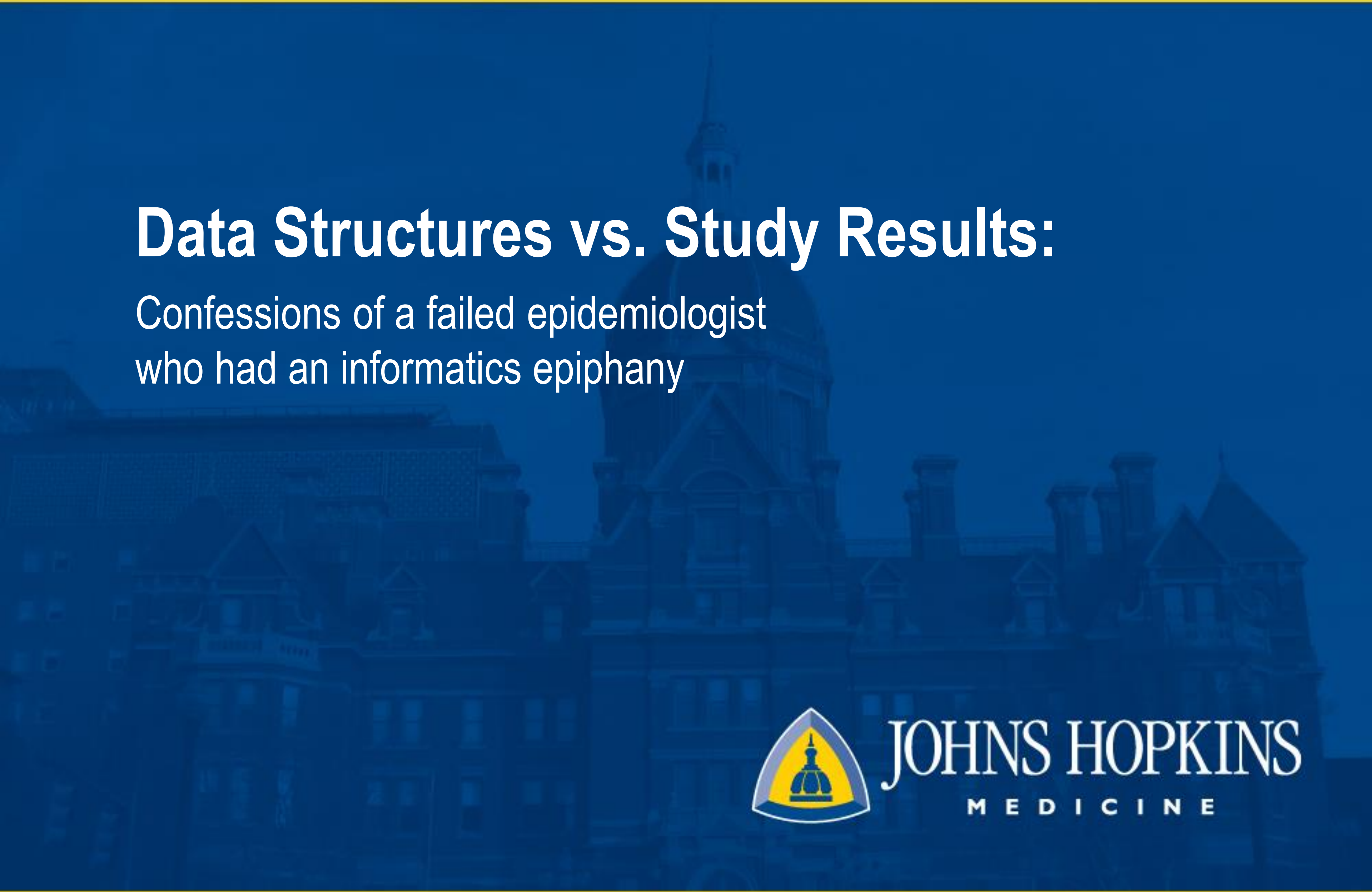


Data Structures vs. Study Results:

Confessions of a failed epidemiologist who had an informatics epiphany

JOHNS HOPKINS
MEDICINE

CG Chute, MD DrPH, Bloomberg Distinguished Professor of Health Informatics

April 7, 2015

Chris's story

- Recognized clinical training as apprenticeship
  - Folklore and anecdote
- Sought methodology training for outcomes research
  - Had not yet heard of "evidence-based medicine"
- DrPH in epidemiology and biostatistics—but no data
- Sought informatics; discovered that data was junk
  - No comparability or consistency, no standards
- Established career in clinical data representation

Where did my training go wrong?

- Set up “Health Professionals’ Follow-up Study” as graduate student
  - Did thesis on Nurses’ Health Study
- Became far more interested in process, data collection, methods, meaning, and data quality
  - Latent informaticist, though I did not know the word
- Rather indifferent to “results” as inferences
  - Not a good sign for a junior epidemiologist

Why did my training go wrong?

- Exposed to 256-byte programmable calculator in HS
- Became an English major in college
- Imbued in computer science
  - All undergrads had computer accounts
  - Daily user of email (campus) since 1973
  - Lots of CS and applied math courses
  - Directed undergraduate computer consulting program
- It was in the water…
  - Musen, Cimino, Lipman, Butte, Kohane, …

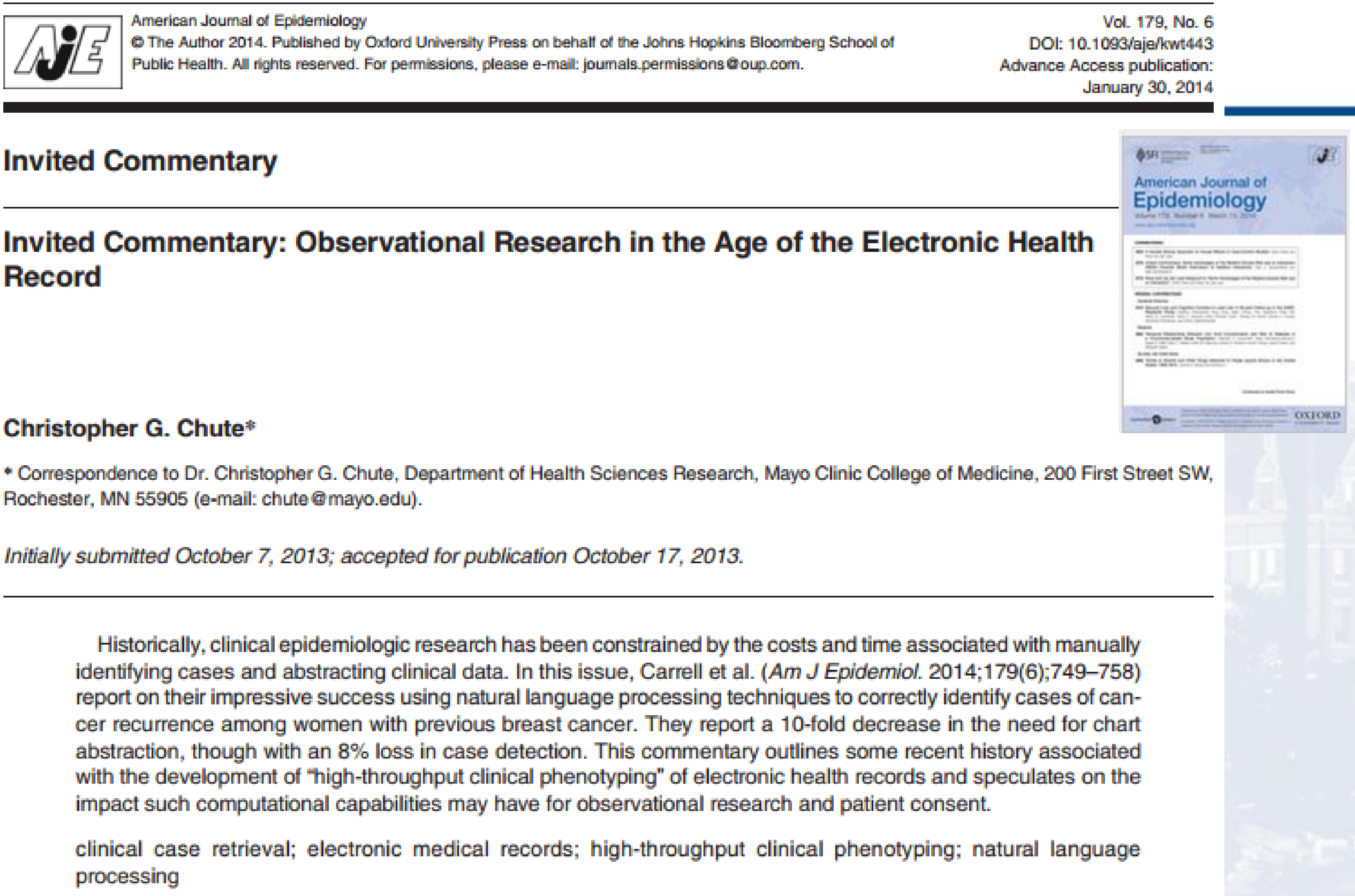
How many boats did I miss?

- Myopic focus on clinical data generated during the process of care
  - Discount survey data
  - Discount reimbursement data
  - Discount vital statistics
  - Discount environment and exposure
  - Discount occupational health

What do I think was going on?

- Healthcare benefits from analyses
- Inferencing methodology is not sufficient
- It’s all about the [clinical] DATA
  - Assume universal healthcare
  - Assume complete data capture and availability
- Data remained heterogeneous, non-comparable
- *Informatics emerged as the only path to truth!*

One of my least-cited articles….



American Journal of Epidemiology
OXFORD

American Journal of Epidemiology
© The Author 2014. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Vol. 179, No. 6
DOI: 10.1093/aje/kwt443
Advance Access publication:
January 30, 2014

Invited Commentary

Invited Commentary: Observational Research in the Age of the Electronic Health Record

Christopher G. Chute*

* Correspondence to Dr. Christopher G. Chute, Department of Health Sciences Research, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905 (e-mail: chute@mayo.edu).

Initially submitted October 7, 2013; accepted for publication October 17, 2013.

Historically, clinical epidemiologic research has been constrained by the costs and time associated with manually identifying cases and abstracting clinical data. In this issue, Carrell et al. (*Am J Epidemiol.* 2014;179(6);749–758) report on their impressive success using natural language processing techniques to correctly identify cases of cancer recurrence among women with previous breast cancer. They report a 10-fold decrease in the need for chart abstraction, though with an 8% loss in case detection. This commentary outlines some recent history associated with the development of "high-throughput clinical phenotyping" of electronic health records and speculates on the impact such computational capabilities may have for observational research and patient consent.

clinical case retrieval; electronic medical records; high-throughput clinical phenotyping; natural language processing

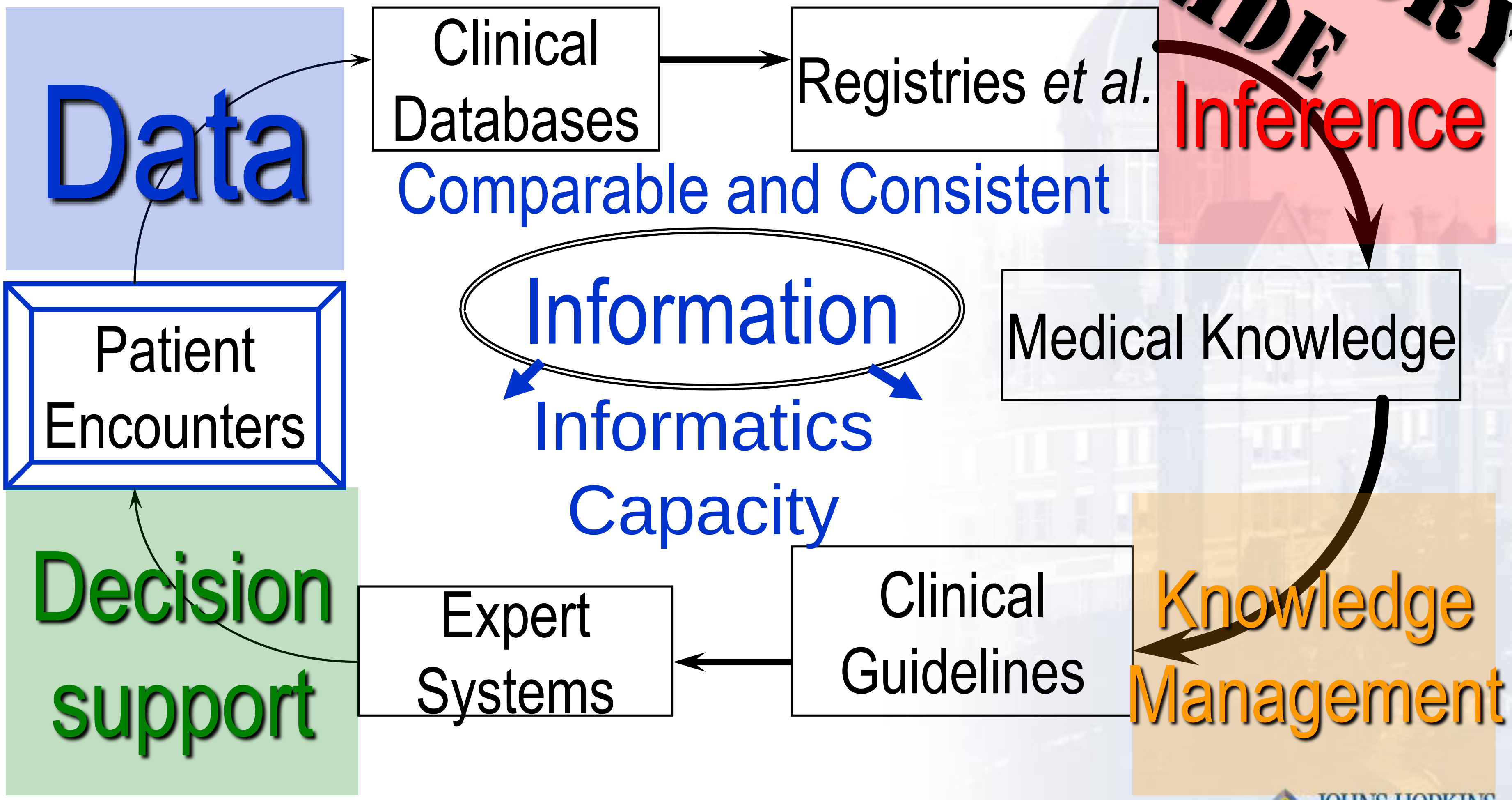
So, what was the epiphany?
What are epiphanati?

Within the biomedical data world:

- Comparable and consistent data is prerequisite
- That rests on semantic coherence
  - Classification, Ontology, Terminology, Value Sets
- Semantics must be bound to context
  - Information models, EHR
- Practice late-binding to application schema

From Practice-based Evidence to Evidence-based Practice

MANDATORY SLIDE



JOHNS HOPKINS MEDICINE

Coherent semantics

NATIONAL LIBRARY OF MEDICINE

UMLS®

How do ***you*** say glucose?

Glucose 葡萄糖 Glükoos Glucose Glucose Glucjkose Glukose

LOINC

the ***lingua franca*** of clinical observation exchange

Γλυκόζη Glucosio 포도당 Glicose Glucosa Glucosa Glucosa

Rx NLM Nav

- IHTSDO delivering
- SNOMED CT®
- the global clinical terminology

go Gene Ontology

ICD-11 *alpha*

International Classification of Diseases

Eleventh Revision

World Health Organization
Geneva

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

ICD-9-CM

International Classification of Diseases, Ninth Revision, Clinical Modification

2010

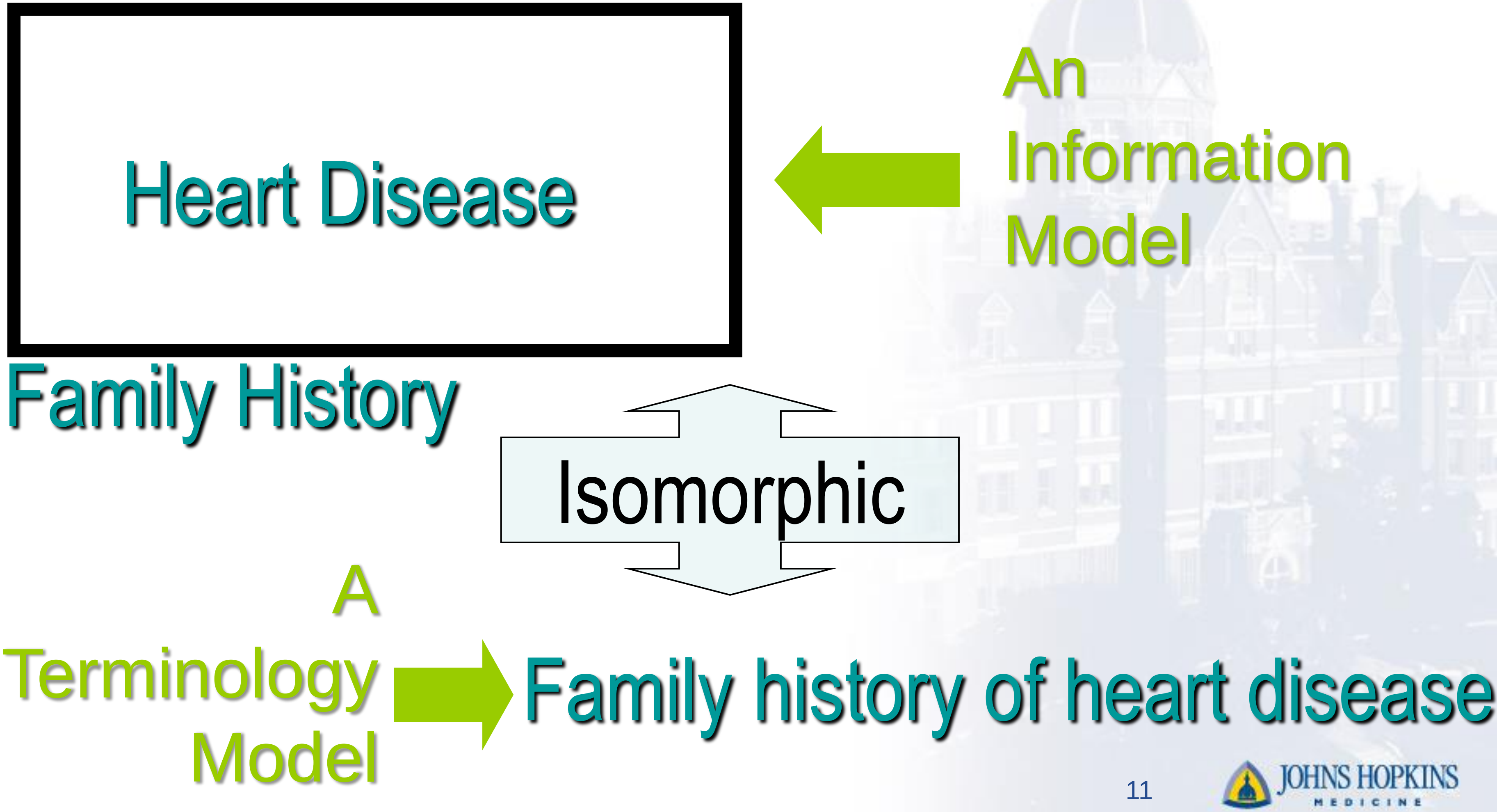
6

HL7® INTERNATIONAL

ISO HEALTH INFORMATICS TC215

CDISC®

Content vs. Structure
Semantics is intertwined with structure



Heart Disease

An Information Model

Family History

Isomorphic

A Terminology Model

Family history of heart disease

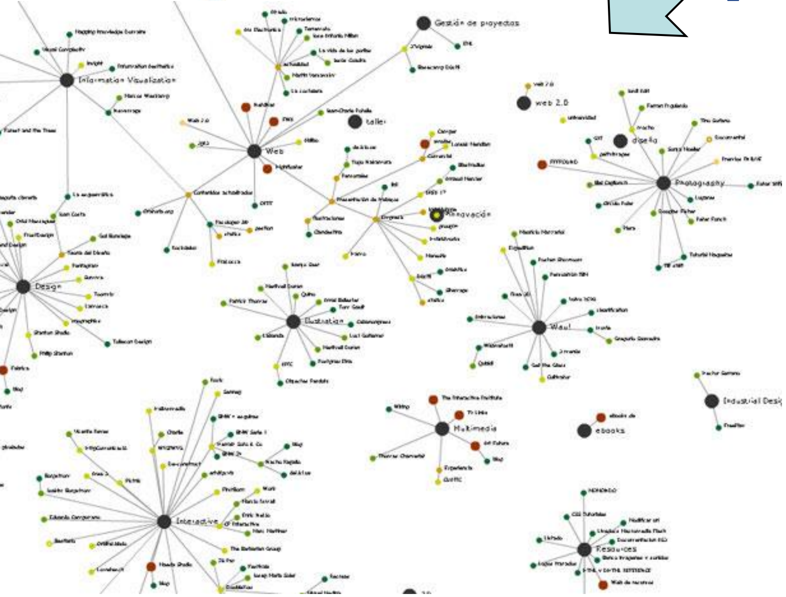
Discrete data elements
Just-in-time model binding

CIMI Archetypes

- Demographics
- Observations
- Medications
- Procedures
- …

Data Marts

- Registries
- Protocols
- Studies
- Cohorts
- …



VS.

What does any of this mean for Hopkins?

- Promote principle of “clinical data as a first-rank resource”
- Pursue the implications
  - Data governance, security, curation
  - Informatics critical mass, development, application
- Propose extension beyond Hopkins to community
  - Population health

The “data lake”

Establish repository of clinical data

- Invoke NOSQL accumulation of data elements
- Leverage Accumulo/Topaz (Armstrong Institute)
- Leverage EPIC data warehouse
- Incorporate departmental data sources
  - Include original content and metadata
  - Capture waveforms and raw signals
  - Integrate claims data
- Incrementally normalize to canonical form

Maryland as a Population Laboratory

Many unique features

- CMS waiver among hospitals
- Successful emergence of CRISP
  - Framework for collaboration
- Goal of federated data repositories
  - Build on “data lake” technologies
  - Participants have secure silos

Where is this going?

- Outstanding opportunity, talent, material
- Hopkins must embrace clinical data
- Collaborate with University resources
- Collaborate with community partners
- Enable unprecedented discovery
- Rewind Chris’s story

Normalized data → Analyses → Evidence → Practice