Integrative Informatics

Isaac S. Kohane
4.27.04

Early Vision

Integrating Genomics and Pharmacology

- RNA expression in NCI 60 cell lines was determined using Affymetrix HU6000 arrays
  - 5,223 known genes
  - 1,193 expressed sequence tags

- The RNA expression data set and Anti-cancer susceptibility data set were merged, using the 60 cell lines the two tables had in common

<table>
<thead>
<tr>
<th>RNA Expression</th>
<th>Common 60 Cell Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>6,000 genes</td>
<td>5,000 anti-cancer agents</td>
</tr>
</tbody>
</table>

Genes and Anti-Cancer Agents

- Threshold $r^2$ was 0.8
- 202 networks
- 834 features out of 11,692 (7.1%)
- 1,222 links out of 68,345,586 (.0018%)
- Only one link between a gene and anti-cancer agent

Population Studies Integrates with Genomics

- ApoE4

  Normal individuals
  Alzheimer's patients
How is digital biology helping generate quantitative physician scientists?

So...

Show Me The Data!

Recurring Themes

- Minimal data model
- Abstraction layers
- Leverage existing data sources

An Institutional View
W3-EMRS: The Problem

- 1993: Sad state-of-the-art
- Within institutions, multiple heterogeneous applications with disparate databases.
- Across merged, collaborating institutions, little commonality in function, data model, or user interface.
- Economic realities: increase in large multi-institutional healthcare delivery systems.

Repository/Warehouse

- Building common repositories may not always be the optimal strategy
  - Time of modelling
  - Time to implement
  - Keeping real-time consistency with legacy apps.
  - Intellectual property
  - Privacy
  - Performance (e.g. in ICU)

The NLM-Funded W3-EMRS Project

- 1994 NLM RFA
  - Challenge to design national EMR
- 1994 Implementation
- 1995 Boston EMR Collaborative
  - Children's, BIH, MGH, MIT

W3-EMRS Solution

- Understanding that "legacy" systems have long lives
- Understanding of importance of a unifying architecture
- Use of Internet technologies for "lightweight" integration

Standard First-Generation W3-based EMR Design

- Single data-model
  - Often implicit
  - Difficult to maintain
  - Unsuitable for unifying multiple heterogeneous systems
- Remote, multi-platform viewing
- Rapid development of view-only function

Second Generation Web-based EMR: W3-EMRS Architecture
Early Success with W3-EMRS

- Hyperbilirubinemia project with Children’s, BWH, BHH.
- Single institution implementation at Children’s.
- 6.872 Final Project by John Halamka --> CareWeb

The Patient View

- Multiple Causes of Inertia Beyond Technology
- Patients have difficulty accessing their medical information
• Highly mobile patients
• No formalized data exchange

Solution: PING

- Personal internetworked notary and guardian
- Personally controlled records
- Ubiquitous
- Open standards
- Open source
- Web interface
- Public key infrastructure
- XML-based data

“Data sponge”
- Independent

Actors use

Clients to communicate with

PING Servers which authenticate and authorize the access to
PING Storage which contain

PING Records which can also contain pointers to
Remote PING Storage which contain
Remote PING Records

- A patient accesses her medical record with any of a number of clients.
  - The most basic client is the generic PING record browser
  - Other types of clients may provide:
    - decision support
    - laboratory value interpretation
    - data acquisition functions

- A patient may choose to share the record with another actor by granting role-based access
PING Example Application

- Regional Immunization Registry
- Larger scale services
- Wide variety of data
  - Results
  - Intervention
- Self-reporting
- Graphical representation of data
- Ability to annotate a record
- New Brunswick and Canada
- PING Response

Consented High-Performance Indexing and Retrieval of Pathology Specimens

CHIRPS
one of 2 Shared Pathology Informatics Networks (SPiN, funded by NCI)

UCLA Medical Center
Harvard Medical Center

But Can We Still Better Leverage What is Already Electronic

But Can We Still Better Leverage What is Already Electronic

Goals

- Maximize access/finding of resources.
- Minimize disclosure (privacy and IP).
- Decentralized architecture.
- Maximize autonomy.
- Meticulous attention to the heterogeneity of consent.
Local view of institution: Loosely Coupled Federation

Network

SPIN Node

Node Tools

Pathology

Clinical

MPI

Institutional Systems

Inside a CHIRPS SPIN node

CHIRPS services (Java)

OS (Win32, Linux) / file services

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS services (Java)

Xerces XML parser

JXTA Core

JXTA Search

Apache/Tomcat/JBOSS

MySQL

CHIRPS Node Tools

Metadata tools

User Review and Control Tools

CSI NLP

CHIRPS NLP

PathReports

SSA XML

Peer discovery

“Rendezvous” Peer SPIN Node (NCI)

SuperPeers advertise at Rendezvous nodes

New Peer checks Rendezvous node for SuperPeers available, then registers with selected one. Or manually enter an address.

Boston-UCLA: Network view

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned

Running CHIRPS software

Running non-CHIRPS software but supporting SPIN protocols

Link established

Link planned
So, where do we go from here?
What could this mean for healthcare

PING Records
Proteomics
Sequencing
and genotyping
Transcriptional Profiling

REFSEQ
FASTA
MAGE-ML
Genbank