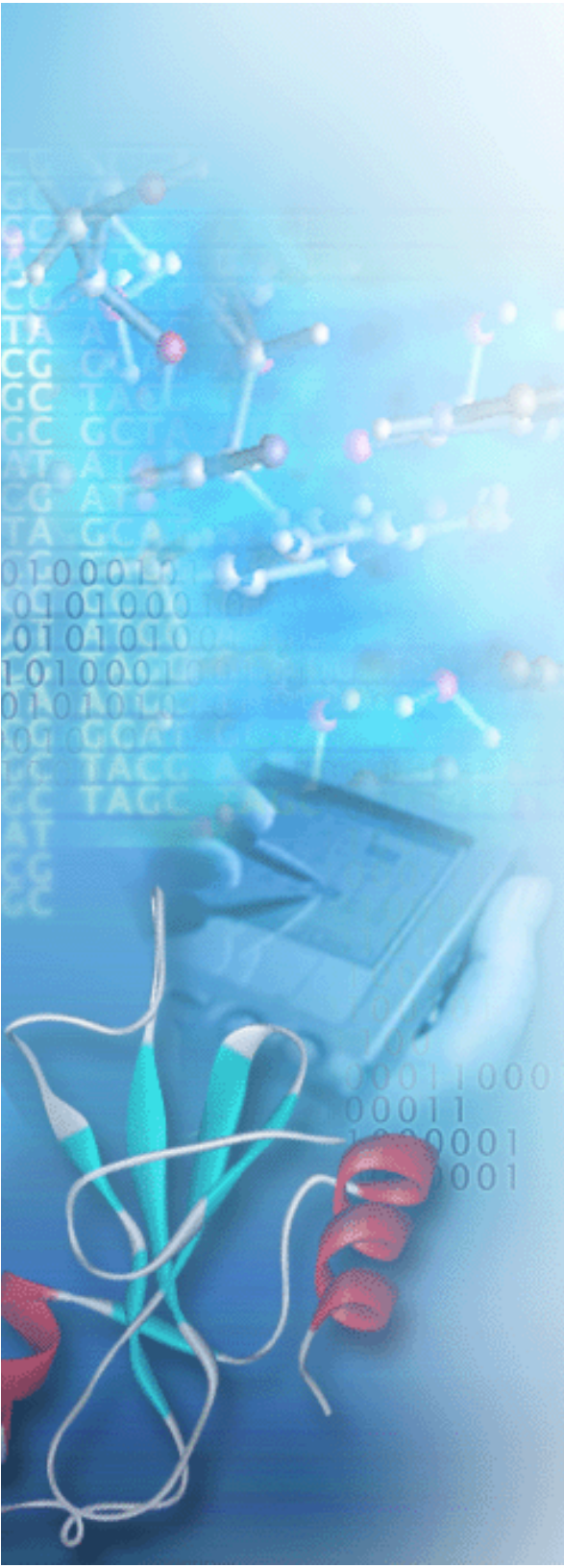


The Integration of Phenotypic and Genotypic Data: Clinical Genomics

NIH BECON/BISTIC Symposium
Bethesda, MD
21 June 2004

Michael Hehenberger, Ph.D.
hehenbem@us.ibm.com

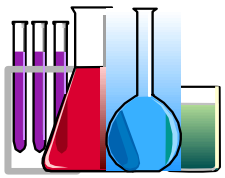
Healthcare technology trend

Whole Body — Organs / Tissues — Cells / Molecules

pre 1930
History &
Physical

Anecdotal to Evidence Based Medicine

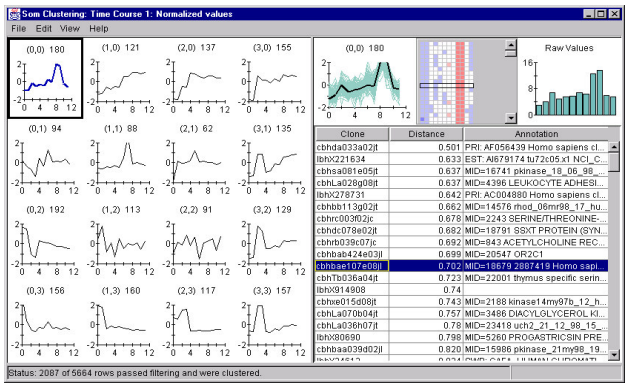
1930-1950
stethoscope,
x-ray,



1950-2000
Lab Tests,
Cat Scan,
MRI

1930-2000
Medical information is
localized, difficult to access,
maintain, transport and compare

2000 - Applied Genomics



Molecular level
understanding
of disease

IBM

IBM has created an Information Based Medicine Unit to address strategic IT needs of the Healthcare and Life Sciences segments

What is Information Based Medicine (I_B_M)?

- Information Based Medicine is the process of improving existing medical and pharmaceutical practices with knowledge generated from the integration of diverse clinical and biomedical data. A key goal of Information-Based Medicine is to improve treatment outcomes by improving the accuracy of diagnostic decisions.

Vision:

- Information Based Medicine will enable all Health Sciences stakeholders (patients, providers and payers of care, biomedical researchers, and the institutions and organizations they represent) to improve the diagnosis and treatment of disease by accelerating the industry towards targeted therapies and personalized healthcare.

Focus Areas:

- IBM's Information Based Medicine EBO Unit will focus on Clinical Genomics and Biomedical Imaging in 2004 to achieve tangible proof of its vision.

IBM®

Genomics and Transcriptomics stimulate the understanding of Gene-Disease Relationships and Biochemical Processes

GENOMICS:
"Scientific discipline of mapping, sequencing, and analyzing genes and chromosomes"

Human Genome Project

How to apply the new knowledge?

STRUCTURAL GENOMICS:
"Construction of high-resolution, genetic maps of an organism"

FUNCTIONAL GENOMICS:
"From Gene Sequence to Gene Function, Transcriptome(*), Protein Structure and Biochemical Processes ..."

Gene-disease relationships:
Uncover genetic origins of diseases. Increased complexity due to multi-gene connections

Target-based drug discovery:
Target Validation of new genes - leading to a systematic drug development approach

Pharmacogenomics:
Understand genetic and other contributors to patient variability in response to drugs

(*) The Transcriptome is defined as the complete collection of transcribed elements of the Genome. In addition to mRNAs, it also represents non-coding RNAs, used for structural and regulatory purposes. Alterations in the structure or levels of expression of any one of these RNAs or their proteins can contribute to disease.

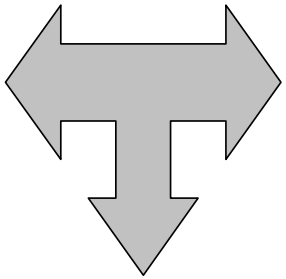
IBM

Clinical Genomics (CG): What is it?

CG addresses the technological challenges behind the acquisition, collection, management, integration and analysis of phenotypic and genotypic data …

Phenotype:

- **EMR / EDC Data**
- **Patient outcome data**
- **Cell Image (e.g. IHC)**
- **Medical Images**
- **Disease Progression**
- **etc…**



Genotype:

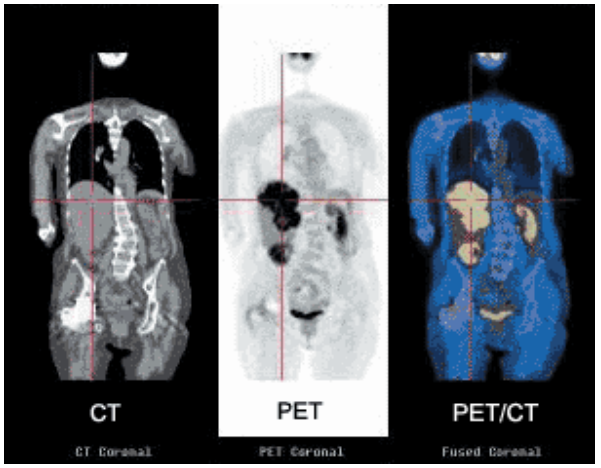
- **Raw sequence**
- **SNP's/Microsatellites**
- **Transcriptome**
- **Proteome**
- **Genealogy**
- **etc…**

- ❖ **To identify and validate novel therapeutic targets**
- ❖ **To conduct more focused clinical research (Pharmacogenomics)**
- ❖ **To revolutionize the ways by which diseases are diagnosed/ treated**

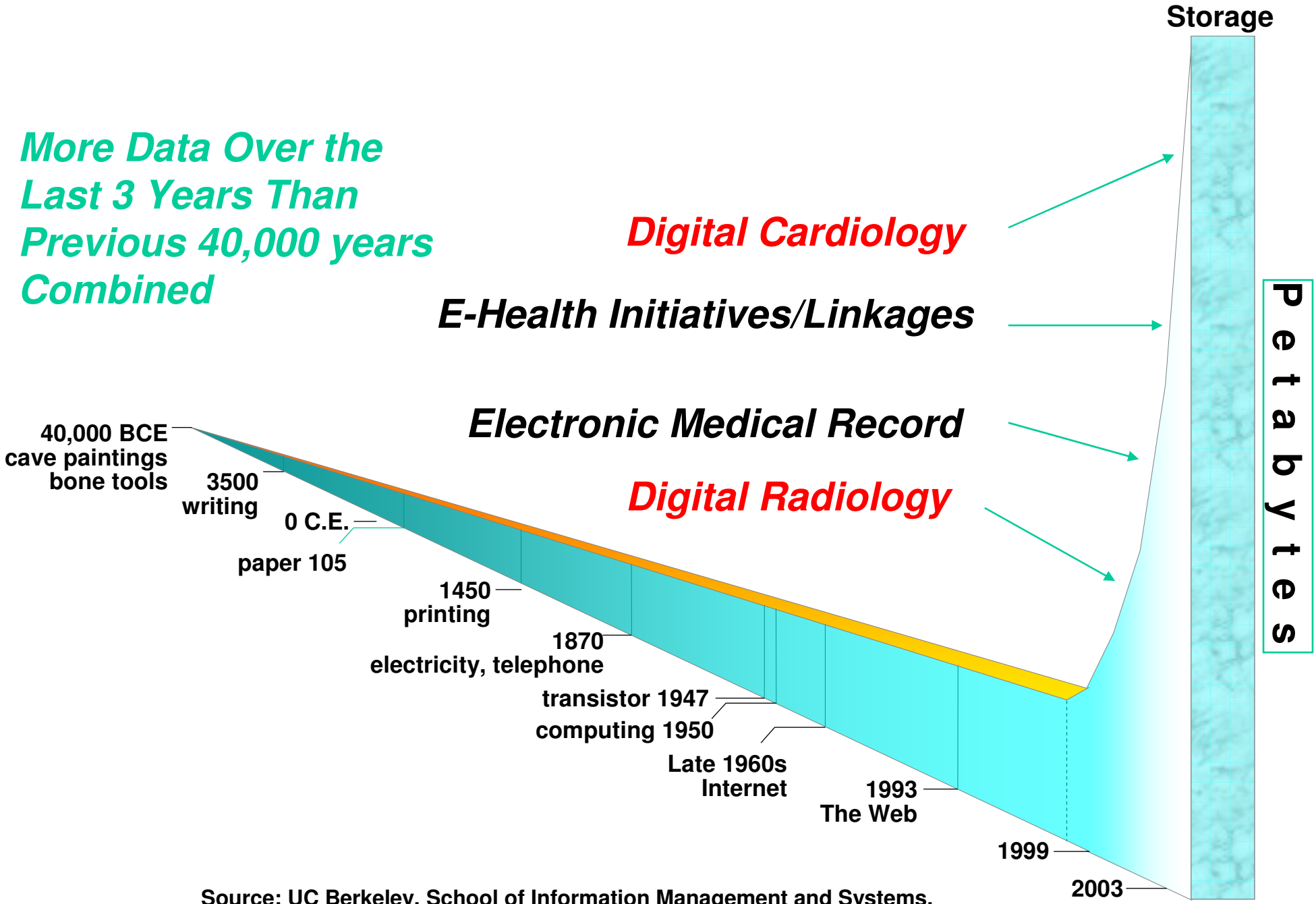
IBM®

We define BioMedical Imaging (BMI) to include both traditional medical imaging modalities as well as R&D imaging modalities.

- **Image based diagnostics**
  - –Traditional Radiology and additional ‘ologies (e.g. Cardiology)
  - –Clinical lab data & histopathology (e.g. microscopy)
  - –New visualization techniques (e.g. CAD, 3D)
- **Image Utilization in R&D**
  - –Research image data (histology)
  - –Functional MRI
  - –Molecular imaging

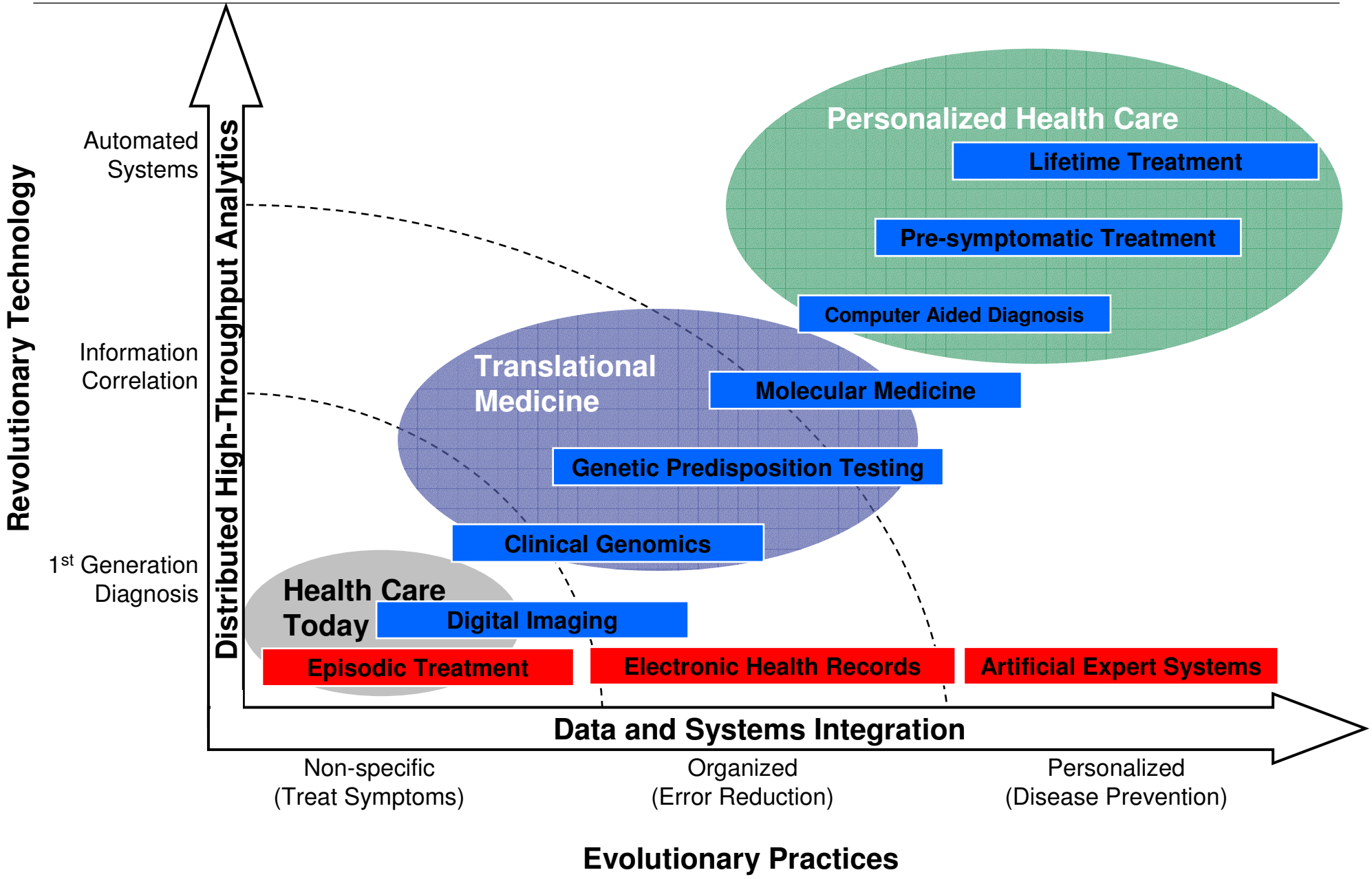


BioMedical Imaging is fueling a Data Explosion!



Source: UC Berkeley, School of Information Management and Systems.

Evolutionary Practices and Revolutionary Technologies are shaping the Future of Information Based Medicine



A possible future Healthcare Scenario

Hospital

Baby

Genotype recorded → *Secure database*

BioBank

Patient scenario in the future

- Improve quality of life
- Increase life span

Physician's office

Personalized immunization

+

Screening schedule

+

Preventive measures

Screening based on biomolecular profiling

Refined treatment

Life

Adult

Drugs

Developed to target particular disease subtypes on particular genetic background

Sources: Science, McKinsey

IBM

CG addresses the needs of AMRC's and Biobanks …

Medical Research

Care Delivery

Patient

Academic Medical Research Centers (AMRC's) and Biobanks focus heavily on the integration of care delivery and research functions to improve understanding of disease, improve quality of care and ultimately move towards personalized medicine

- **Collect, access or mine all patient data (including blood and tissue samples) securely for research**
- **Patient records not well organized – difficult to find patient history/need to build enterprise clinical data warehouses**
- **Compliance with regulatory requirements, particularly patient privacy regulations (e.g. HIPAA in US)**
- **Emerging role of genomics requires patient records to be enhanced with new molecular profiling data**
- **Leverage data standards to enable linkages and associations between disparate data types (e.g. phenotype and genotype)**
- **IT departments not equipped to deal with complexities of the data integration challenge faced by translational research**
- **Establish leadership position to attract gov't and industry funding as well as attract top medical researchers**

IBM

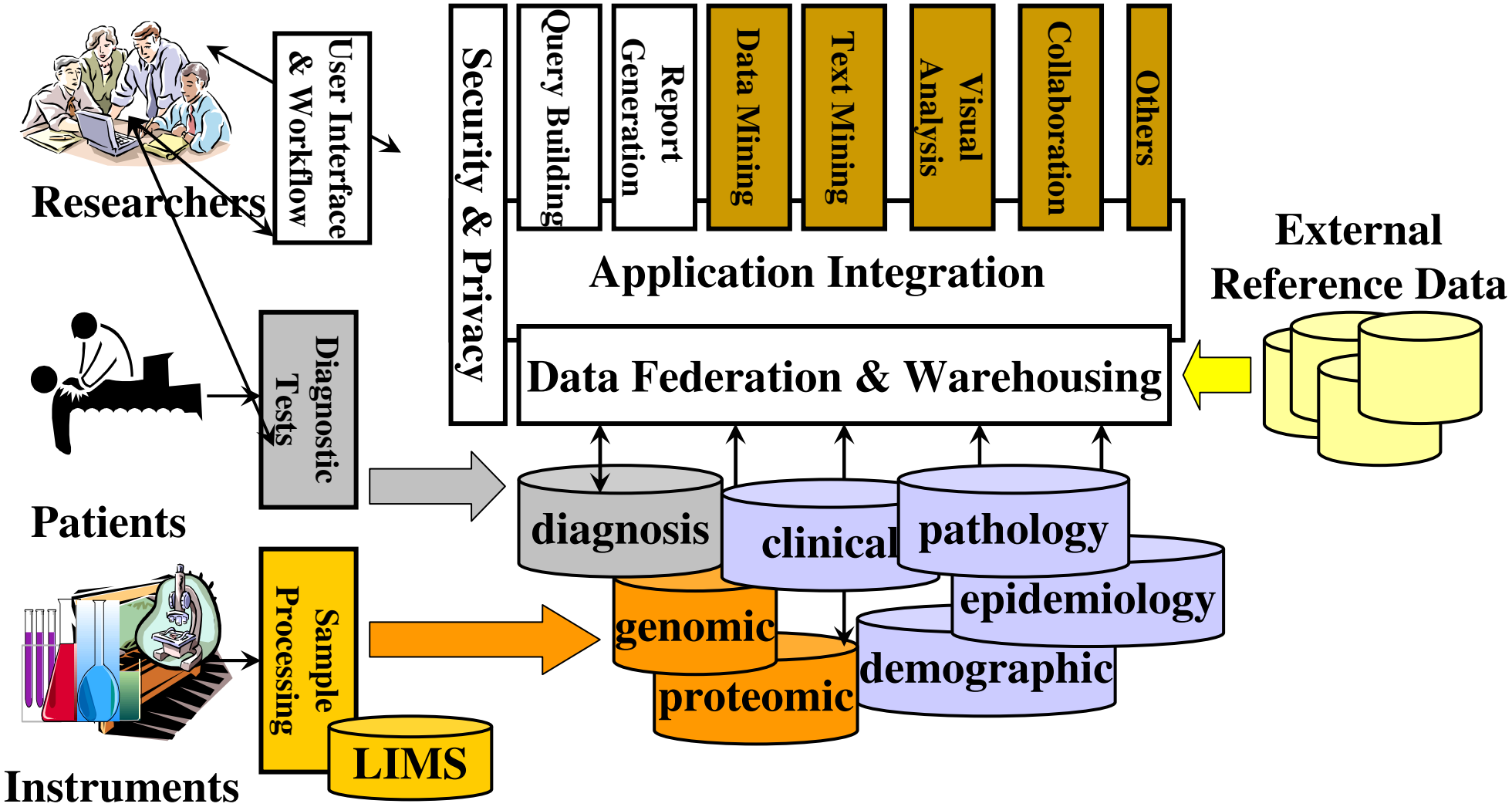
… and the emerging needs of Pharma R&D:

Bio-Pharmaceutical R&D organizations are adding targeted treatment capabilities requiring investment in new technologies and IT infrastructures for blood and tissue sample management, and patient databases that incorporate genotypic data, and query, analysis, and mining capabilities.

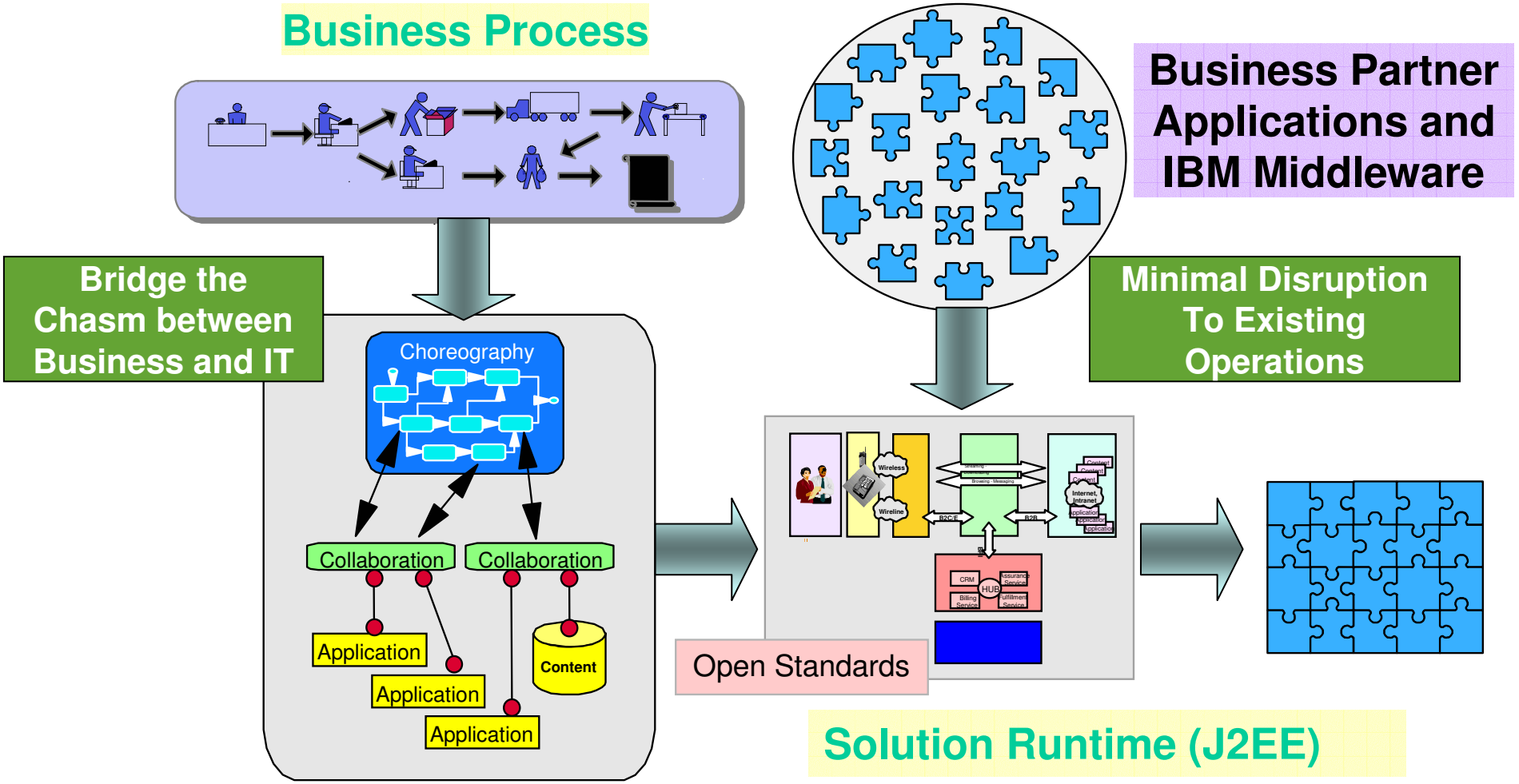
- Identify and validate specific drug targets associated with chosen therapeutic areas
- Collect, access or mine patient data (including blood and tissue samples) securely for research
- Recruit patient populations characterized by Biomarkers identified by diagnostic tests
- “Rescue” drugs that failed due to side effects in subsets of patient populations characterized by shared Biomarkers
- Manage blood and tissue samples collected during clinical trials
- Compliance with regulatory requirements, particularly patient privacy regulations (e.g. HIPAA in US)
- Leverage data standards to enable linkages and associations between disparate data types (e.g. phenotype and genotype)
- Pharma IT departments must deal with complexities of the data integration challenge faced by need to combine patient data with genomics, biomedical image and literature data

IBM

Data Integration and Clinical Data Warehousing are the first steps in the creation of a Clinical Genomics system.



IBM's Approach to Solution (SW) Development is based on the Understanding of Business Processes/Workflow, Open Standards, Data & Application Integration

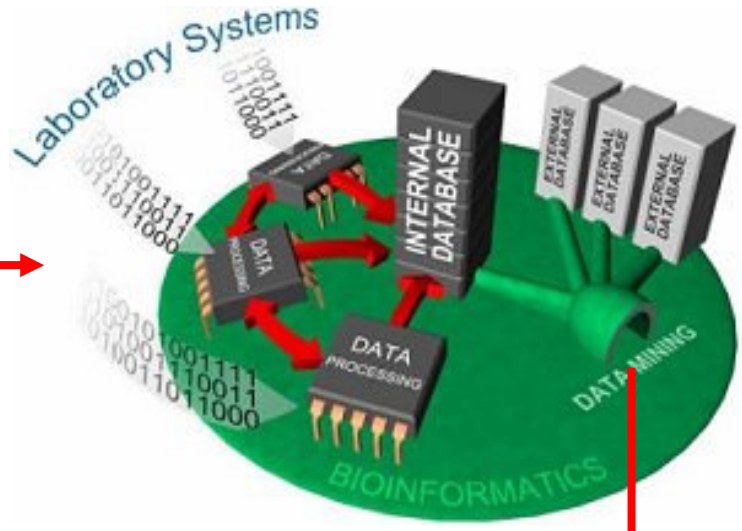


Case Study: IBM/Mayo Clinic Collaboration

Applied Genomics Data Analysis

Genomic data (DNA) – GeneChip array data (RNA)
Protein data

Clinical Data
Signs
Symptoms
Laboratory
Radiology
Etc.



Databases
Genome
Proteome
Disease
Tumors
Drugs

Phase I

Optimized, individualized healthcare

A major Challenge was to enable Mayo's non-IT specialists to ask complex questions to the DB

Find all patients with:

- Coronary artery disease
- Diabetes Mellitus *("diabetes")*
- Nonalcoholic steatohepatitis *(a form of liver disease)*
- Who had a breast biopsy at Mayo *(a procedure)*
- In ZIP code 55901, 55902, 55903, 55904 *(local region)*
- Between 45 and 65 years of age *(certain age)*
- Who are female *(female gender)*
- And are alive *(vital status)*

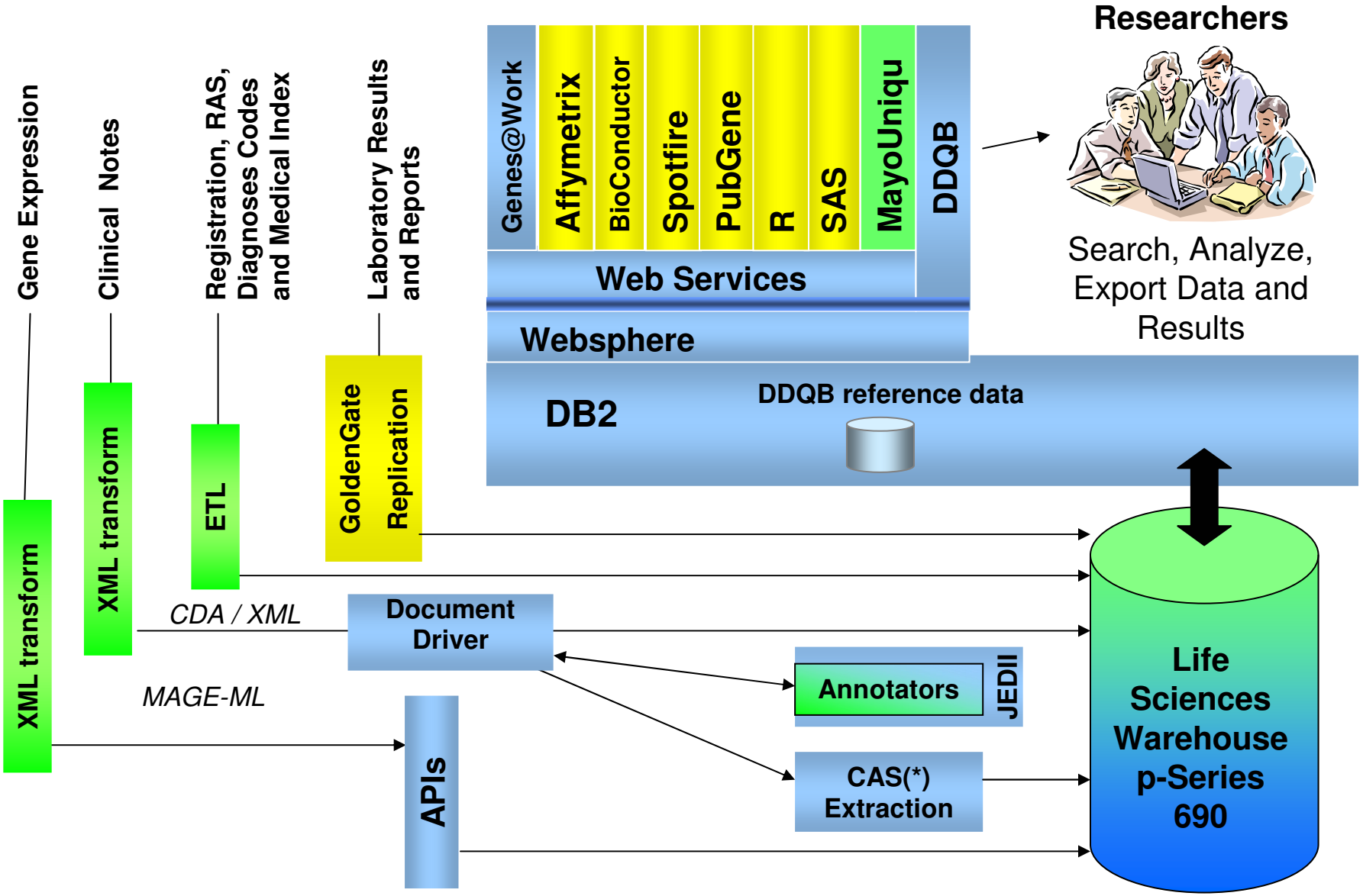
A Second Query Example …

"Find all living female patients with diabetes with a good quality microarray experiment"

- **Diabetes Mellitus** (*Diagnosis Codes, Medical Index & Clinical Notes*)
- Serum **Glucose** > 150 mg/dL (*Results*)
- **Microarray** data exist (*Storage & Retrieval*)
  - **cRNA labeling efficiency** > 90% *(Analysis)*
- Between 45 and 65 years of **age at first diagnosis** (*Demographics combined with Diagnosis Codes, Medical Index & Clinical Notes*)
- Who are **female and alive** (*Demographics*)

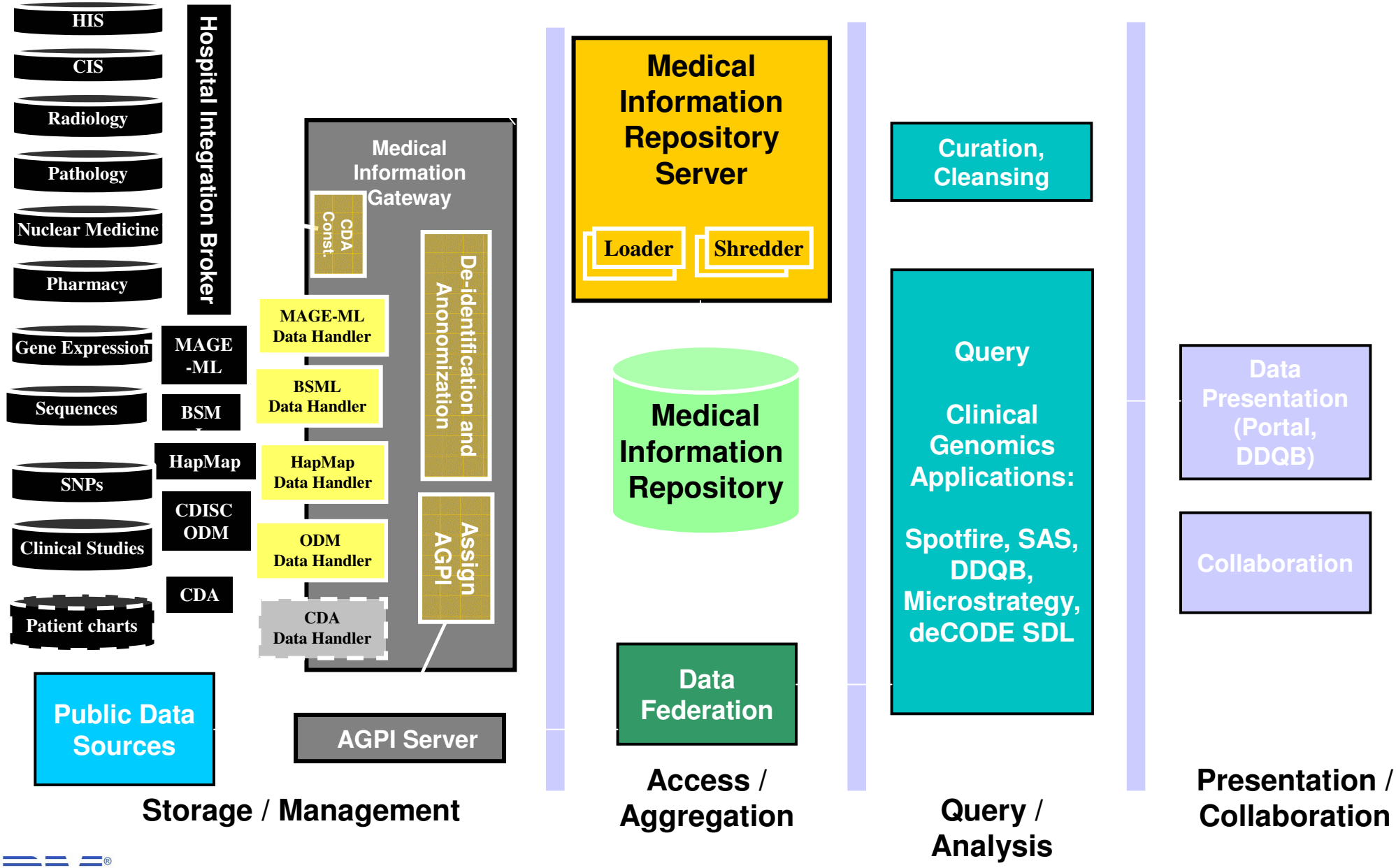
Mayo-IBM Project Conceptual Overview

Architecture Phase I & II

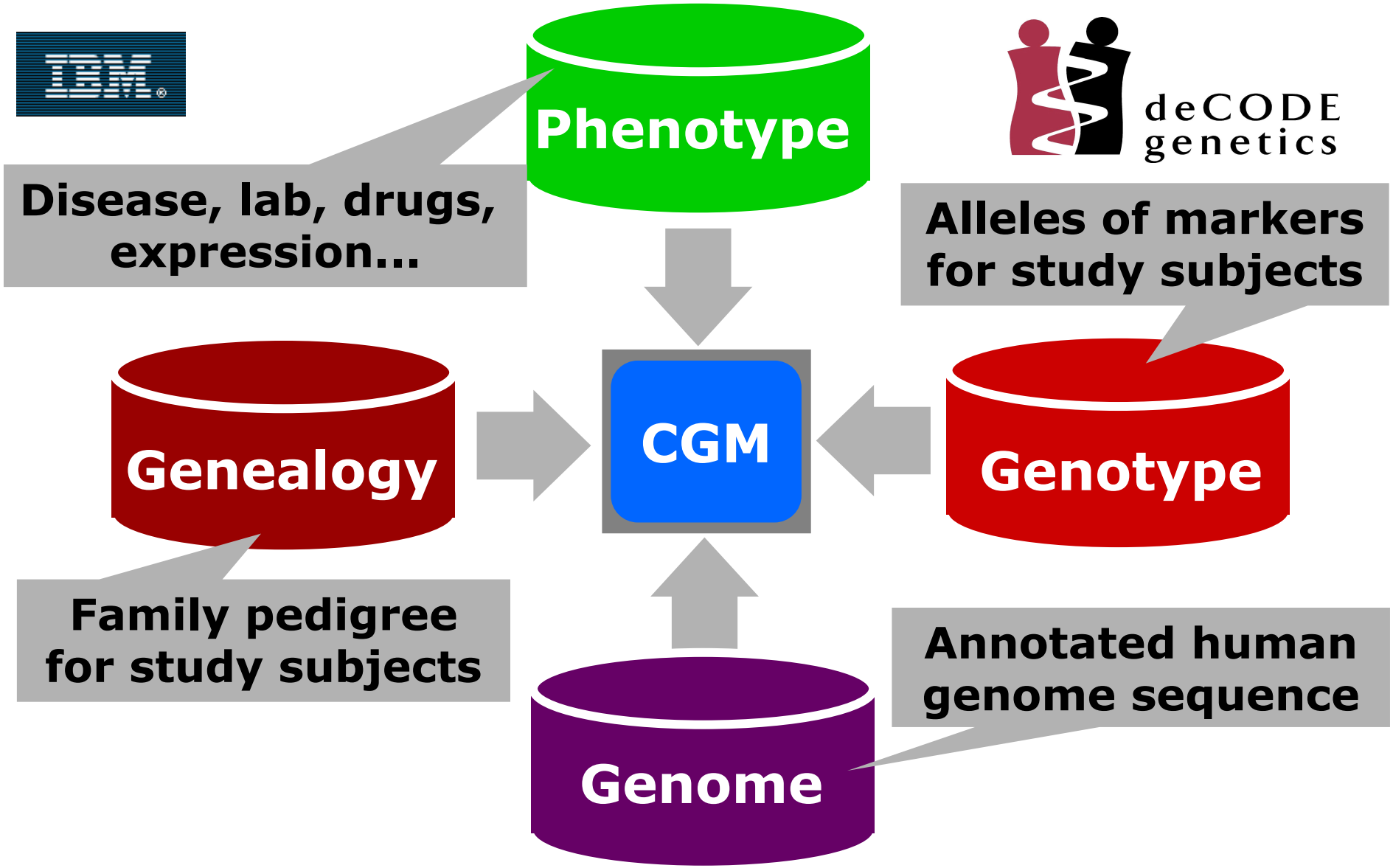


(*) CAS: Common Annotation Scheme

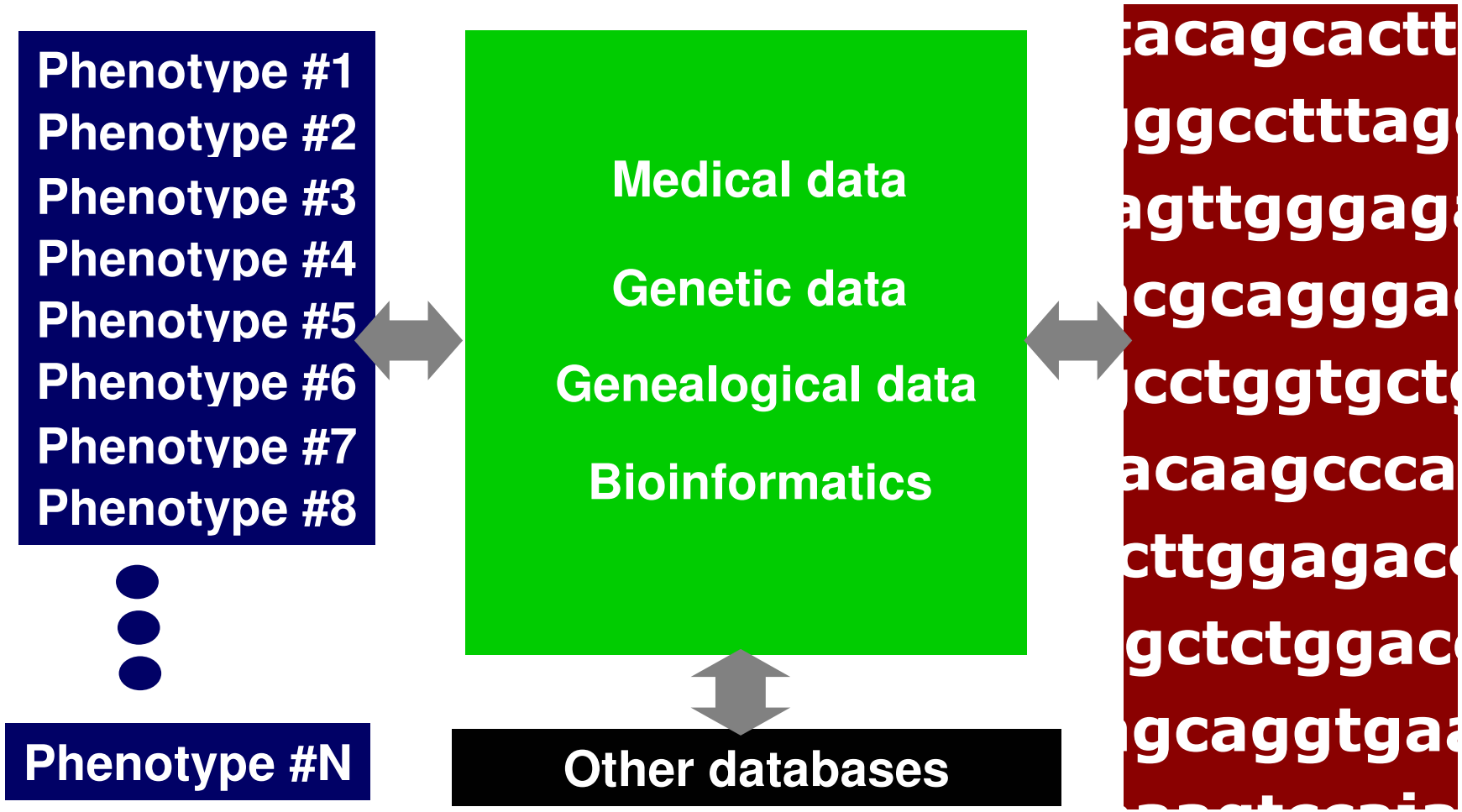
The Clinical Genomics Solution concept includes Data Sources, de-identified and fed into the Medical Information Repository, and the Curation, Cleansing, Query, Analysis, and Presentation of Data



The deCODE Genetics Approach: Combine phenotypic, genetics and genomics data with analysis tools

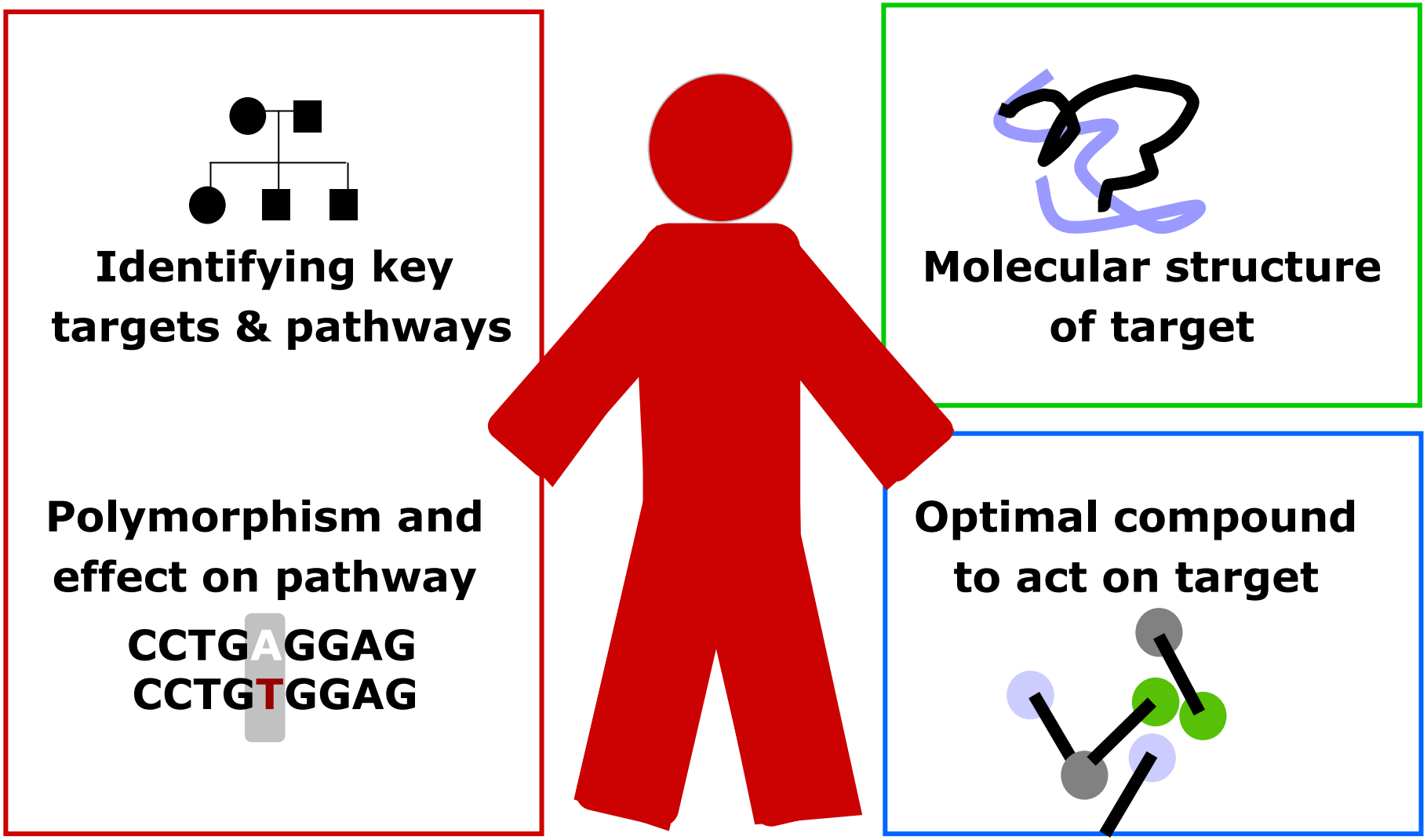


deCODE: Scan the human genome for links between diseases and genes



e.g. 21 Sep 2003: Nature Genetics is publishing deCODE's research leading to the finding that "The Gene encoding phosphodiesterase 4D confers risk of ischemic stroke".

From Disease-Gene Relationships to Target-based Drug Discovery to Lead Identification and Optimization ...



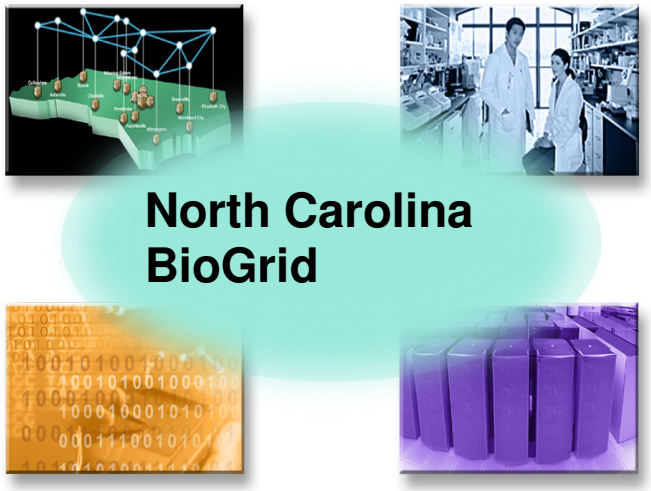
To realize the I_B_M Vision, we need to focus on both Clinical Genomics and on Diagnostic Images stored in Medical Archives

- **Clinical Genomics will enable individualized treatments of patients**
  - **Diagnostic tests for Wellness and Disease**
  - **Preventive Medicine**
- **Advanced Diagnostics will enable "evidence based medicine"**
  - **Integration of IT into medical devices**
  - **Diagnostic information management**
  - **Standards for medical information**
  - **Privacy, security, access control**
  - **Grid Computing**
  - **Advanced database management**
    - **National / Regional Medical Archives**

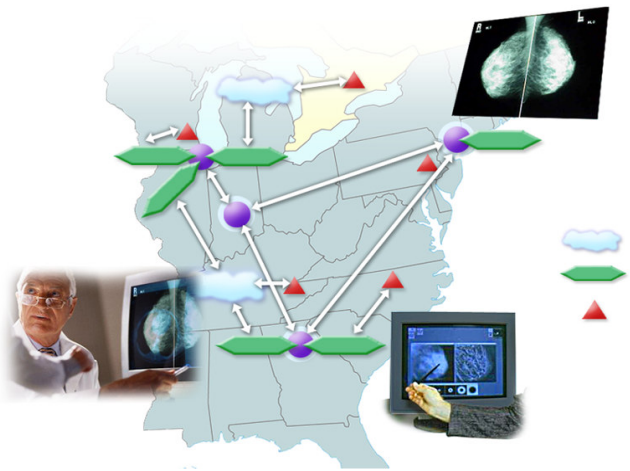
IBM®

Projects in Grid based medical research

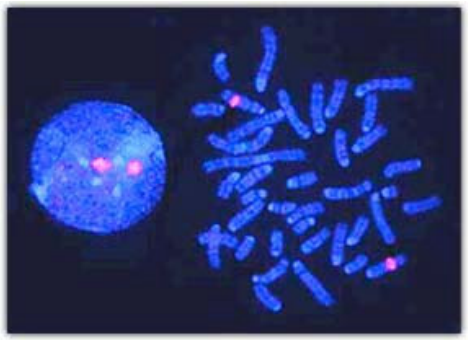
eDiamond Project



North Carolina BioGrid



National Digital Mammography Archive



Smallpox Grid Project

Conclusion: Information Based Medicine is the integration of all forms of biomedical data, across the research to delivery continuum, and imaging is a key data source that drives the need for integration …