Final Presentation

of New Technologies & Future Markets

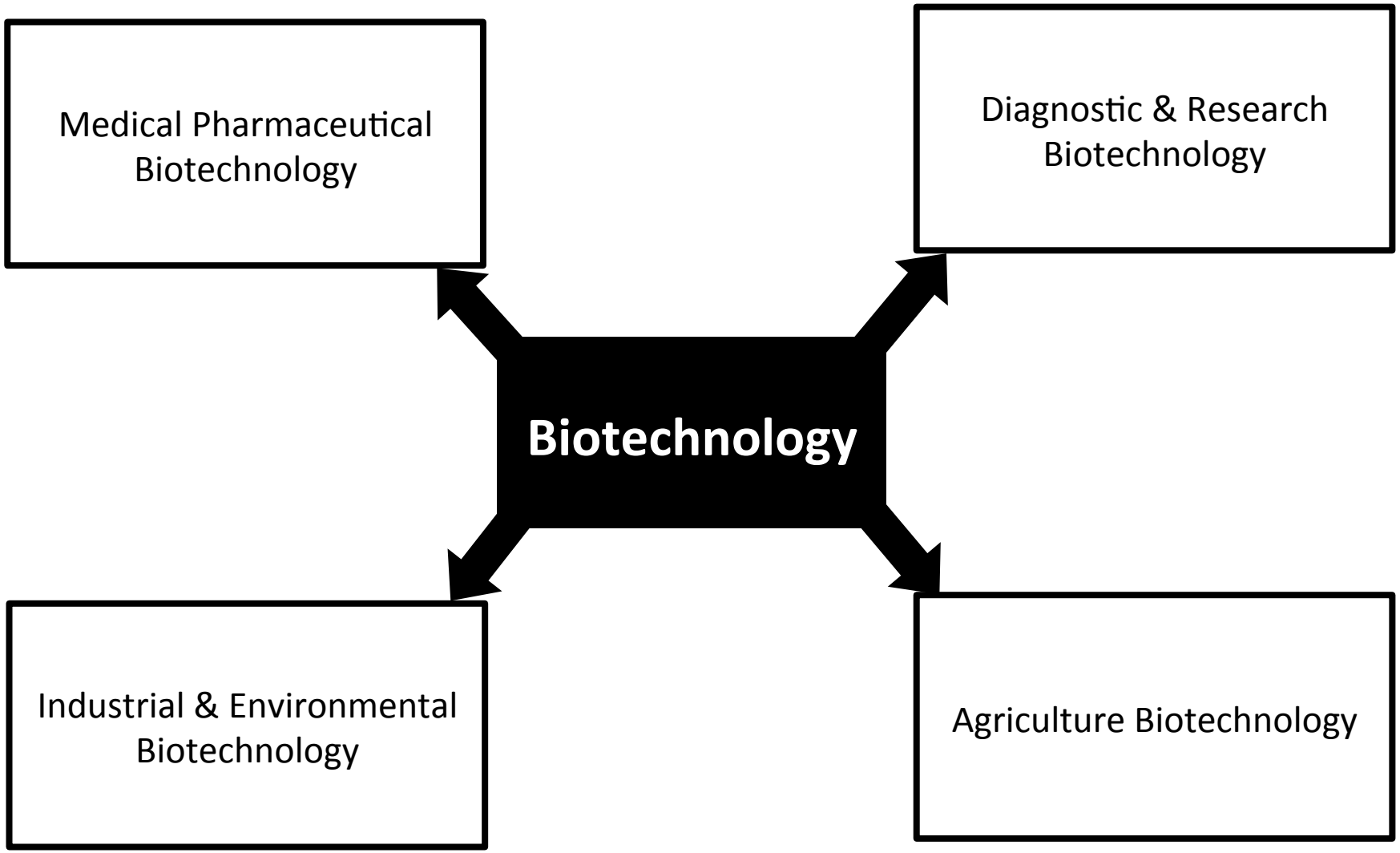
-- Knowledge Management

Du, Yang Lee, Jay-Yoon Fang, Yan
M.S. in Biotechnology Innovation and Computation
Carnegie Mellon University
December 5, 2011

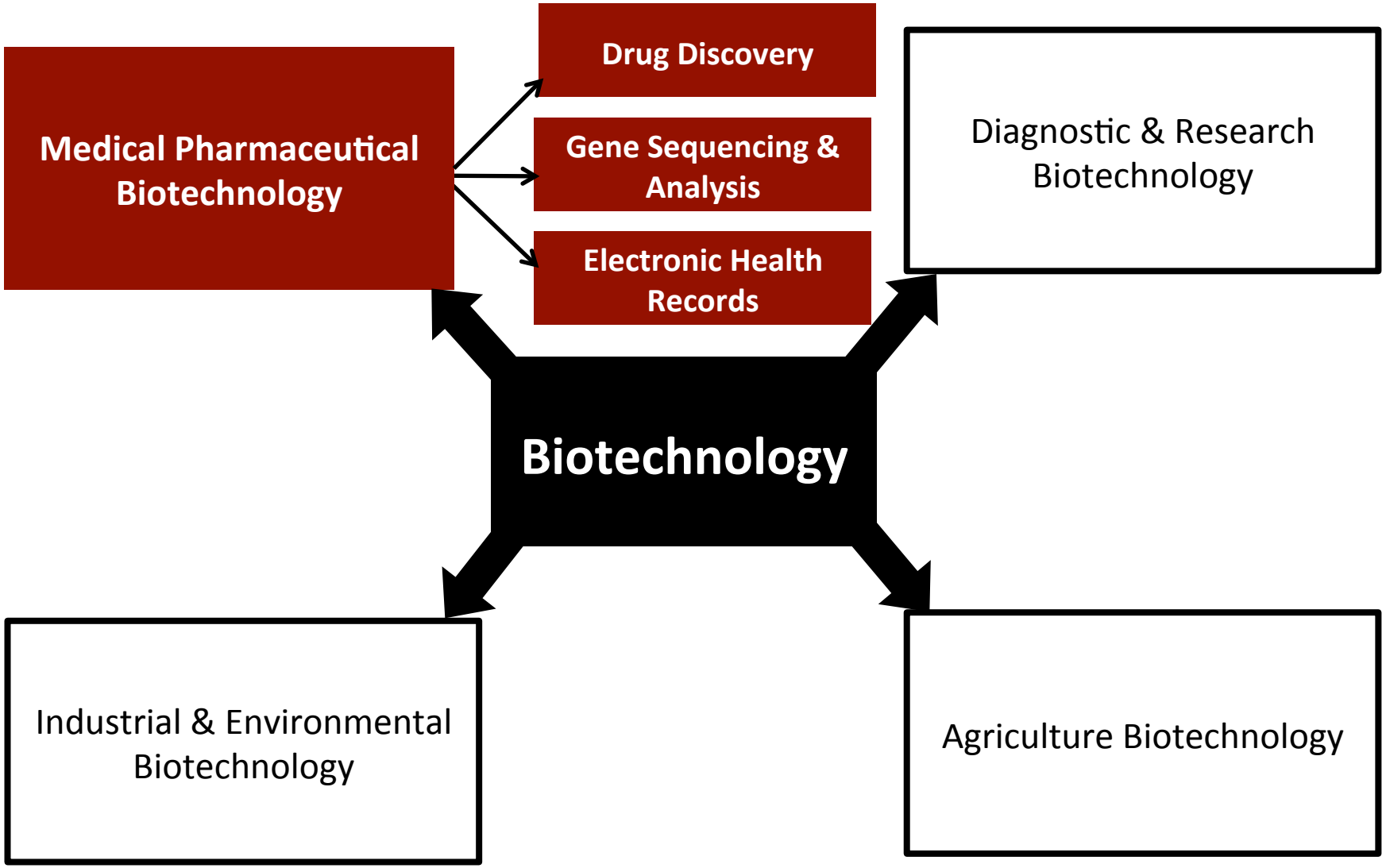
Agenda

- Research Domain
- Problems in the Industry and Causes
- Possible Solutions & Drivers to the Problem
  - New Sources of Data
  - Data Integration Technologies and Analysis
    - Current
    - Future
- Scenarios

Topic Selection



Topic Selection

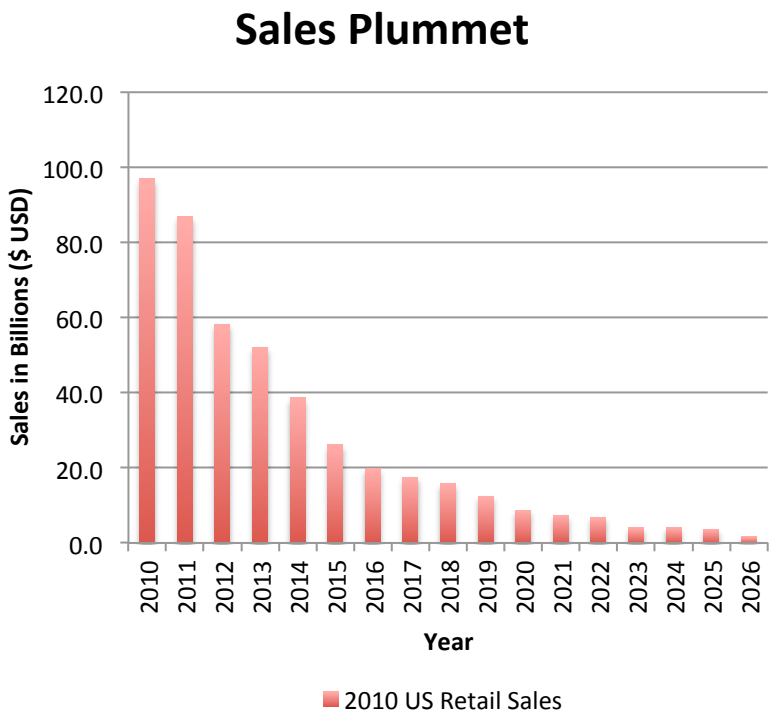


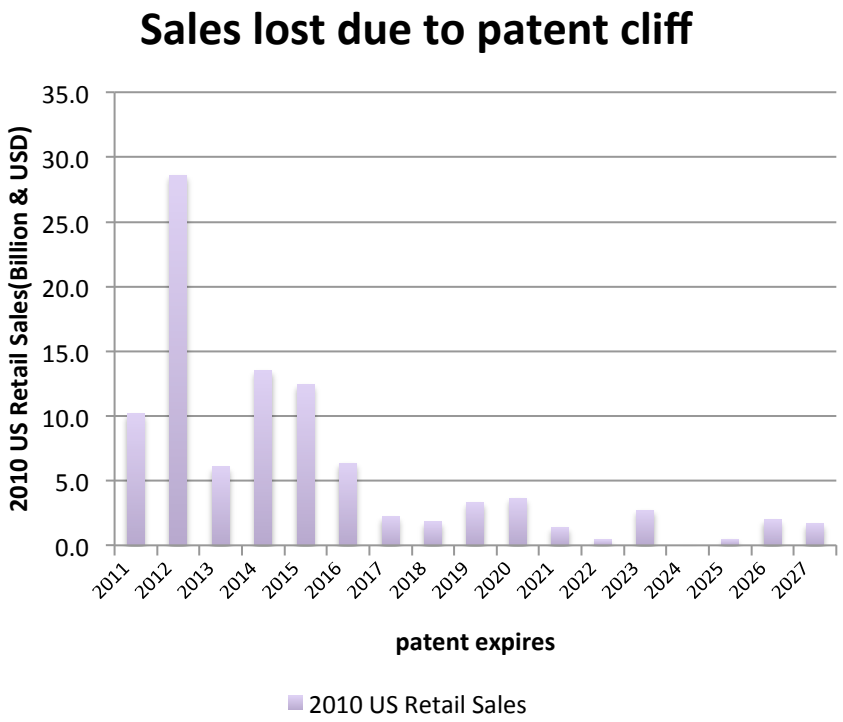
Problem Statement

Pharmaceutical companies are facing many pressures that are threatening their current business model, ultimately the existence of individual companies

Problems: Patent Cliff

- **Expected sales loss**
  - By 2015, 73% of Drug Sales in 2010 are gone.
- **Not enough drugs in clinical trials to compensate**

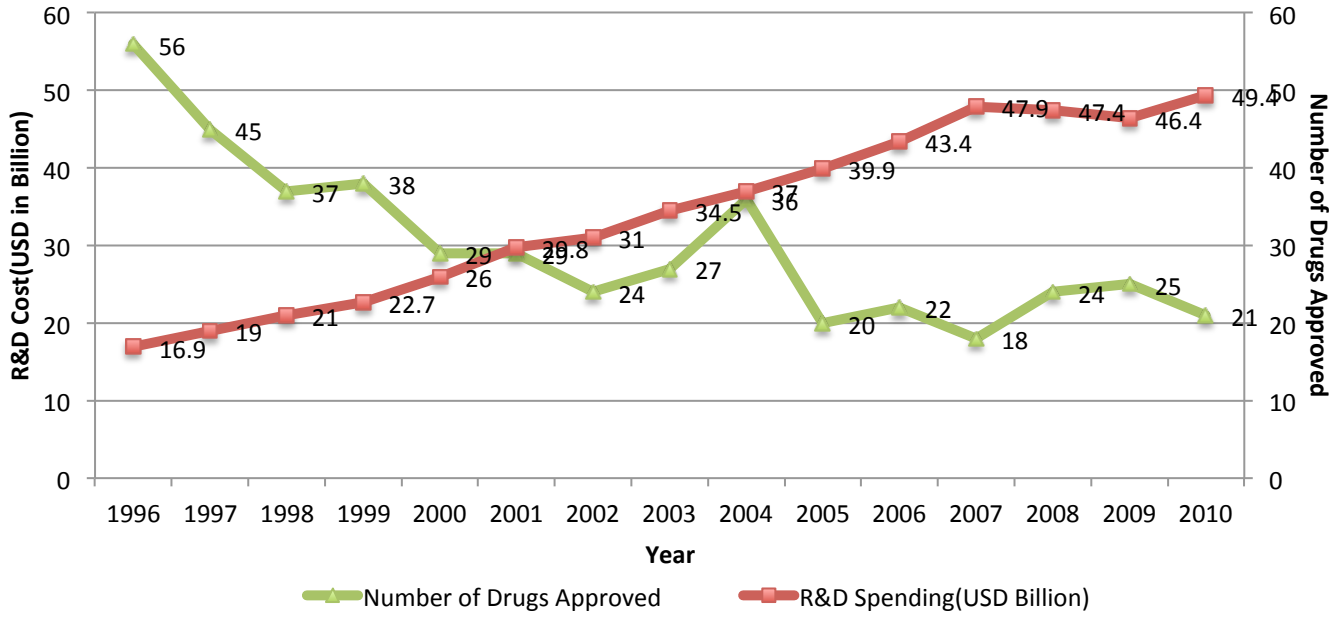




FDA. (2011). Orange Book. Retrieved from http://www.fda.gov/Drugs/InformationOnDrugs/ucm129662.htm

Problems: Innovation Gap

- **R&D spending rising**:
  - almost $50 billion a year in 2010
- **Number of drugs constantly decreased:**
  - staying around 20 from 2005 to now.



Mullard, A. (2011). 2010 FDA drug approvals. *Nature Reviews Drug Discovery , 10*, 82-85.

PhRMA. (2007, Feb). *innovation.org.* Retrieved Nov 2011, from Drug Discovery and Development: http://www.phrma.org/sites/default/files/159/rd_brochure_022307.pdf

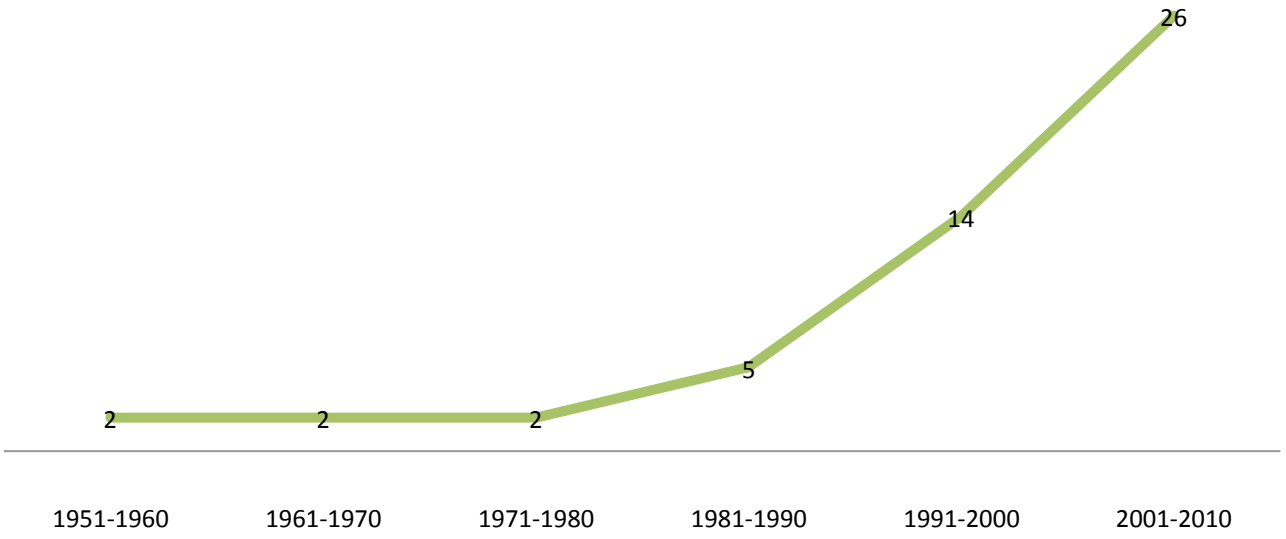
Problems: Inefficiency in Drug Development in Detail

- Out of 1,000 drugs, one drug goes into clinical trial
- 66% of drugs fail in Phase III.
- Average cost in developing drugs: $0.8-1 billion
- Average time for development of drugs: 10-15 years

Mehta, S. S. (2011). *Commercializing Successful Biomedical Technologies.* Cambridge: Cambridge University Press.
PhRMA. (2007, Feb). *innovation.org.* Retrieved Nov 2011, from Drug Discovery and Development: http://www.phrma.org/sites/default/files/159/rd_brochure_022307.pdf

Problems: Drug Withdraw

Number of the Drugs withdrawn from the market



Loss from withdraw is more than just loss in Sales

e.g. While their sales were $2.5 Billions, Merck experienced $28.8 billion decrease in market value in 2004 by withdrawing Vioxx from the market.

Rotthoff, W. K. (2007, May). *Clemson Libraries.* Retrieved Nov 2011, from Product Liability Litigation: An Issure of Merck and Lawsuits over Vioxx: http://etd.lib.clemson.edu/documents/1181251809/umi-clemson-1184.pdf

Why are we having this difficulty?

Even though we have large amounts of research generated everyday, researchers are unable to efficiently convert this into new drugs.

Issues: Data Explosion

The Amount of Data Have Increased Dramatically

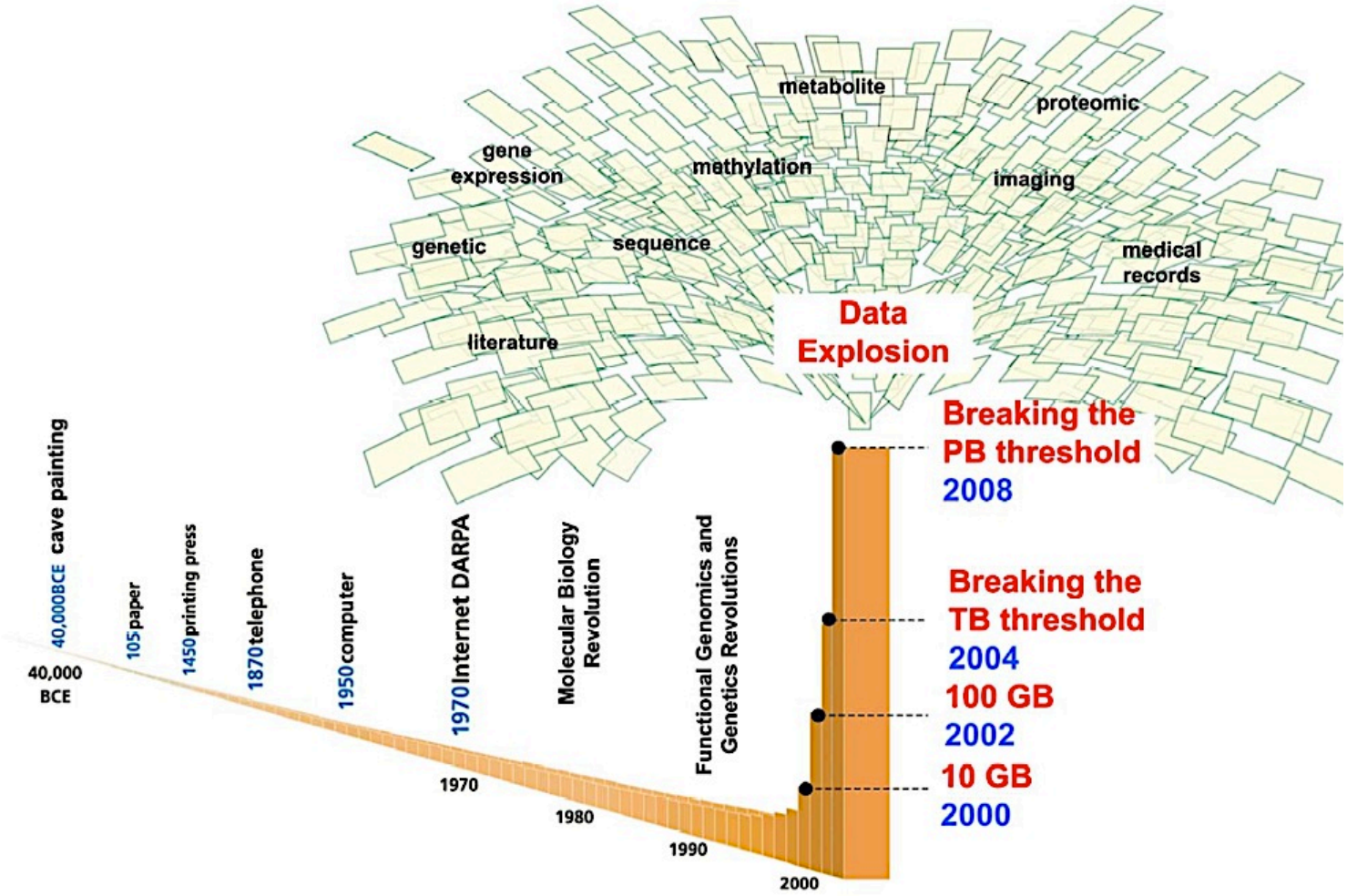
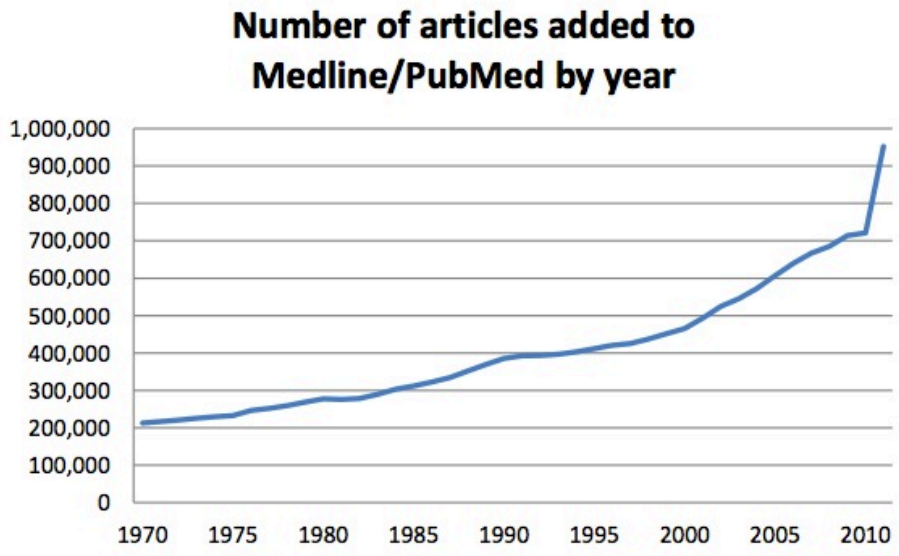


Image: Friend, S. (2010). Case Study: Sage Bionetworks. Retrieved from http://sagebase.org/info/NewsInfoDownloads/WellcomeFriend0617.pdf

Issues: Data Explosion

- MEDLINE and PUBMED are currently growing at an annual rate of approximately 5%, adding an average of 2,900 new records daily as of 2011.
- By year 2000 “rough draft‟ of human genome was completed and this led to huge volume of bio-data accumulation.



U.S. National Library of Medicine. (2011). Statistical Reports on MEDLINE/PubMed Baseline Data. Retrieved from http://www.nlm.nih.gov/bsd/licensee/baselinestats.html

Obstacles in Data Integration

1. **Data are in different sources**
   - Many different databases: Bimolecular Interaction Network Database (BIND) , Kyoto Encyclopedia of Genes and Genomes (KEGG), etc.
   - Some of the data are even unreliable
2. **Data are unstructured**
   - Published Literature, Patent Information
3. **Data are unavailable in e-format**
   - Lab Notes, Some failure test reports

Cai, J. (2011, November 18). Phone Interview

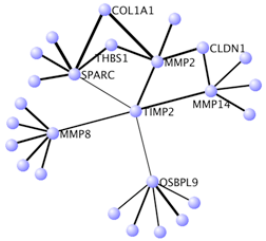
Issues: Biological Complexity of Disease Pathways

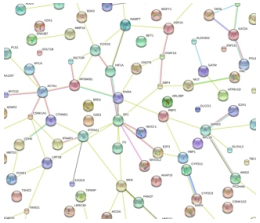
Disease is a highly complex biological process

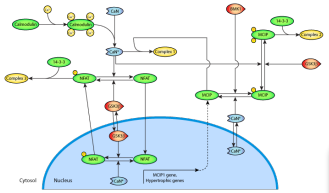
1. Lack of information on the structure of target molecule and its function
2. Lack of knowledge in target's mechanism of action(MOA) on the system view
3. Some disease related targets are non-drugable.
4. Lack of information sharing
   - Chasm between different disciplines
   - Have no access to all of the previous research result

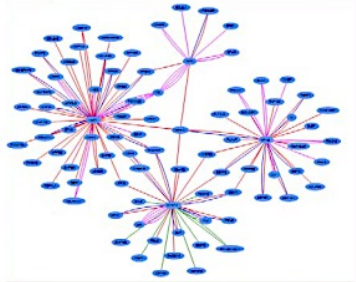
Nicklas, B. (2009). Relieving the first bottleneck in the drug discovery pipeline: using array technologies to rationalize membrane protein production. Vol. 6. p 501-505
Ausiello, D. (2010). The Complexity of Drug Discovery – New Models for the Future. Retrieved from http://www.itmat.upenn.edu/symposium_2010/slides/ITMAT_HealthCareIndustryEconomics10_20_2.pdf
Klein, F. (2008). Bottlenecks in Drug Discovery. Bio-IT World Magazine. Retrieved from http://www.bio-itworld.com/issues/2008/oct/image-analysis.html

Issues: Biological Complexity of Disease Pathways









Genomics Network

Non-coding RNA Network

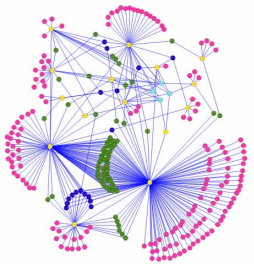
Signaling Network

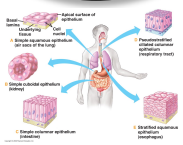
Protein Network

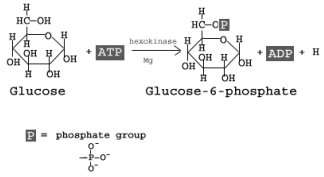
Transcriptional Network

A Single Issue

Phosphorylation

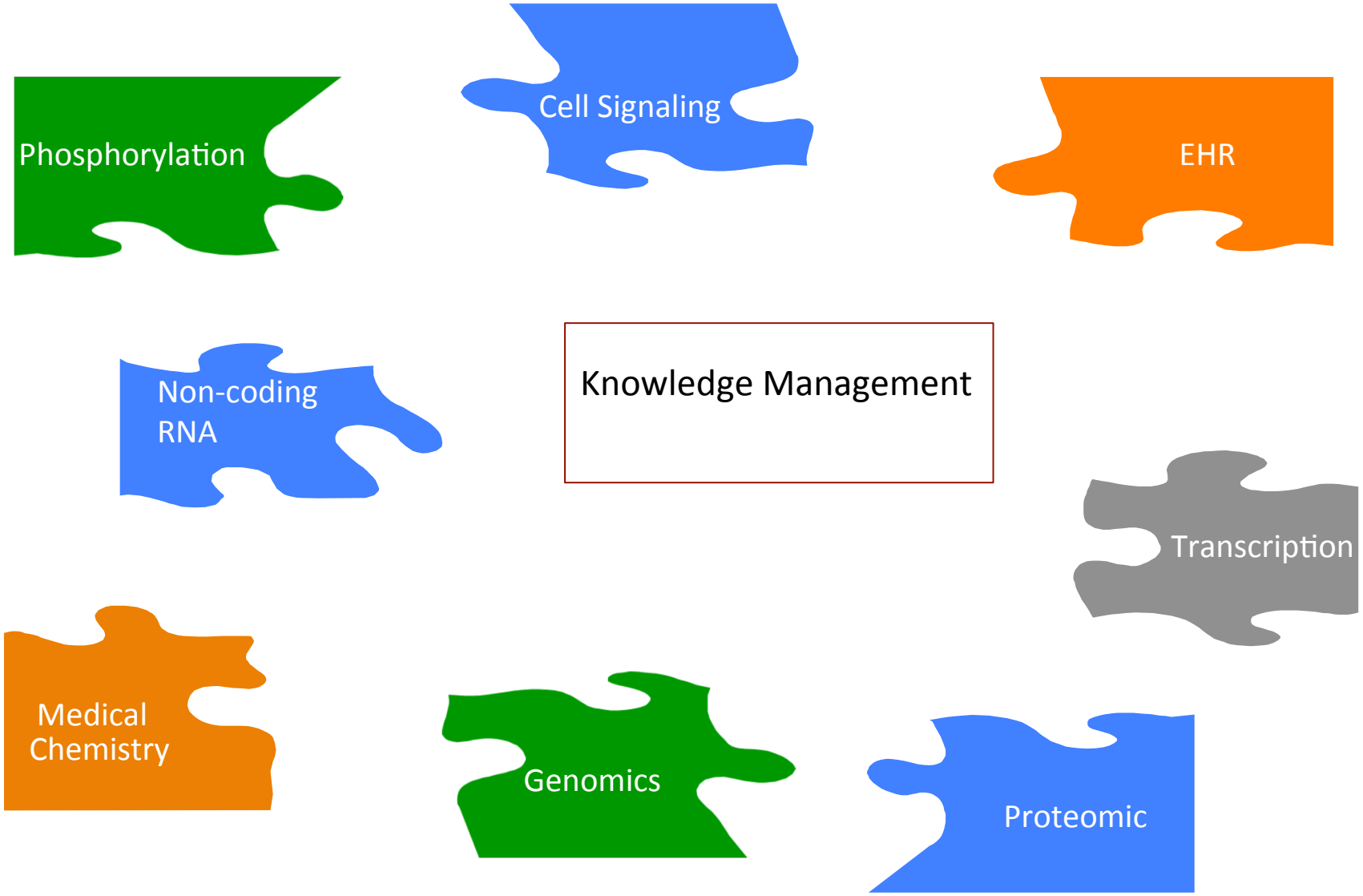




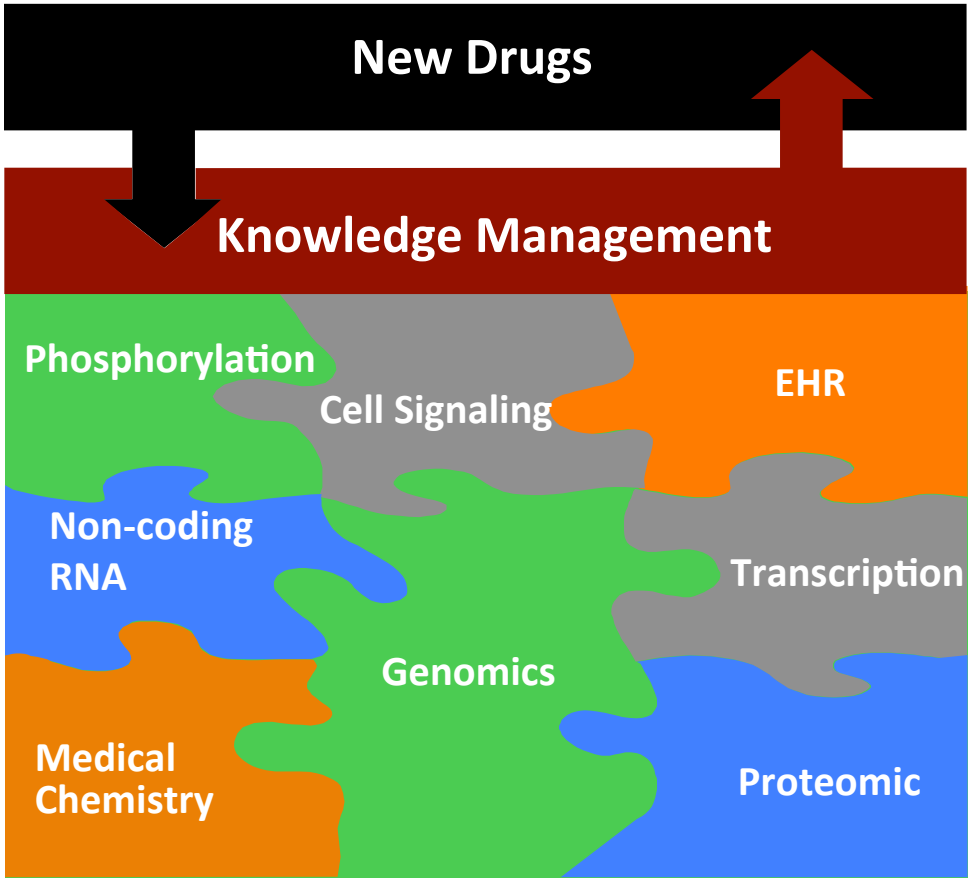


Friend, S. (2010). Case Study: Sage Bionetworks. Retrieved from http://sagebase.org/info/NewsInfoDownloads/WellcomeFriend0617.pdf

Why Knowledge Management is Needed



Why Knowledge Management is Needed



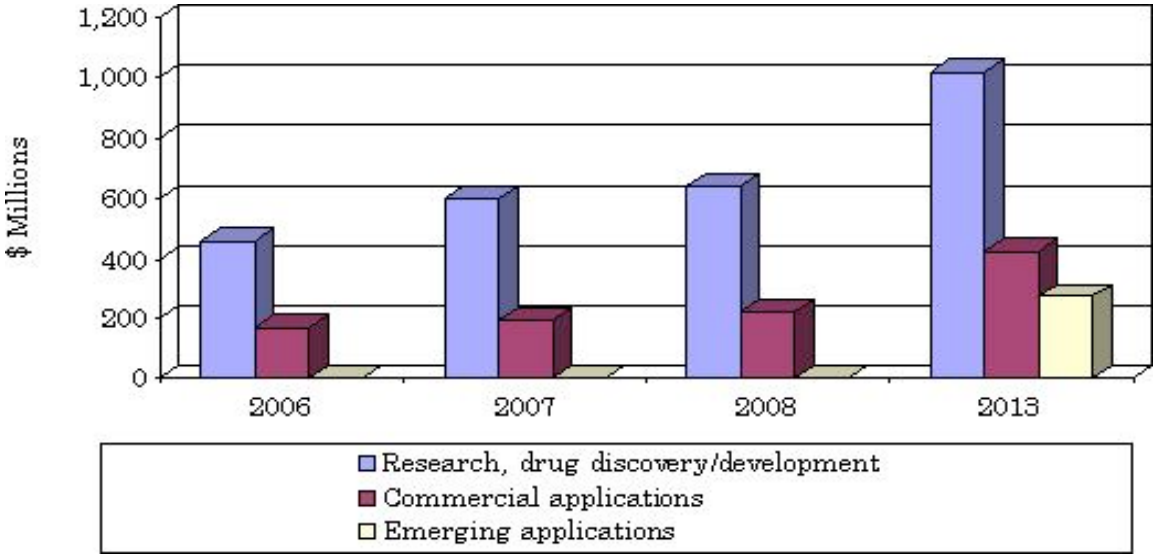
New Data Sources: Gene Sequencing Technology is Growing.

- **Competing Companies**

  ComHelicos Biosciences, Pacific Biosciences, Complete Genomics, Illumina, Sequenom, ION Torrent Systems, Halcyon Molecular, NABsys, IBM, and GE Global are now all competing.

- **Market**

  $794.1 million in 2007. This is expected to reach $862.5 million in 2008 and $1.7 billion in 2013, a compound annual growth rate (CAGR) of 14.7%.



Bcc Research. (2008). BCC Research Reports on DNA Sequencing Market

New Data Sources: Price Reduction of Gene Sequencing Technology

- **Revolution in the sequencing Technology**
  - As the sequencing technology evolved to the next generation, the price dropped significantly starting from 2008
- **Cheaper Gene Sequencing**

  ➔ **Increase in number of Genomic data**

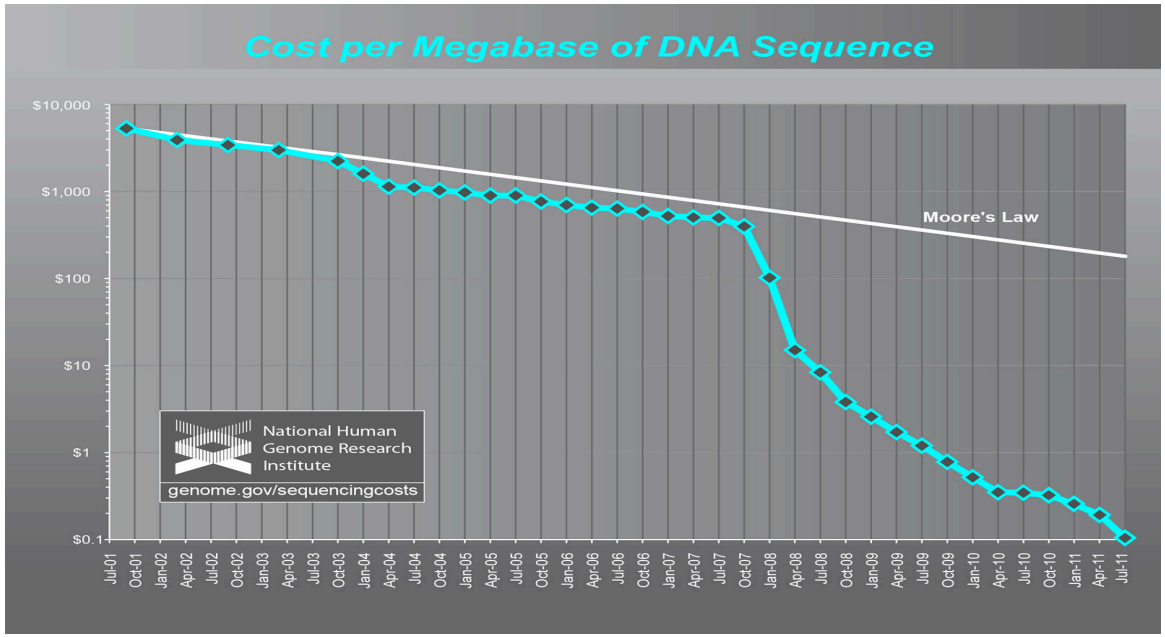
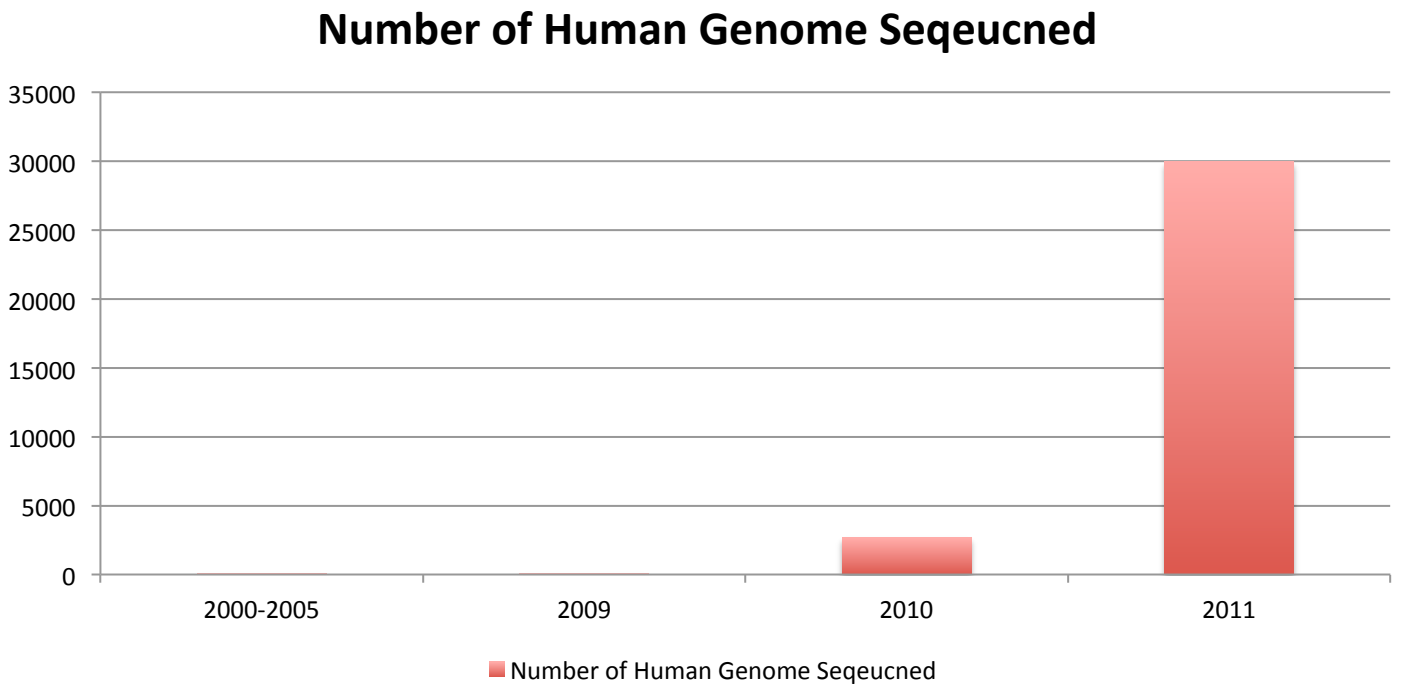


Image: Saenz, A. (2011). Costs of DNA Sequencing Failing Fast – Look At These Graphs!. Retrieved from http://singularityhub.com/2011/03/05/costs-of-dna-sequencing-falling-fast-look-at-these-graphs/

New Data Sources: Increase in Genomic Data



By 2020 :

It is expected that every single human genome to be sequenced in a twelve-month stretch.

Nature Survey. (2010). Genomes by the thousand. Nature. Vol. 467. p 1026-1027. Retrieved from http://www.nature.com/news/2010/101027/pdf/4671026a.pdf

New Data Sources: Increase in Genomic Data

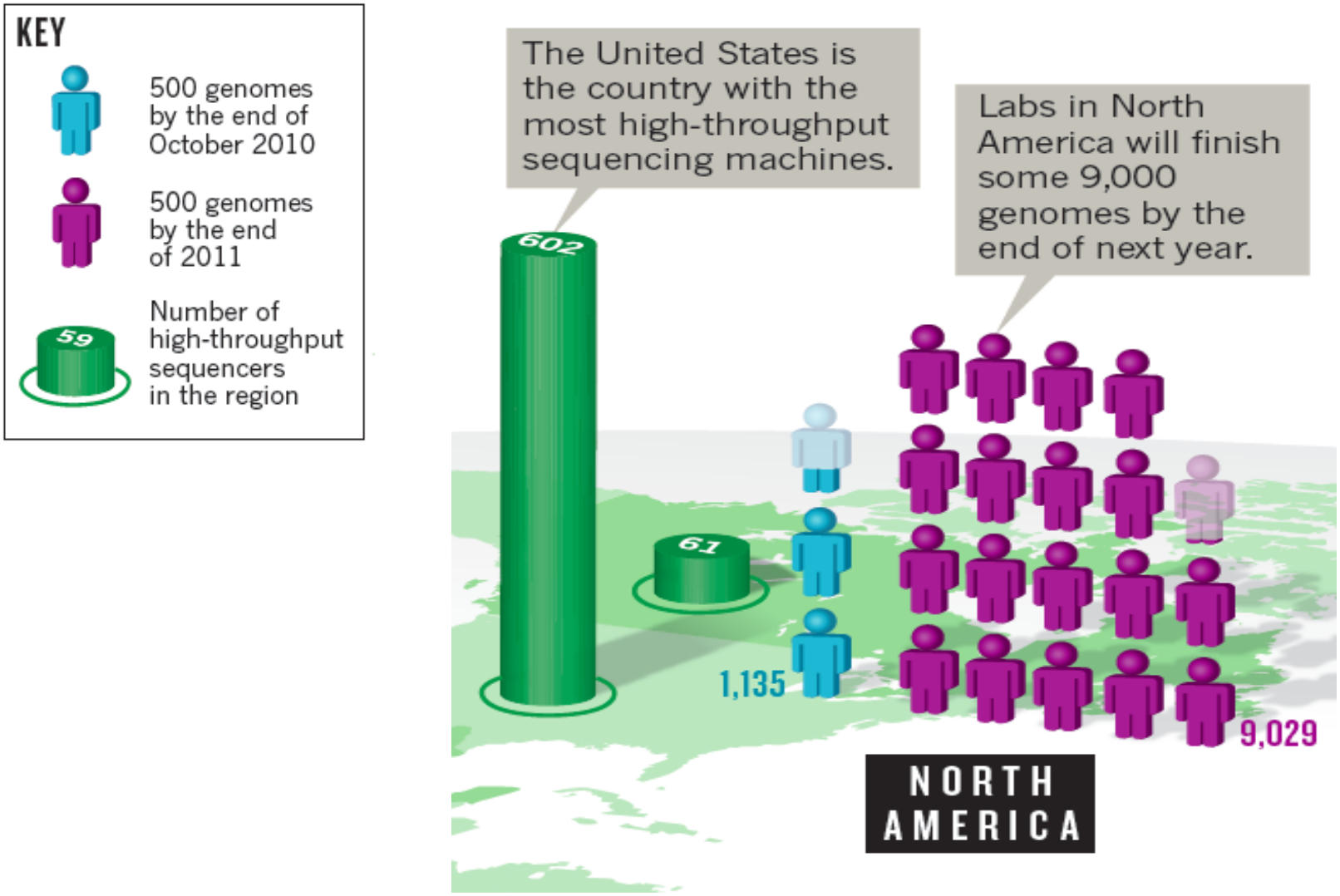


Image: Nature Survey. (2010). Genomes by the thousand. Nature. Vol. 467. p 1026-1027. Retrieved from http://www.nature.com/news/2010/101027/pdf/4671026a.pdf

New Data Source: Increase in Genomic Data

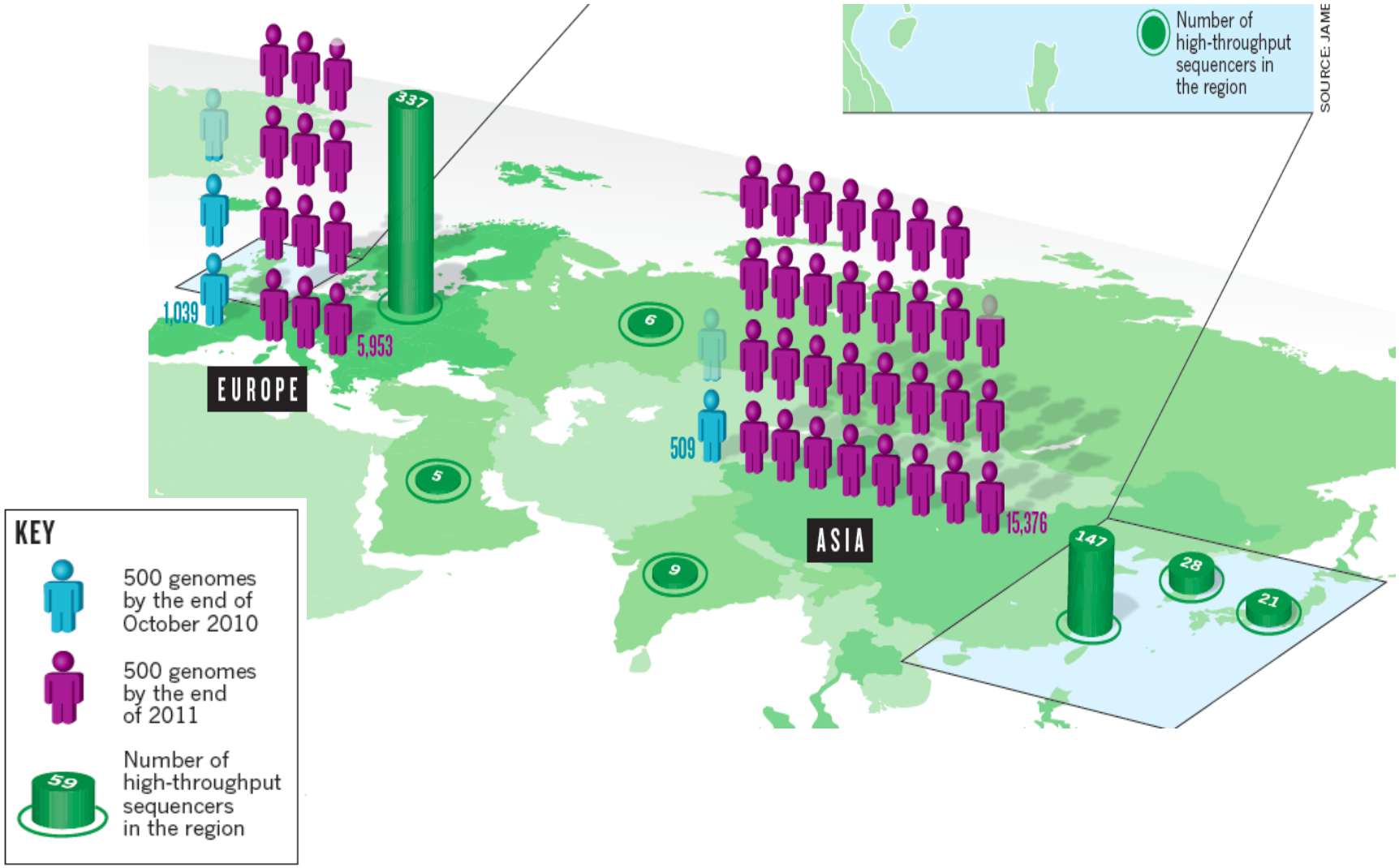


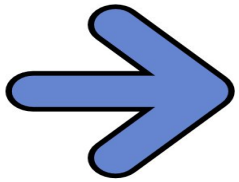
Image: Nature Survey. (2010). Genomes by the thousand. Nature. Vol. 467. p 1026-1027. Retrieved from http://www.nature.com/news/2010/101027/pdf/4671026a.pdf

Genomic Data & EHR in Knowledge Management

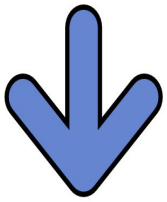
- **Comparative analysis**
  - Combined data source of genome & medical record will enable richer genome analysis
- **Treatment**
  - With more data, predictive medicines and lifestyle management are possible
- **Prevention of disease side effects**
  - Genome and medical record can stratify patient groups to increase efficacy and reduce side effects
- **Drug development**
  - Deeper understanding of disease unveils the disease pathway

New Data Sources: Electronic Healthcare Records

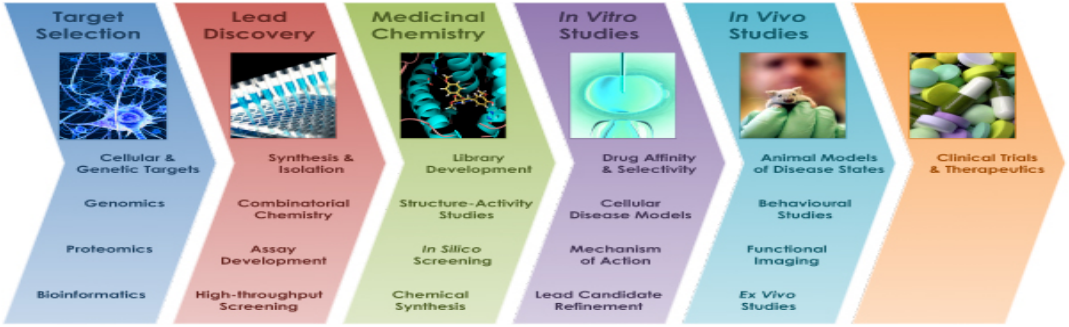
Electronic Healthcare Records



Patient Data



Drug Discovery Process



New Data Sources: Trend of EHR

Will EHR prevail ?

Support

EHR has so many advantages that can improve the health care service

Government incentive program:
up to $2 millions for single hospital
or up to $63,750 to individual professional

Government incentive program:
up $3.2 millions loss in Medicare reimbursement for single hospital after 2017 if failing in showing meaningful use of EHR

Oppose

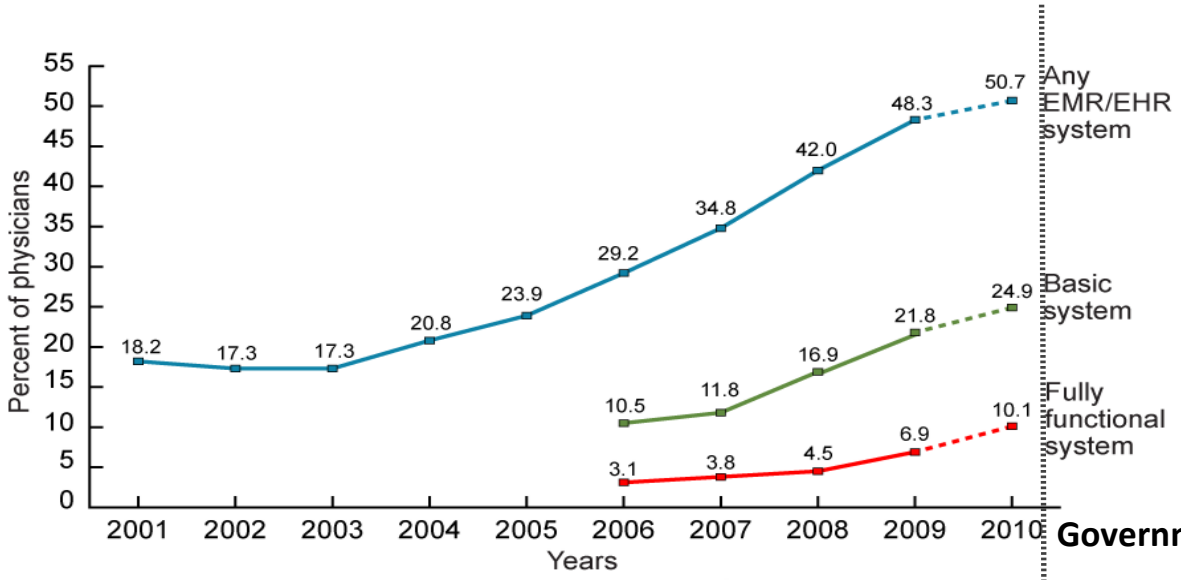
Substantial learning curves for end-users

Startup and maintenance cost can be excessive ($123,750 to $225,000 for three-physician in 2 years after implementation)

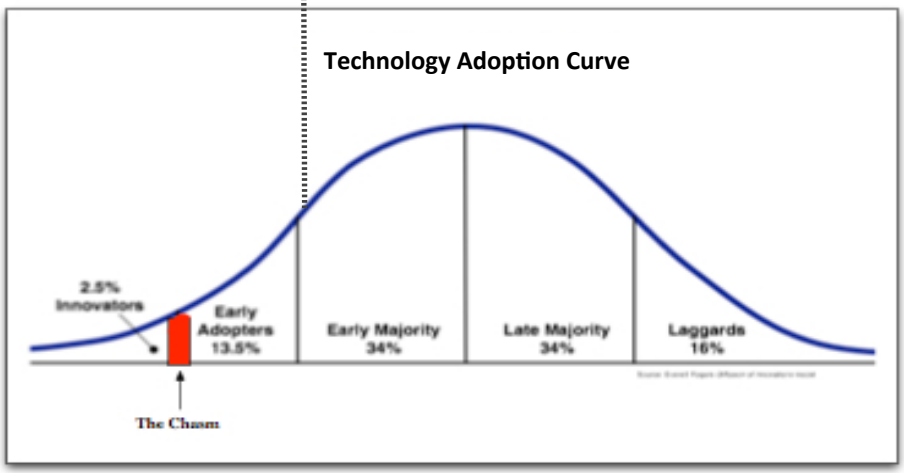
Price Water House Coopers. (2009). Rock and a Hard Place. Retrieved from http://www.cabarruschamber.org/clientuploads/PWC_HealthIndustries_ElectronicHealthRecords.pdf

New Data Sources: Trend of EHR

Percentage of office-based physicians with EHR :
United States, 2001-2009 and preliminary 2010



Government Incentive Program starts at 2011



What Do We Need for Knowledge Management?

We need **Data Extraction** and **Data Integration** technologies to leverage existing data sources and emerging data sources (e.g. Electronic Healthcare Records and DNA Sequencing Data)

Data Extraction **Data Integration**

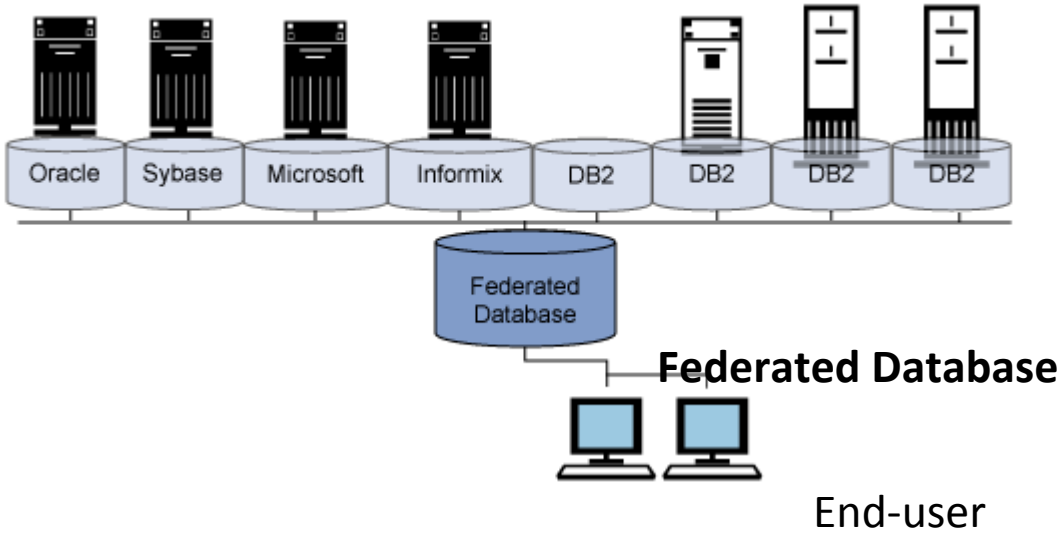
Current Data Integration Technologies: Data Warehouse

Data Warehouse

Data Warehouse Rationale

1.Data warehouse extracts data from transaction systems and store it into data warehouse
2.Data from the transaction systems is analyzed, “cleaned”, and restructured
3.The formatted files are then transferred and loaded into the warehouse
4.The data is then available for analysis and reports

Current Data Integration Technologies: Federated Database



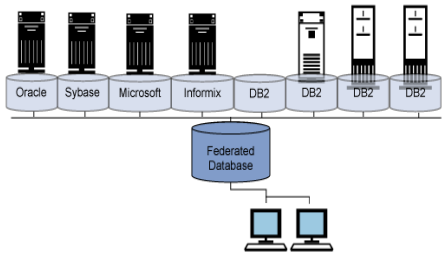
Federated Database Rationale

1.End-user submits a single query
2.Federated DB system decomposes the query into sub-queries
3.Data is retrieved from heterogeneous databases management systems (DBMS)
4.Result sets are merged and delivered to end-user

Current Data Integration Technologies: Comparison

Data Warehouse

Federated Database



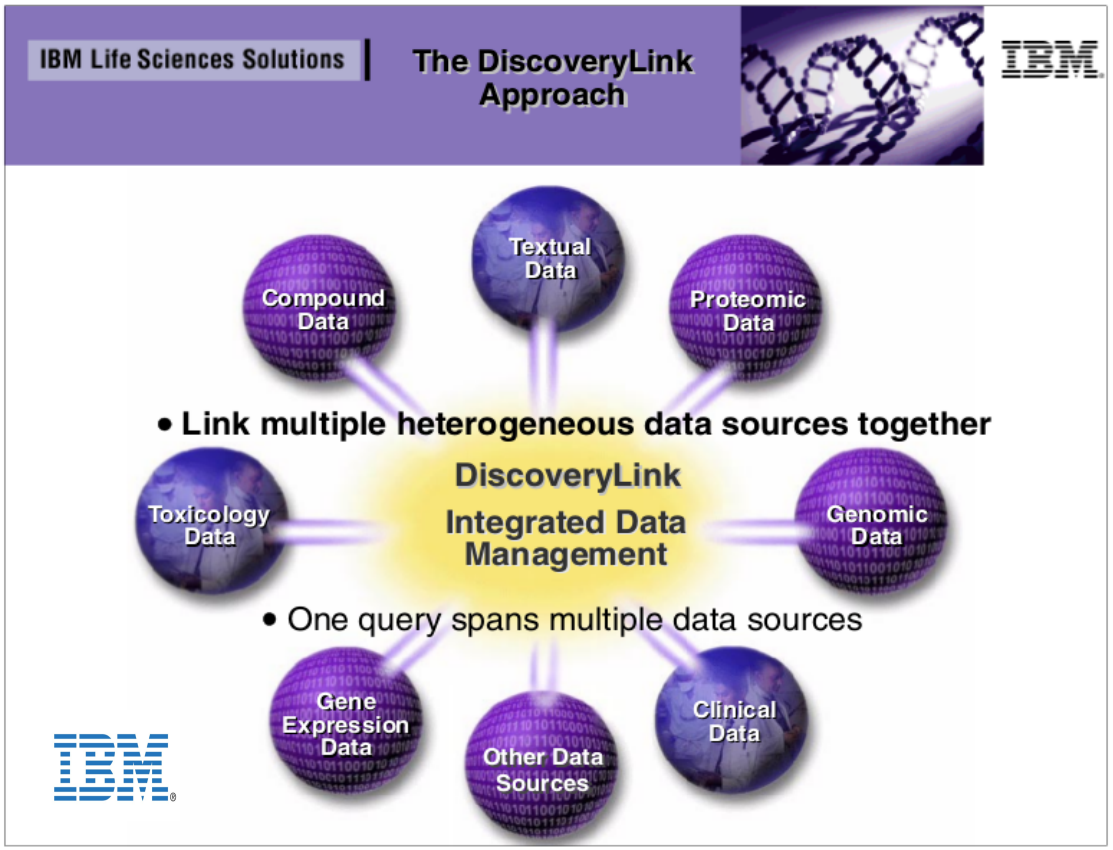
Difficult to build and expensive to maintain	✗ 👍	Much less effort is required to develop and support
Data not exacted into data warehouse cannot be queried	✗ 👍	Data remains in original sources and data is always as current as original data source
Eliminate network bottlenecks, low response times, and temporarily unavailable source	👍 ✗	Sensitive to network bottleneck and low response time and temporarily unavailable sources

👍 **Common Strengths:**
Comparing to single DBMS, data warehouse and federated database can link to various data sources rather than just single one.

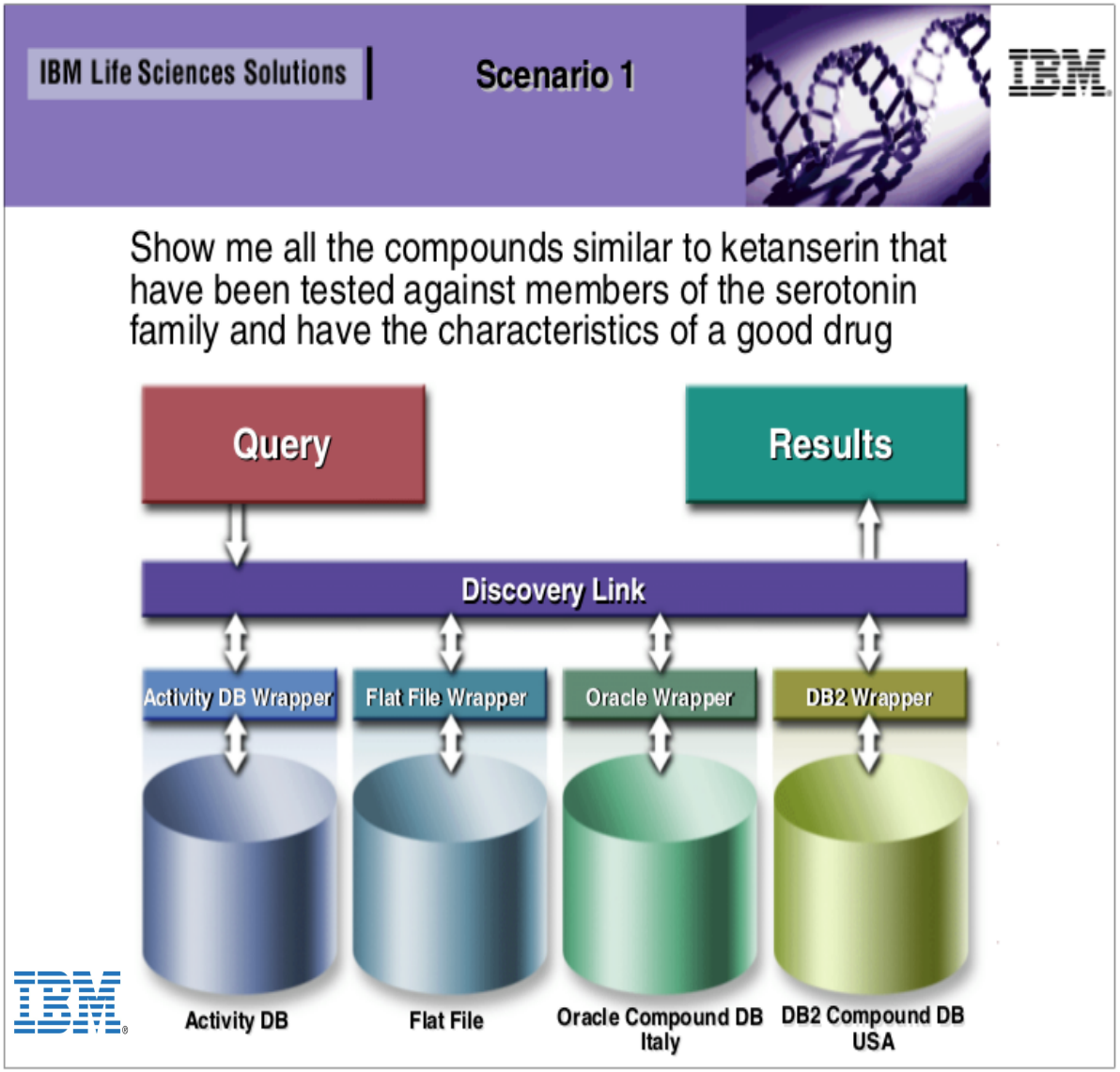
✗ **Common Issues:**
Both of them do not have or have very limited capability to integrate unstructured and semi-structured data with heterogeneous forms.
They do not contain semantic knowledge and usually fail to capture the detail and richness of relationships between concepts in multiple sources.

DiscoveryLink

IBM DiscoveryLink can link multiple heterogeneous data sources as the data integration and management solution for Life Sciences.



DiscoveryLink



Strength

Integrate data from heterogeneous data formats

Weakness

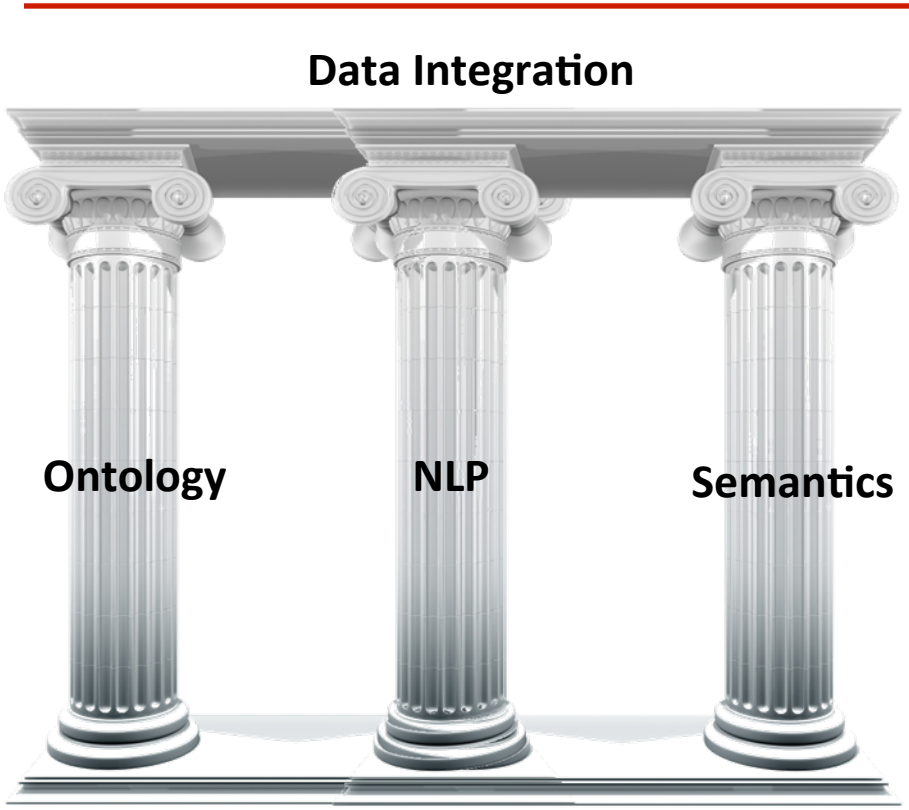
It cannot solve the problems of semantic integration

Current Data Integration Technologies: Issues

Can we link up all the data with heterogeneous formats?

Even after we link up all the data, can we generate valuable knowledge from pieces of information?

Future Knowledge Management Approaches



Open PHACT

DiscoveryLink

Semantic Web Technology

Concept Web Alliance

Question Answering Systems

Lilly Singapore Center for Drug Discovery Data Architecture Project

Open PHACT
(Open Pharmacological Concepts Triple Store)

Scenarios

- **In 2 years**
  - <$700 DNA sequencing technology
  - Nutrition and prescription customization based on genetic information, past medical history
  - Gene sequencing would start to take portion of microarray chips market away
  - Coverage of EMR increase rapidly because of incentive programs and penalties
  - M&A will continue to happen
  - Drug discovery processes are going to be outsourced more
  - Launch QA 2.0 – early adoption age

- **In 5 years**
  - DNA data will start to be integrated with EHR
  - Stratified medicine and personalized medicine becomes more popular
  - QA 2.0 will keep growing along the technology adoption curve

- **In 10 years**
  - High computing power which will enable mobile medical monitoring
  - Personalized medicine and preventive medicine will become pervasive
  - Gene therapies- even in the conception level
  - Knowledge management technologies will revolutionize pharmaceutical industry