Surveillance from Clinical Data: Tradition, Standards, Text

FDA Sentinel Network Public Meeting

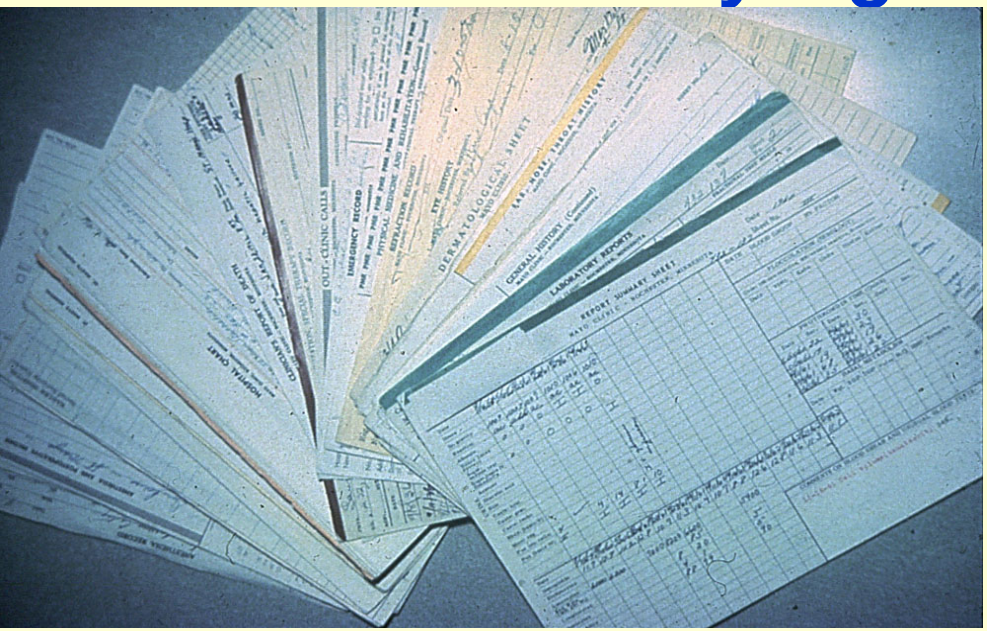
7 March 2007

CG Chute, MD DrPH

Chair, Biomedical Informatics

Mayo Clinic College of Medicine

Mayo: A Century-Long Tradition of Studying Patient Outcomes

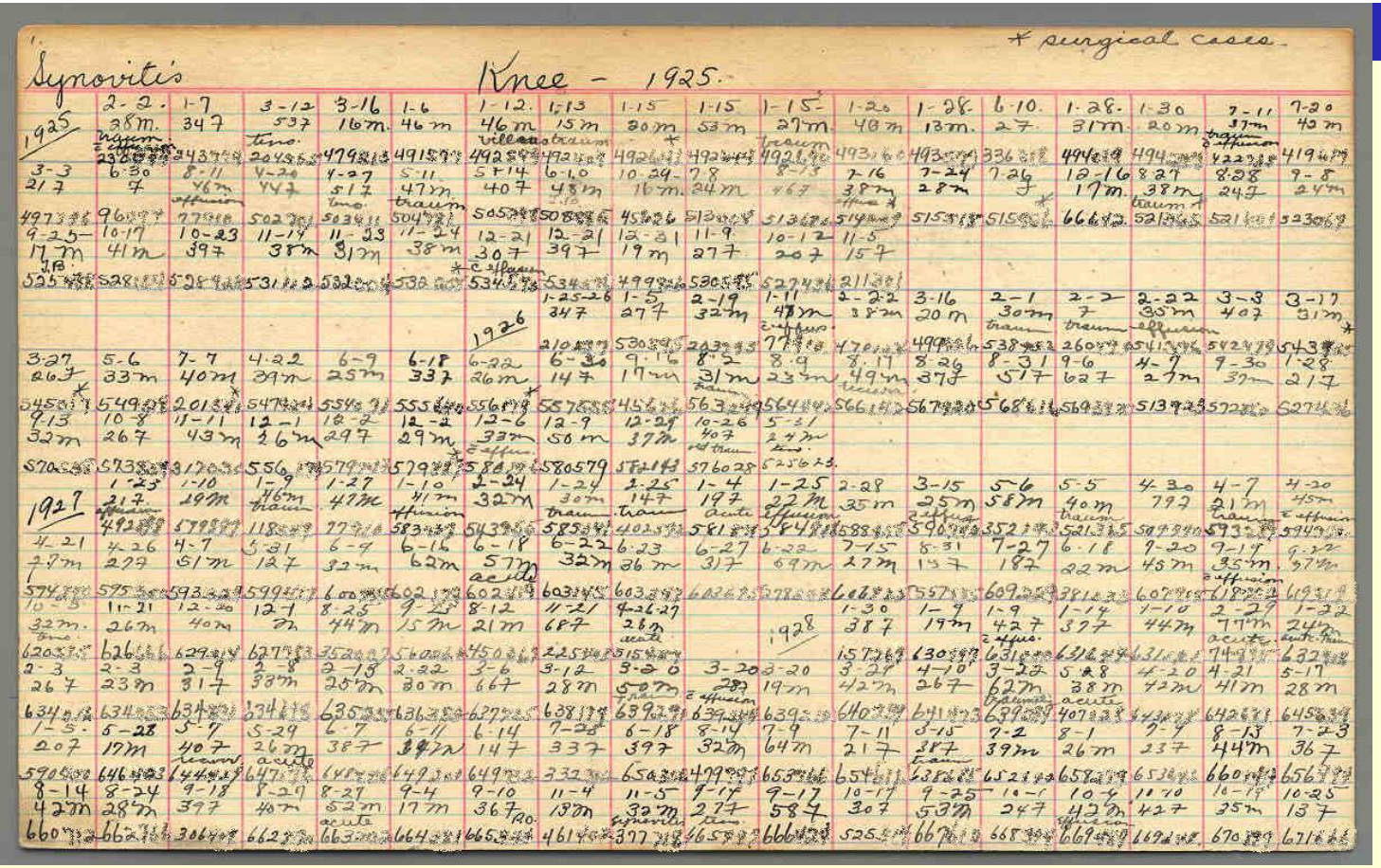


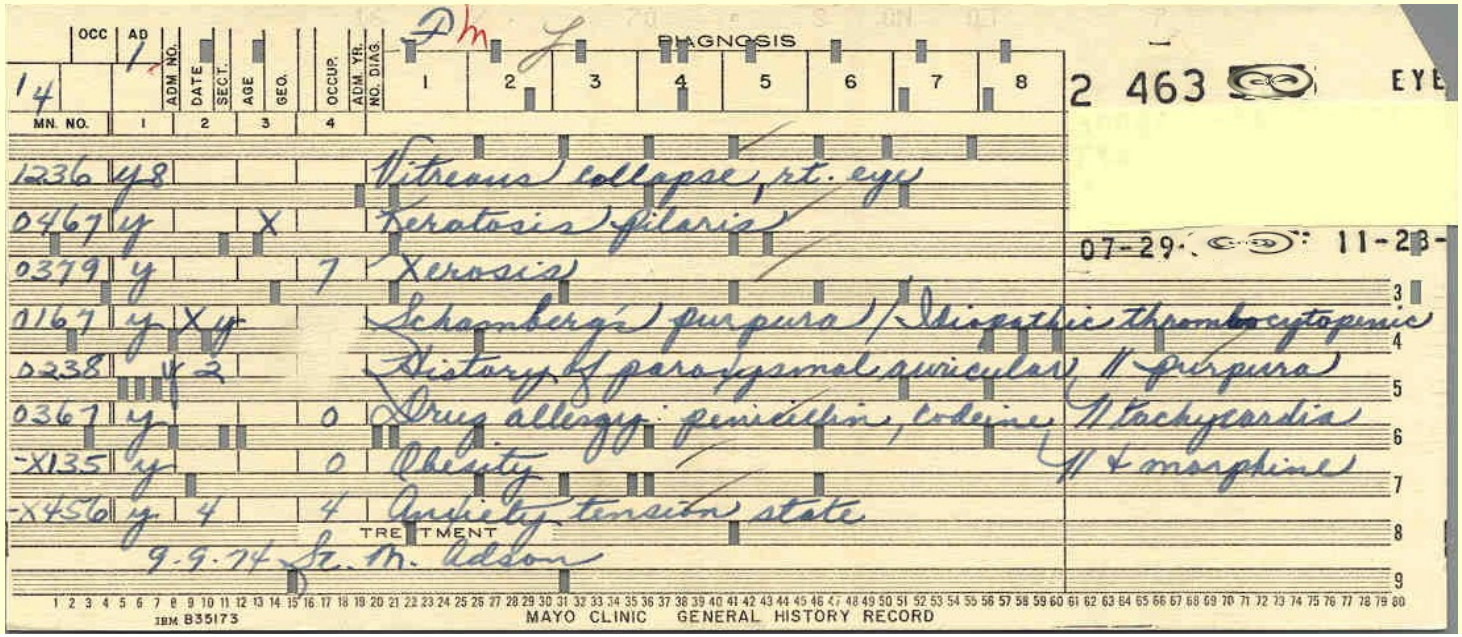
Demographics
Diagnoses
Procedures
Narratives
Laboratories
Pathology…

High-Volume Data Storage

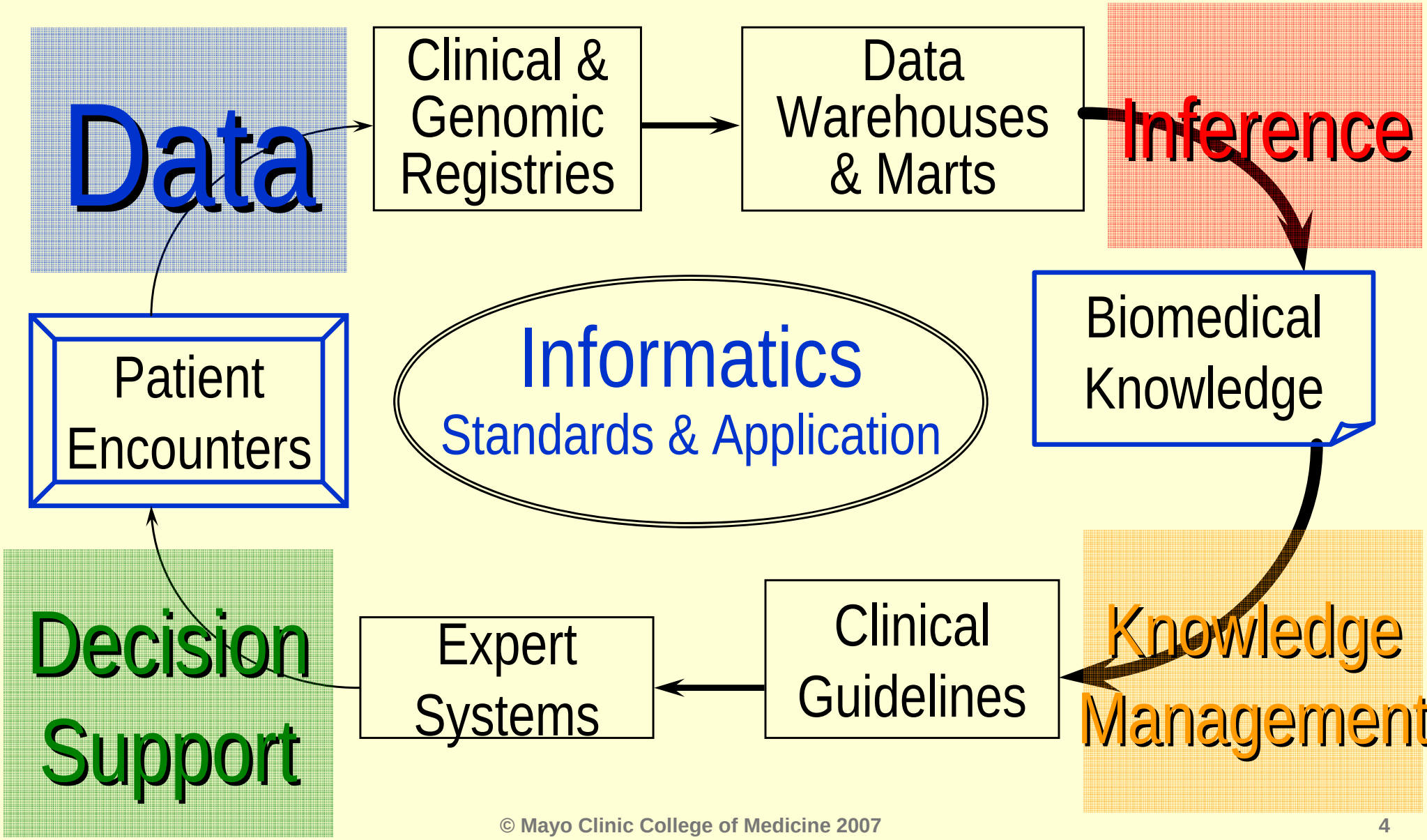
© Mayo Clinic College of Medicine 2007

Semantic Organization





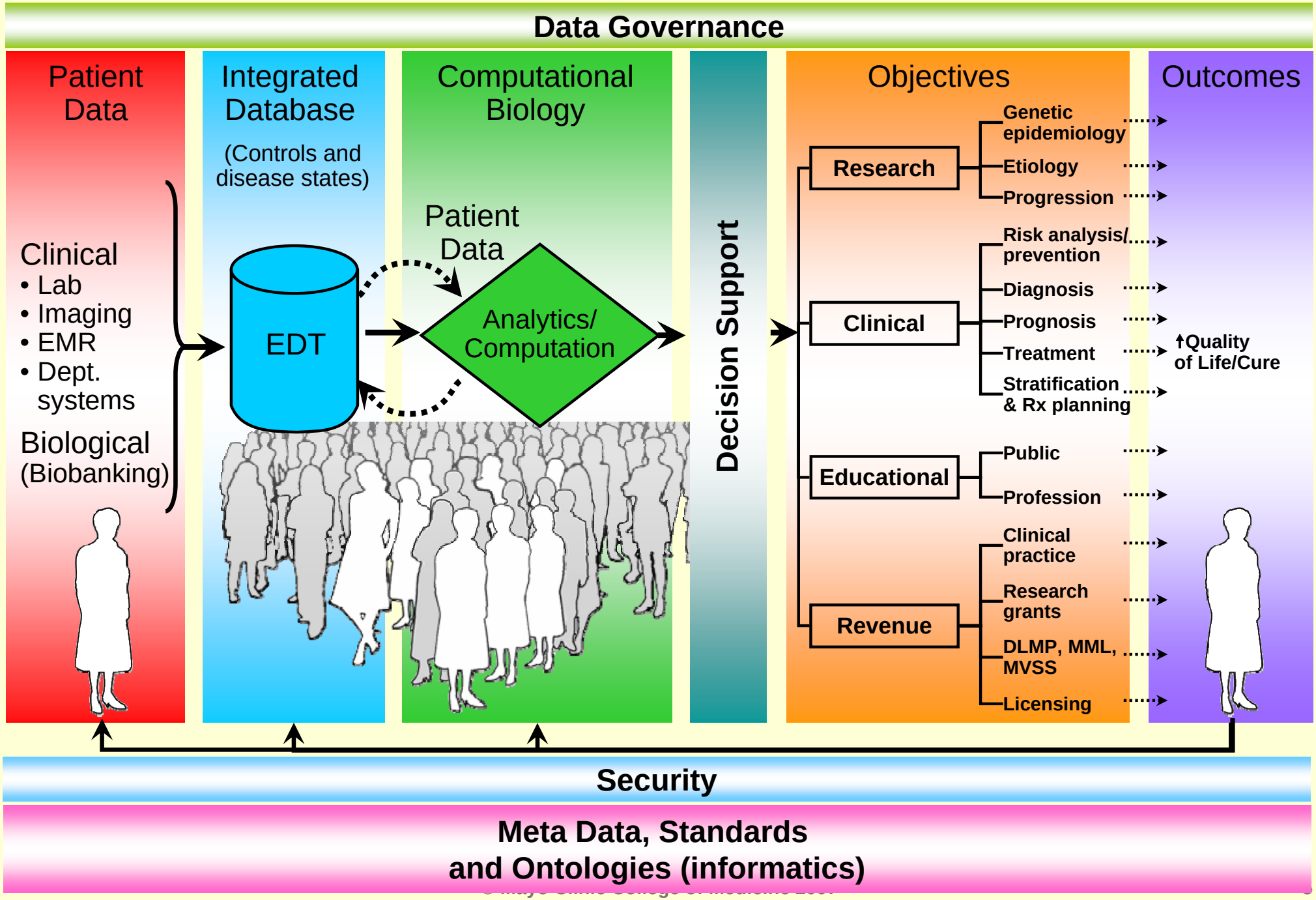
From Practice-based Evidence to Evidence-based Practice



© Mayo Clinic College of Medicine 2007

MAYO CLINIC
Biomedical Informatics

Informatics & Standards Foundations



Features of the Mayo-IBM Data Trust

- Formal ETL (Extract-Transform-Load) clinical and genomic data across Mayo enterprise
  - Achieved for core subset of clinical data in MN
  - Data "pipeline" from research cores in place
  - Scope is Mayo's National Network: MN, FL, AZ
- Plans to extend to regional practice in 5-state area around Rochester, MN
- Novel component of sophisticated NLP (Natural Language Processing)
  - Currently ~20M clinical notes

© Mayo Clinic College of Medicine 2007

NLP (Natural Language Processing)

- Adjunct to “medications list”
  - Parse drug names (RxNorm)
  - Experimental status mentions (start, stop, change)
- Detailed diagnostic and event data
  - Entities tagged with SNOMED codes
  - Far richer than reimbursement codes
- Capacity for longitudinal tracking below “ICD” radar
- Capacity for association studies

© Mayo Clinic College of Medicine 2007

NLP Infrastructure

- Built on IBM's UIMA (Unstructured Information Management Architecture)
  - Now Open-Source (Eclipse)
  - Third general design in open standards (OASIS)
- Current version ~16 annotators in UIMA pipeline
  - Tokenizer, part of speech, phraser, coders, etc.
  - Most are open source, *all can be*
- Potential for shared, open-source NLP in clinical space

© Mayo Clinic College of Medicine 2007

The Genomic Era

- The genomic transformation of medicine far exceeds the introduction of antibiotics and aseptic surgery
- The binding of genomic biology and clinical medicine will accelerate
- The implications for *shared semantics* across the basic science and clinical communities are unprecedented

© Mayo Clinic College of Medicine 2007

Pharmacogenetics Research Network (PGRN)

(PGRN) is a nationwide collaboration of scientists studying the effect of genes on people's responses to a wide variety of medicines.

The following institutes contribute support to the Pharmacogenetics Research Network:

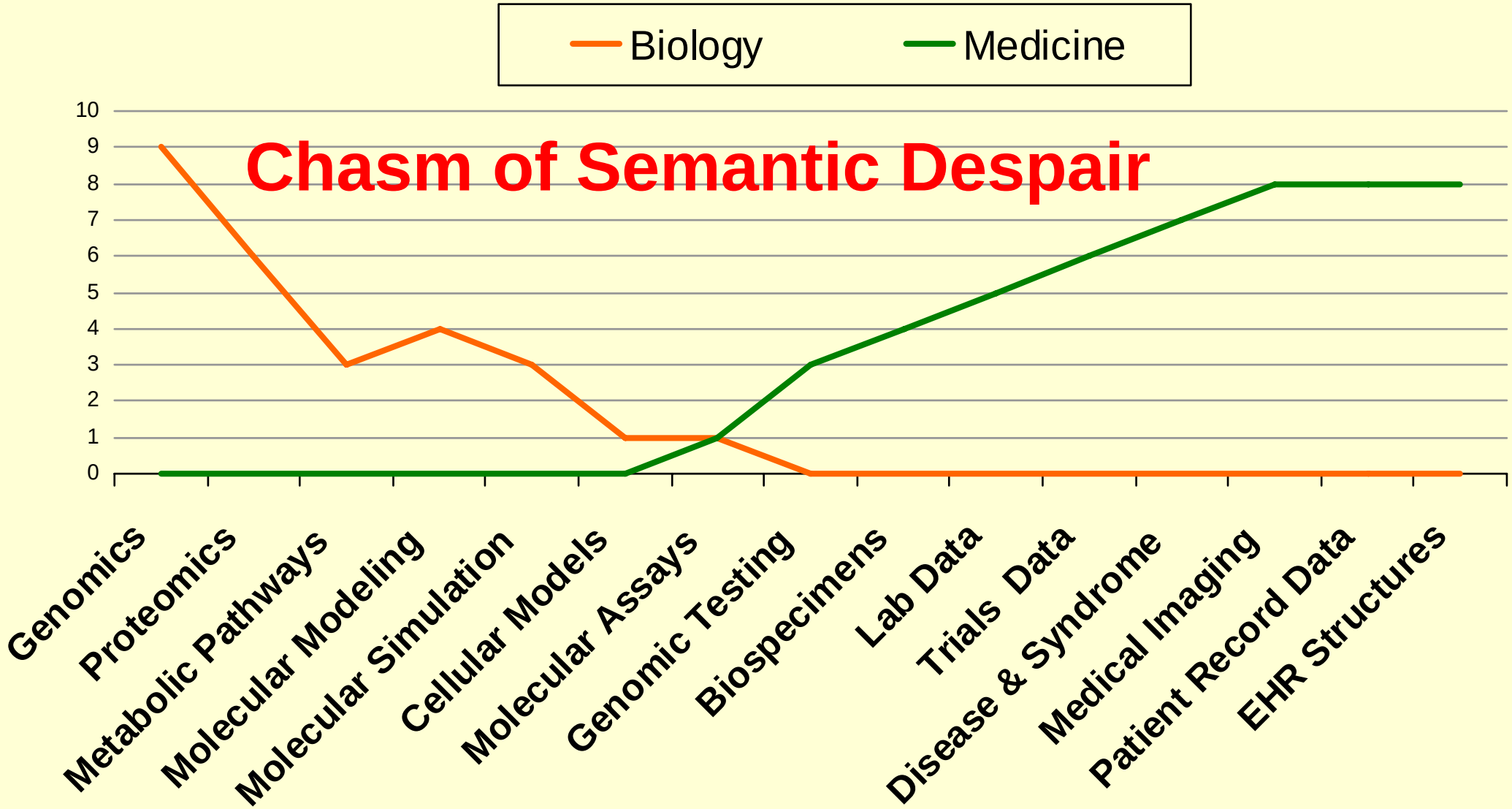
NIGMS, NHLBI, NCI, NIDA, NIEHS, NHGRI, NLM, NIMH, ORWH

PharmGKB

The Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB) is an integrated knowledge base for pharmacogenetics linking phenotypes and genotypes.

© Mayo Clinic College of Medicine 2007

Communities of Scientific Endeavor
Silos, Continuity, Synergy



© Mayo Clinic College of Medicine 2007

NIH CTSA
Clinical and Translational Science Awards

- A national consortium that will transform how clinical and translational research is conducted
- Ultimately enabling researchers to provide new treatments more efficiently and quickly to patients.
- Today 12 academic health centers
  - 52 planning grants
- Anticipated by 2012, about 60 institutions
- Re-engineer the clinical research enterprise, one of the key objectives of the NIH Roadmap

© Mayo Clinic College of Medicine 2007

CTSA and Sentinel Surveillance

What can or should the CTSA network undertake that cannot be easily done outside the network?

- Promote shared standards for clinical repository design and implementation
  - Models, vocabularies, ontologies
  - Support aggregation and surveillance across the network
- Deploy and enhance NLP resources
  - Funding augmentation for purpose
- Leverage pharmacogenomics agenda
  - Couple events/non-events with genotype

© Mayo Clinic College of Medicine 2007