About Recombinant

Best of breed software/analytics company focused 100% on secondary uses of clinical and pre-clinical data

Core Competencies

• Healthcare data warehousing & integration services
• Translational research & quality reporting solutions
• Data strategy, governance & compliance consulting
• Open Source implementations & extensions

Core Values

• Pragmatism
• Effective Communication
• Trust
Translational Medicine Defined

Non-profit research
Pharma R&D

Clinical Trials

Adoption of
evidence based medicine

Basic Research
- Diseases
- In Vitro
- Animal Studies
- Genomics

Clinical Research
- Mechanisms
- Markers
- Drugs
- Interventions

- Human Research
- Participants

- Quality Improvement
- Feasibility
- Strategy
- Effectiveness
- Safety
- Barriers

Health
- Patient Populations
- Practice Workflows
- Policies

Bench to Bedside Translation
Laboratory to Human

Bench to Community Translation
Evidence to Practice
• Recombinant Customers
  • 11 – Year Ago
  • 27 – Today

• Recombinant Employees
  • 21 – Year Ago
  • 42 – Today
  • 50 projected by end of 2010

• Healthcare requires expertise
  • Matvey Palchuk, MD, MS
    Director of Medical Informatics
  • D. Doug Liu, PhD
    Chief Scientist
  • David Fram, PhD
    Research Director
2010 Business Drivers

• **Research/Quality convergence**
  - Quality activities require research approaches
  - Cross-over of tools (e.g. GPP, i2b2 for quality)
  - Registries

• **Bio-banking / genomics**
  - Lower costs for genetic research drive clinical data
  - Big studies/big data sets

• **ARRA driving new data demands**
  - HIE, Interoperability, accountable care, shared decision making, permissions

• **Public data use**
  - Profiles, pre-competitive sharing, benchmarking
Target vision - incremental approach
Research Study Process

1. Feasibility
2. Find Collaborators
3. Authorize and Fund
4. Launch Study
5. Analyze/Publish

Multi-site Cohort Discovery/Analysis
- i2b2 / SHRINE

Executive Research
- Approved data sets
- REDCap
- Biobanks

Researcher Collaboration
- Profiles Research Networking Software

Protocol Approval
- Funding requests
- Multi-site eIRB
What is the benefit to each side?

Benefits

Research To Quality
- Access rich research registry and study data
- Translate findings to improve care model
- Bridge genomics personalized medicine
- Leverage Informatics expertise and tools

Quality to Research
- Increased data availability
- Increased data quality/source knowledge from use in PTO
- Access to patient cohorts
- Quality study collaboration

Translational Approach
AMCs are Building Toward Personalized Medicine Vision
Trust?

“On the Internet, nobody knows you’re a dog.”
Data Trust Conceptual Model
Open source initiatives
Open source initiatives

• i2b2
• Pentaho ETL/BI
• Hudson
• Indivo PHR
• REDCap Electronic Data Capture
• Gene Pattern
• caTissue
• Profiles Research Networking Software
• Lucene search
### Open source data management

#### Benefits of Open Source:
- Sometimes free
- Shared ideas
- Interoperability
- Science friendly (share)
- Evidence based directive
- Metrics/Benchmarks
- Anyone can contribute
- Grants
- Innovation
- Dissemination

#### Challenges of Open Source:
- Distrust by IT/CIOs
- Support model
- Marketing/sales
- Solution providers
- Commercial FUD
- Completeness
- Predictability
- Financial limits

---

**What is in the best interest of patients?**
Data Access Problem

Unapproved Research System Load

Unmanaged Data

Researchers

Repeat Queries

Bottleneck

Data Analyst

Approval

IT Project Mgmt.

Hospital IT

IRB

PHI Controls

System Load

Unapproved Research

Clinicians

Bottleneck

Data Access Problem

Repea

Requests

Custom

Data

Approval

Off-peak batch queries in live systems

Core Applications

Priority

Claims

Labs

Meditech
The Solution: Managed self-service
Step 1: Data Collection

Institutional Data Sources
- EMR
- Claims
- EMPI
- Schedule
- Etc.

Step 2: Clinical Data Warehouse (PHI)
Integration ETL (data extracts and HL7 feeds)

Step 3: Create "Limited Data Set" (destroy identifiers)

Step 4: i2b2 research mart (Limited Data Set)

Step 5: i2b2 queries

Step 6: eIRB protocol request

Step 7: IRB approval

Step 8: Re-issue query to data mart with PHI

Step 9: i2b2 research mart with patient identifiers

Step 10: Research data (PHI) to researcher

Data Center (Honest Broker and server support)
i2b2 History

1999 RPDR initially established at Partners

2000-2004 RPDR and tool evolution at Partners

2004 i2b2 grant to “port RPDR” as an open source framework

2006 First beta release of i2b2

2008 Strong adoption

2009 De-facto CTSA standard CRDW

2010 i2b2 renewal

* RPDR (Research Patient Data Registry) was the precursor to i2b2
United States
- Beth Israel Deaconess Hospital, Boston, MA
- Boston University School of Medicine, Boston, MA
- Brigham and Women's Hospital, Boston, MA
- Case Western Reserve Hospital
- Children's Hospital of Boston, MA
- Children's Hospital of Cincinnati, OH
- Children's Hospital of Denver, CO
- Children's Hospital of Philadelphia, PA
- Cleveland Clinic, Cleveland, OH
- Duke Medical College
- GE Research, Niskayuna, NY
- Group Health Cooperative
- Harvard Medical School, Boston, MA
- Kaiser Permanente, Oakland, CA
- Kimmel Cancer Center, Philadelphia, PA
- Maine Medical Center, Portland, ME
- Marshfield Clinic, Wisconsin
- Massachusetts General Hospital, Boston, MA
- Morehouse School of Medicine, Atlanta, GA
- Ohio State University Medical Center, Columbus, OH
- Oregon Health & Science University, Portland, OR
- Renaissance Computing Institute, Chapel Hill, NC
- Tufts New England Medical Center, Boston, MA
- University of California, Davis, CA
- University of California, Irvine, CA

United States (continued)
- University of California, San Francisco, CA
- University of Cincinnati, Cincinnati, OH
- University of Massachusetts Medical School, Worcester, MA
- University of Michigan Medical Center, Ann Arbor, MI
- University of Pennsylvania School of Medicine, Phila, PA
- University of Rochester Medical Center, Rochester, NY
- University of Texas Health Sciences Center SW, Dallas, TX
- University of Texas Health Sciences Center, Houston, TX
- University of Texas Health Sciences Center, San Antonio, TX
- University of Washington, Seattle, WA
- Utah Health Science Center, Salt Lake City, UT
- Veterans Administration Boston and Utah
- Weil Medical College of Cornell, NYC, NY

International
- Georges Pompidous Hospital, Paris, France
- University of Goettingen, Goettingen, Germany
- University of Pavia, Pavia, Italy
- University of Seoul, Seoul, Korea

Pharmaceutical
- Johnson and Johnson
Star Schema: i2b2

- Entity Attribute Value (EAV) Data Model
Grid projects: SHRINE and caBIG
Research systems
Research Focused Interfaces

Central Data
- Meta Data Repository
- Reference Repository
- Benchmark Repository
- Protocol Viewer
- Data Dictionary Viewers
- File repository

Data Spokes
- Search and Biomarker
- i2b2 De-identified
- Phenotype Genotype Database (PGD)
- Study Marts
- Recruitment Mart
- Data Capture Marts

Web Toolbox
- Hypothesis pathway
  - i2b2 cohort explorer
  - GenePattern
  - tranSMART
  - Profiles (PRNS)
- Study pathway
  - CER Tools
  - NLP Pipeline Manager
  - Consent/Sample/Experiment Accrual
- Recruitment and Enrollment
- Data Capture
  - REDCap/EDC
  - Encore
How Informatics Can Potentiate Precompetitive Open-Source Collaboration to Jump-Start Drug Discovery and Development

ED Perakslis¹, J Van Dam¹ and S Szalma¹
Best Practices Award

• Recombinant Solution Wins 2010 Bio-IT World Best Practices Award in Knowledge Management

• Implementation at Centocor R&D, Inc., a subsidiary of Johnson & Johnson

• tranSMART designed to share drug research and development data across Johnson & Johnson
tranSMART
tranSMART primary components

Search

Data Analysis
Subset Definitions
Set Value for Query

Set Value

Searches by value can be constrained by the high/low flag set by the performing laboratory or by the values themselves.

- No Value
- By high/low flag
- By numeric value

Please select operator:
- LESS THAN (<)

Please enter value:
- 1.5

Histogram of ... \Medical History \Pulmonary Function Test \FEF in L/Second (PFFE F) for all

H1: (1.574, 23)

Histogram of ... \Medical History \Pulmonary Function Test \FEF in L/Second (PFFE F) for subset
Analysis of concepts found in Subsets

Analysis of Medical History\Pulmonary Function Test\FEF in L/Second (PFFE) for subsets:

Comparison of Medical History\Pulmonary Function Test\FEF in L/Second (PFFE)

Histogram of Medical History\Pulmonary Function Test\FEF in L/Second (PFFE)

- Subset 1
  - Trial55
  - Mean: 0.93
  - Median: 0.96
  - IQR: 0.65
  - SD: 0.35
  - Data Points: 32

- Subset 2
  - Trial55
  - Mean: 2.34
  - Median: 2.05
  - IQR: 1.04
  - SD: 0.7
  - Data Points: 37

The results are significant at a 95% confidence level.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Subset</th>
<th>Trial</th>
<th>Sex</th>
<th>Age</th>
<th>Race</th>
<th>FEF_in_L/Second_(PFEF)</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>65476</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>46</td>
<td>Caucasian</td>
<td>1.21</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65478</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>57</td>
<td>Black or Afr...</td>
<td>1.4</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65479</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>47</td>
<td>Other</td>
<td>1.43</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65482</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>38</td>
<td>Black or Afr...</td>
<td>1.15</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65483</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>39</td>
<td>Caucasian</td>
<td>0.93</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65492</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>39</td>
<td>Caucasian</td>
<td>0.51</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65500</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>40</td>
<td>Black or Afr...</td>
<td>0.77</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65506</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>48</td>
<td>Caucasian</td>
<td>1.27</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65511</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>34</td>
<td>Caucasian</td>
<td>0.62</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65514</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>66</td>
<td>Caucasian</td>
<td>1.06</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65525</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>58</td>
<td>Caucasian</td>
<td>0.82</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65533</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>58</td>
<td>Caucasian</td>
<td>0.29</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65534</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>47</td>
<td>Black or Afr...</td>
<td>1.24</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65536</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>48</td>
<td>Black or Afr...</td>
<td>1.3</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65538</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>66</td>
<td>Caucasian</td>
<td>0.72</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65541</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>56</td>
<td>Black or Afr...</td>
<td>1.18</td>
<td>Female</td>
<td>N</td>
</tr>
<tr>
<td>65542</td>
<td>subset1</td>
<td>Trial55</td>
<td>F</td>
<td>47</td>
<td>Black or Afr...</td>
<td>0.95</td>
<td>Female</td>
<td>N</td>
</tr>
</tbody>
</table>
Integration of Omics and i2b2

1. Select Cohort(s)

2. Initialize Analytical Pipeline

3. Interactive Analysis Wizard

4. Compile Results Across datasets

5. Present results to scientist

Workflow Engine

i2b2 Clinical Data set

Reference Data (Gene Annotation)

Gene Expression

SNP Chip
Heat Map

Adenomatous ADP-AGP
Aldo-1 Antigen_AEP
Alpha-2 Macroglobulin_AGM
Alpha-2 Macroglobulin_AFP
Adipoprotein A1_AAG
Adipoprotein A2_AAP
Adipoprotein A3_AAP
Adipoprotein A4_AAP
Beta-2 Microglobulin_B2M
Brain-Derived Neurotrophic Factor_BDNF
C Receptor_Protein_CRP
CD40L, CD40L
Calcitonin_CALC
Cancer Antigen 125_C125
Cancer Antigen 15-3_C153
Cancer Antigen 19-9_C199
Carcinoembryonic Antigen_CEA
Complement C3_C3
Creatine Kinase-MB_CK-MB
Erythropoietin_EPO
Endothelin-1_ET1
Erythropoietin_EPO
FGF Basic_FGF2
Fatty Acid Binding Protein_FABP
Ferritin_FTH1
Fibrinogen_GAP
CPSF_CSF3
CPSF_CSF3
Cystathionine S-Transferase_CST1
Growth Hormone_GH
Haptoglobin_HP
ICAM-1_ICAM1
Interferon IFI
IFN-antiviral_JIFNG
IFN-antiviral_JIFNG
IL-10_IL-10
IL-12p40_IL-12
IL-12p70_IL-12
IL-12p35_IL-12
IL-15_IL-15
IL-16_IL-16
IL-17A_IL-17A
IL-18_IL-18
IL-19_IL-19
IL-20_IL-20
IL-21_IL-21
Profiles Research Networking Software
Open Source Initiative
• Research networking & expertise mining tool
• Developed as part of Harvard Catalyst, The Harvard Clinical and Translational Science Center
• Deployed at UCSF as an alpha site
• Now being released as open source software
• Recombinant is an authorized support provider for Profiles Research Networking Software
• Working with Griffin Weber to implement two beta sites

Profiles Research Networking Software was developed under the supervision of Griffin M Weber, MD, PhD, with support from Grant Number 1 UL1 RR025758-01 to Harvard Catalyst, The Harvard Clinical and Translational Science Center from the National Center for Research Resources and support from Harvard University and its affiliated academic healthcare centers
Professional Open Source Model

- End Users
- Community
- Licensing
- Innovation
- Awareness
- Implementation
- Support
John David Halamka, M.D.

Title: Associate Professor of Medicine
Institution: Beth Israel Deaconess Medical Center
Department: Medicine
Division: Emergency Medicine
Address: Beth Israel Deaconess Med Ctr Information Systems, 6th Fl 1135 Tremont St Roxbury Crossing, MA 02120
Telephone: 617735-3012
Fax: 617735-8015
Email: jhalamka@caregroup.harvard.edu

Other Positions:
Title: Chief Information Officer
Institution: Harvard Medical School
Department: IT

Narrative
John D. Halamka, MD, MS, is Chief Information Officer of the CareGroup Health System, Chief Information Officer and Associate Dean for Educational Technology at Harvard Medical School, Chairman of the New England Health Electronic Data Interchange Network (NEHEN), CEO of MA-Share, Chair of the Healthcare Information Technology Standards Panel (HITSP), Chief Information Officer of the Harvard Clinical Research Institute and a practicing Emergency Physician.

Publications
Co-Author Networks: Map View

John Halamka, M.D.

**Co-Author Networks**

*Red markers* indicate the co-authors of

*Blue lines* connect people who have published papers together.

**Halama's...**

- Keywords (MeSH):
  - Patient Identification Systems
  - Radiology
  - Medical Record Linkage
  - Regional Medical Programs
  - Computer Security
  - See all (15) keywords

**Co-Authors**

- Adige, Benjamin
- Kohane, Isaac
- Mandl, Kenneth
- Nigam, Daniel

See all (15) people

**Similar People**

- Safran, Charles
- Kohane, Isaac
- Donnell, C
- Middleton, Blakeford

See all (20) people

**Department**

- Arbit, Robert
- Buehlin, Jonathan
- Epstein, Stephen

Search for (1336) people

**Neighbors**

- Armstrong, Elizabeth
- de Silva, Regis
- Sharpe, Miles

**My Network**

- Halamka, John
- Kohane, Isaac
- Mandl, Kenneth

**Search History**

- Halamka, John
- Lander, Eric
- Cepko, Constance

**Institution**

- Select

- Search
- Clear

- More Search Options
Co-Author Networks: Network View

John David Halamka, M.D.

Co-Author Networks: Network View

Halamka's...

Keywords(MeSH)
- Patient Identification Systems
- Radio
- Medical Record Linkage
- Regional Medical Programs
- Computer Security

See all (154) keywords

Co-Authors
- Adica, Benjamin
- Kohane, Isaac
- Mandi, Kenneth
- Nigrin, Daniel

See all (15) people

Similar People
- Satran, Charles
- Kohane, Isaac
- Bennett, G
- Middotin, Blackford

See all (20) people

Department
- Antesia, Robert
- Burofin, Jonathan
- Epstein, Stephen

Search for(13860) people

Neighbors
- Armstrong, Elizabeth
- de Silva, Regis
- Shore, Miles
Quality Reporting Overview
Quality Reporting Architecture

Source Data
- Labs
- Meds
- Claims
- CPM
- Sched
- EMPI

Data Warehouse

Quality Mart

Report Server

Quality Reports

Quality Dashboard

Electronic Medical Records

Printed Reports

Feedback/QA

Outreach: Phone, mail, etc.
Avoid Reinventing the Wheel
Operations

- Denominators (panels, locations)
- Practice snapshots
- EHR adoption
- Visits/utilization
Chronic Disease Management

- Diabetes
- Asthma
- CAD
- CHF

<table>
<thead>
<tr>
<th>Measures</th>
<th>My Value (%)</th>
<th>Goal Average (%)</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>% A1C control</td>
<td>&lt;70% (7.2)</td>
<td>7% (7.2)</td>
<td>&gt;70%</td>
</tr>
<tr>
<td>% BP control</td>
<td>&lt;140/90 mmHg</td>
<td>90%</td>
<td>&gt;140/90</td>
</tr>
<tr>
<td>% LDL control</td>
<td>&lt;100 mg/dL</td>
<td>70%</td>
<td>&gt;100</td>
</tr>
<tr>
<td>% CHD patients on statin therapy</td>
<td>80%</td>
<td>80%</td>
<td>&gt;80%</td>
</tr>
<tr>
<td>% Stroke patients on aspirin therapy</td>
<td>70%</td>
<td>70%</td>
<td>&gt;70%</td>
</tr>
</tbody>
</table>

Total # of CAD Patients: 21

Operations

- Preventative Medicine
- In Patient/Hospital
- Ad-hoc queries
- Report Library

Chronic Disease Mgmt.
Data Needed for Typical Report

1. Thru date: Displaying the period allows the user to understand how up to date the data is and how the group is being measured.

2. Patient, MRN, Age: Data from EMPI or other central patient data sets.

3. Next Visit: Scheduling system (e.g. IDX).


5. Data collected from EHR or claims.

6. Definitions: Definitions provide details into what the data means and how numerators or denominators were calculated. For example – What are my patients? Who is considered to be diabetic? What is considered on time?

7. Graphs: Summary performance for the provider across all patients based on with color highlights against categories of patients based on current outcomes. Potentially includes comparisons with peers.

8. Diabetes patients identified by claims, problem lists, constrained by provider’s panel of patients based on enrollment (payors) and visits (non-payors).
### Dashboard

#### System Risk - Current Risk Total 2.5%

<table>
<thead>
<tr>
<th>Goal (%)</th>
<th>Current (%)</th>
<th># Cases</th>
<th>Date Thru</th>
<th>Current System Risk</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>79th</td>
<td>78th</td>
<td>850</td>
<td>3/12/10</td>
<td>1.0 %</td>
<td>My Score</td>
</tr>
<tr>
<td>79th</td>
<td>77th</td>
<td>30</td>
<td>3/12/10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### PQRI - % of Clinic Physicians submitting claims with CPT II codes attached

<table>
<thead>
<tr>
<th>Goal</th>
<th>Current</th>
<th>Reporting Physicians</th>
<th>Date Thru</th>
<th>Current System Risk</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;75%</td>
<td>85%</td>
<td>550 of 830</td>
<td>3/12/10</td>
<td>0.5%</td>
<td>My Practice</td>
</tr>
</tbody>
</table>

#### Core Measures - # Hospitals at core measure composite target

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Goal</th>
<th>At Target</th>
<th>Composite Score</th>
<th>Period</th>
<th>Current System Risk</th>
<th>Details by Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>4/4</td>
<td>2/4</td>
<td>N/A</td>
<td>Q4-09</td>
<td>1.0%</td>
<td></td>
</tr>
</tbody>
</table>

#### PQRI Measure

<table>
<thead>
<tr>
<th>Category</th>
<th>Current Period Value (10/01/2010 thru 02/12/2010)</th>
<th>% Participation</th>
<th>% Performance</th>
<th>Quality Information Available?</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c Poor Control in Diabetes Mellitus</td>
<td>90% (n=20) 90% (n=10) 90% (n=20) 90% (n=10)</td>
<td>90%</td>
<td>90%</td>
<td>more</td>
</tr>
<tr>
<td>(LDL-C) Control in Diabetes Mellitus</td>
<td>85% (n=20) 85% (n=10) 85% (n=20) 85% (n=10)</td>
<td>85%</td>
<td>85%</td>
<td>more</td>
</tr>
<tr>
<td>Asthma Assessment</td>
<td>85% (n=20) 85% (n=10) 85% (n=20) 85% (n=10)</td>
<td>85%</td>
<td>85%</td>
<td>N/A</td>
</tr>
<tr>
<td>Inquiry Regarding Tobacco Use</td>
<td>56% (n=240) 66% (n=165) 66% (n=240) 66% (n=185)</td>
<td>66%</td>
<td>66%</td>
<td>more</td>
</tr>
</tbody>
</table>
Bending the curve?

2010

2013?
What’s coming – ACO analytics

SPORT trial – Jim Weinstein et. al
Thank You