Interventional Informatics: Data-Driven Approaches to Discovery and Practice

Philip R.O. Payne, PhD, FACMI

Professor and Chair, College of Medicine, Department of Biomedical Informatics
Professor, College of Public Health, Division of Health Services Management and Policy
Director, Translational Data Analytics @ Ohio State, Discovery Themes Initiative
Associate Director for Data Sciences, Center for Clinical and Translational Science
Co-Director, Bioinformatics Shared Resource, Comprehensive Cancer Center

The Ohio State University
Wexner Medical Center
COI/Disclosures

- **Federal Funding:** NCI, NLM, NCATS, AHRQ

- **Additional Research Funding:** BAH, SAIC, Rockefeller Philanthropy Associates, AcademyHealth, Pfizer, Hairy Cell Leukemia Foundation

- **Academic Consulting:** CWRU, Cleveland Clinic, University of Cincinnati, Columbia University, Emory University, Virginia Commonwealth University, University of California San Diego, University of California Irvine, University of Minnesota, Northwestern University, Stonybrook University, University of Kentucky

- **International Partnerships:** Soochow University (China), Fudan University (China), Clinica Alemana (Chile), Universidad de Chile (Chile)

- **Other Consulting/Honoraria:** American Medical Informatics Association (AMIA), Institute of Medicine (IOM), Geisinger Health System

- **Editorial Boards:** Journal of the American Medical Informatics Association, Journal of Biomedical Informatics, eGEMS, BMC Medical Informatics and Decision Making (Section Editor, Healthcare Information Systems)

- **Study Sections:** NLM (BLIRC), NCATS (formerly NCRR)

- **Corporate:** Signet Accel LLC (Founder), Signet Innovations LLC (Advisor), Aver, Lumeris
Let’s Start with a Few Definitions...

def·i·ni·tion
n. 1.
The teacher gave definitions of the new words.
Defining Biomedical Informatics (BMI)

“The field that is concerned with the optimal use of information, often aided by the use of **technology and people** (researchers, practitioners, users, etc.), to improve individual health, health care, public health, and biomedical research” (Bill Hersh, 2010)
A “Working” Central Dogma for BMI: Enabling Translation and Systems Thinking

Data

Information

Knowledge

Test = HbA1c
MRN = 1234
Date = 6/15/2016
Normal is < 5.7%

6.5%

Send Alert to PCP via EHR

MRN = 1234
Test = HbA1c
Date = 6/15/2016
Normal is < 5.7%

Send Alert to PCP via EHR
A Historical Perspective on the Evolution of BMI as a Scientific Field

- **Basic Science**
  - Standards and data representation
  - Knowledge engineering
  - Cognitive and decision science
  - Human factors and usability
  - Computational biology

- **Applied Science**
  - Clinical Decision Support Systems (CDSS)
  - Clinical Information Systems (incl. EHRs)
  - Consumer-facing tools (incl. PHRs)
  - Bio-molecular data analysis “pipelines”

- **At the Intersection of Basic and Applied Science**
  - Information Retrieval (IR)
  - Text Mining and Natural Language Processing (NLP)
  - Visualization
  - Image Analysis

---

AI in Medicine

Computers in Medicine

Medical Informatics

Biomedical Informatics

An Evolving Nomenclature...
How Do We Differentiate BMI from the Computational and Information Sciences?
A Roadmap for Today’s Discussion

Emerging Trends In Healthcare And Informatics
Informatics As The Intervention
A Vision For Data-driven Discovery And Practice

“Information liberation + new incentives = rocket fuel for innovation”
– Aneesh Chopra (The Advisory Board Company)
# Evolving HIT and Data Landscape

## The Advent of the Printing Press and the 1st Information Age

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Before HIT and Big Data</th>
<th>After HIT and Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost</strong></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Data sets generated and/or curated on a need basis</td>
<td>Data production and storage costs decreasing in excess of Moores Law</td>
</tr>
<tr>
<td><strong>Ubiquity</strong></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Proprietary data situated in vendor or project-specific repositories and formats</td>
<td>Data becoming a renewable resource enabled by diverse re-use scenarios</td>
</tr>
<tr>
<td><strong>Reproducibility</strong></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Errors of transcription and omission very common</td>
<td>Linked public data enables the creation of “commons” model</td>
</tr>
</tbody>
</table>

## Growth in HIT and Big Data in the Healthcare Information Age

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Before HIT and Big Data</th>
<th>After HIT and Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost</strong></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Data sets generated and/or curated on a need basis</td>
<td>Data production and storage costs decreasing in excess of Moores Law</td>
</tr>
<tr>
<td><strong>Ubiquity</strong></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Proprietary data situated in vendor or project-specific repositories and formats</td>
<td>Data becoming a renewable resource enabled by diverse re-use scenarios</td>
</tr>
<tr>
<td><strong>Reproducibility</strong></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Errors of transcription and omission very common</td>
<td>Linked public data enables the creation of “commons” model</td>
</tr>
</tbody>
</table>
A New Initiative on Precision Medicine

Francis S. Collins, M.D., Ph.D., and Harold Varmus, M.D.

President Obama has long expressed a strong conviction that science offers great potential for improving health. Now, the President has announced a research initiative that aims to accelerate progress toward a new era of precision medicine (www.whitehouse.gov/precisionmedicine). We believe that the time is right for this visionary initiative, and the National Institutes of Health (NIH) and other partners will work to achieve this vision.

The concept of precision medicine — prevention and treatment strategies that take individual variability into account — is not new. Blood typing, for instance, has been used to guide blood transfusions for more than a century. But the prospect of applying this concept broadly has been dramatically improved by the recent development of large-scale biologic databases (such as the human genome sequence), powerful methods for characterizing patients (such as proteomics, metabolomics, genomics, diverse cellular assays, and even mobile health technology), and computational tools for analyzing large sets of data. What is needed now is a broad research program to encourage creative approaches to precision medicine, test them rigorously, and ultimately use them to build the evidence base needed to guide clinical practice.

The proposed initiative has two main components: a near-term focus on cancers and a longer-term aim to generate knowledge applicable to the whole range of health and disease. Both components are now within our reach because of advances in basic research, including molecular biology, genomics, and bioinformatics. Furthermore, the initiative taps into converging trends of increased connectivity, through social media and mobile devices, and Americans' growing desire to be active partners in medical research.

Oncology is the clear choice for enhancing the near-term impact of precision medicine. Cancers are common diseases; in the aggregate, they are among the leading causes of death nationally and worldwide, and their incidence is increasing as the population ages. They are also especially feared, because of their lethality, their symptoms, and the often toxic or disfiguring therapies. Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes — and to give all of us access to the personalized information we need to keep ourselves and our families healthier.
1. Speed and Agility Can and Will Become Critical
2. Health Data Will Become Ubiquitous and Extend Beyond the Clinic and Hospital
3. Data-driven Approaches to Population Health Will Create True Value for Patients, Providers, and Payers
4. Disruption Will Come From Outside of the Health and Life Science Communities
5. Open Innovation Will Reshape Healthcare Delivery, Workforce Development, and Economics
6. Biomedical and Health Informatics Sub-disciplines and Pre-fixes Will Become Obsolete
A Unique Confluence of Trends and Environmental Factors That Will Define the Future of Healthcare

- Creating data “liquidity” and generating value for research and practice
- Major changes in culture, incentives, and business model(s)
- Advent of the “HIT and Big Data Age”
- Open Science
- Evolving HIT and “Big” Data Landscape
- Healthcare Transformation
Without feedback from precise measurement, invention is doomed to be rare and erratic. With it, invention becomes commonplace

– Bill Gates (2013 Gates Foundation Annual Letter)
Rethinking The Role of BMI: Informatics as the Interventional When Improving Health and Wellness
Three Example Biomedical Informatics Projects Underway at OSU with Interventional Potential

1) **TOKEN**: high throughput in silico hypothesis discovery for translational medicine
2) **SPHERE**: interactive and participatory decision support
3) **CIELO**: supporting open innovation and team science
Accelerating Hypothesis Generation, Testing, and Translation Using Existing Data Resources
TOKEn (1): An in silico Hypothesis Discovery Pipeline for Adaptive and Combination Therapies

- Translational Ontology-anchored Knowledge Discovery Engine (TOKEn)
- in silico hypothesis discovery platform
- TOKEn platform initial verified and validated in the context of the NCI-funded Chronic Lymphocytic Leukemia Research Consortium (CLL-RC)
- Current efforts have focused on extending the TOKEn framework to create a novel scoring metric to identify and prioritize combination therapies in terms of their shared drug concepts and proximity to disease-related concepts
  - Leveraging melanoma as an experimental context

ABSTRACT

Objective: The conduct of investigational studies that involve large-scale data sets presents significant challenges related to the discovery and testing of novel hypotheses capable of supporting in silico discovery science. The aim of what we term Constructive Knowledge Discovery in Databases (CKDD) methods is to provide a practical means of scaling hypothesis discovery and testing approaches for large data sets. This approach utilizes a knowledge-anchored approach to knowledge discovery and testing that models such as that described in this report will ideally be capable of supporting the synthesis of novel knowledge that serves to link clinical phenotype and biomolecular variables in that data collection.

Methods: The authors have developed a method known as T2, the Translational Ontology-anchored Knowledge Discovery Engine (TOKEn). This model utilizes a specific CKDD approach known as Constructive Knowledge Discovery in Databases (CKDD) methods to identify and prioritize potential hypotheses related to the meaningful semantic relationships between variables within in large-scale and heterogeneous biomedical data sets.

Results: The authors have verified and validated TOKEn in the context of a translational research studies recently evaluated by the NCI-funded Chronic Lymphocytic Leukemia Research Consortium. Such studies have provided a rich source of regulatory and functional data that have been used to generate novel hypotheses capable of informing the design of novel therapeutic paradigms.

Conclusion: The TOKEn project represents a potentially useful and systematic approach to knowledge synthesis in silico science. The model the authors have developed is capable of supporting the synthesis of novel knowledge that serves to link clinical phenotype and biomolecular variables in that data collection.

INTRODUCTION

The conduct of basic science, clinical, and translational research is extremely complex, involving a variety of actors, processes, resources, and information types that ideally should be integrated as a system. In particular, the translational research process often involves the development of novel hypotheses concerning relationships between phenotype and biomolecular variables in that data collection.

The TOKEn project: knowledge synthesis for in silico science

Philip R O Payne, Tara B Borlawsky, Omkar Lele, Stephen James, Andrew W Graessle

TOKEN (2): Leveraging Constructive Induction (CI) for Discovery “Big Data” Science

Targeted Metadata
- Drug 1 – Variable 1
- Drug 1 – Variable 2
- Drug 2 – Variable 1
- Drug 2 – Variable 2

Mapped Concepts
- Concept 1
- Concept 2
- Concept 3
- Concept 4
- Concept 5

Intermediate Concepts
- Concept 6
- Concept 7
- Concept 8

Example of Induced “Fact”
- Drug 1 – Variable 1
- Concept 1
- Concept 6
- Concept 7
- Concept 4
- Drug 2 – Variable 1

Constructive Induction
TOKEN (3): Applying the TOKEN “Pipeline” to Discovery Rational Combination Therapies for Melanoma

**TOKEN (4): Exemplary Results**

- Top ten ranked non-BRAF inhibitor drugs hypothesized for use in combination with BRAF inhibitor drugs.

<table>
<thead>
<tr>
<th>Non-BRAF inhibitor drug</th>
<th>DCS</th>
<th>Overlap score</th>
<th>Distance score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGRO100</td>
<td>28.83</td>
<td>14.46</td>
<td>14.37</td>
</tr>
<tr>
<td>Peginterferon-alfa-2a</td>
<td>28.66</td>
<td>13.93</td>
<td>14.74</td>
</tr>
<tr>
<td>Interferon_alfacon-1</td>
<td>28.45</td>
<td>13.87</td>
<td>14.58</td>
</tr>
<tr>
<td>Interferon_Alfa-2b</td>
<td>25.82</td>
<td>11.24</td>
<td>14.58</td>
</tr>
<tr>
<td>AZD-8330</td>
<td>25.77</td>
<td>14.68</td>
<td>11.10</td>
</tr>
<tr>
<td>Trabectedin</td>
<td>25.33</td>
<td>11.92</td>
<td>13.41</td>
</tr>
<tr>
<td>Trametinib</td>
<td>25.02</td>
<td>14.99</td>
<td>10.04</td>
</tr>
<tr>
<td>ZEN-012</td>
<td>23.09</td>
<td>11.40</td>
<td>11.70</td>
</tr>
<tr>
<td>PI-88</td>
<td>23.01</td>
<td>11.09</td>
<td>11.93</td>
</tr>
<tr>
<td>ABT-510</td>
<td>22.31</td>
<td>10.99</td>
<td>11.33</td>
</tr>
</tbody>
</table>

DCS = Drug Combination Score  
Log-2 transformed values shown
Improving Outcomes with Interactive and Participatory Clinical Decision Support
SPHERE CDSS Platform (1): Instrumenting the EHR to support risk profiling and patient-centered decision making

SPHERE CDSS Platform (2): Embedding Decision Support and Visualization Tools in Existing EHR Workflow

Interactive Risk Visualization

Patient-Centered Decision Making
SPHERE CDSS Platform (3): Impacting Decision Making and Clinical Outcomes in At-Risk Populations

One-year changes in CVH: Control clinic (n=109)

- Average age was 74 years (eligible patients ≥ 65)
- Intervention clinic was 35% black (control clinic 19% black)
- Improvements seen in the intervention clinic – but not control clinic – for diabetes and body mass index

One-year changes in CVH: Intervention clinic (n=160)

Pragmatic RCT Design (Clinic-Based Randomization)
Creating an Open Innovation Platform for Collaborative and Team-based Science
CIELO: Enabling Open Science and Collaborative Innovation in Healthcare Research

**Project Objectives:**

1) Provide members of the healthcare research community with access to an easy to use and highly flexible data commons platform

2) Reduce time and cost of research while enhancing the reproducibility and cumulative benefit of data analyses

3) Evolve and meet emerging community needs

**Blue-Sky:** not grounded in the realities of the present: visionary <blue–sky thinking> (Merriam Webster Dictionary)
Building on Standards and Enabling a Simplified Data Commons "Life Cycle"

- Designed explicitly for the health research community
- Builds on Git and provides a simplified user experience targeting a broad user community
- Support bundling of data and code and facilitates the ability to 'show' your methods in a comprehensive manner
- Semantic and social annotation and search reduces search space and enabled rapid discovery and interaction
Creating an Open Science Community-of-Practice

Open Innovation Cycle

Content Mgmt.
Submission and “Packaging” of Data and Analytic Tool Bundles

Social Search
Identification and Adoption/Adaptation of Data and Analytics Bundles

Enhanced Semantic Annotation, Search, and Distributed Execution

Scaling and Public Access

Sharing of Data With Metadata Curation

Delivering Resources to the Research Community

Simplifying the Discover and Use of “Bundles”

Expanding our “Community of Practice”

Join Our Beta Community:
http://cielo.bmi.osumc.edu

Philip R.O. Payne, PhD, FACMI
philip.payne@osumc.edu
@prpayne5
www.slideshare.net/prpayne5
Placing These Projects in a Broader Context...

**in silico** Discovery of Knowledge

**Driving Biological and/or Clinical Problems**

**Data Resources**

**CIELO**
An Open Innovation Platform for Team Science

**TOKEN**
Accelerating Evidence Generation and Validation

**SPHERE**
Applying EBM and Measuring Impact

Creating Data Liquidity to Fuel Future Discovery

Virtuous Cycle
A Roadmap for Today’s Discussion

- Emerging Trends In Healthcare And Informatics
- Informatics As The Intervention
- A Vision For Data-driven Discovery And Practice

“No Outcome, No Income”
– Eric Topol
Anticipating and Embracing Evolution in Technologies and Information Needs Is Critical...
**Structural Evolution:** Where Does a Contemporary Informatics Model Fit in the Emerging Academic Healthcare Enterprise?

<table>
<thead>
<tr>
<th>Scholarly Home</th>
<th>Traditional Model</th>
<th>Emerging Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Departments and Divisions</td>
<td>Multi-disciplinary Centers and Institutes</td>
</tr>
<tr>
<td>Revenue</td>
<td>Tuition, Grant, and Service Revenue</td>
<td>Technology Transfer Revenue, Public-Private Partnerships, Contracts, Multi-Center Consortia</td>
</tr>
<tr>
<td>Culture</td>
<td>Separation of Science and Service</td>
<td>Service as Science:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Institutional</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Community</td>
</tr>
<tr>
<td>Dissemination</td>
<td>Publications and Presentations</td>
<td>Commercialization, Translation into Healthcare Delivery Organizations</td>
</tr>
</tbody>
</table>

**How To Achieve Balance?**
Scientific Evolution: What Should Our Scholarly Foci Be?

1) **Fully embrace interdisciplinary:**
   - Structure
   - Function
   - Competency-based Training

2) **Pursue emerging (or remerging) research foci:**
   - Data science
   - Health services and quality improvement
   - Decision science and support (in the context of “Big Data”)
   - Human factors and workflow
   - Integrating patients and communities into the healthcare and research “fabric”

3) **Engage with health system(s):**
   - Analytics
   - Workflow and human factors
   - Transformation

4) **Adapting strategies from the private sector**
   - Identify and place disproportionate emphasis on “blue oceans”
   - Behave like a start-up (speed, agility, “real artists ship”)
   - Rapid translation of technologies to the “market”
Educational Evolution: Creating an Informatics Capable Workforce

Olga Strachna was considering medical school, or perhaps a Ph.D., after graduating with a biological sciences degree from Pittsburgh’s Carnegie Mellon University in 2009. But neither avenue felt quite right.

Then she discovered a brand new master’s program in health informatics at Weill Cornell Medical College in her hometown, New York City, designed to turn out pros who will work on the leading edge where health care meets information technology.

Since the program focuses on how to implement electronic health records and how to crunch – and make use of – data on quality of care and insurance reimbursements, it satisfied her varied interests in tech, research and health.

"This field is exploding," says Charles Friedman, director of the health informatics program at the University of Michigan—Ann Arbor, which enrolled its first master’s class in the fall of 2012. "Access to health information on the Web is taking off at a meteoric pace. It’s creating enormous employment opportunities."

A number of universities have launched similar programs to fill that pipeline, responding to rapid changes forcing health providers to greatly expand their use of technology. The Affordable Care Act is ensuring, for example, that systems be implemented for transmitting patients’ test results to their EHRs, so their physicians can use them effectively at the point of care.

There are now more than 70 advanced degree programs in the field, according to the American Medical Informatics Association. Most combine the necessary technical instruction with courses in medical practice and on-the-job training.

Two Final Thoughts (1): Behaving Like A High Performance System and Evolving Gracefully Requires A Systems Approach

- Three characteristics of a high performance system:
  1) Leverage data to identify problems and opportunities
  2) Design reproducible solutions
  3) Implement those solutions

Mastering the art of designing and implementing solutions is the greatest challenge facing the field of Biomedical Informatics
Two Final Thoughts (2): Now is the Time for Applying Biomedical Informatics in our Health Systems

Technology as the Primary Driver For Research, Education, And Practice and The Pursuit of the “Triple Aim”
Acknowledgements

Collaborators:
- Peter J. Embi, MD, MS
- Albert M. Lai, PhD
- Randi Foraker, PhD
- Kun Huang, PhD
- John C. Byrd, MD
- William E. Carson, MD
- Omkar Lele, MS, MBA
- Tara Borlawsky-Payne, MA
- Marcelo Lopetegui, MD, MS

Funding:
- NCI: R01CA134232, R01CA107106, P01CA081534, P50CA140158, P30CA016058
- NCATS: U54RR024384
- NLM: R01LM009533, T15LM011270
- AHRQ: R01HS019908
- Rockefeller Philanthropy Associates
- Hairy Cell Leukemia Foundation
- Academy Health – EDM Forum

Laboratory for Knowledge Based Applications and Systems Engineering (KBASE)

http://u.osu.edu/kbase/