



# Open Data Cloud for Biomedical Informatics: A Public/Private Partnership to Fast Track Evidence- based Medicine

*Dr. Chalapathy Neti*

*Executive Architect, Information Agenda for Healthcare*

*Co-Strategist, Worldwide Healthcare Research*

*Dr. Shahram Ebodallahi, Research Staff Member*

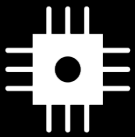
*IBM Research*

# The World is also becoming Smarter

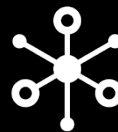
***FLATTER***



***SMALLER***



***INSTRUMENTED***



***INTERCONNECTED***



***INTELLIGENT***

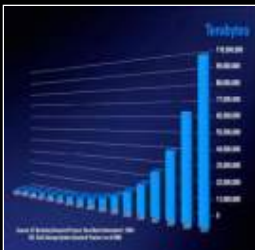
# The Information Explosion...

**15 Petabytes**  
*of new information  
generated every day*



**New Intelligence...**  
*Smarter Enterprises are  
Using Information to Make  
More Intelligent Choices*

## Volume

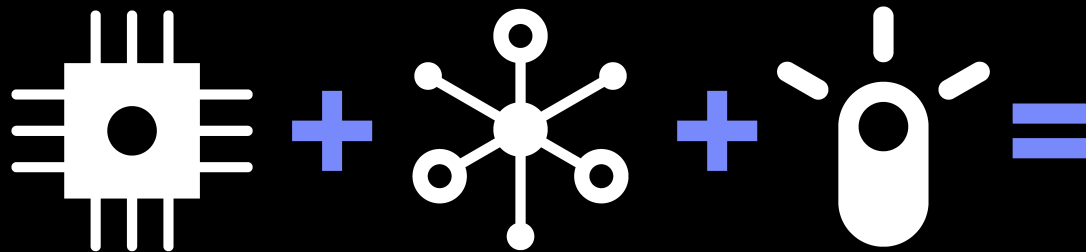


## Variety



## Velocity





An opportunity for health systems  
to think and act in new ways.

**Optimize best  
healthcare practices.**

**Deliver excellent,  
individualized health and  
care experiences.**

**Create new healthcare  
service models  
and markets.**

# IOM EBM Roundtable goal...

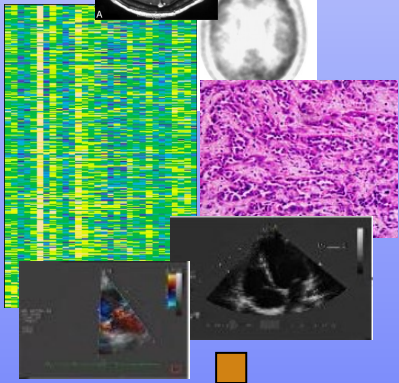
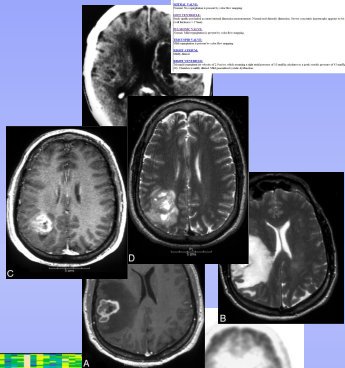
By the year 2020, *ninety percent of clinical decisions* will be supported by accurate, timely, and up-to-date clinical information, and will reflect the best available evidence

*IOM Roundtable Charter*



# Vision: data → information → evidence → decision

## Patient Data

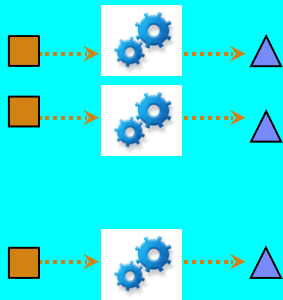


## Information Extraction

Text Analytics

Image Analytics

Gene Analytics



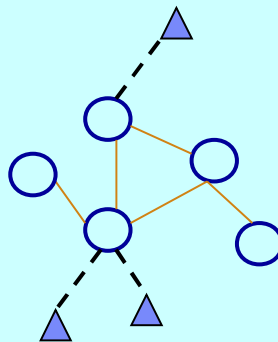
Analytics

## Information Organization and Integration

Ontologies

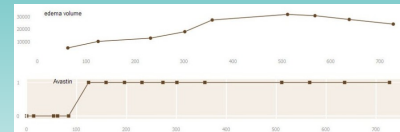
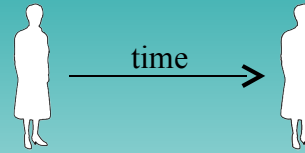
Representation

Re-purposing

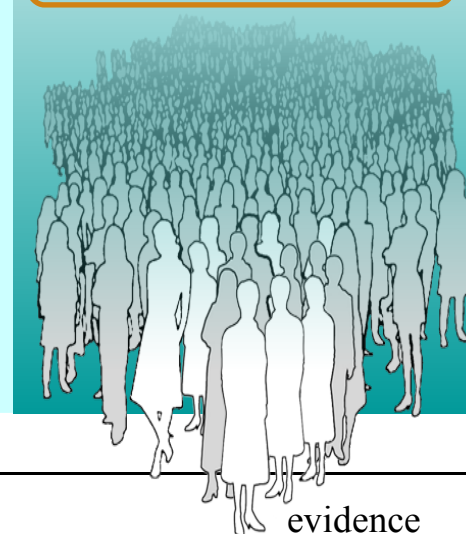


## Mining for Evidence

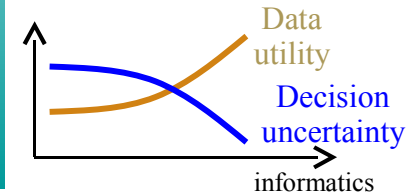
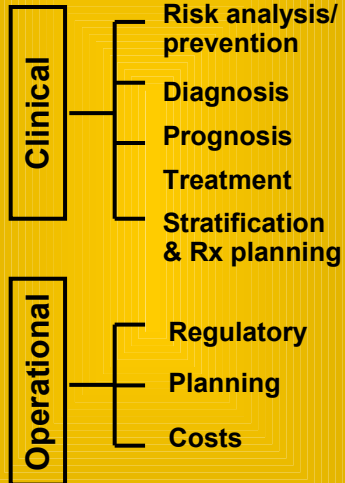
### Time-based Analysis



### Population-based Analysis



## Evidence-based Decision



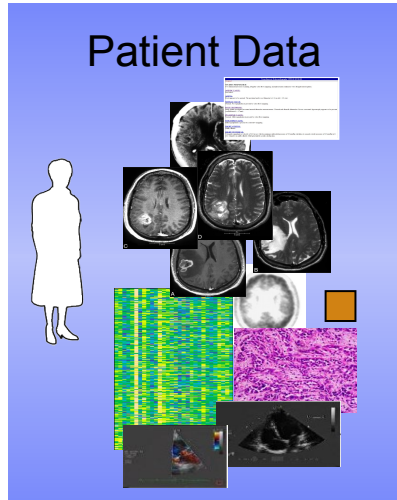
data

information

evidence

decision

# Data



## ■ Current Status:

- Data Collection: **Many ongoing initiatives** for data collection and sharing
  - ADNI, i2b2, caBIG, PhysioNet, ...
  - funds allocated to medical data collection: several \$100M
- Data usage and utility
  - Largely by Individual labs and researchers
  - Results of research activities and independent developments reported via the regular scientific publication channels



i2b2



PhysioNet  
the research resource for  
complex physiologic signals



## ■ Issues:

- **Lack of comprehensive multi-modal longitudinal data sets** for many diseases
- **Ease of access to the larger scientific community** for evidence and insight generation



- **Background:**
  - Low power of clinical measures
  - Need for direct evidence of “disease progression” and measures of “disease modification”
- Overall goals: to provide methods and data which facilitate AD trials
- Long term goal is to develop “validated surrogate markers” for early detection and to monitor progression of AD
- Immediate goals are
  - To develop improved standardized biomarkers for trials
  - To validate these biomarkers
  - By correlating with clinical progression/autopsy
- Funding \$67 mil: \$40 mil NIH, \$27 M private

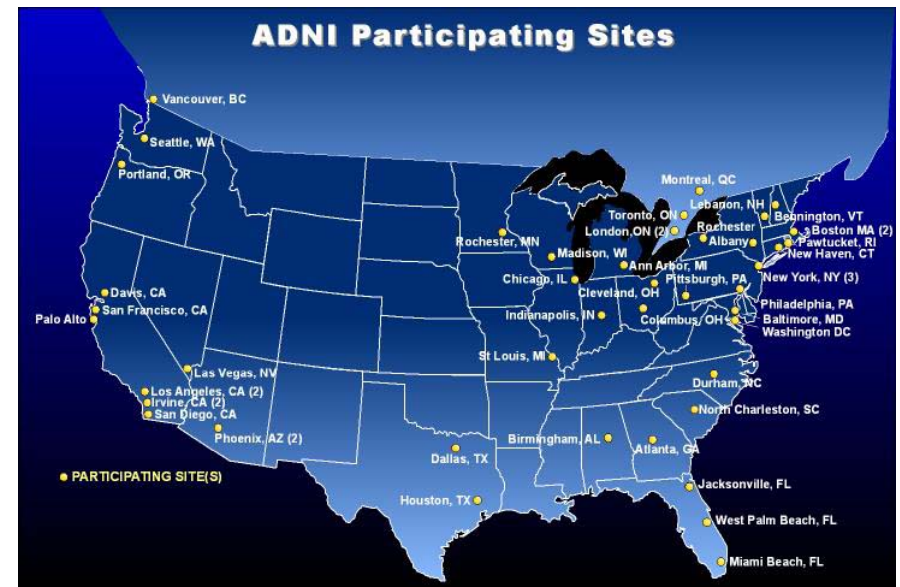
### Data

- 200 NORMAL 3 yrs
- 400 MCI 3 yrs
- 200 AD 2 yrs
- Visits every 6 months

- 57 sites
- Clinical, blood, LP
- Cognitive Tests
- 1.5T MRI

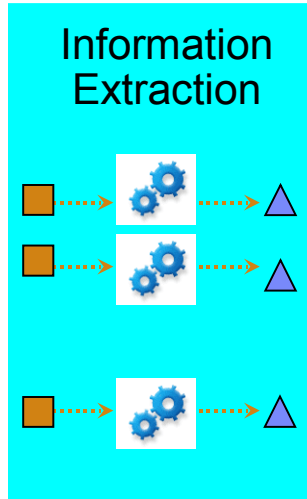
### Some also have

- 3.0T MRI (25%)
- FDG-PET (50%)
- PiB-PET (approx 100)





# Analytics for Information Extraction



geWorkbench

LONI Pipeline

ITK/VTK

MedLEE

- Purpose:
  - Tools for quantification, localization, and characterization of concepts of interest in multi-modal patient data (eg. Hippocampal volume from brain imaging data, dosage of drug from clinical notes, ...)
- Current Status:
  - Many analytic tools are available today
    - Text analytics: MedLEE, MedTAS, ...
    - Image analytics: ITK/VTK, XIP, Slicer, LONIPipeline, ...
    - Gene analytics: geWorkbench, ...
- Issues:
  - Lack of comprehensive benchmarking and validation of the analytics for clinical applications
    - Which analytic tool to trust?
  - Slow pace of analytics tool development

# Information Organization

## ■ Purpose:

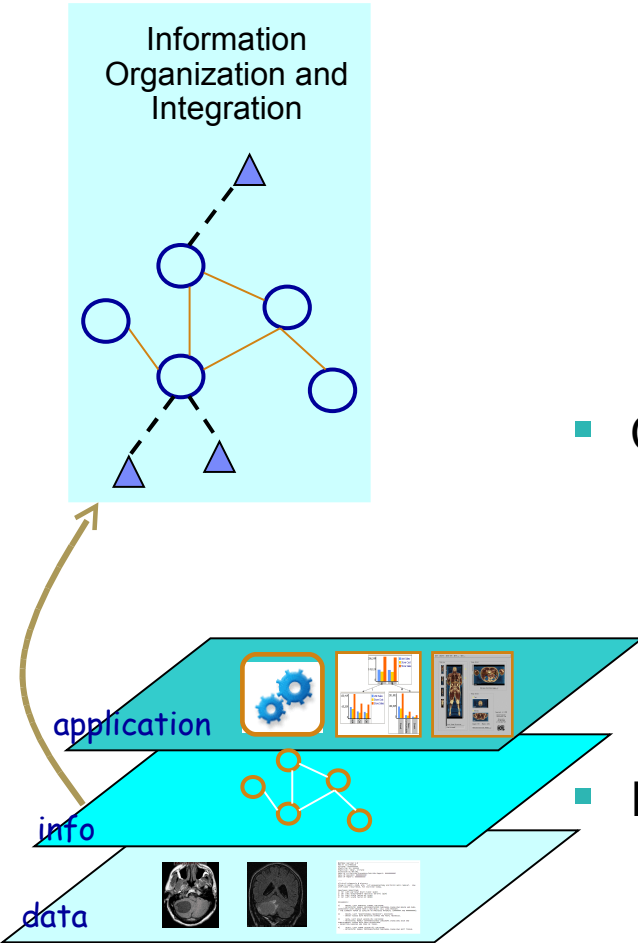
- **Creating an information model** anchored in the underlying multi-modal patient data
- **Summarizing** the characteristics of the different concepts of interest as obtained from various multi-modal data sources

## ■ Current Status:

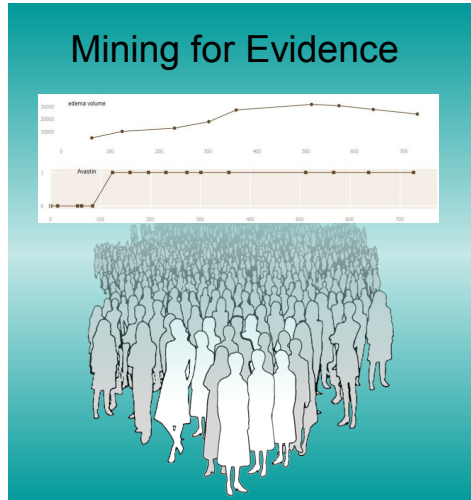
- Many taxonomies to capture key concepts encountered by the clinicians (RadLex, GO, UMLS, ...)
- Many ongoing efforts in linking various taxonomies

## ■ Issues:

- **Lack of flexible and extensible information models** of the longitudinal multi-modal patient data for different purposes and use case scenarios



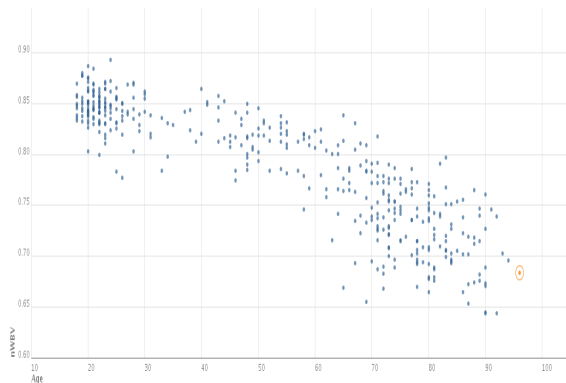
# Evidence mining



- Purpose:
  - Cross-cohort and cross-time information analysis
  - Mine insight and evidence from the large pool of information extracted from patients' multi-modal data
- Current Status:
  - Statistical tools are available today to analyze and mine for insight over large pool of information but with **limited use of cross-modal and multi-modal mining**

Visualizations : whole brain volume by age

Creator: Anonymous  
Tags: health brain, aging, disease, Alzheimer's MRI, neuroimaging.

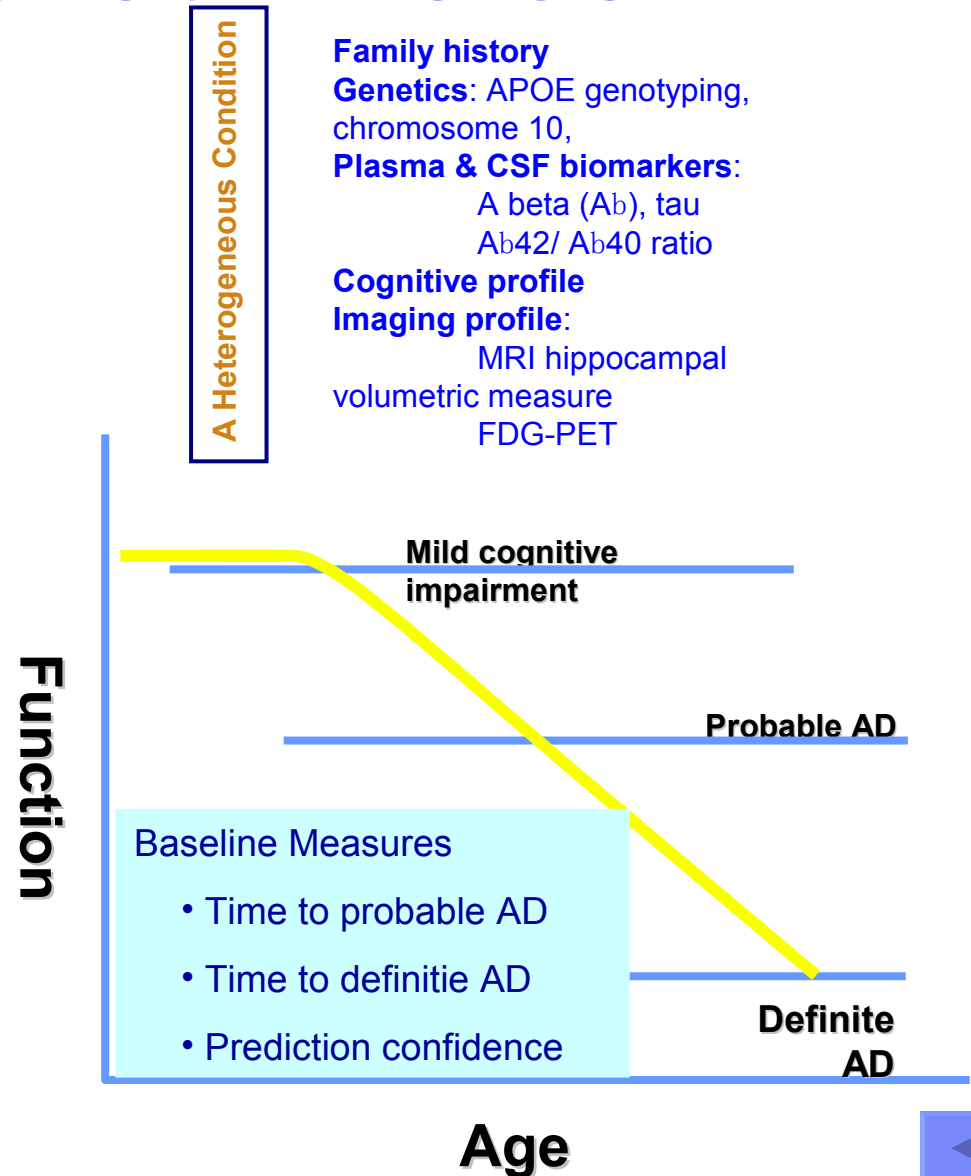


## Issues:

- Lack of an **open and extensible analytics infrastructure for open collaboration**
- Lack of **consumable and composable analytics** for the general medical researcher for ease of hypothesis generation and validation

# Example: Mild Cognitive Impairment → Alzheimer's

- **History:** Parent forgetful for few years before death
- **Cognitive Profile:**
  - Increased forgetfulness:
  - Test of mental status:** Delayed recall
- **Imaging Profile:**
  - MRI:** Mild hippocampal atrophy
  - FDG PET:** Reduced metabolism medial temporal lobe



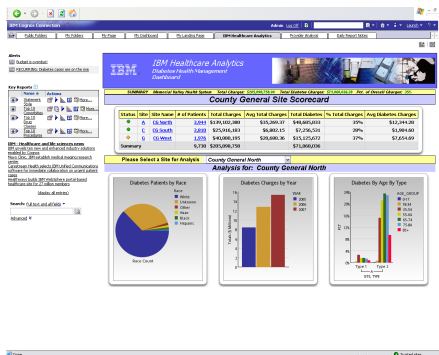
\* Pieter J. Visser, 2006

# Improving clinical and operational decisions

## Evidence-based Decision



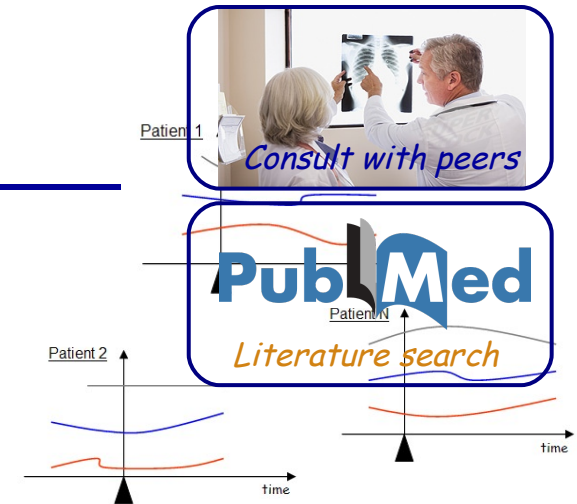
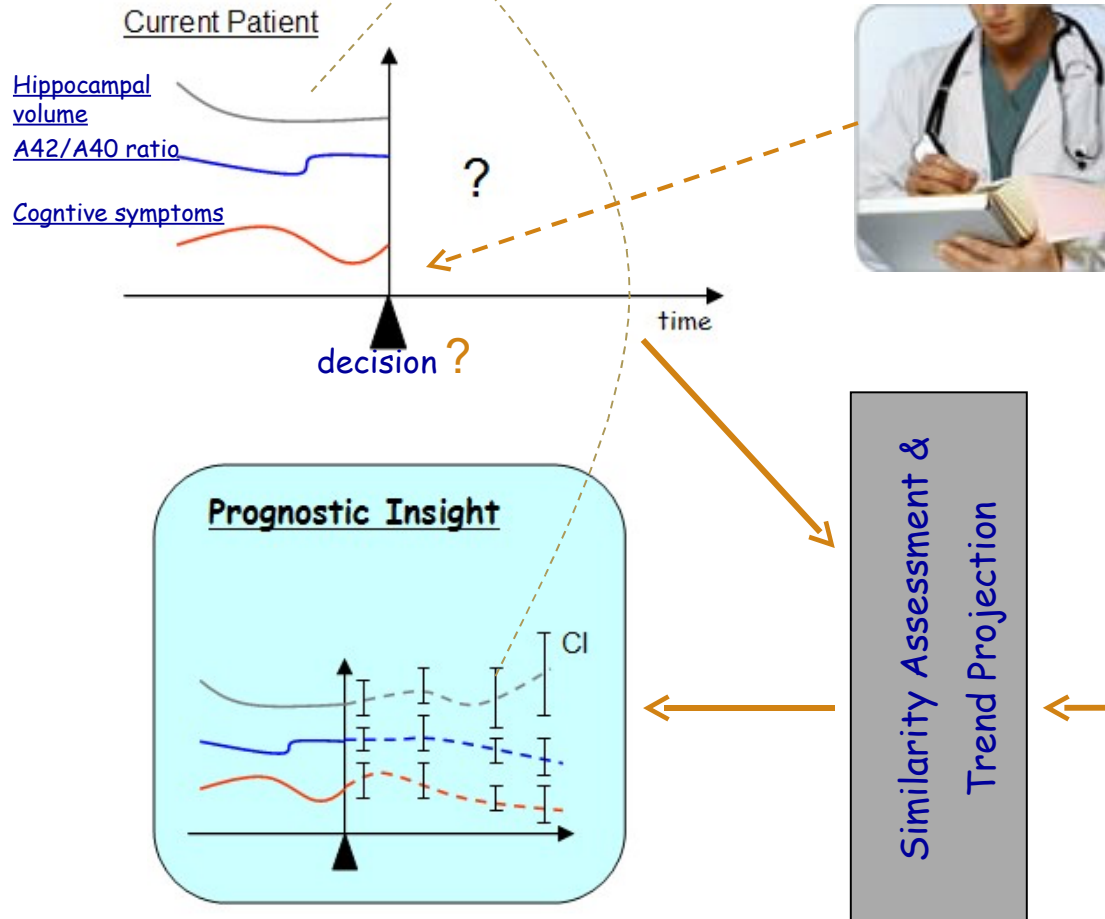
- Purpose:
  - Use insight and evidence derived from multi-modal longitudinal data to improve decisions
    - Time to decision
    - Quality of decision
- Current Status:
  - Using evidence published in journal articles
- Issues:
  - Lack of physician dashboard populated with validated and easy to use tools
  - Inability for **quick decision validation based on evidence obtained from large cohort of “similar” patients**





# Example: "Patient like this"

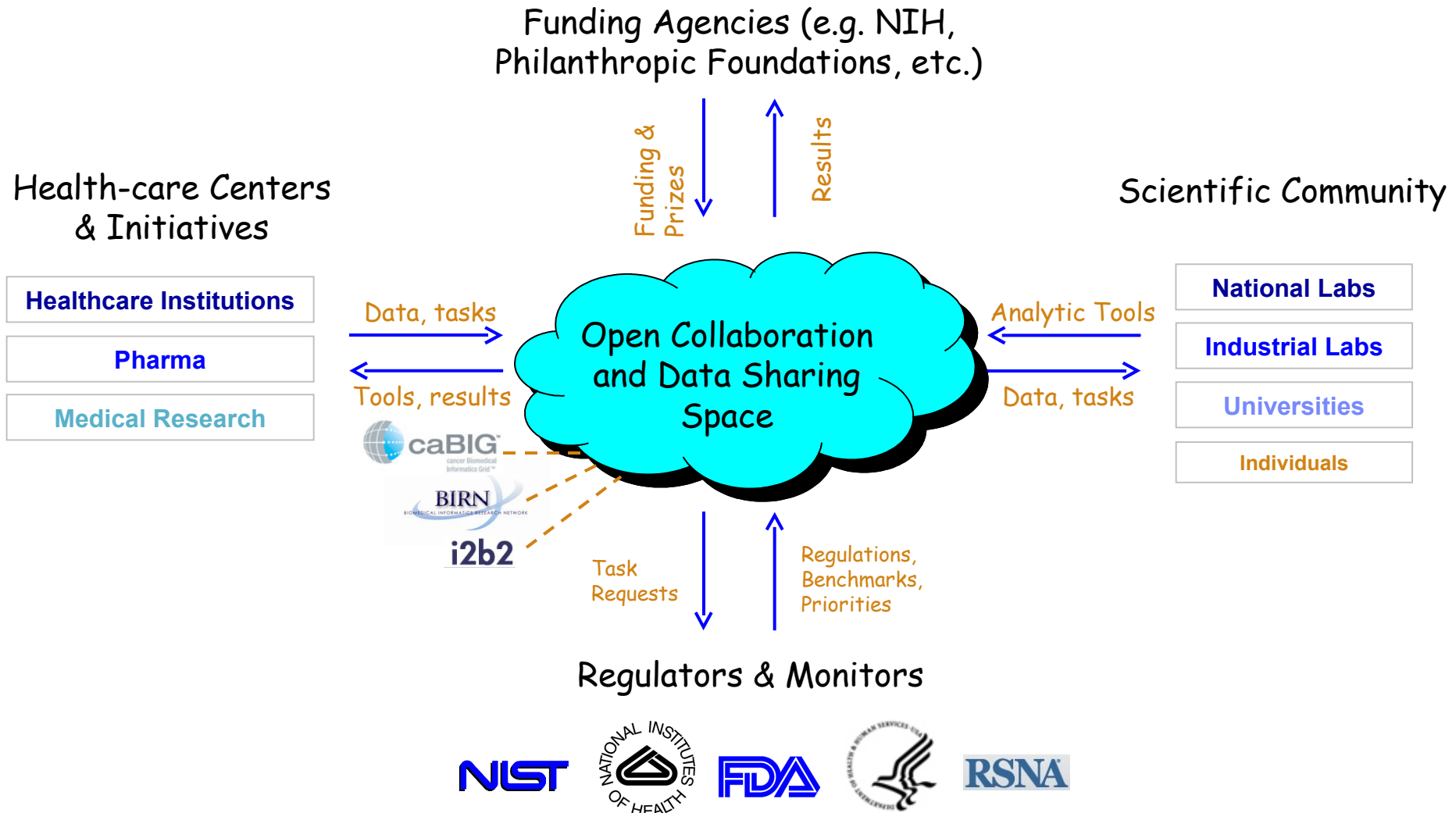
\* In general, more data is needed for increased number of "bio-markers"



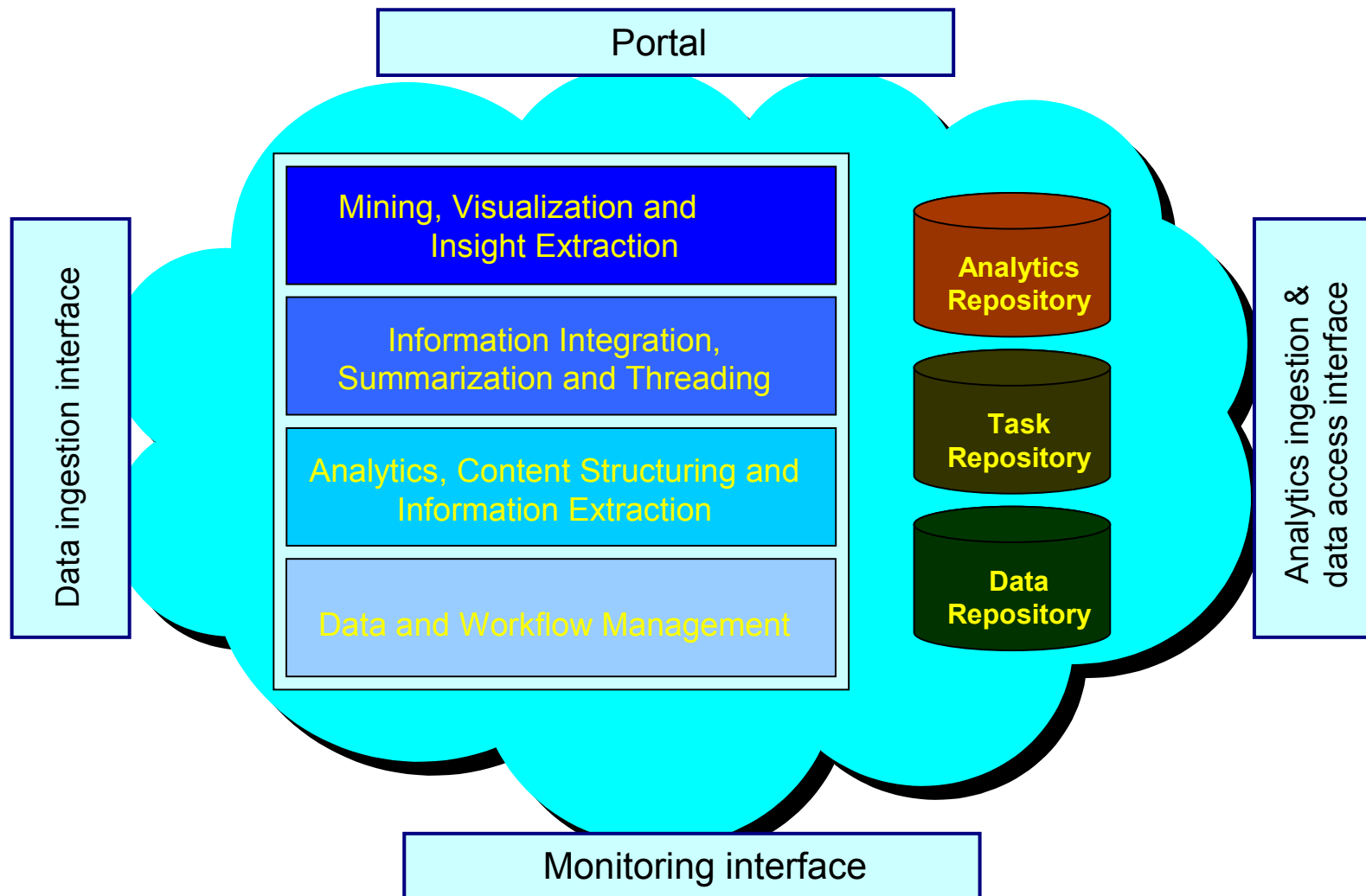
Measure of Improvement

- Time to decision
- Quality of match

# Proposal: An Open Collaborative Model for Fast Tracking Information-Based Medicine



# Proposed Architecture



# Value to Different Constituents

## **Health-care Institutions**

- ROI not justifiable for single-handed data collection
- Minimal investment results in benefit from the efforts of the larger scientific community
- No clinical institution is capable of hiring and maintaining a large scientific resource required for making substantial progress

## **Scientific Community**

- Grand challenge problems have substantial societal and business benefits
- Reduces barrier for entry in grand problem solving arena
- Common data sets supported and approved by funding agencies could facilitate funding for participants

## **Commercial Entities**

- Provides a platform for developing scalable content management, information integration, and decision intelligence technologies
- Cloud computing applied to health-care
- Referenceable context for showcasing value of IT platforms

## **Regulators and Funding Agencies**

- Assess and monitor progress by community
- Prioritize tasks and scientific research to match current directions and issues of importance
- Regulate and approve analytics and tools for clinical use

Thank  
You