Conquering Diseases

The *Conquering Diseases Biorepository, led by Dr. Craig Lilly,* seeks to

- To provide a pathway that lets our patients support translational research in their diseases
- Welcome all of our patients to participate in clinical trials at UMass sites

We will do this by

- Creating a data repository of de-identified clinical information and linked patient derived samples that is available to qualified UMass researchers
- Creating a separate database of individuals who desire contact about participation in clinical research into their disease
Approach

- School IS retained the services of Recombinant Data Corporation to assist with strategy and planning.
- Project team includes UMassMed, UMMHC and RDC staff.
- The project team was tasked to identify and analyze existing data sources and data types from UMMHC and UMASSMED and map the information to the i2b2 schema.
- The deliverable for this phase of the project is a detailed project plan and budget for a clinical research data repository to support the BioRepository and Clinical Research Data Repository.
- Strategy is a phased approach with well defined, on-time deliverables to ensure a culture of success.
Recombinant Data Corporation

- Data Warehousing and Reporting for Healthcare
- Quality & Pay-for-performance Reporting
- Translational Research
- Major Projects at Partners HealthCare Systems
  - 1999 - The Research Patient Data Registry
  - 2005 - Clinical Quality Reporting
- Offers expertise in Data Warehouse Strategy and
- Planning, Organizational and Technical Infrastructure,
- Data Modeling, System Design, Implementation
- Planning and Management
“i2b2 (Informatics for Integrating Biology and the Bedside) is an NIH-funded National Center for Biomedical Computing based at Partners HealthCare System. The i2b2 Center is developing a scalable informatics framework that will bridge clinical research data and the vast data banks arising from basic science research in order to better understand the genetic bases of complex diseases. This knowledge will facilitate the design of targeted therapies for individual patients with diseases having genetic origins. The i2b2 is funded as a cooperative agreement with the National Institutes of Health.”

http://www.i2b2.org
The i2b2 Hive and Workbench is written to accommodate three strata of users:

- Clinical investigators who want to use the software in as “shrink-wrapped” a way as possible,
- Bioinformatics scientists who want the ability to customize the flow of data and interactions, and
- Bio-computational software developers who want to develop new software capabilities that can be integrated easily into the computing environment.
i2b2
Informatics for Integrating Biology & the Bedside

- Project Management
- File Repository
- Ontology Management
- Data Repository (CRC)
- Annotating Genomic Data #1
- Natural Language Processing
- Identity Management
- Workflow Framework
- Correlation Analysis
- De-Identification of Data
- PFT Processing
- Annotating Genomic Data #2
- Annotating Imaging Data

UMASS Medical School
MICARD

• Servers have been acquired and installed in the data center.
• We have successfully developed a prototype application using a small sample of UMass data.
• We have developed a detailed project plan for a Phase 2 implementation that includes Patient Demographics, Diagnosis, Lab and Pharmacy data.
• Integration of Bio-Sample data pending selection of vendor
• Projected during of implementation is three months.
Implementation

There are 5 primary functions that the implementation will address. These are:

1. Security and Auditing
2. Data Preprocessor
3. Metadata Files
4. Report Generators
5. Data Delivery
Security and Auditing

Security and Auditing (SA) functions require the development of software, policies, and administrative procedures that will ensure that the system complies with HIPAA and other regulations. The goal is to create a set of barriers that together provide the highest level of security possible. The SA will include the following elements:

• **A process for encrypting (and decrypting) personal health information.**
  – This process will include safeguards for encryption keys, auditing to ensure that encryption algorithms are effective, and testing to ensure that the process does not excessively degrade system performance.

• **Processes and procedures to control access to the database**
  – These processes are for authorizing new users and authenticating existing users. It includes login screens, the ability to manage accounts (authorize/suspend/renew) and for ensuring that accounts are not shared or otherwise compromised.

• **Software and Processes to provide auditing of system use**
  – The audit system will track system usage and include reporting to track system usage and detect usage that may indicate a breach of security. Processes and procedures are an important aspect of this task.
Data Preprocessor

The Data Preprocessors take feeds from data sources and prepare them for import to the i2b2 data repository. These processors are responsible for encryption, inter-source mapping, normalization, and data cleaning. RDC will work closely with UMass IT personnel to develop systems that allow reliable, timely and verifiable updates.
The Metadata files contain the ontologies for navigating concepts such as Medications, Laboratory tests, Diagnoses, Procedures, and Patient Demographics. Multiple ontologies may be supported for each concept, and a key task is to develop data mappings when more than one ontology exists for a single type of data. For example, the ICD9 ontology is used for codifying diagnoses in the medical claims domain, but it will not necessarily map to a Problem list that is maintained in a clinical system where pre-diagnostic assessment of a patient may be recorded.
Report Generators

A minimal set of Reports for analyzing the data from an overall perspective will support several key functions, including quality control and security auditing.
Data Delivery

The Data Delivery mechanism is both a set of processes and procedures for obtaining approval for use of data as well as a mechanism for securely executing the delivery of data within those processes and procedures. The specific technical infrastructure will depend on researchers’ preferences and security considerations.
System Implementation

The System Implementation task is an overarching management task that includes all tasks related to the actual deployment of the system. This includes training, documentation, and testing.
Projected Timelines

- Contract (1 m)
- Phase 2 (3 m)
- Stabilization (3m)
- Phase 3
  - MIAME
  - HTS
  - ????