Mobilizing Computable Biomedical Knowledge

Charles P. Friedman, PhD
Josiah Macy, Jr. Professor
Chair, Department of Learning Health Sciences
Professor of Information and Public Health
University of Michigan
October 18, 2017

M SPH
MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN
school of information

Main Menu

- Preliminaries
- The LHS and Infrastructural Services to Support It
- The “Keystone” Role of Persistent Computable Knowledge
  - And the Concept of a “K2P” Service
- Doing this at Scale: Vision of a Computable Knowledge Ecosystem…
- Goals and Plans for This Meeting

A Definition for Purposes of This Meeting

Knowledge: The result of an analytical and/or deliberative process that holds significance for an identified community.

Main Menu

- Preliminaries
- The LHS and Infrastructural Services to Support It
- The “Keystone” Role of Persistent Computable Knowledge
    - And the Concept of a “K2P” Service
- Doing this at Scale: Vision of a Computable Knowledge Ecosystem…
- Goals and Plans for This Meeting

Properties of a Health System That Can Learn & Improve

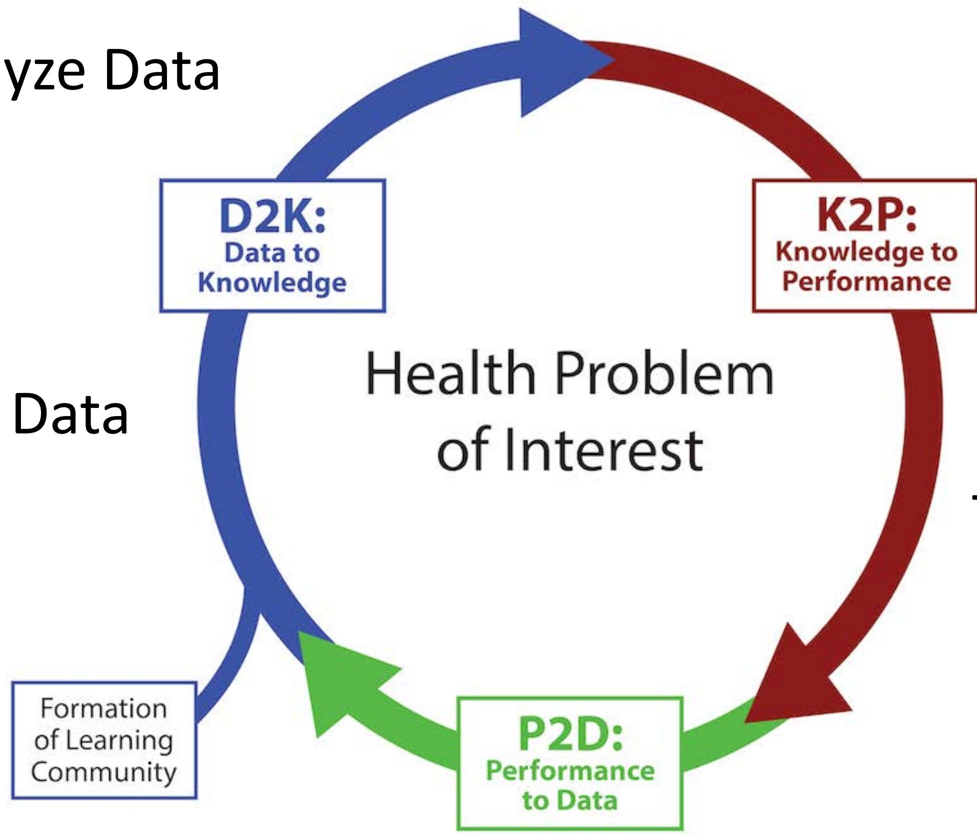
- ✓Every participating patient's characteristics and experience are available to learn from
- ✓Best practice knowledge is immediately available to support decisions
- ✓Improvement is continuous through ongoing study
- ✓An infrastructure enables this to happen routinely and with economy of scale
- ✓All of this is part of the culture

Learning Cycles
Better Health Requires a Flow:
D2K -> K2P -> P2D

Interpret Results

Analyze Data

Communicate
Tailored Change
Strategies

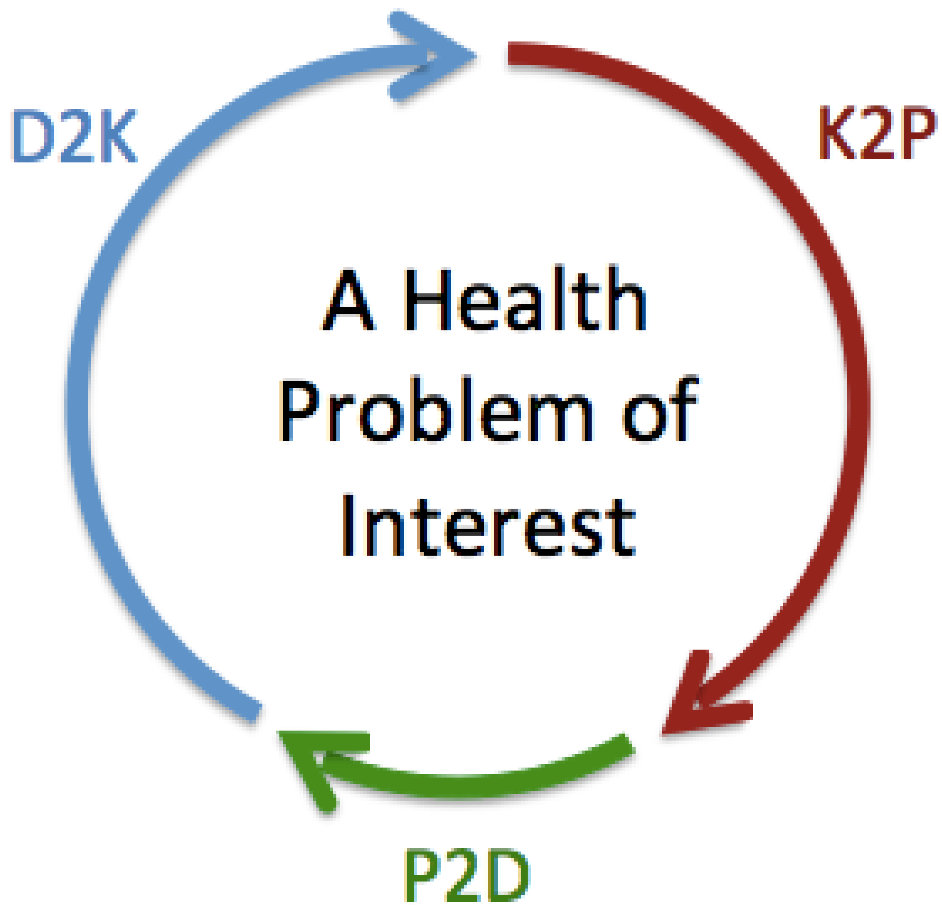


Assemble Data

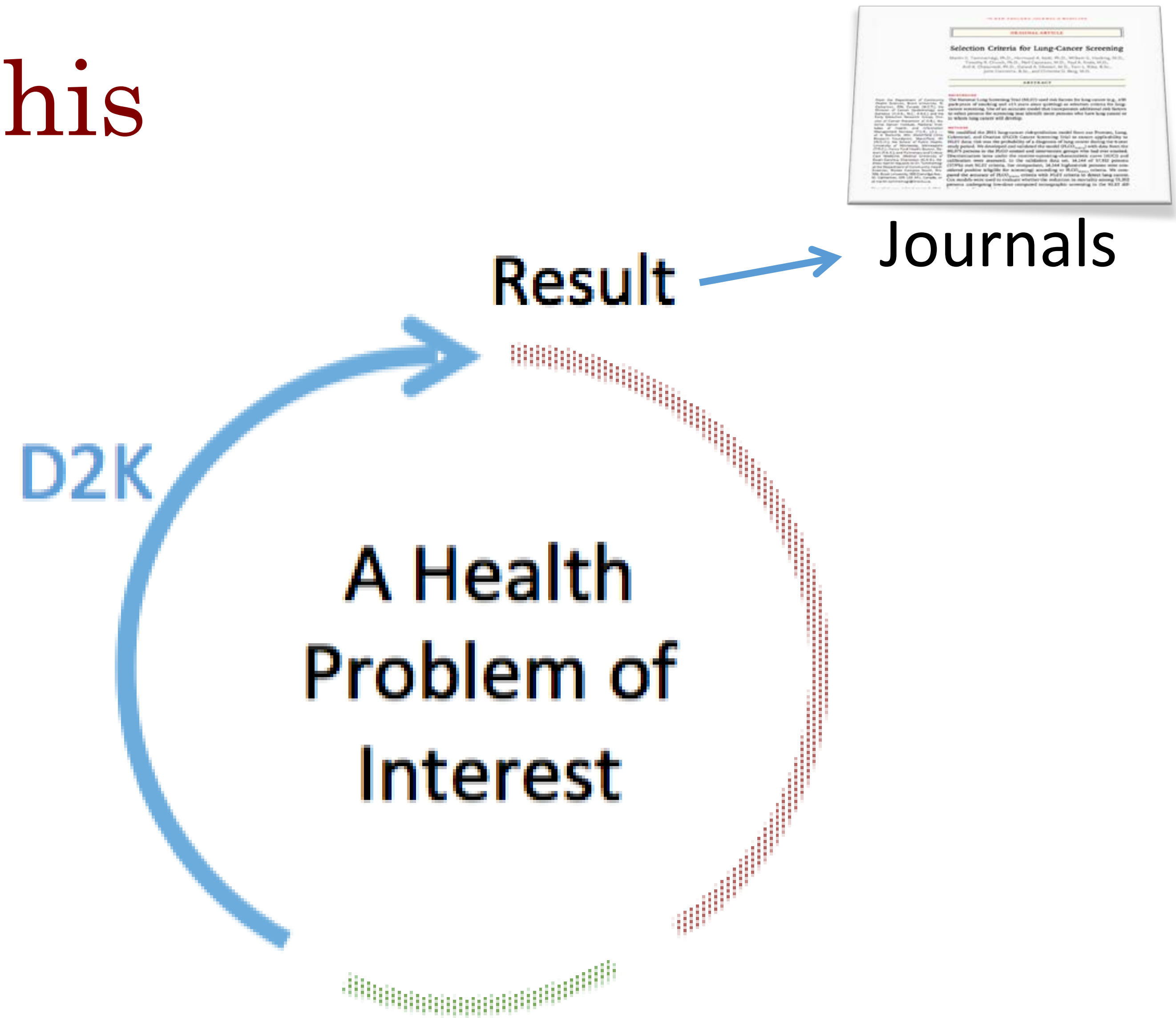
Take Action

Record Actions

Better Health Requires This

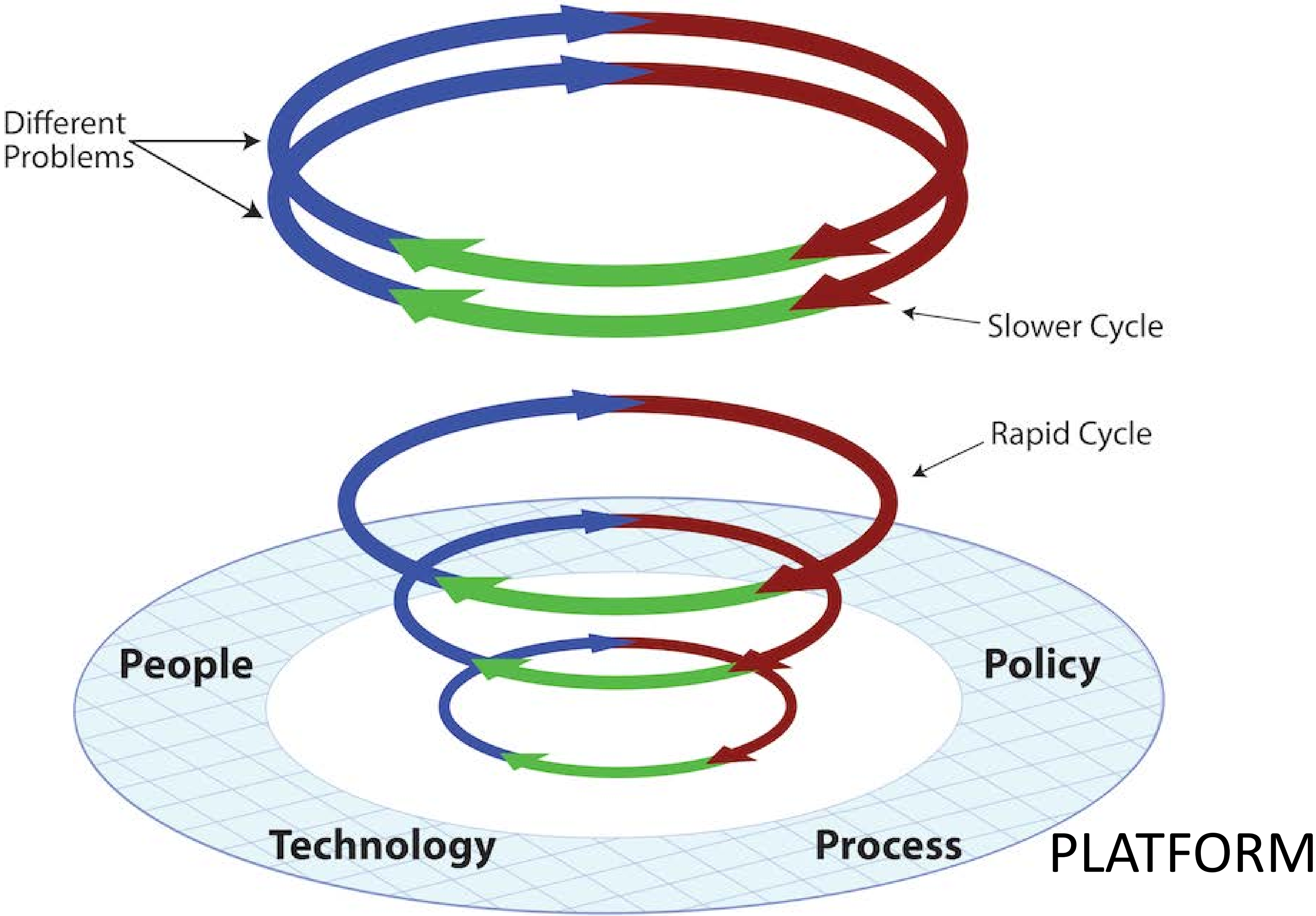


Not This

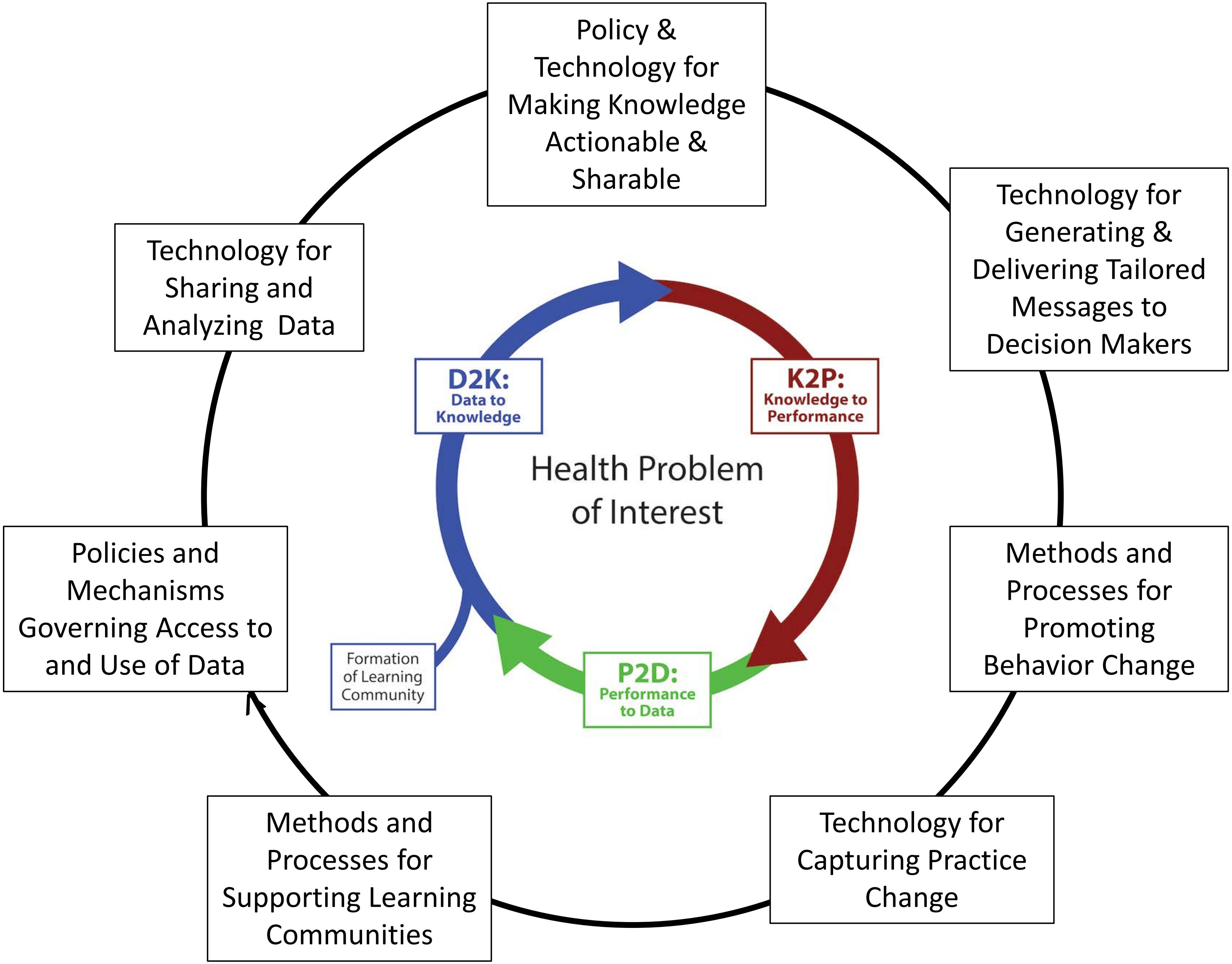


Selection Criteria for Lung-Cancer Screening
Journals
Result
D2K
A Health Problem of Interest

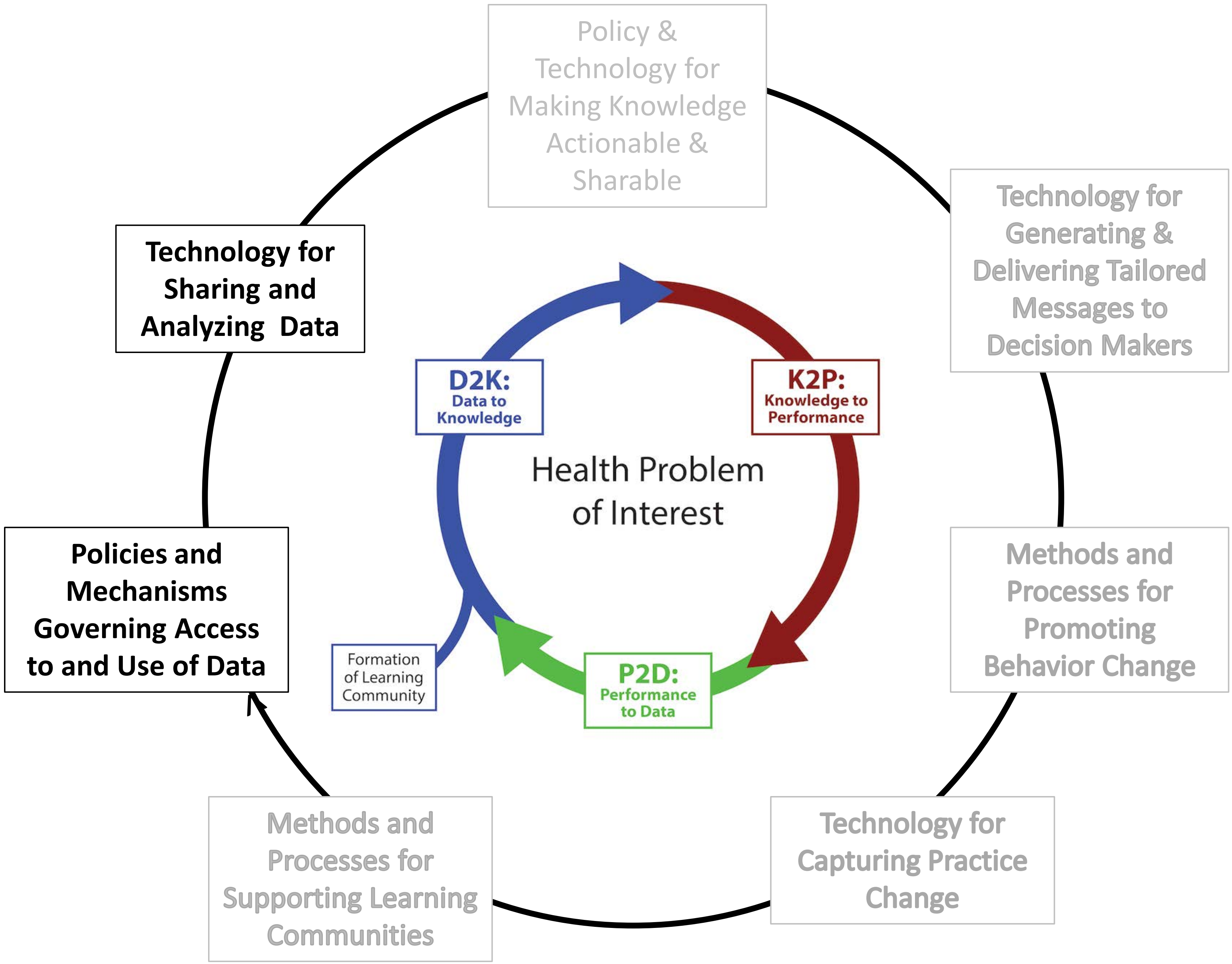
LHS Infrastructure: A *Platform* Supporting Multiple Simultaneous Learning Cycles



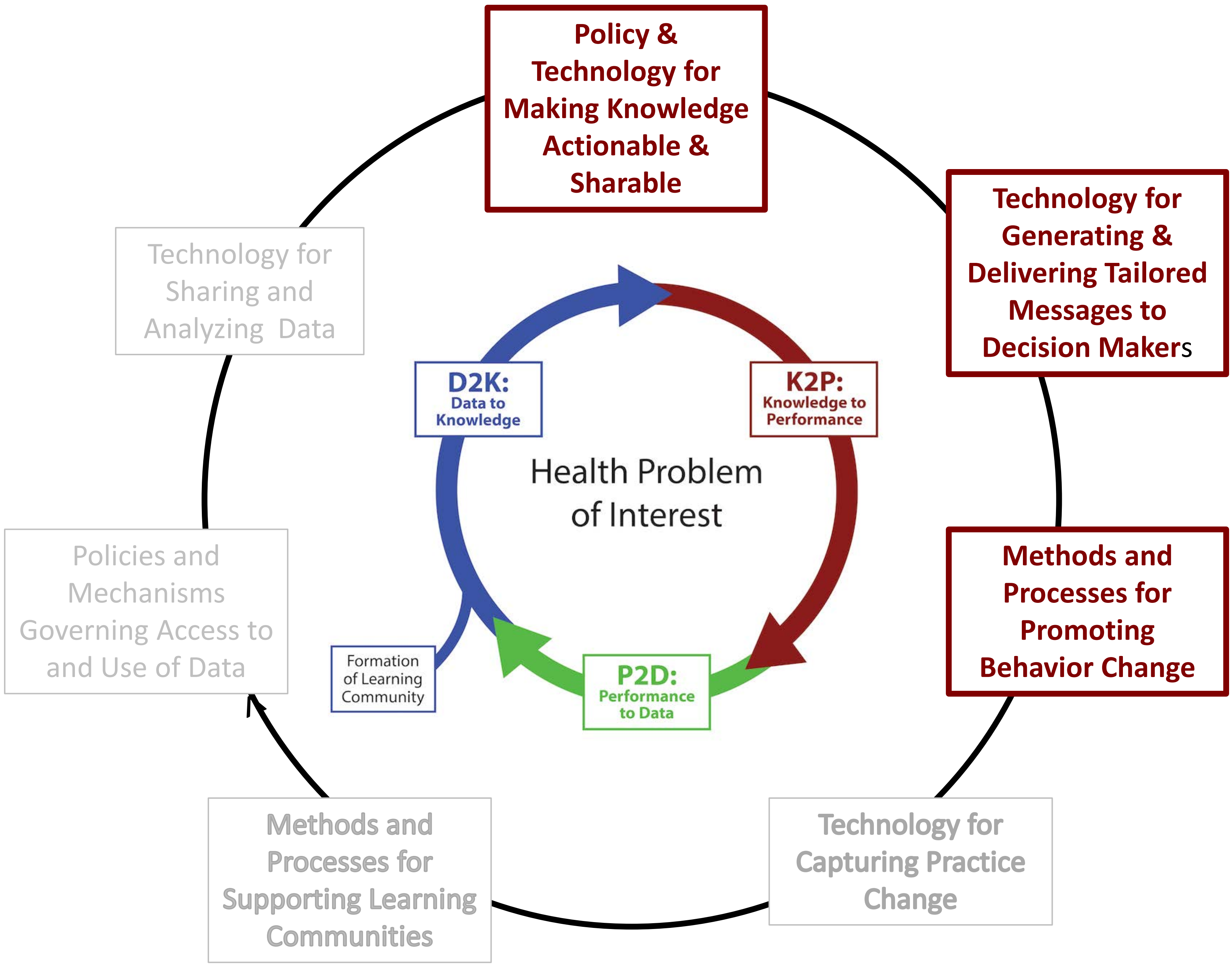
LHS Platform as a Set of Integrated Services



In Relative Terms, What Exists Today



And Our Focus is Here, at K2P…

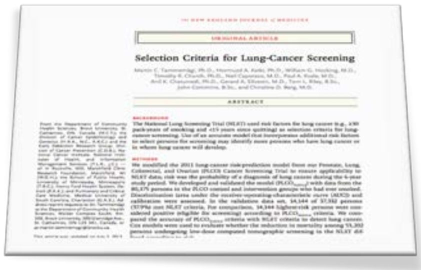


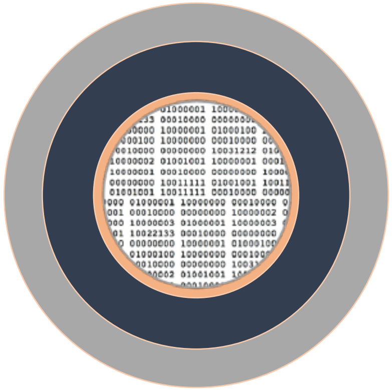
Main Menu

- Preliminaries
- The LHS and Infrastructural Services to Support It
- The "Keystone" Role of Persistent Computable Knowledge
  - And the Concept of a "K2P" Service
- Doing this at Scale: Vision of a Computable Knowledge Ecosystem…
- Goals and Plans for This Meeting

Persistent Knowledge

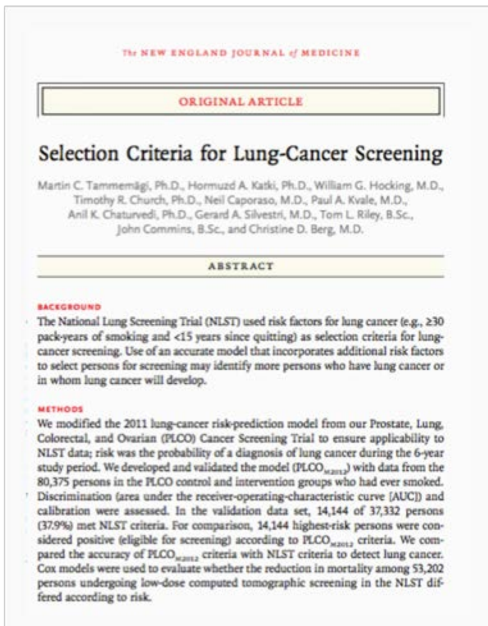
- Knowledge: The result of an analytical and/or deliberative process that holds significance for an identified community.
- Persistence: A representation exists at any point in time
- Persistent ≠ Static
- Persistent knowledge can be represented in two ways:
  - human readable
  - machine-executable


Selection Criteria for Lung-Cancer Screening



Two Complementary Ways to Represent Knowledge

Present: Human readable in words & pictures


The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Selection Criteria for Lung-Cancer Screening

Martin C. Tammemägi, Ph.D., Hormuzd A. Katki, Ph.D., William G. Hocking, M.D., Timothy R. Church, Ph.D., Neil Caporaso, M.D., Paul A. Kvale, M.D., Anil K. Chaturvedi, Ph.D., Gerard A. Silvestri, M.D., Tom L. Riley, B.Sc., John Commins, B.Sc., and Christine D. Berg, M.D.

ABSTRACT

BACKGROUND

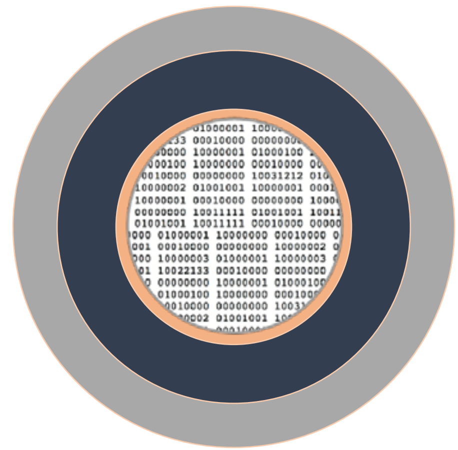
The National Lung Screening Trial (NLST) used risk factors for lung cancer (e.g., ≥30 pack-years of smoking and <15 years since quitting) as selection criteria for lung-cancer screening. Use of an accurate model that incorporates additional risk factors to select persons for screening may identify more persons who have lung cancer or in whom lung cancer will develop.

METHODS

We modified the 2011 lung-cancer risk-prediction model from our Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial to ensure applicability to NLST data; risk was the probability of a diagnosis of lung cancer during the 6-year study period. We developed and validated the model ($PLCO_{M2012}$) with data from the 80,375 persons in the PLCO control and intervention groups who had ever smoked. Discrimination (area under the receiver-operating-characteristic curve [AUC]) and calibration were assessed. In the validation data set, 14,144 of 37,332 persons (37.9%) met NLST criteria. For comparison, 14,144 highest-risk persons were considered positive (eligible for screening) according to $PLCO_{M2012}$ criteria. We compared the accuracy of $PLCO_{M2012}$ criteria with NLST criteria to detect lung cancer. Cox models were used to evaluate whether the reduction in mortality among 53,202 persons undergoing low-dose computed tomographic screening in the NLST differed according to risk.

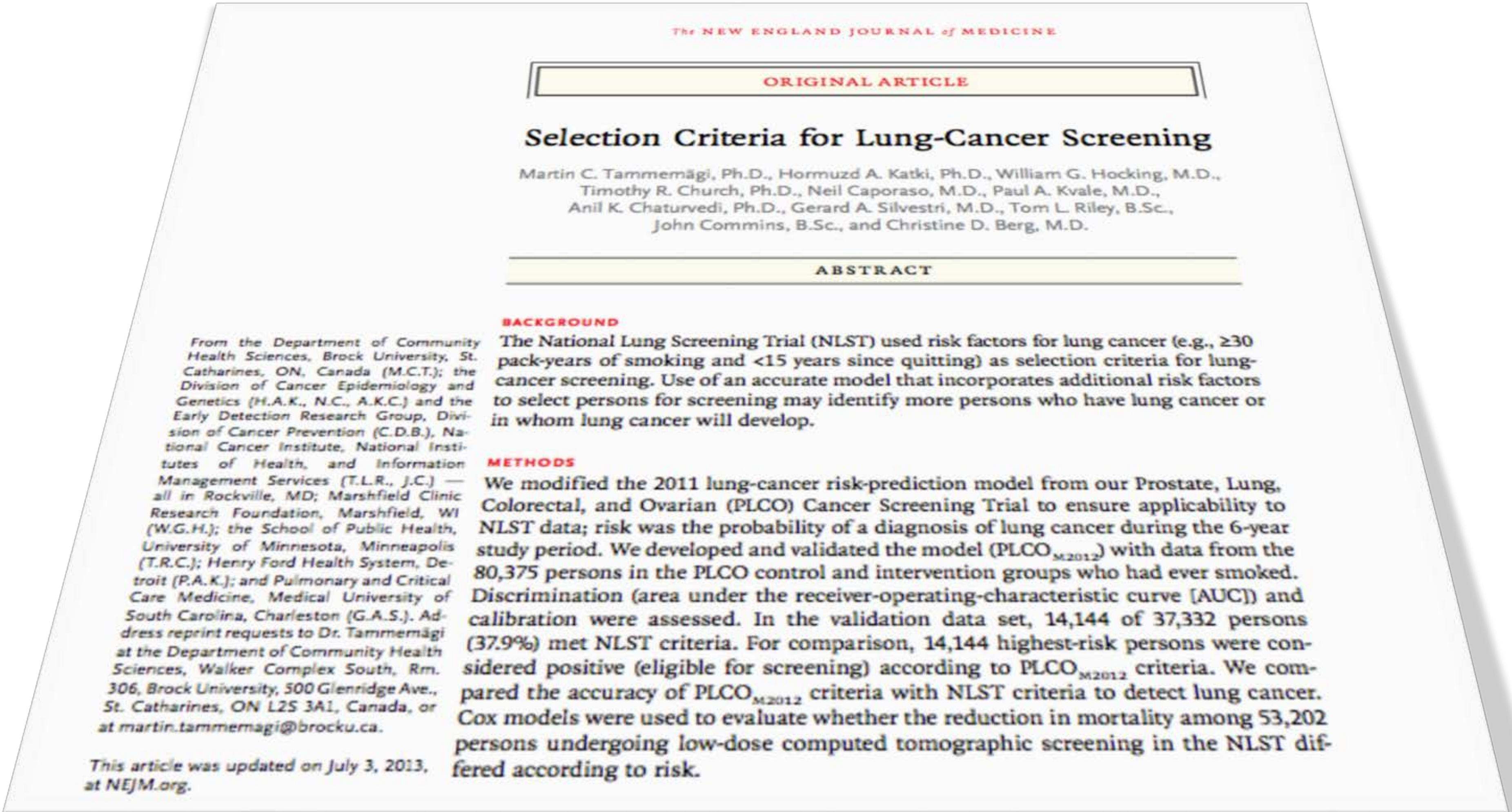
Library Holdings: Books & Journals

Future: Computable (machine-executable) in code



Library Holdings: Will add Digital Knowledge Objects

Example: Human Readable Knowledge



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Selection Criteria for Lung-Cancer Screening

Martin C. Tammemägi, Ph.D., Hormuzd A. Katki, Ph.D., William G. Hocking, M.D., Timothy R. Church, Ph.D., Neil Caporaso, M.D., Paul A. Kvale, M.D., Anil K. Chaturvedi, Ph.D., Gerard A. Silvestri, M.D., Tom L. Riley, B.Sc., John Commins, B.Sc., and Christine D. Berg, M.D.

ABSTRACT

BACKGROUND

The National Lung Screening Trial (NLST) used risk factors for lung cancer (e.g., ≥30 pack-years of smoking and <15 years since quitting) as selection criteria for lung-cancer screening. Use of an accurate model that incorporates additional risk factors to select persons for screening may identify more persons who have lung cancer or in whom lung cancer will develop.

METHODS

We modified the 2011 lung-cancer risk-prediction model from our Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial to ensure applicability to NLST data; risk was the probability of a diagnosis of lung cancer during the 6-year study period. We developed and validated the model ($PLCO_{M2012}$) with data from the 80,375 persons in the PLCO control and intervention groups who had ever smoked. Discrimination (area under the receiver-operating-characteristic curve [AUC]) and calibration were assessed. In the validation data set, 14,144 of 37,332 persons (37.9%) met NLST criteria. For comparison, 14,144 highest-risk persons were considered positive (eligible for screening) according to $PLCO_{M2012}$ criteria. We compared the accuracy of $PLCO_{M2012}$ criteria with NLST criteria to detect lung cancer. Cox models were used to evaluate whether the reduction in mortality among 53,202 persons undergoing low-dose computed tomographic screening in the NLST differed according to risk.

From the Department of Community Health Sciences, Brock University, St. Catharines, ON, Canada (M.C.T.); the Division of Cancer Epidemiology and Genetics (H.A.K., N.C., A.K.C.) and the Early Detection Research Group, Division of Cancer Prevention (C.D.B.), National Cancer Institute, National Institutes of Health, and Information Management Services (T.L.R., J.C.) — all in Rockville, MD; Marshfield Clinic Research Foundation, Marshfield, WI (W.G.H.); the School of Public Health, University of Minnesota, Minneapolis (T.R.C.); Henry Ford Health System, Detroit (P.A.K.); and Pulmonary and Critical Care Medicine, Medical University of South Carolina, Charleston (G.A.S.). Address reprint requests to Dr. Tammemägi at the Department of Community Health Sciences, Walker Complex South, Rm. 306, Brock University, 500 Glenridge Ave., St. Catharines, ON L2S 3A1, Canada, or at martin.tammemagi@brocku.ca.

This article was updated on July 3, 2013, at NEJM.org.

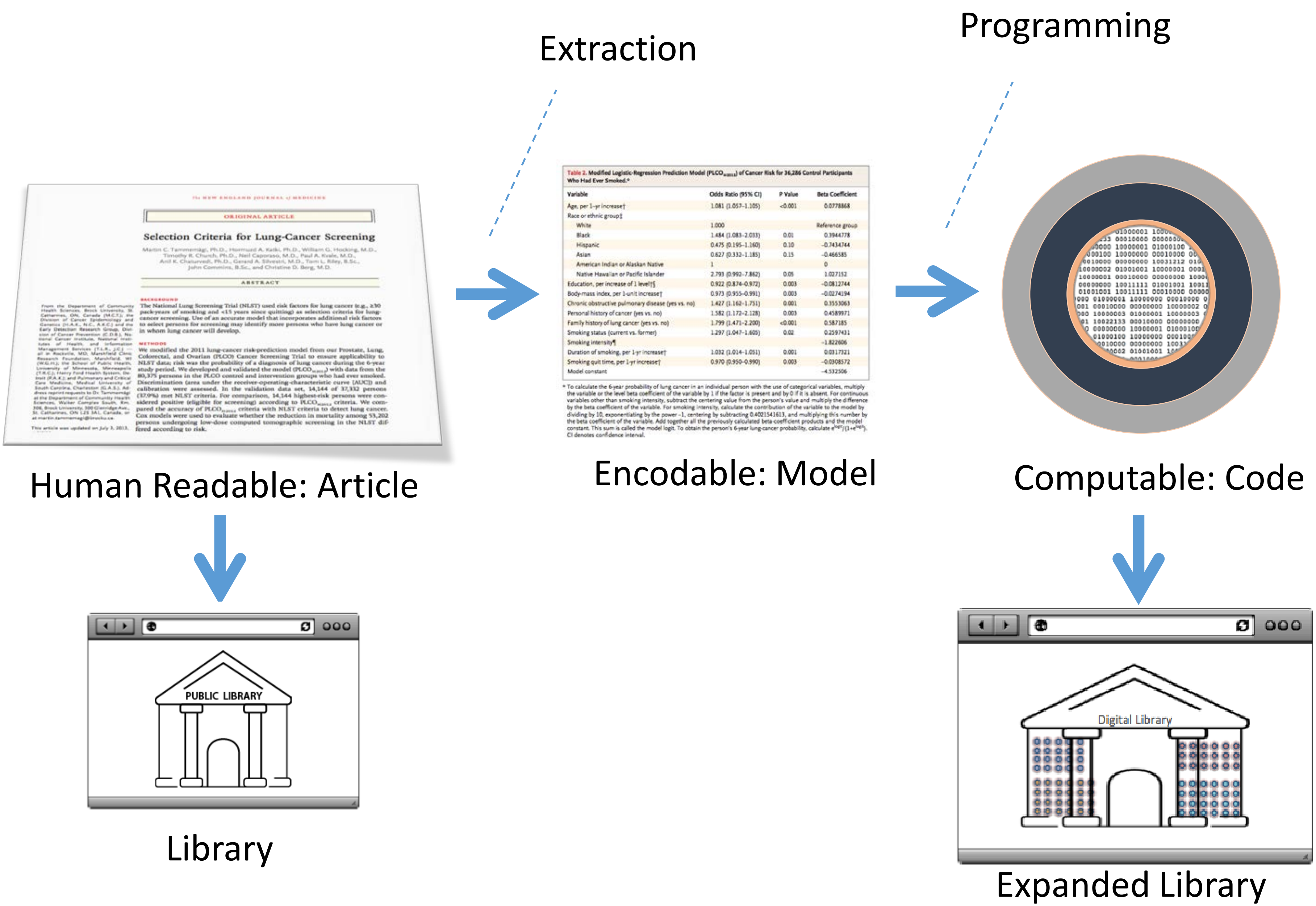
The New Knowledge is Expressed in a Model

Table 2. Modified Logistic-Regression Prediction Model ($PLCO_{M2012}$) of Cancer Risk for 36,286 Control Participants Who Had Ever Smoked.*

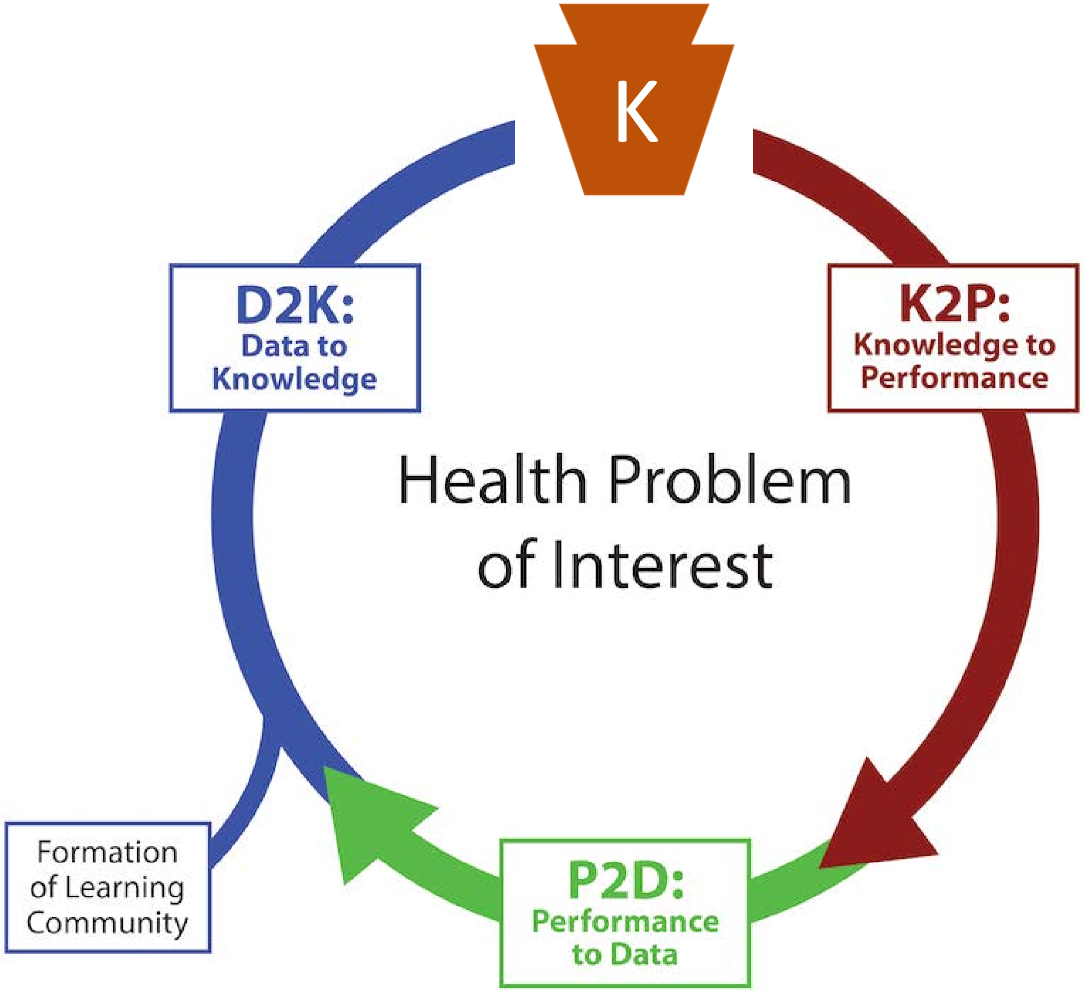
Variable	Odds Ratio (95% CI)	P Value	Beta Coefficient
Age, per 1–yr increase†	1.081 (1.057–1.105)	<0.001	0.0778868
Race or ethnic group‡			
White	1.000		Reference group
Black	1.484 (1.083–2.033)	0.01	0.3944778
Hispanic	0.475 (0.195–1.160)	0.10	−0.7434744
Asian	0.627 (0.332–1.185)	0.15	−0.466585
American Indian or Alaskan Native	1		0
Native Hawaiian or Pacific Islander	2.793 (0.992–7.862)	0.05	1.027152
Education, per increase of 1 level†§	0.922 (0.874–0.972)	0.003	−0.0812744
Body-mass index, per 1-unit increase†	0.973 (0.955–0.991)	0.003	−0.0274194
Chronic obstructive pulmonary disease (yes vs. no)	1.427 (1.162–1.751)	0.001	0.3553063
Personal history of cancer (yes vs. no)	1.582 (1.172–2.128)	0.003	0.4589971
Family history of lung cancer (yes vs. no)	1.799 (1.471–2.200)	<0.001	0.587185
Smoking status (current vs. former)	1.297 (1.047–1.605)	0.02	0.2597431
Smoking intensity¶			−1.822606
Duration of smoking, per 1-yr increase†	1.032 (1.014–1.051)	0.001	0.0317321
Smoking quit time, per 1-yr increase†	0.970 (0.950–0.990)	0.003	−0.0308572
Model constant			−4.532506

* To calculate the 6-year probability of lung cancer in an individual person with the use of categorical variables, multiply the variable or the level beta coefficient of the variable by 1 if the factor is present and by 0 if it is absent. For continuous variables other than smoking intensity, subtract the centering value from the person's value and multiply the difference by the beta coefficient of the variable. For smoking intensity, calculate the contribution of the variable to the model by dividing by 10, exponentiating by the power −1, centering by subtracting 0.4021541613, and multiplying this number by the beta coefficient of the variable. Add together all the previously calculated beta-coefficient products and the model constant. This sum is called the model logit. To obtain the person's 6-year lung-cancer probability, calculate $e^{logit}/(1+e^{logit})$.
CI denotes confidence interval.

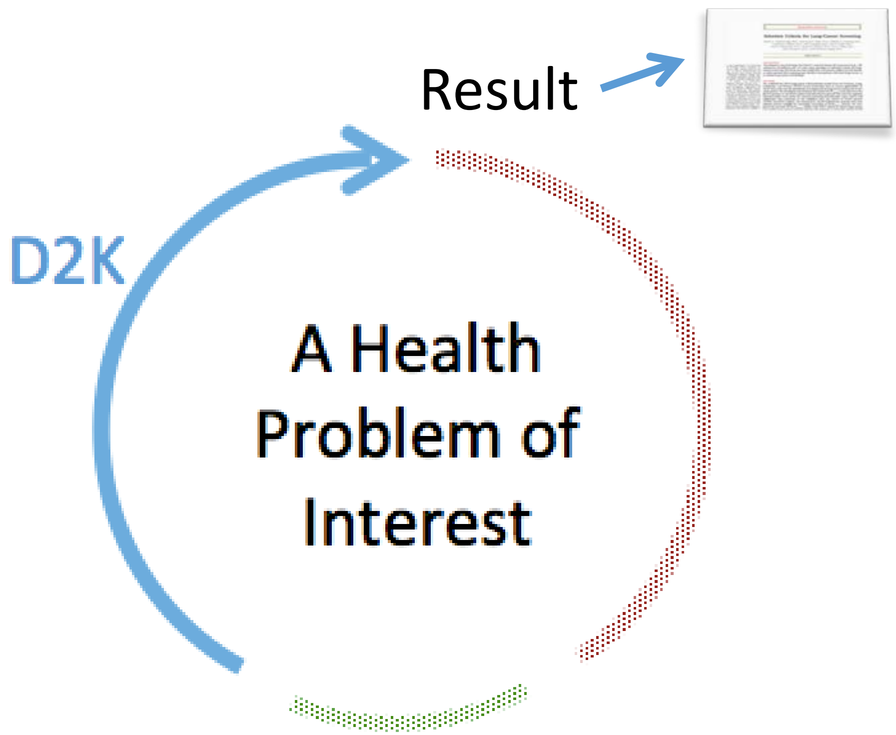
Envisioning An Extended Publication Pipeline



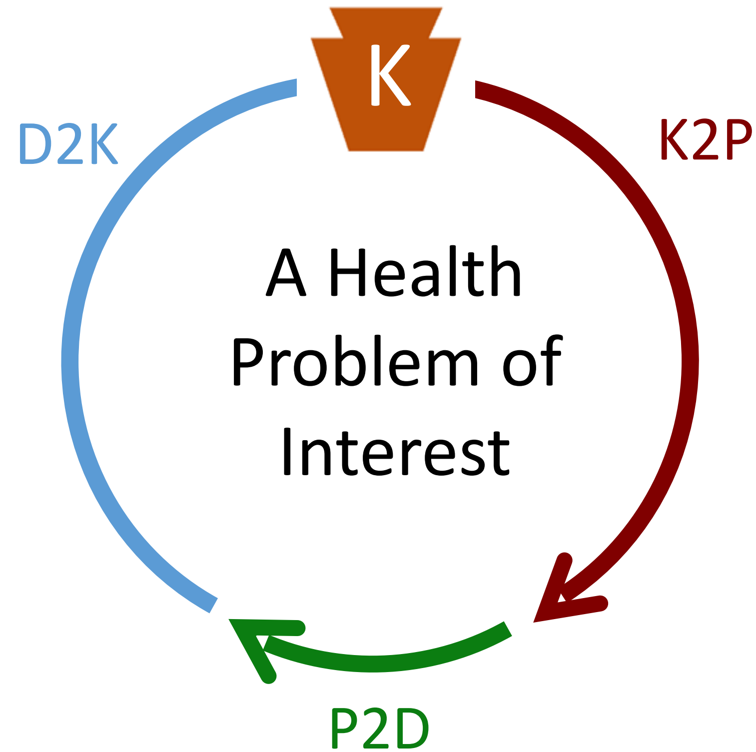
Computable, Persistent Knowledge is the LHS "Keystone"



The Keystone Enables Discovery Systems to Become Learning Systems

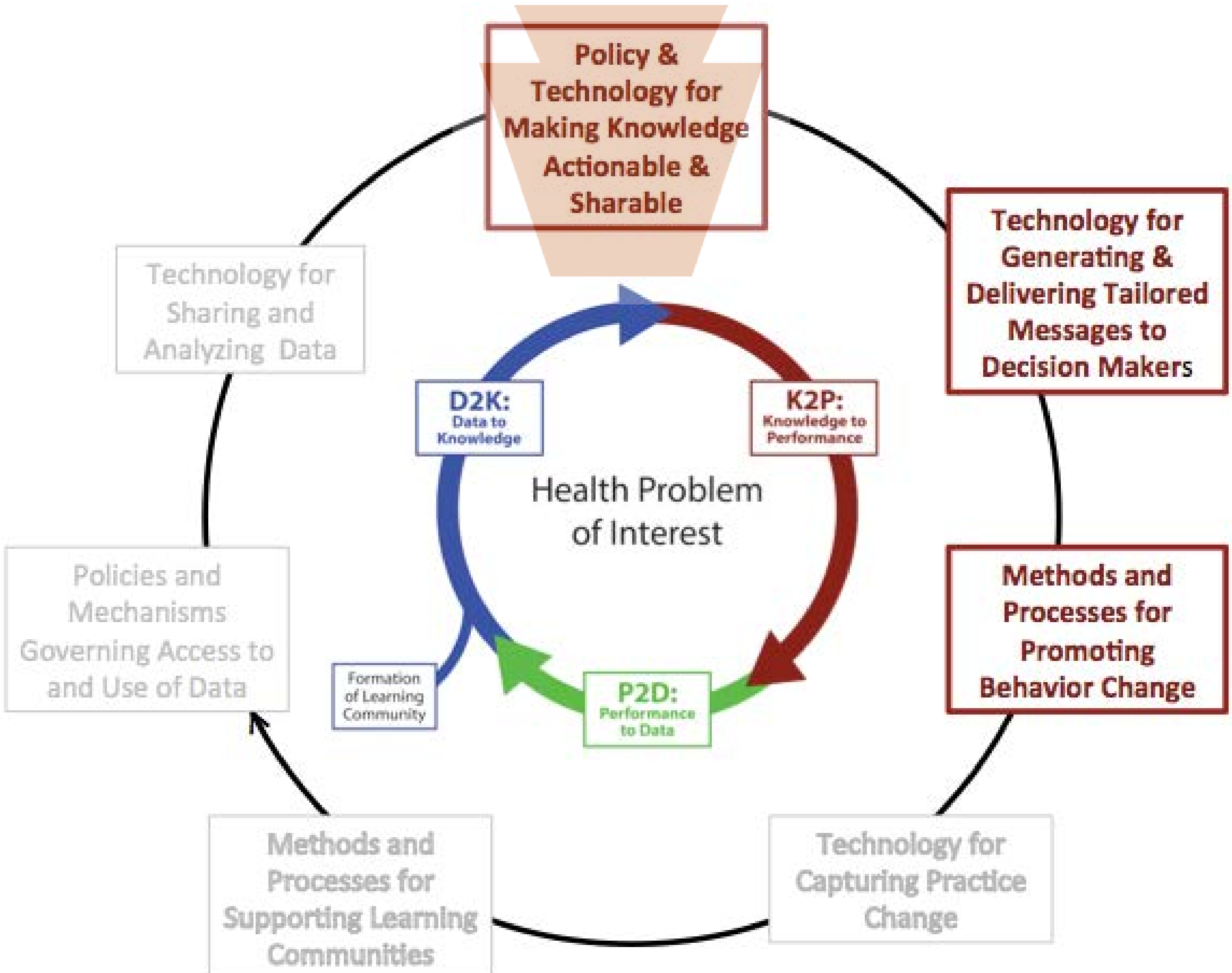


Discovery System



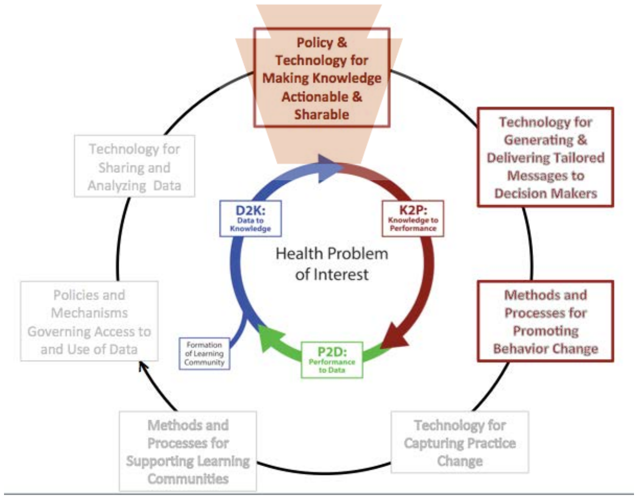
Learning System

Folding in the Concept of Infrastructure…
The LHS Requires a "K2P" Service



Minimum Requirements for a K2P Service

- Representation of knowledge in machine-executable forms
- Capability for rapid knowledge revision as the system learns
- Modular linking of related knowledge
- Sharing of knowledge across an ecosystem
- Scalable computation and delivery of tailored messages to inform practice



K2P Use Cases: CDS & Beyond

- Clinical: Bringing advice, generated from computable knowledge, to inform decisions of providers, consumers, and managers
- Research: Enhancing the scientific record, computable phenotypes, analytic "packages"
- Public Health: Event detection objects; rapid response deployment
- Education: Learning analytic objects, preparation for practice in an environment of ubiquitous knowledge

Main Menu

- Preliminaries
- The LHS and Infrastructural Services to Support It
- The “Keystone” Role of Persistent Computable Knowledge
  - And the Concept of a “K2P” Service
- Doing this at Scale: Vision of a Computable Knowledge Ecosystem…
- Goals and Plans for This Meeting

Much Has Been Said About *Data* FAIRness This Meeting is about ***Knowledge*** FAIRness

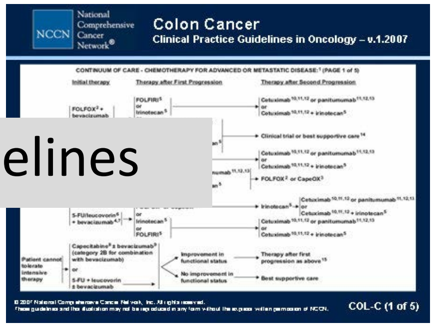
Making Knowledge:

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

FROM: https://www.force11.org/group/fairgroup/fairprinciples

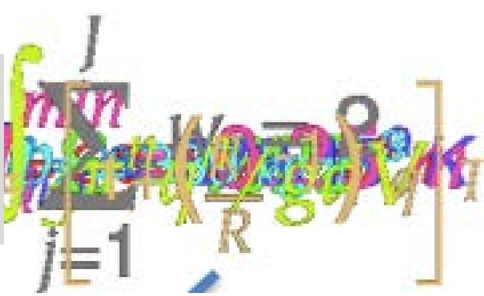
Approach to Knowledge FAIRness: Machine-executable Knowledge Objects

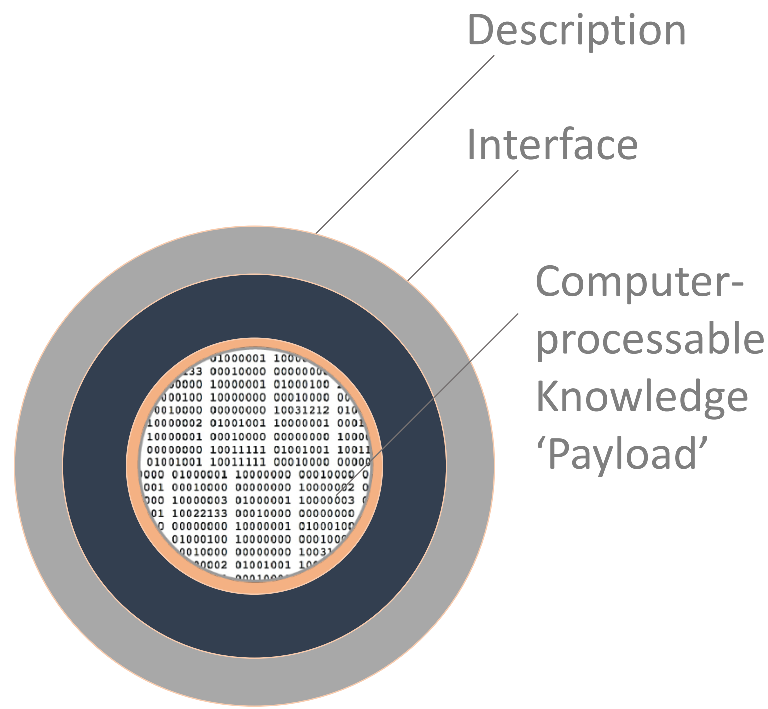
Guidelines



Articles

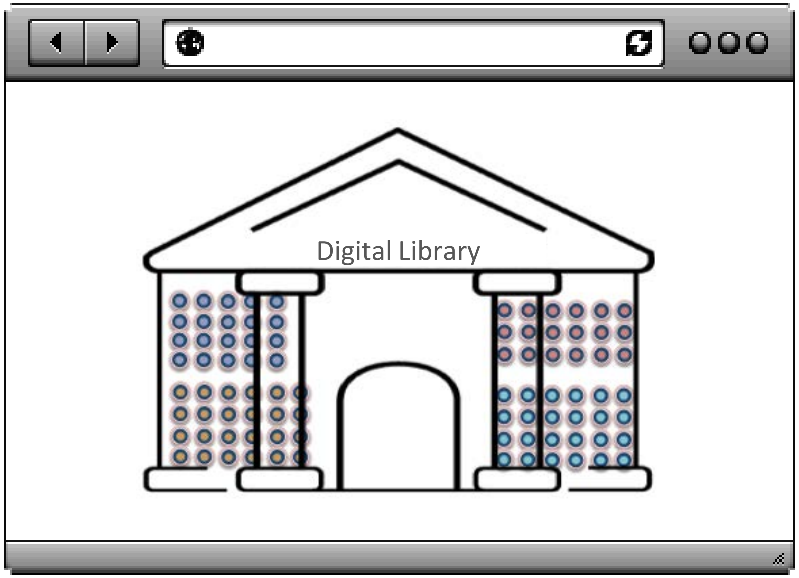
Local Analytical Results





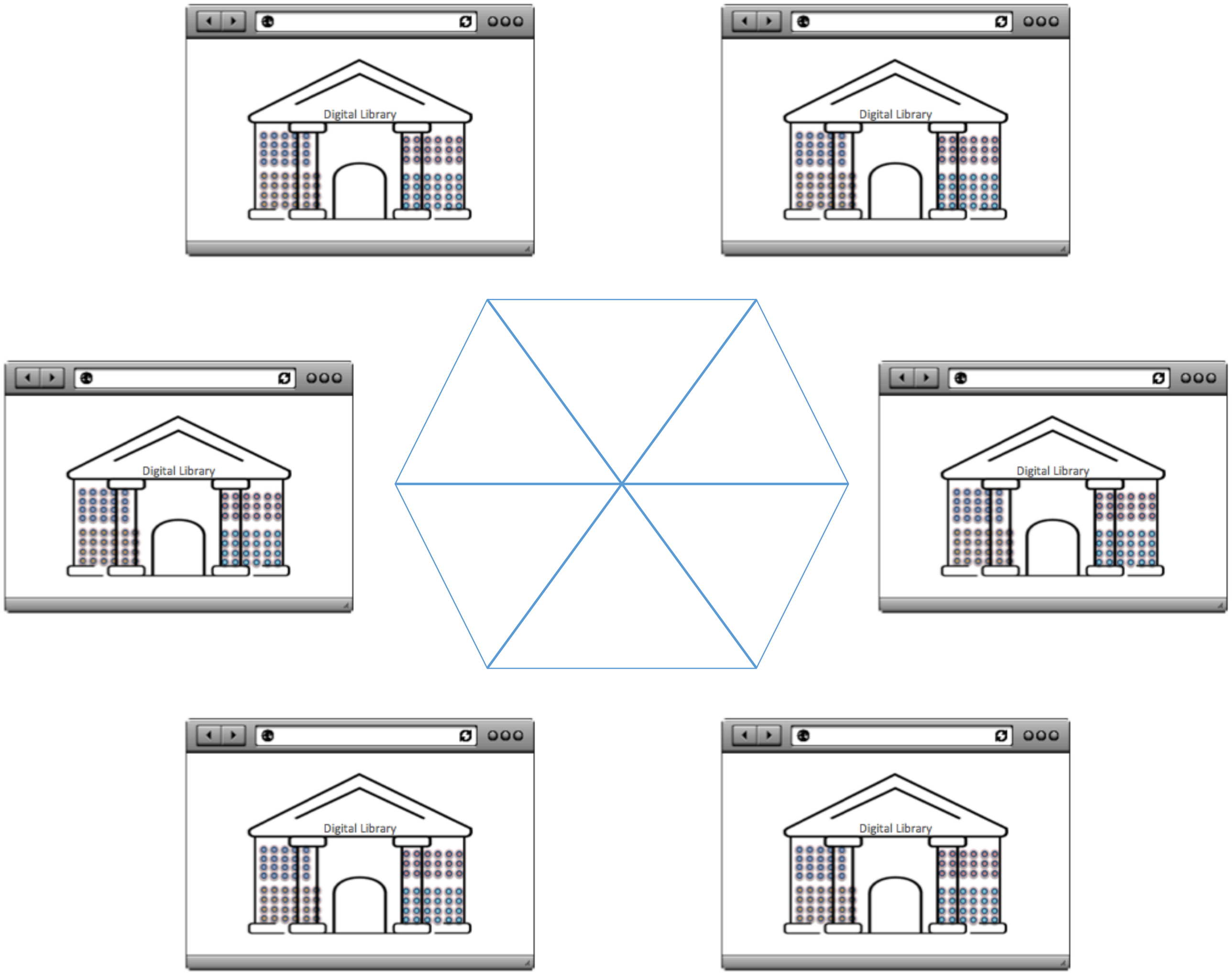
Knowledge Objects

And Digital Libraries to Manage and Share Computable Knowledge



Capability to curate and manage
online collections
of knowledge objects

And Networks of Digital Libraries to Enable a Computable Knowledge Ecosystem



Main Menu

- Preliminaries
- The LHS and Infrastructural Services to Support It
- The “Keystone” Role of Persistent Computable Knowledge
  - And the Concept of a “K2P” Service
- Doing this at Scale: Vision of a Computable Knowledge Ecosystem…
- Goals and Plans for This Meeting

Thanking the Planning Committee

- Julia Adler-Milstein
- Jane Blumenthal
- Milton Corn
- Chris Dymek
- Peter Embi
- Bob Greenes
- Ken Mandl
- Dan Masys
- Blackford Middleton
- Mark Musen

Meeting Goals

- To begin exploring a set of issues – not primarily about technology – that need to be understood to advance a community interested in computable biomedical knowledge .

- To begin building the core group that anticipates and works in support of establishing a larger computable biomedical knowledge community

Four Organizing Themes

A briefing paper on each has been distributed:

1. Knowledge infrastructure requirements
2. Establishing a trusted system
3. Metadata
4. IP and Copyright

The Plan: Following This Talk

Today

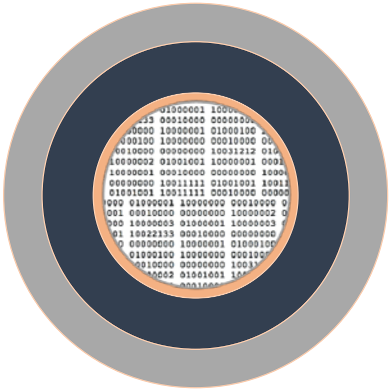
- Panel: Three ongoing efforts
- Briefing sessions keyed to the four themes (and briefing papers)
- Small group discussions keyed to the four themes and focused on specific questions
- Report out
- Dinner at the Gandy Dancer

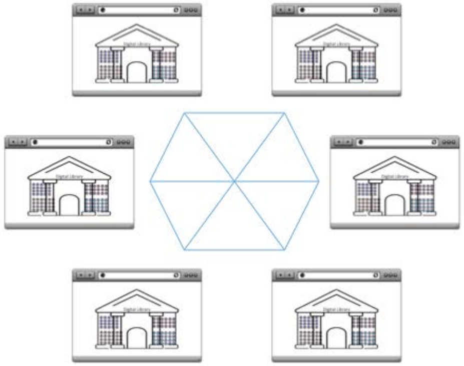
Tomorrow

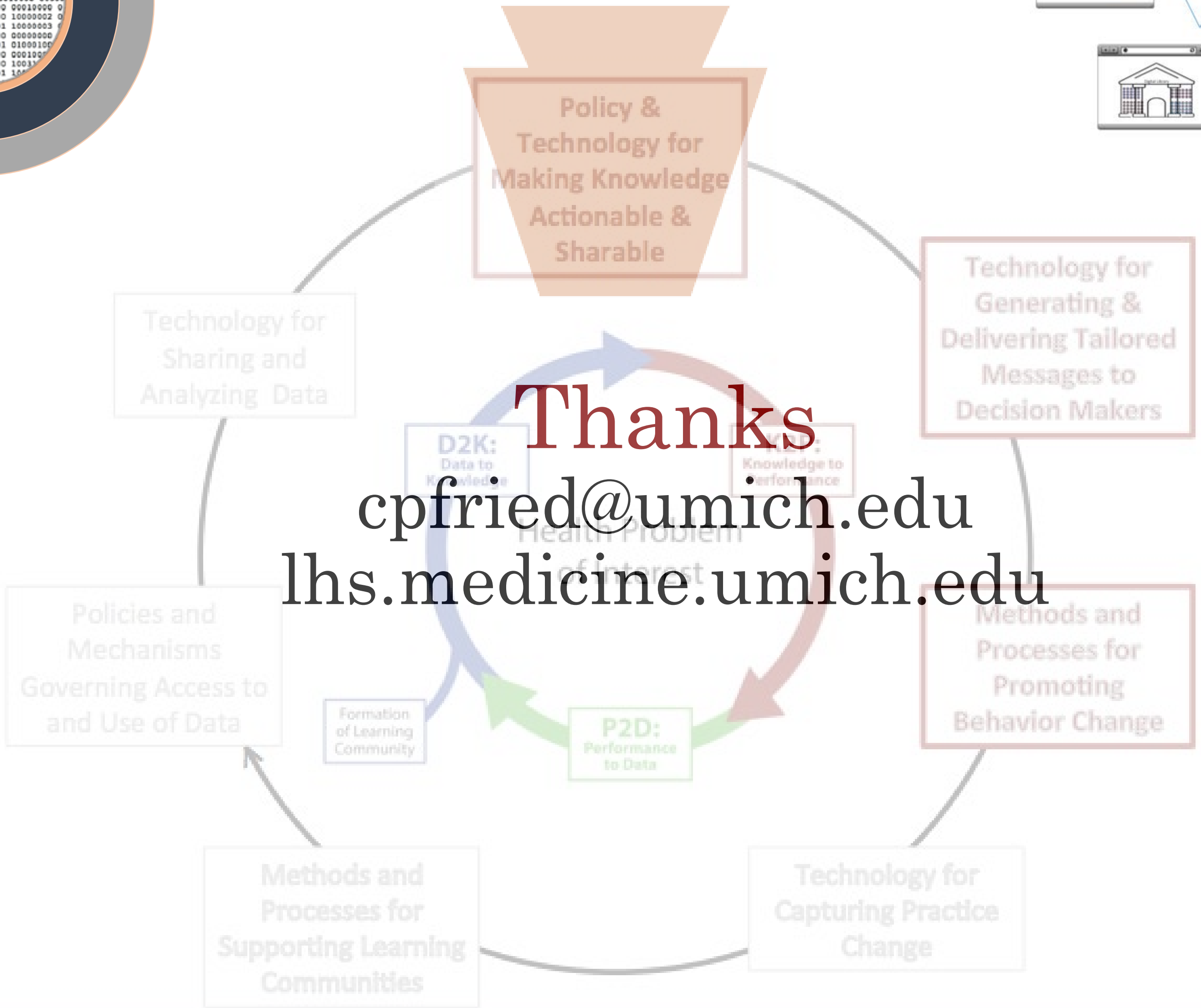
- Small groups resume around four themes
- Cross-fertilization groups
- Synthesizing discussion
- Moving forward from here

What Success Might Look Like

- One or more articles
- An open meeting in mid-2018
- A nascent organization or association with an existing organization
- Computable meta-knowledge (putting what we learned in computable form)
- Funding
- Opening a public library of computable knowledge







Thanks

cpfried@umich.edu

lhs.medicine.umich.edu