UNIVERSITY *of* MARYLAND
SCHOOL OF MEDICINE

Big Data and the Research HARBOR (Helping Advance Research By Organizing Resources)

J. Kathleen Tracy, PhD
Director, Clinical and Translational Research Informatics Center (CTRIC)
Associate Director, Center for Health Informatics and Bioimaging (CHIB)

Universities, Innovation & Big Data

- Natural incubators for innovation
- Diverse disciplines
- Semi-common infrastructure

Big Data, Big Science

- Novel, creative approaches
- Knowledge integration across disparate fields
- Collaboration across multiple areas of expertise that cross major disciplinary, scientific, and administrative boundaries

Copyright 2014

tagxedo.com

Opportunities and Challenges for Healthcare and Biomedical Research

Opportunities	**Challenges**
Greater access to more diverse data	The 4 V's
Identification of previously unknown relationships (e.g., unintended effects of txs, differential response to txs for sub-groups of patients)	New methods
Identification of a greater array of targets for cost-savings	Analytic tools
Personalization of care and greater engagement with patients as partners	Protection of patient privacy
Building the next generation of scientists and care providers	Agility

Big Data: Opportunities and Challenges

What are some impacts of *Big Data*?

- After Haiti’s 2010 earthquake, Columbia University tracked the movements of 2 million refugees by the SIM cards in their cell phones and were able to determine where health risks would likely develop.

Our Unique Capacities

- Affiliation with UMMS
- Proximity and affiliation with Baltimore VA Hospital
- Diversity of patient populations
- Depth and breadth of scientific expertise of the University of Maryland System

RESEARCH
HARBOR

Helping **A**dvance **R**esearch **B**y **O**rganizing **R**esources

Research HARBOR

The UM Research Harbor is an interactive, web-based platform that provides one-stop shopping for research support needs. Through a centralized hub, researchers and their staff at the University of Maryland can access our data warehouse, identify and access research support resources, tools and services, find experts, access regulatory support, learn about educational and training opportunities, and much more.

Researchers Area	**Community Area**
Data Warehouse	Community resources
Regulatory support (CICERO, WebRPR)	COMPASS Volunteer Registry
Research Support Services	COMPASS Community Partners Registry
Research Tools	
Educational Resources	
Specialty scientific “ports”	

Personalized Medicine	SOM	CTRIC	UMMS	IGS
API	iPad	Desktop Mobile	iPad	API

- Teleform
- RedCap
- OpenClinica
- R-studio
- NGS Exome, Genome, RNA-seq, ChIP-seq

HARBOR

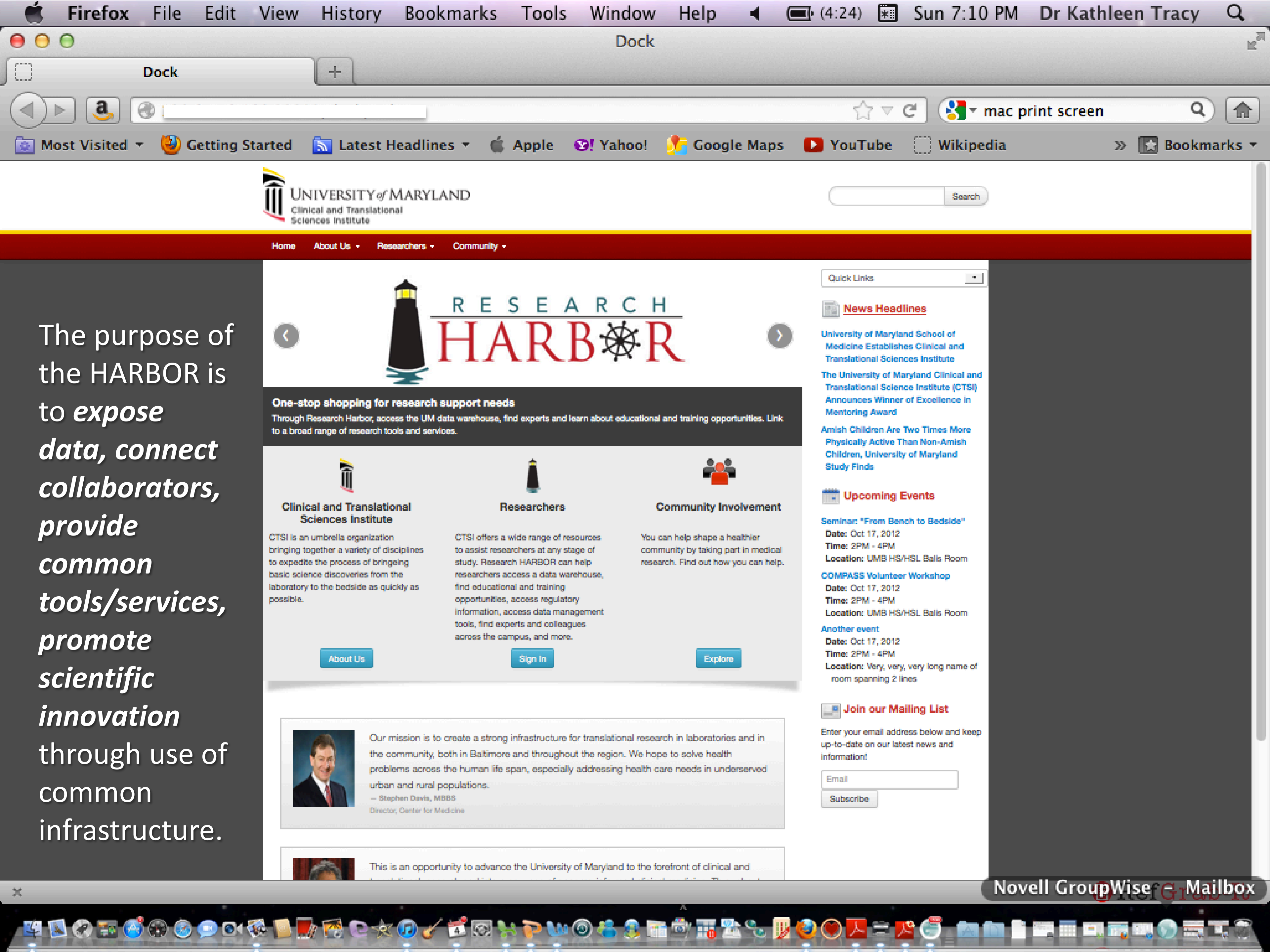
- Recruitment
- Curated Catalog
- Personalized Med.
- Data Depot & Sharing
- Secondary Sources
- Research Compliance
- Standardized Capture
- Volunteer Registry

Personalized Medicine	SOM	CTRIC	UMMS	IGS
BioBank	CICERO Coeus	PIs COMPASS	CDR EPIC	HMP DACC NGS

HARBOR Data Warehouse

- Data catalog of data sources and related documentation for data that have been placed in Warehouse
- Provides ability to search for and compile identified (with IRB approval) and de-identified datasets from a variety of data sources
  - UMMS clinical data repository
  - Observational registries (e.g., cancer, anesthesiology, sports medicine)
  - COMPASS Volunteers
  - Geospatial data
- Provides tutorial information related to appropriate end use of data from the Warehouse

The purpose of the HARBOR is to ***expose data, connect collaborators, provide common tools/services, promote scientific innovation*** through use of common infrastructure.



One-stop shopping for research support needs

Through Research Harbor, access the UM data warehouse, find experts and learn about educational and training opportunities. Link to a broad range of research tools and services.

Clinical and Translational Sciences Institute

CTSI is an umbrella organization bringing together a variety of disciplines to expedite the process of bringeing basic science discoveries from the laboratory to the bedside as quickly as possible.

Researchers

CTSI offers a wide range of resources to assist researchers at any stage of study. Research HARBOR can help researchers access a data warehouse, find educational and training opportunities, access regulatory information, access data management tools, find experts and colleagues across the campus, and more.

Community Involvement

You can help shape a healthier community by taking part in medical research. Find out how you can help.

Our mission is to create a strong infrastructure for translational research in laboratories and in the community, both in Baltimore and throughout the region. We hope to solve health problems across the human life span, especially addressing health care needs in underserved urban and rural populations.

— Stephen Davis, MBBS
Director, Center for Medicine

This is an opportunity to advance the University of Maryland to the forefront of clinical and

News Headlines

University of Maryland School of Medicine Establishes Clinical and Translational Sciences Institute

The University of Maryland Clinical and Translational Science Institute (CTSI) Announces Winner of Excellence in Mentoring Award

Amish Children Are Two Times More Physically Active Than Non-Amish Children, University of Maryland Study Finds

Upcoming Events

Seminar: "From Bench to Bedside"
Date: Oct 17, 2012
Time: 2PM - 4PM
Location: UMB HS/HSL Balls Room

COMPASS Volunteer Workshop
Date: Oct 17, 2012
Time: 2PM - 4PM
Location: UMB HS/HSL Balls Room

Another event
Date: Oct 17, 2012
Time: 2PM - 4PM
Location: Very, very, very long name of room spanning 2 lines

Join our Mailing List

Enter your email address below and keep up-to-date on our latest news and information!

Big Data, Research HARBOR & Innovation

PERSPECTIVE

ACCESS TO PATIENT-LEVEL TRIAL DATA

level playing field for all stakeholders. What is sometimes labeled as "free riding" may ultimately pay dividends for innovative companies and for public health. It is ironic that the organizations that most resist wider access to data are the ones that stand to benefit so much from greater transparency.

The views expressed in this article are those of the authors and do not necessarily reflect those of the European Medicines Agency or any of its committees.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the European Medicines Agency, London.

This article was published on October 21, 2013, at NEJM.org.

1. Zarin DA. Participant-level data and the new frontier in trial transparency. N Engl J Med 2013;369:468-9.
2. European Medicines Agency. Publication and access to clinical-trial data: draft policy 0070. June 24, 2013 (http://www.ema.europa.eu/docs/en_GB/document_library/Other/2013/06/WC500144730.pdf).
3. Pharmaceutical Research and Manufacturers of America, European Federation of Pharmaceutical Industries and Associations. Principles for responsible clinical trial data sharing (http://phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf).
4. Rabinowitz J, Werbeloff N, Stauffer V, Mandel F, Kapur S. More efficient drug discovery trials in schizophrenia: insights from 29 placebo-controlled randomised controlled trials (RCTs) from the NEWMEDS repository. Presented at the 24th Congress of the European College of Neuropsychopharmacology, Paris, September 3–7, 2011 (http://conferenceservices.elsevier.nl/11ecnp/index.cfm?fuseaction=CIS2002&hoofdnav=Search&content=zk.results_all&topicselected=*&searchtext=rabinowitz&what=FREE%20TEXT&selection=ALL&abstrnbr=S.11.03).
5. Catalyzing the transformation of health-care innovation: a working paper from the MIT Center for Biomedical Innovation. Cambridge: Massachusetts Institute of Technology, January 2010 (http://cbi.mit.edu/wp-content/uploads/2011/04/MIT-NewDigs-White-Paper-20100118.pdf).

DOI: 10.1056/NEJMp1310771

The Randomized Registry Trial — The Next Disruptive Technology in Clinical Research?

Michael S. Lauer, M.D., and Ralph B. D'Agostino, Sr., Ph.D.

Related article, p. 1587

The randomized trial is one of the most powerful tools clinical researchers possess, a tool that enables them to evaluate the effectiveness of new (or established) therapies while accounting for the effects of unmeasured confounders and selection bias by indication. Randomized trials, especially huge megatrials, have transformed medical practice. Thanks to randomized trials, we no longer, for example, treat acute myocardial infarction with lidocaine and nitrates. Instead we use rapid revascularization, anticoagulants, and antiplatelet agents, and during long-term follow-up we routinely prescribe statins, beta-blockers, and angiotensin-converting–enzyme inhibitors. But the reputation of randomized trials has suffered of late,[1] owing to reasonable concern about excess complexity, expense, and time required to recruit study participants, as well as inadequate representativeness. What good are trials if the results aren't applicable to real-world patients and if, because of excessive expense, they can be used to answer only a tiny fraction of our important clinical questions?

One possible solution is to look to observational registries for answers. Over the past 20 to 30 years, a number of professional societies, government agencies, private corporations, and independent researchers have established high-quality registries that collect standardized data from patients seen in a variety of settings. In cardiovascular medicine, for example, registries in the United States and abroad have collected vast amounts of data from patients with acute coronary syndromes, stable coronary disease, and heart failure, as well as from patients with rare diseases such as hypertrophic cardiomyopathy and patients referred for surgery, percutaneous invasive procedures, and device implantation. Investigators and public health officials use registries to describe practice patterns and trends, to identify outliers, and to detect safety signals. They often use registries to assess comparative effectiveness, too, but are forced to admit that purely observational findings may not be internally valid owing to the absence of randomization.

As debates about comparative-effectiveness research have intensified over the past few years, we find ourselves in a kind of intellectual trap: yes, in theory we would like to conduct more randomized trials, but in practice they are too complex and difficult to apply to many clinical questions. And, yes, in theory we could answer many questions at

Examples of *Big Data*

Healthcare insurance organizations collect billing, procedure and diagnostic code data from all providers in their networks on a daily, weekly, and monthly basis to assess costs, meet regulatory requirements, determining whether the organization qualifies for additional reimbursement based on more cost-effective standard of care. Insurance organizations also mine these data to better care for patients and contain costs.

- **Creating tailored reminders to female patients that have not received a Pap smear in the last 3 years.**

- ***Creating an automated outbound calling system for girls age 9-26 who started the HPV vaccination series but did not complete all 3 doses.***

Big Data in Healthcare

- Rich information for health policy
  - How does spending differ by location? Patient mix?
  - What are the trends in disease prevalence?
  - What are the trends in treatment choices?
  - How do disease, treatments, outcomes, etc. vary from region to region? By gender? By type of insurance coverage? By provider?
  - Which providers are better/worse in quality and cost?

- Support for performance improvement
  - Transparent reporting of provider and payer results
  - Data can be used by providers to drive their QI efforts

Big Data in Healthcare (cont.)

- Leading causes of illness and hospitalization
- Rates of accidents, infections and cancer
- Geographic differences in incidence of diseases, such as diabetes or heart disease
- Ethnic, sex or socioeconomic variations in illness
- Most expensive diagnoses and procedures
- Role of prevention on illness and costs
- Unique genetic variations in response to treatments, outcomes
- Comparative effectiveness studies

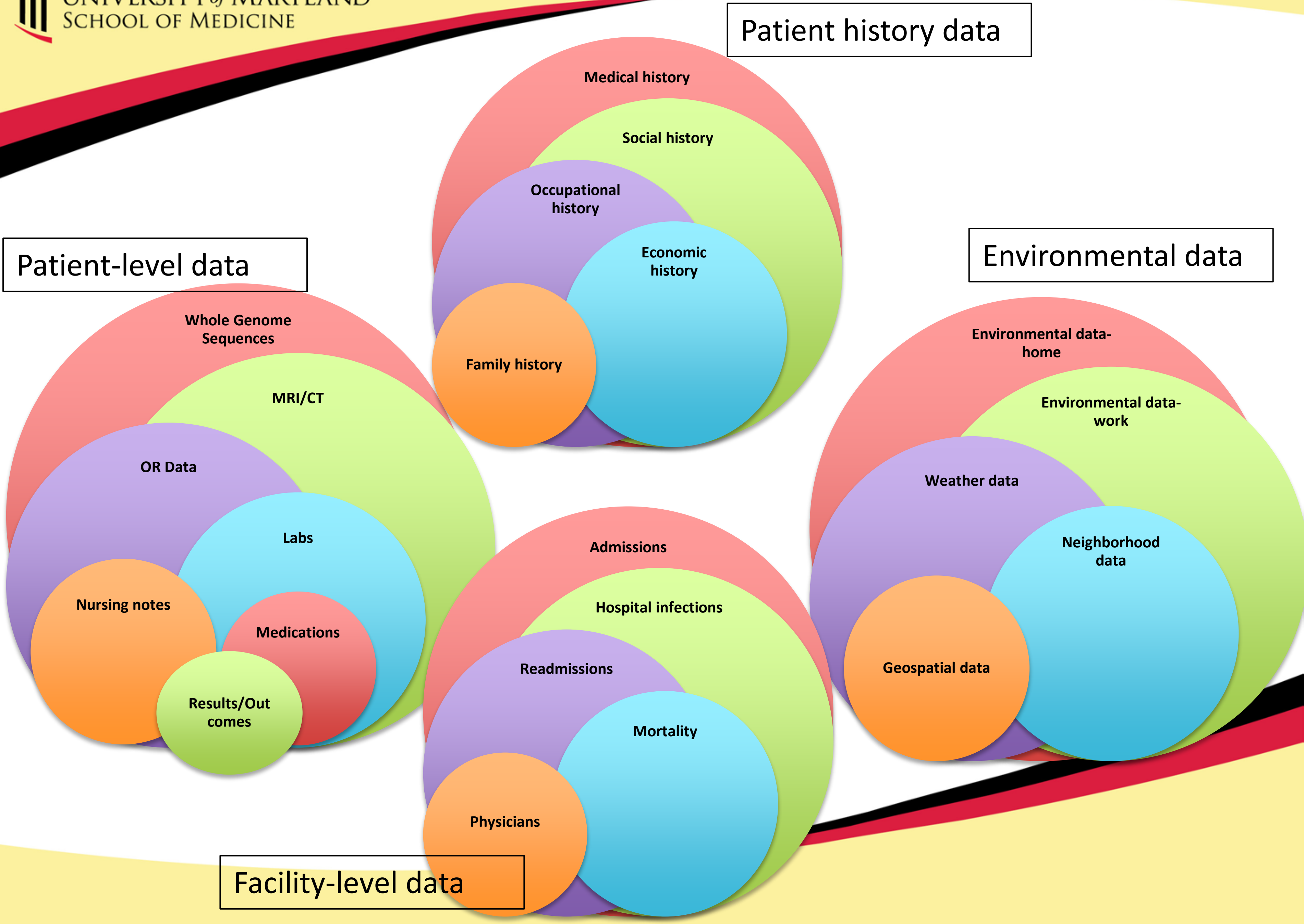
Links to Other Data and Initiatives

- Quality – CMS, state reports, regional collaboratives
- State or National vital statistics – to assess mortality rates
- Hospital discharge datasets – for additional data detail and measures
- Health Information Exchanges – integrate claims and clinical (EMR) data
- Proprietary data marts that contain claims data for different payors, healthcare delivery systems, patient populations

In-Depth Example: Colectomy

- Average duration in OR = 182 minutes (Pandit & Carey, 2006)
- Data diversity and complexity
  - Patient-level data
  - Provider-level data
  - Facility-level data
  - Geospatial data

Copyright 2014 tagxedo.com



Patient history data

- Medical history
- Social history
- Occupational history
- Economic history
- Family history

Patient-level data

- Whole Genome Sequences
- MRI/CT
- OR Data
- Labs
- Nursing notes
- Medications
- Results/Outcomes

Environmental data

- Environmental data-home
- Environmental data-work
- Weather data
- Neighborhood data
- Geospatial data

Facility-level data

- Admissions
- Hospital infections
- Readmissions
- Mortality
- Physicians

What to do?

- Use NLP to extract meaningful, standardized data elements
- Identify patterns and develop predictive algorithms
- Develop tools and methods to analyze highly diverse and complex data across multiple levels of measurement

Ex. of Big Data and Building a Comparative Effectiveness Trial

- Leverage HARBOR data (e.g., registries) for recruiting participant groups
  - Random selection
- Use of historical health system data rather than self-reported history to characterize participant history
- Use routine clinical data for monitoring at least secondary outcomes

“Do what you can, with what you have, where you are.” Theodore Roosevelt