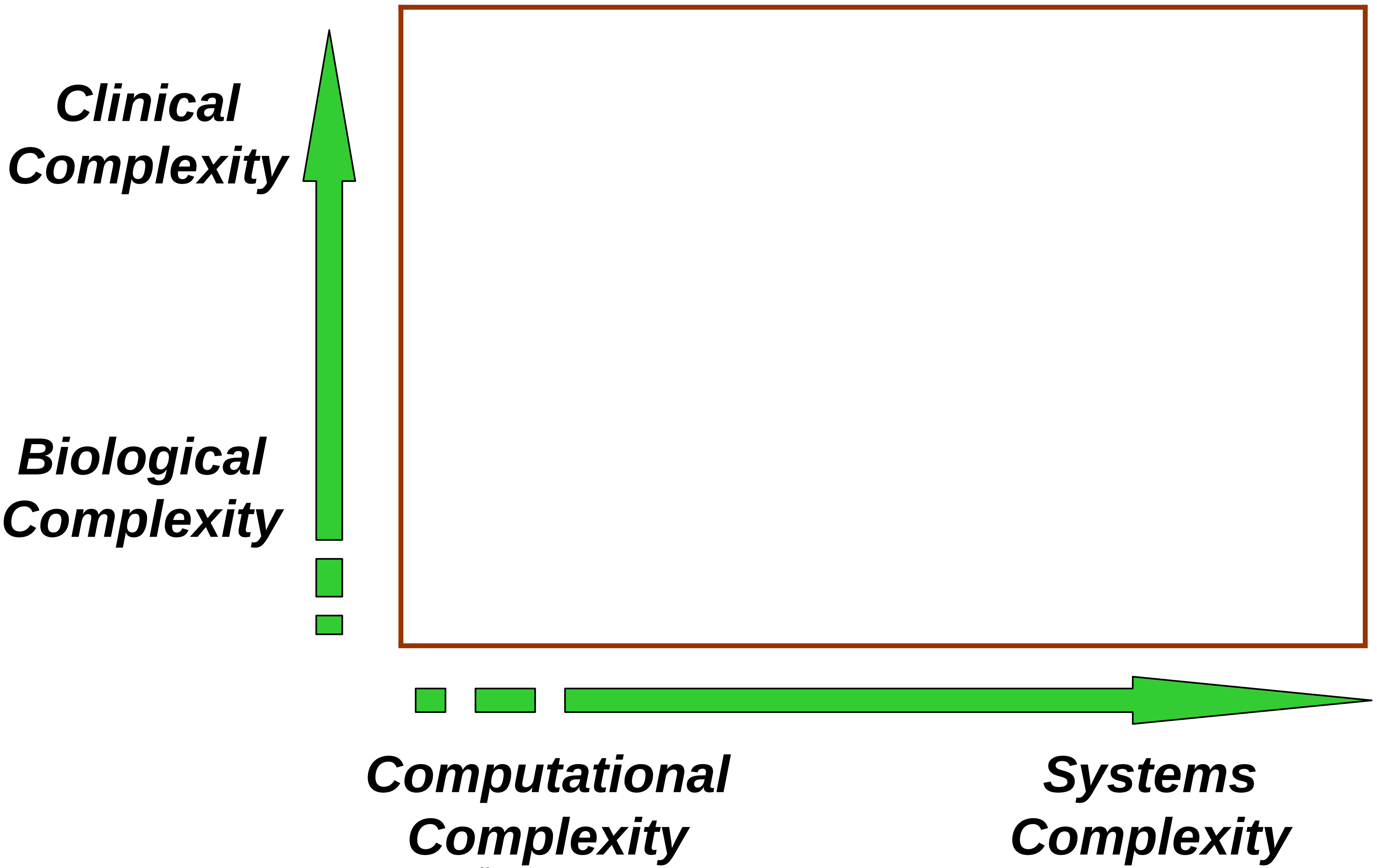
  Use the Web Ontology language . . ***OWL***

- *How can one create a "collisionless" schema?*

  Find a good example from a related field

© cfdewey 2008

The biomedical information platform



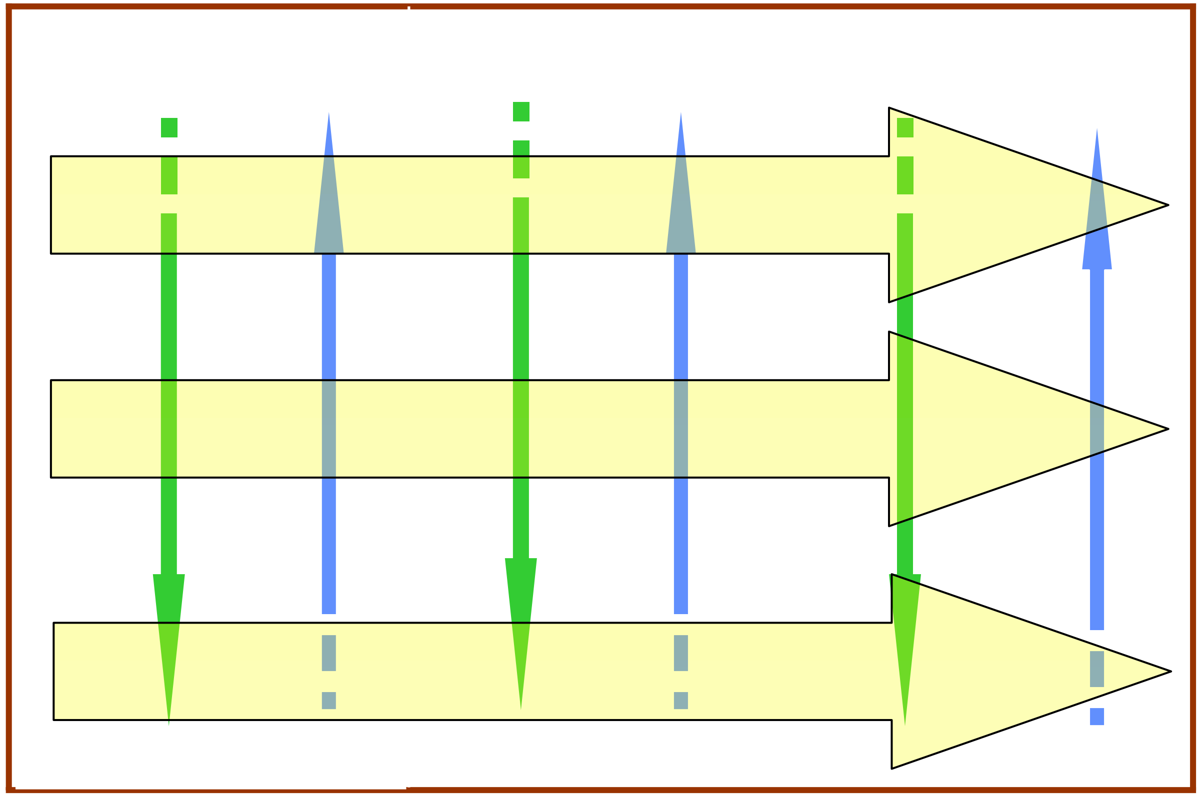
© cfdewey 2008

The biomedical information platform

Clinical
Applied
Use Cases
Drug Trials

Systems Knowledge
Pathways & Models

Data Normalization
Experimental
Curated & Derived



Standards
Interoperability
Data Integrity
Curation

Platform
Knowledge
Representation

Interface
Semantic Web
Access & Query
Creation & Sharing

© cfdewey 2008

The Syllabus

© cfdewey 2008

Scope of Applications (1 week: HY)

Types and characteristics of biological and medical data

- *Distributed data systems*
- *The life cycle of scientific data*
- *Current challenges*

Examples from liver fibrosis

- *Gel Electrophoresis*
- *Microarrays*
- *FACS and other methods*
- *Creating biological pathways*
- *Designing new experiments*
- *Integrating information from the literature*

© cfdewey 2008

Fluorescence Activated Cell Sorter (FACS) (Flow cytometry)

Quantitative, multiple parameter analysis of large numbers of individual cells

- Cell surface markers
- Intracellular proteins
- Ca++ mobilization

12 colors and 15 parameters
Sorting: 60,000 cell/second

Ref.: Jianmin Chen, MIT

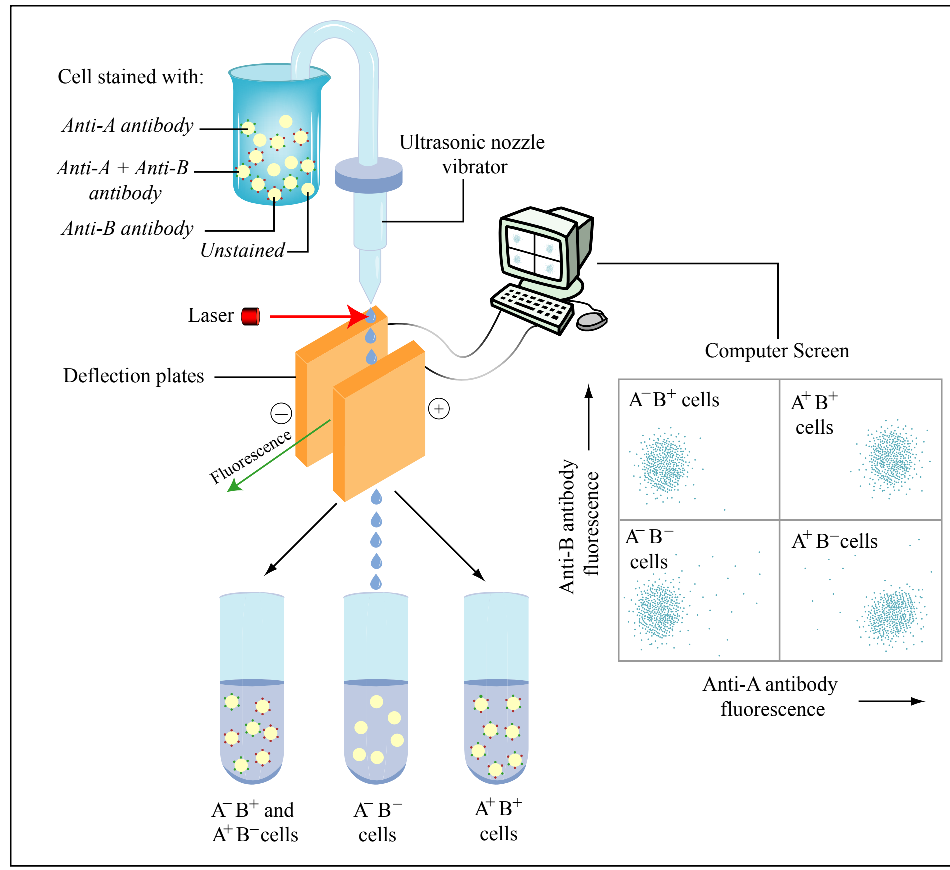


Figure by MIT OpenCourseWare

© cfdewey 2008

Integrating biomedical data (4 weeks: SSB)

All you ever wanted to know about databases

- *Relational model*
- *Database schema*
- *Query methods using SQL*

Trees and graphs

XML

- *Schema for XML relations*
- *Querying XML*

Data integration without semantics

© cfdewey 2008

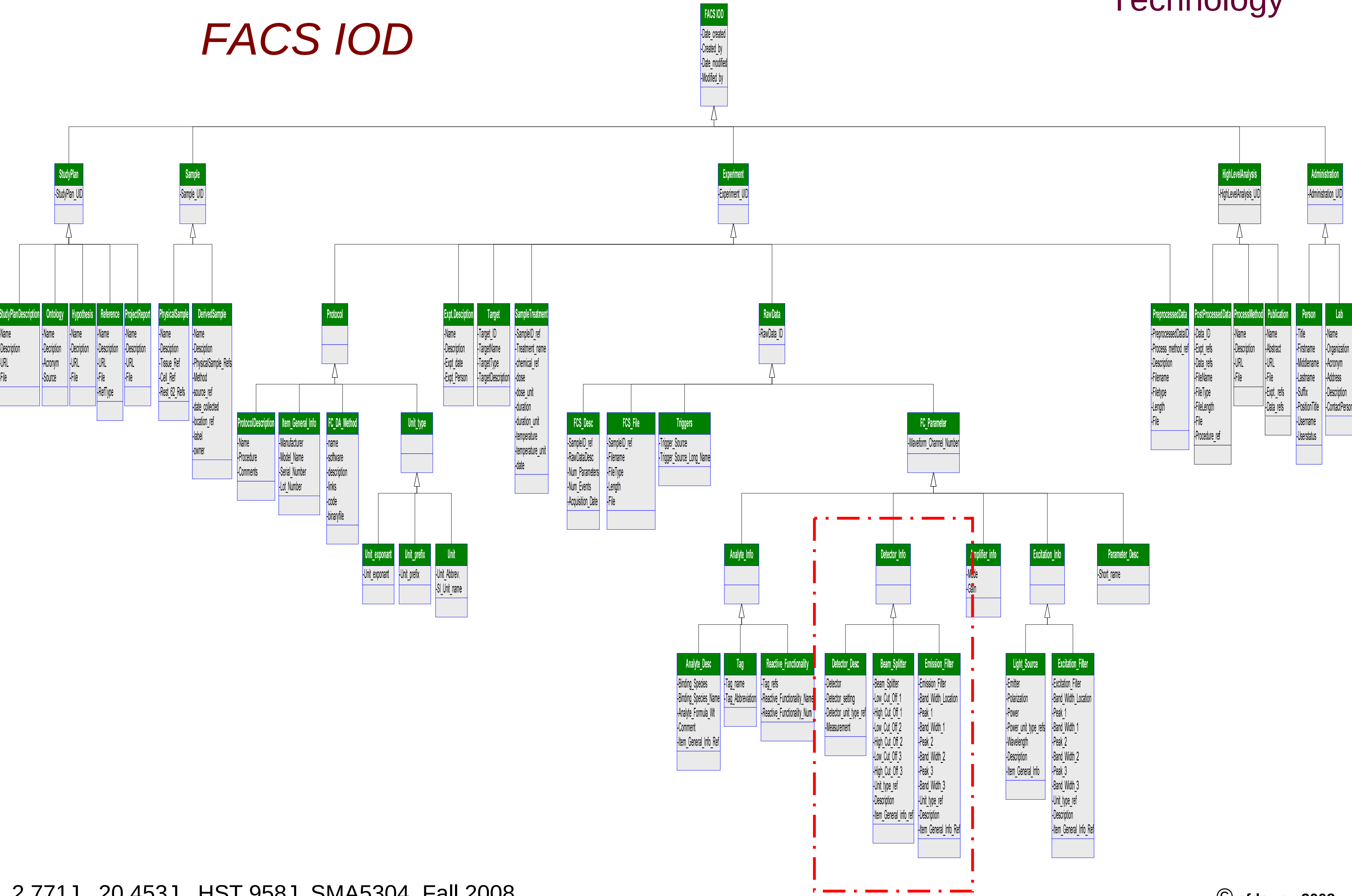
An example – Fluorescence Activated Cell Sorting (FACS)

- *Illustrate the semantics issues*
- *Give an example of how to express semantics*
- *Show how XML can be used in the description*

© cfdewey 2008

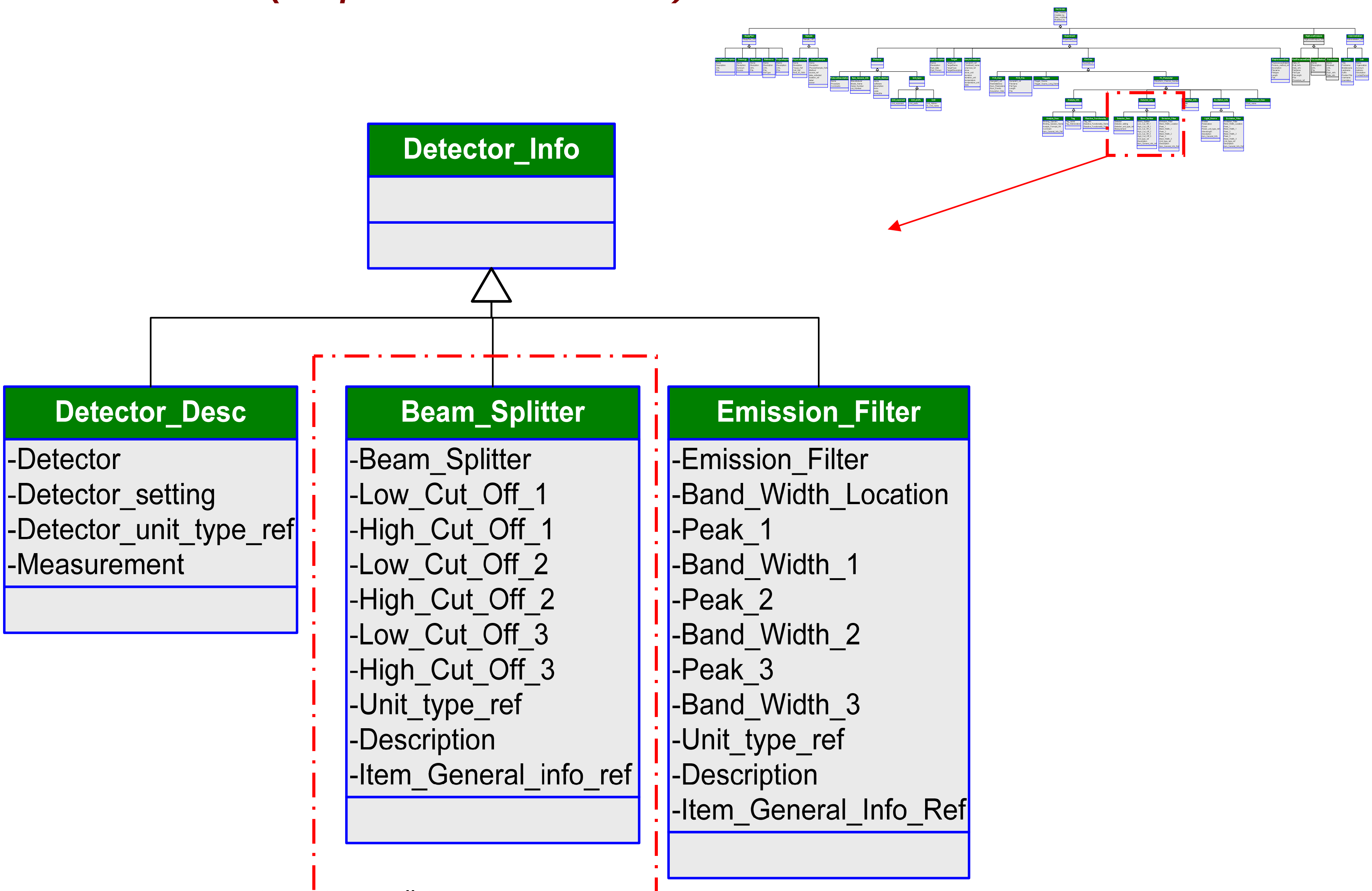
Biomedical Information Technology

FACS IOD



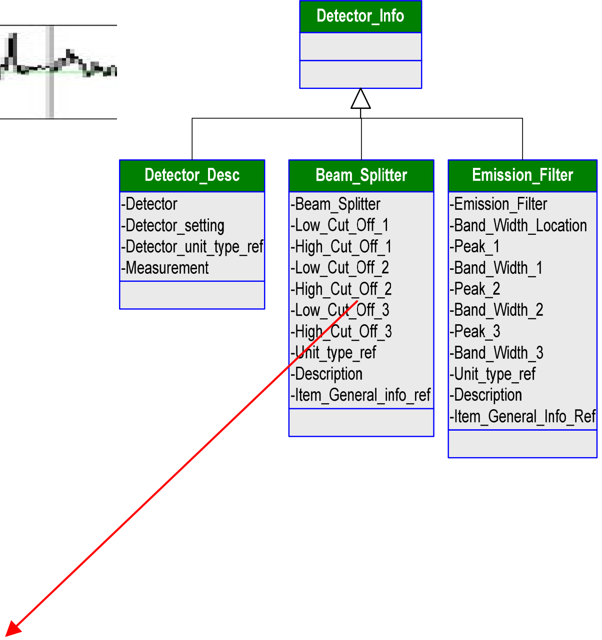
© cfdewey 2008

FACS IOD (Expanded Portion)



© cfdewey 2008

XML schema




```
<simpleType name="Beam_Splitter_Simple_Type">
    <restriction base="dicom:Bd_64_Type">
        <enumeration value="Mirror"/>
        <enumeration value="Dichroic_Reflect_Low"/>
        <enumeration value="Dichroic_Reflect_High"/>
        <enumeration value="Block_All"/>
        <enumeration value="Other"/>
    </restriction>
</simpleType>
<!--xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx-->
<complexType name="Beam_Splitter_Type">
    <sequence>
        <element name="Beam_Splitter" type="filters:Beam_Splitter_Simple_Type"/>
        <element name="Low_Cut_Off_1" type="filters:Wavelength_Type" minOccurs="0"/>
        <element name="High_Cut_Off_1" type="filters:Wavelength_Type" minOccurs="0"/>
        <element name="Low_Cut_Off_2" type="filters:Wavelength_Type" minOccurs="0"/>
        <element name="High_Cut_Off_2" type="filters:Wavelength_Type" minOccurs="0"/>
        <element name="Low_Cut_Off_3" type="filters:Wavelength_Type" minOccurs="0"/>
        <element name="High_Cut_Off_3" type="filters:Wavelength_Type" minOccurs="0"/>
        <element name="Description" type="dicom:Bd_64_Type" minOccurs="0"/>
        <element name="Item_General_Info" type="item:Item_General_Info_Type" minOccurs="0"/>
    </sequence>
    <attribute name="Prefix" type="units:Prefixes_Type" fixed="nano"/>
    <attribute name="Unit" type="units:Si_Unit_Name_Type" fixed="meter"/>
</complexType>
```


© cfdewey 2008

Ontologies in Biology (2 Weeks: SSB & CFD)

Definition and application of ontologies

- *Standards (OWL, RDF)*
- *Examples and usage*
- *Unique identifiers (LSID)*

Database approach to ontology storage

Querying ontologies with SPARQL

- *Integrating ontologies and XML query processing*
- *Role of ontologies in systems biology*

Creating relational databases from ontologies

- *OWLdb*
- *Ontology-based querying*

© cfdewey 2008

RDF – a step beyond XML

- *Lexical triples*
- *Can be used to describe relationships*
- *Used to express the data in an OWL repository*
- *XML used to transmit information stored in an OWL repository*

Images: W3C

© cfdewey 2008

Biological pathways (2 weeks: CFD & SSB)

Modeling and computing biological pathways

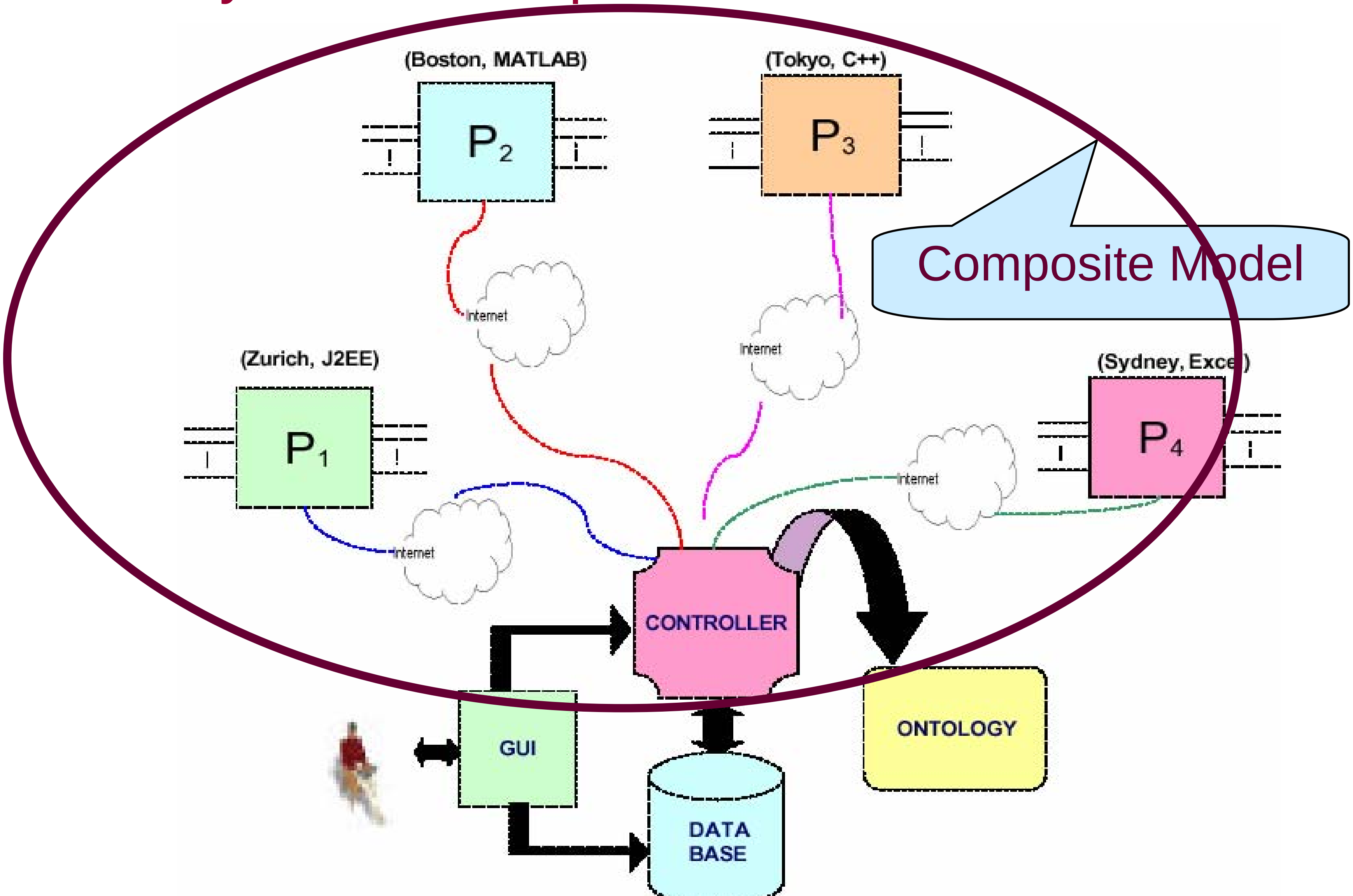
- *SBML, CellML, MML*
- *Cell Designer*
- *Cytosolve*

Biological pathway databases

Molecular network comparisons

© cfdewey 2008

The CytoSolve computational network



© cfdewey 2008

Biological and medical data integration (1 week: TC & CFD)

SWAN

- *An advanced architecture for sharing data*
- *Application to Alzheimer disease*
- *Generalization*
- *Workflow and useability*

Building a distributed biological pathway system

© cfdewey 2008

Grand Challenges (1 week: CFD)

Predicting drug efficacy by modeling

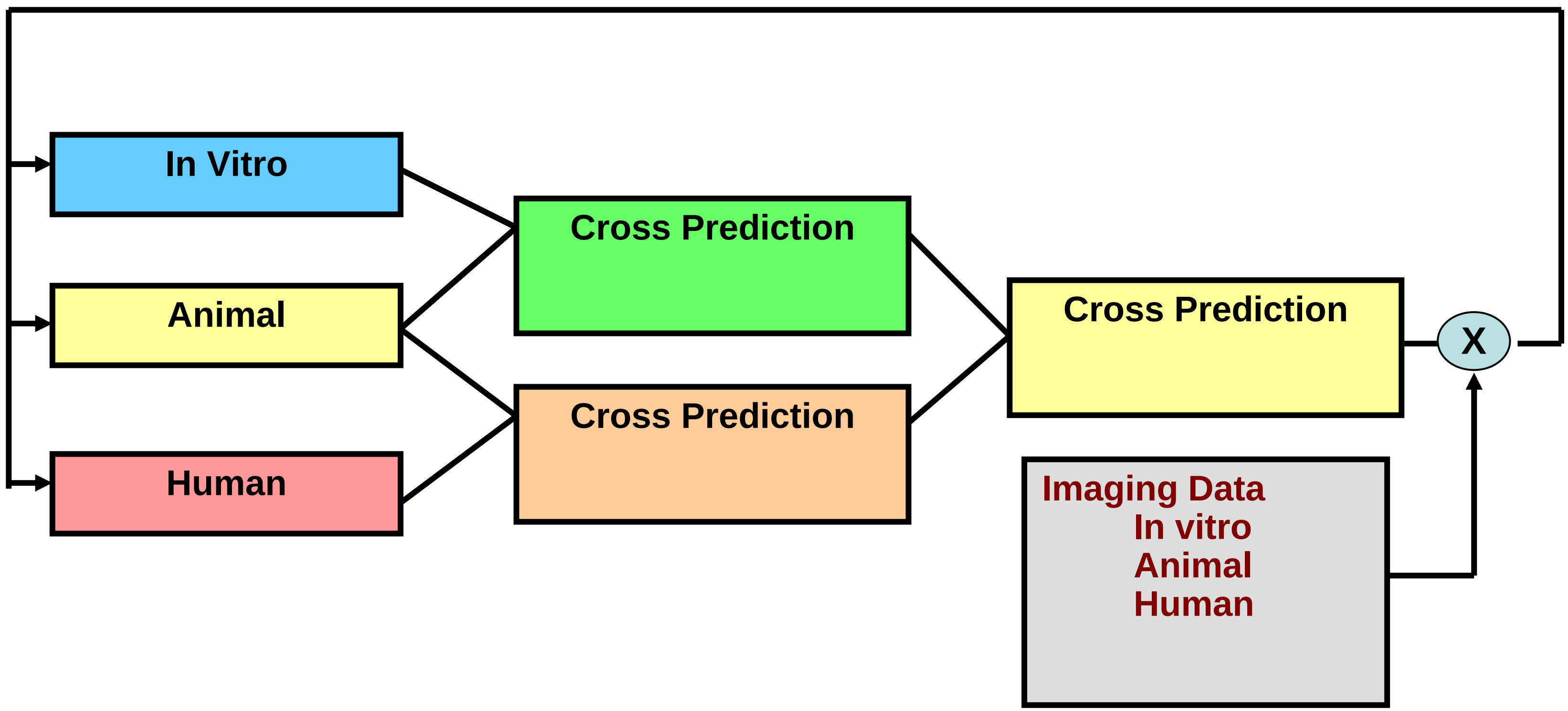
- *Current technology*
- *Future technology*
- *Ecamples*

Revolutionizing the drug discovery pipeline

- *New paradigms*
- *New challenges with multiple drugs*
- *Integration issues and opportunities*

© cfdewey 2008

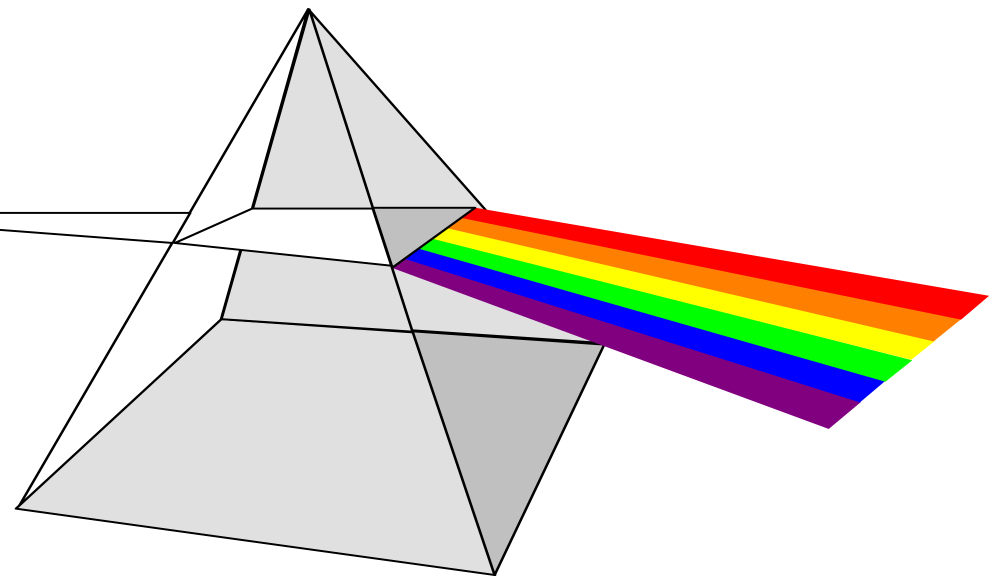
Modeling as the preclinical accelerator



A new paradigm in drug toxicity and efficacy

© cfdewey 2008

Computational and Systems Biology



© cfdewey 2008