A Portable Neuroinformatic System in a Neurological Research Environment

Yilong Ma PhD

Center for Neurosciences

North Shore LIJ Health System
New York University School of Medicine
Manhasset, New York, USA

The Feinstein Institute for Medical Research
North Shore-Long Island Jewish Health System

Biorep.org

Introduction

- Brain imaging is a revolutionary technology to study central nerve system in humans and animals. Changes in brain structure and function can been measured in exquisite detail and under living physiological conditions.
- This innovation has advanced our understanding of normal brain functions and shed important lights into molecular bases and viable experimental therapeutics in a wide variety of neuropsychiatric disorders.
- Data management and sharing are paramount due to complexity and high cost in brain imaging research. It is necessary to construct neuroinformatic tools to effectively store and manage vast amounts of data in a typical neurological research environment.
- This poses a big challenge in both database design and data sharing mechanisms. International efforts are gaining momentum to develop comprehensive information systems covering all aspects of brain imaging for basic and clinical applications.

Background

- We have been developing a Laboratory Informatics Management System (LIMS) database and software system. It resulted from a biorepository created in 2000 to support the New York Cancer Project. This included sample collection of DNA and Plasma of patients or their relatives as well as a repository for other specimen types: Serum, Tissue, Urine, Cells, and RNA (30000+ samples)
- This framework has been expanded to include a Clinical Informatics Management System (CIMS) with brain image data. We have built a system that links Clinical, Sample and Analysis data. Our goal is to have an Enterprise System that is very secure and easy to use.

Objective

Modern technology has helped us create a new database model that is flexible for both Clinical and Sample Data (Dr. Robert Lundsten).

Enterprise System

- Custom Web Based Application
- Barcode Printing
- Clinical, Sample, or Analytical Data Collection
- Security
- Fail-over if Hardware goes down

Technology

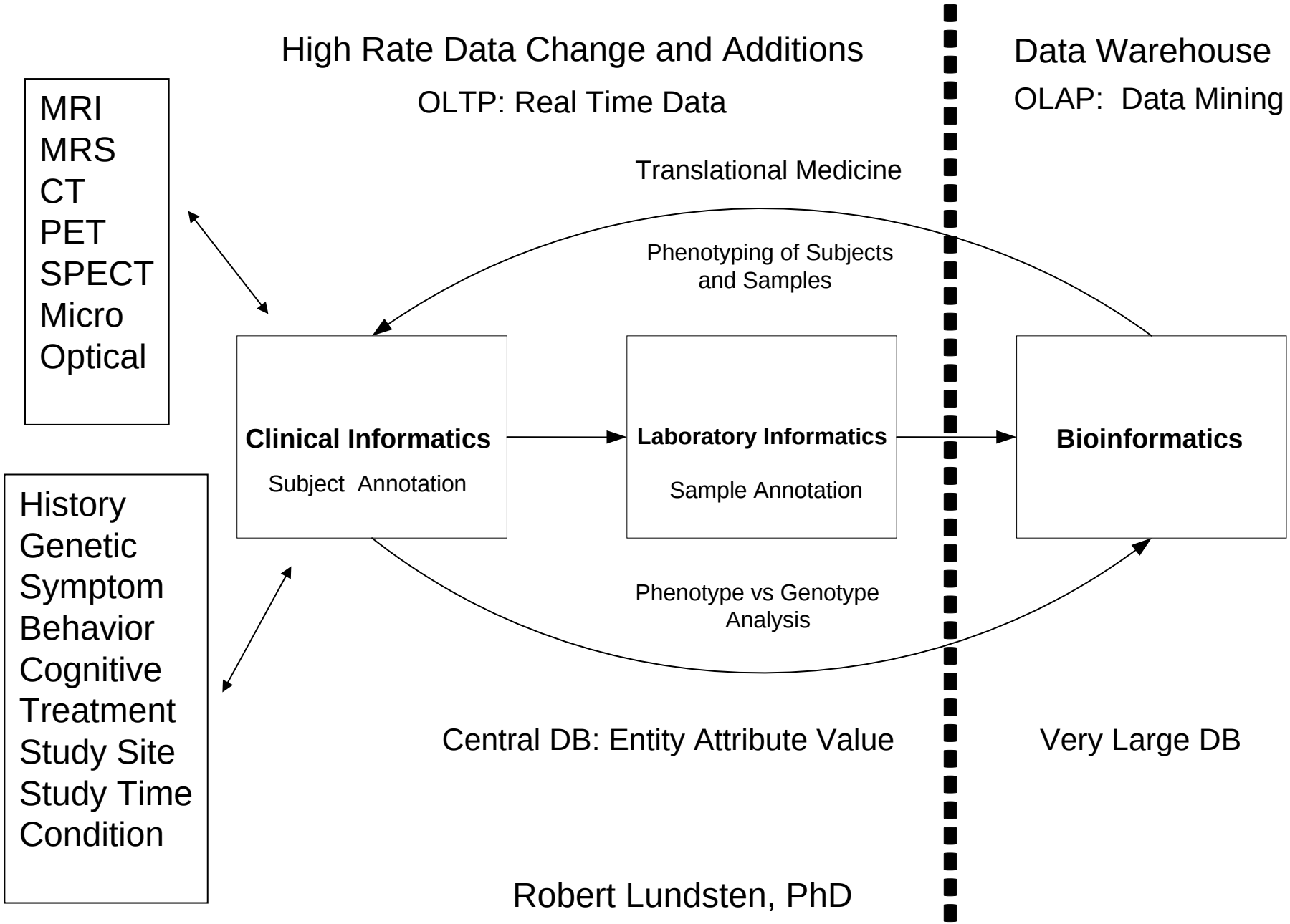
- Past Timeline
  - Feb 2000 — Digitrax Custom Label Printing
  - July 2000 — Microsoft Access 97 LIMS (coding using VBA)
  - Nov 2000 — Microsoft Access 2000
  - Dec 2000 — Barcode Printing in Access using ZPL
  - May 2001 — SQL Server 7.0
  - June 2002 — SQL Server 2000 Enterprise System and Visual Studio 6.0
  - Jan 2003 — Microsoft Visual Studio.NET 2003 Development Environment
  - Jan 2004 — Microsoft ASP.NET Web Based Development
  - Feb 2006 — SQL Server 2005 and Visual Studio .NET 2005 on Dell Server Rack

Hardware Framework

- A large symmetrical multi-processor (SMP) Unisys ES 7000 computer expandable to 256 GB of RAM and 16 Itanium processors with a 64 bit Microsoft Windows operating system (Microsoft Windows Server 2003 Enterprise Edition 64) with Microsoft SQL Server 2005 as the RDBMS.
- Data is stored directly through 4 host bus adapters to a Clariion CX300 RAID disk array from EMC. There is also an assortment of 32 bit applications created by our software development group running on eight Dell Power Edge Servers and two Dell Power Vault disk arrays.

Database Design

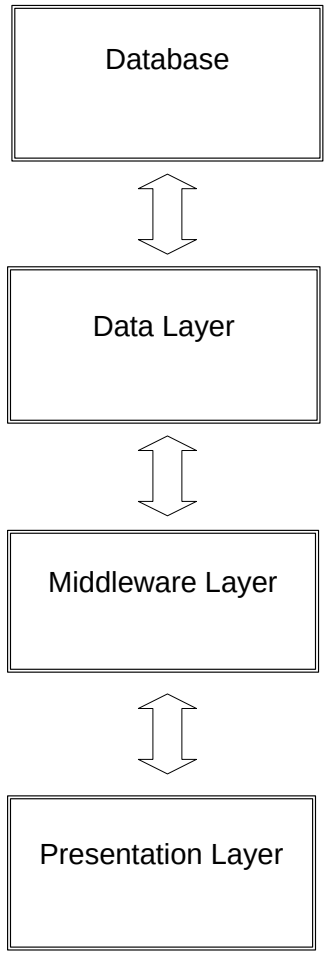
- There are 2 types of data that are stored on our server. Online Transaction Processing (OLTP) is a type of data that handles real time transactions that allows for editing. Online Analytical Processing (OLAP) offers a Data Mining environment with data that is static.
- The OLTP environment records inserts, updates, and deletes of data. It is a single database - Central Data Base) built with the idea about semantic data capture in an EAV (Entity Attribute Value) model. EAV allows for minimal or no database change when a new study enters into the database. This model is more efficient than the standard relational model because it allows for different criteria to be stored without changing the database structure or tables. The Clinical and Sample database is OLTP. User transactions are recorded as part of our Audit system.
- OLAP is the data transferred from OLTP that is cleaned and not changing. This is the data used for data production and reporting in a strictly data mining environment.


High Rate Data Change and Additions
OLTP: Real Time Data
Data Warehouse
OLAP: Data Mining
MRI
MRS
CT
PET
SPECT
Micro
Optical
Translational Medicine
Phenotyping of Subjects and Samples
Clinical Informatics
Subject Annotation
Laboratory Informatics
Sample Annotation
Bioinformatics
History
Genetic
Symptom
Behavior
Cognitive
Treatment
Study Site
Study Time
Condition
Phenotype vs Genotype Analysis
Central DB: Entity Attribute Value
Very Large DB
Robert Lundsten, PhD

Software Design

- Object Oriented Programming

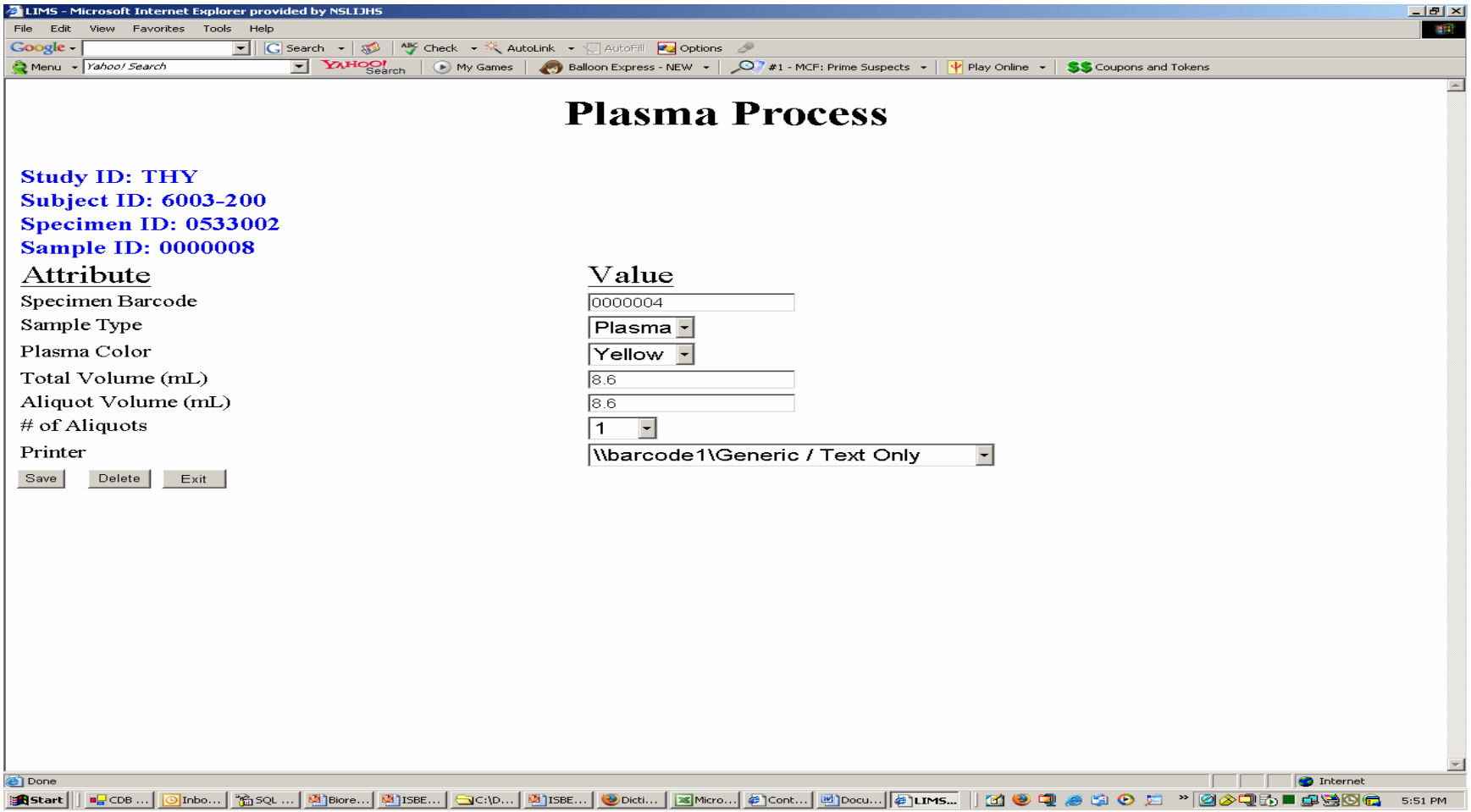
The separation of each object allows for simple programming at a quicker pace. Each layer handles its own inputs and outputs. The Middleware Layer is central to this model. It is like a policeman regulating direction of traffic. Adjacent to this layer is the Data Layer and Presentation Layer. The Presentation Layer builds a control to be displayed to the user. The Data Layer talks to the database via Stored Procedures and Views to insert, update, delete, or retrieve data.



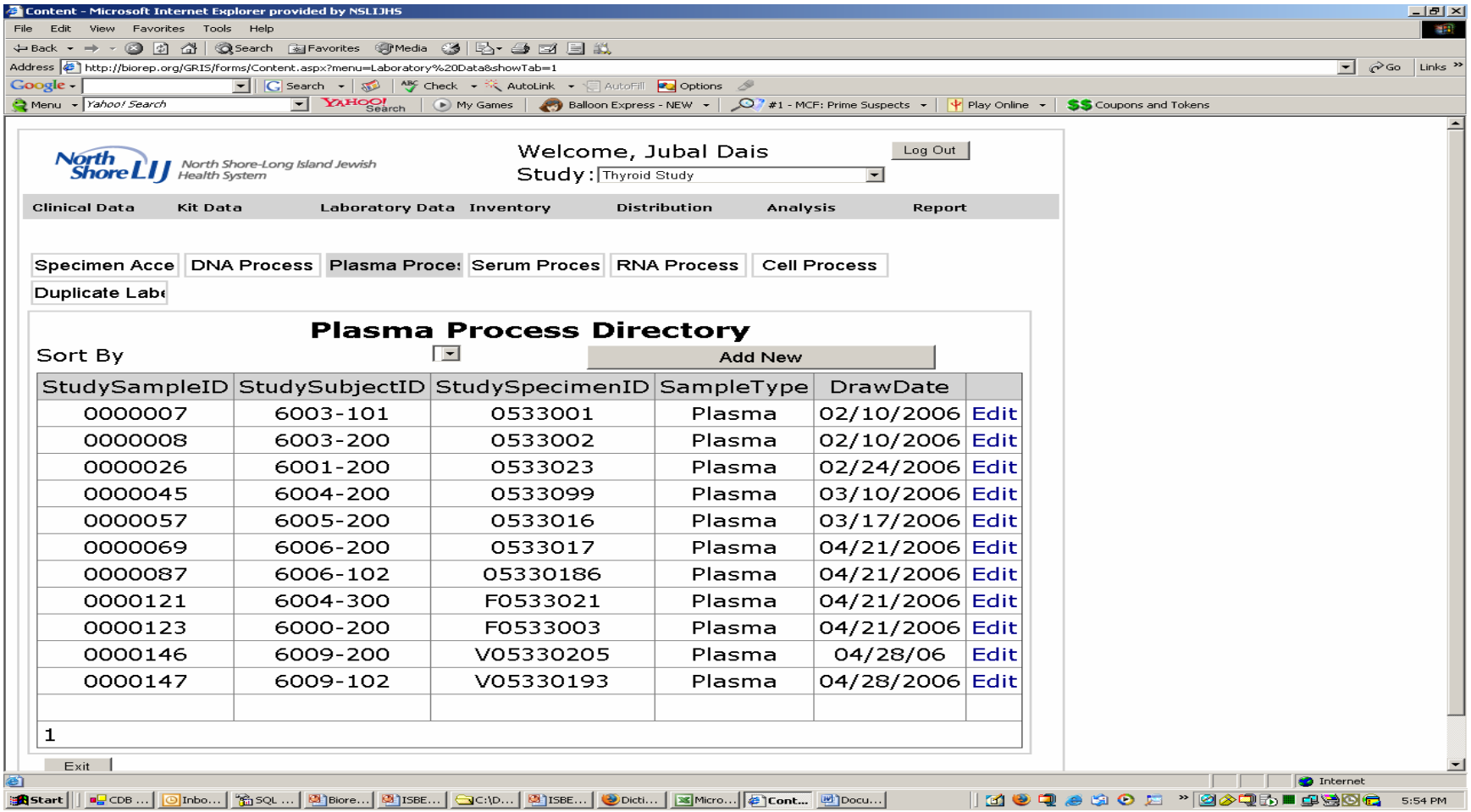
Sample EAV Table

SampleAttribute ID	SampleID	AttributeID	SampleAttribute Value
1	0000001	A000001	10 ml
2	0000001	A000002	4/20/2006
3	0000001	A000003	Red
4	0000001	A000004	2 Tubes received
5	0000002	A000001	4 ml
6	0000002	A000002	4/25/2006
7	0000002	A000003	Orange
8	0000002	A000004	1 Tube received

Dynamic Control



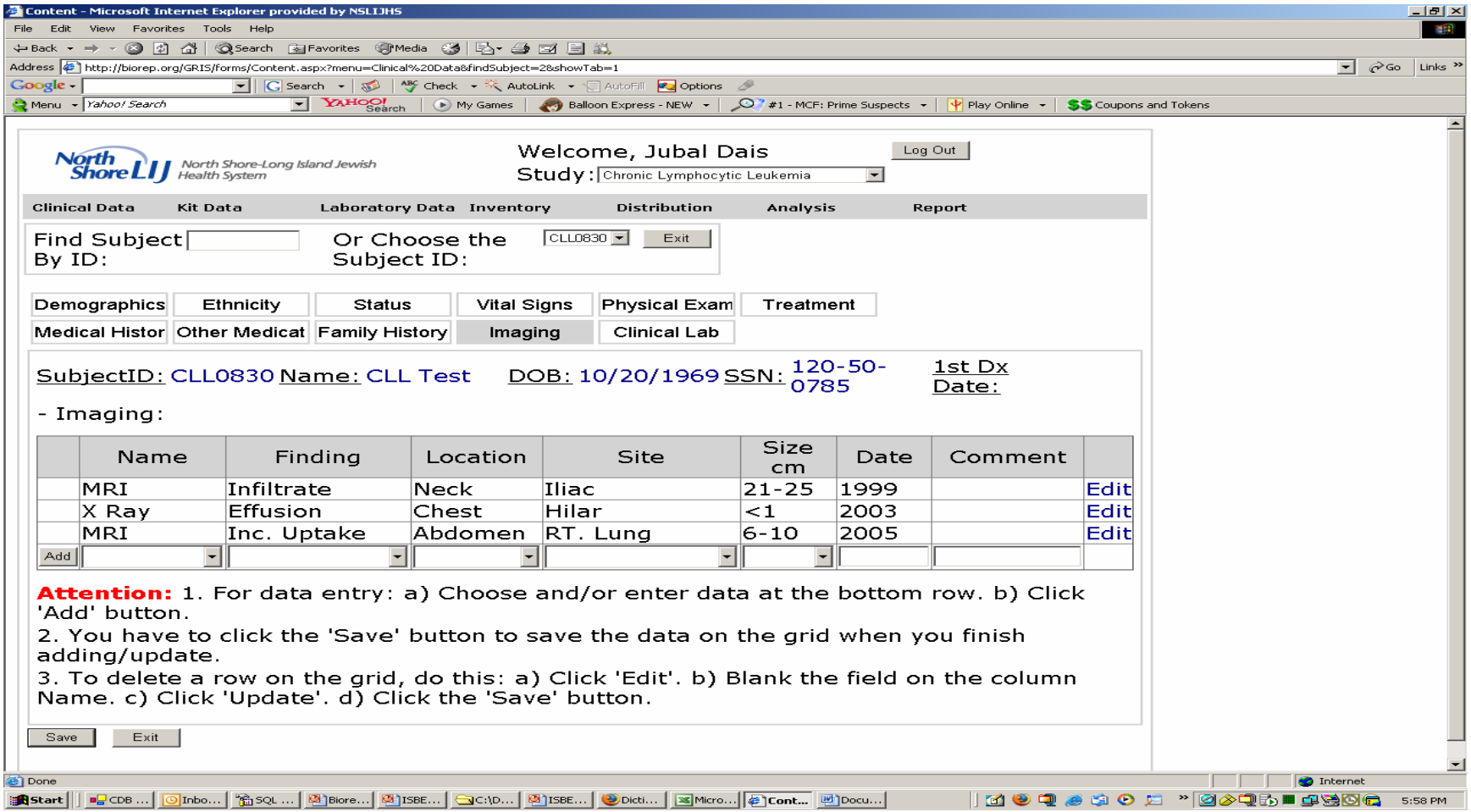
Results: LIMS



StudySampleID	StudySubjectID	StudySpecimenID	SampleType	DrawDate	
0000007	6003-101	0533001	Plasma	02/10/2006	Edit
0000008	6003-200	0533002	Plasma	02/10/2006	Edit
0000026	6001-200	0533023	Plasma	02/24/2006	Edit
0000045	6004-200	0533099	Plasma	03/10/2006	Edit
0000057	6005-200	0533016	Plasma	03/17/2006	Edit
0000069	6006-200	0533017	Plasma	04/21/2006	Edit
0000087	6006-102	05330186	Plasma	04/21/2006	Edit
0000121	6004-300	F0533021	Plasma	04/21/2006	Edit
0000123	6000-200	F0533003	Plasma	04/21/2006	Edit
0000146	6009-200	V05330205	Plasma	04/28/06	Edit
0000147	6009-102	V05330193	Plasma	04/28/2006	Edit

1

Results: CLIMS



Database Application

- This system has evolved into a significant neuroscience resource comprising thousands of brain images with various neuropsychiatric disorders.
- We have used this database to map local or system abnormality in anatomy, hemodynamics, metabolism and biochemistry at resting and activated conditions.
- We have established a set of signature makers that describe the neuropathology of each disorder, natural courses of disease progression as well as treatment response of promising medications and neurosurgical interventions such as deep brain stimulation and cellular-based novel therapies.

Conclusion

- We have developed a web-based and flexible neuroinformatic platform that can greatly increase the productivity of translational medical research in the context of multi-center cooperations.
- This tool is being designed with easy link to other national or global comprehensive neuroinformatic systems.
- It can serve as a prototype of neuroinformatic tools for patient-oriented brain imaging research.

Challenges

- Large scale SNP genotyping is carried out on several genotyping platforms, including the Illumina Sentrix Bead array and Illumina HapMap300 Infinium 2 array.
- SNP genotype production has grown dramatically with 8-10 million SNP genotypes being generated per day; the accumulation of 3 billion or more SNP genotypes over the next year.
- The difficulties in managing and manipulating these very large datasets have forced the creation of a data center capable of high performance data management.
- Management of research subject annotation is also quickly becoming a high performance computing issue.