Duke | FORGE
HEALTH DATA IN ACTION

Quantitative Science to Optimize the Value of Cohort Data

Robert M Califf MD

Vice Chancellor for Health Data Science, Duke University

Advisor, Verily Life Sciences (Alphabet)

Board Chair, People Centered Research Foundation

Dictionary Definition of “Value” as a Noun

- the regard that something is held to deserve; the importance, worth, or usefulness of something.
  - Not just ”academic”: of theoretical interest but no practical importance
- a person's principles or standards of behavior; one's judgment of what is important in life.

Duke | FORGE HEALTH DATA IN ACTION

The Doctor and the Computer

In summary, the Seattle project represents an implementation of an approach that illustrates how doctors and patients can gain from carefully collected and computerized clinical experience. Predictions were that many such projects would be flourishing by 1980. The time course has been slower because of the difficulty of characterizing the complexity of chronic illness rather than because of problems with computer technology. **In the future, data banks will provide a reference library for each patient with chronic disease. Proper interpretation and use of computerized data will depend as much on wise doctors as any other source of data in the past.**

ROBERT M, CALIFF, MD

ROBERT A. ROSATI, MD

Cardiovascular Division, Department of Medicine

Duke University Medical Center

Durham, North Carolina

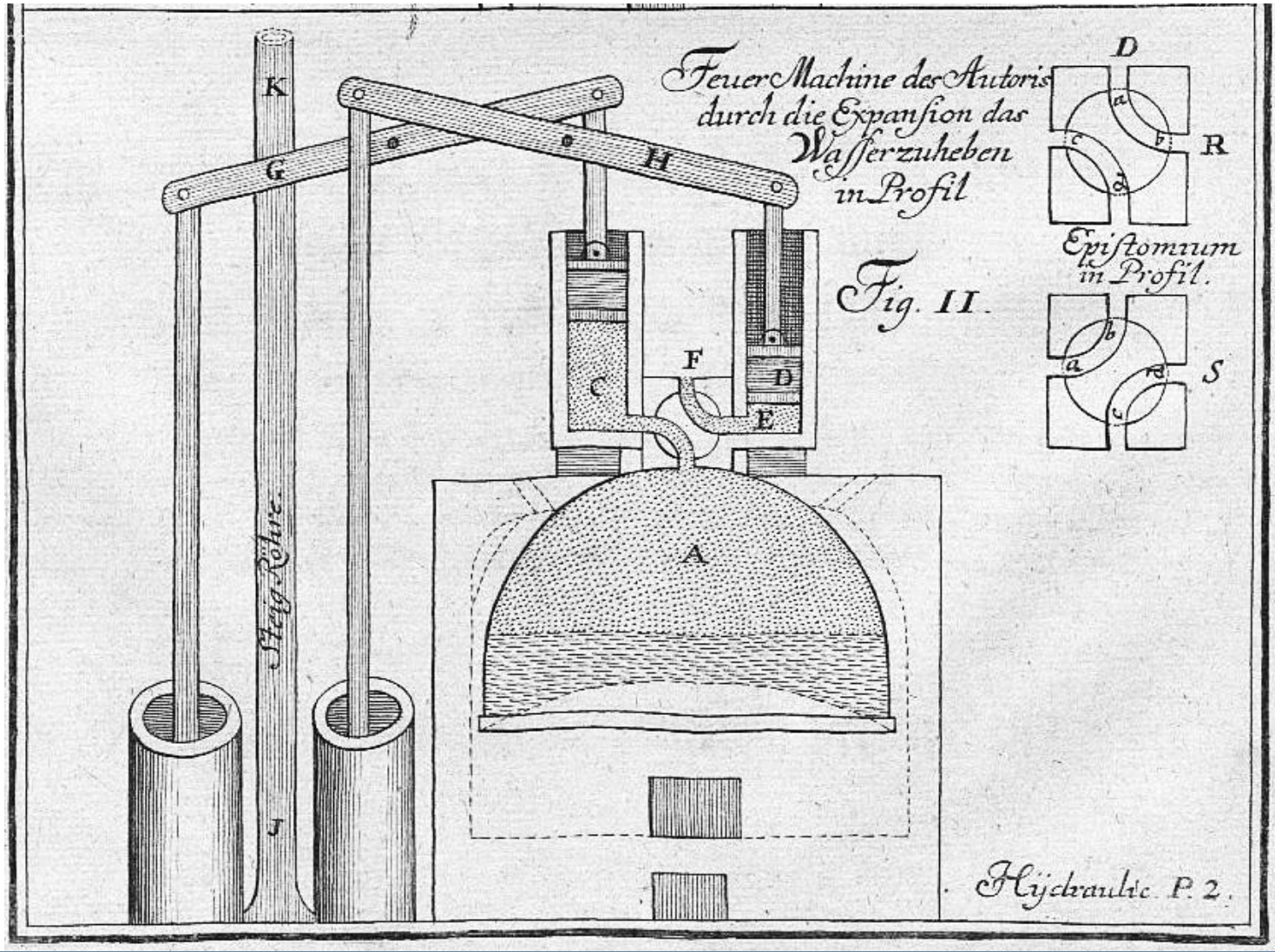
THE WESTERN JOURNAL OF MEDICINE **October 1981**

Duke | FORGE
HEALTH DATA IN ACTION

Four industrial revolutions

FIRST

Water and steam power mechanize production.



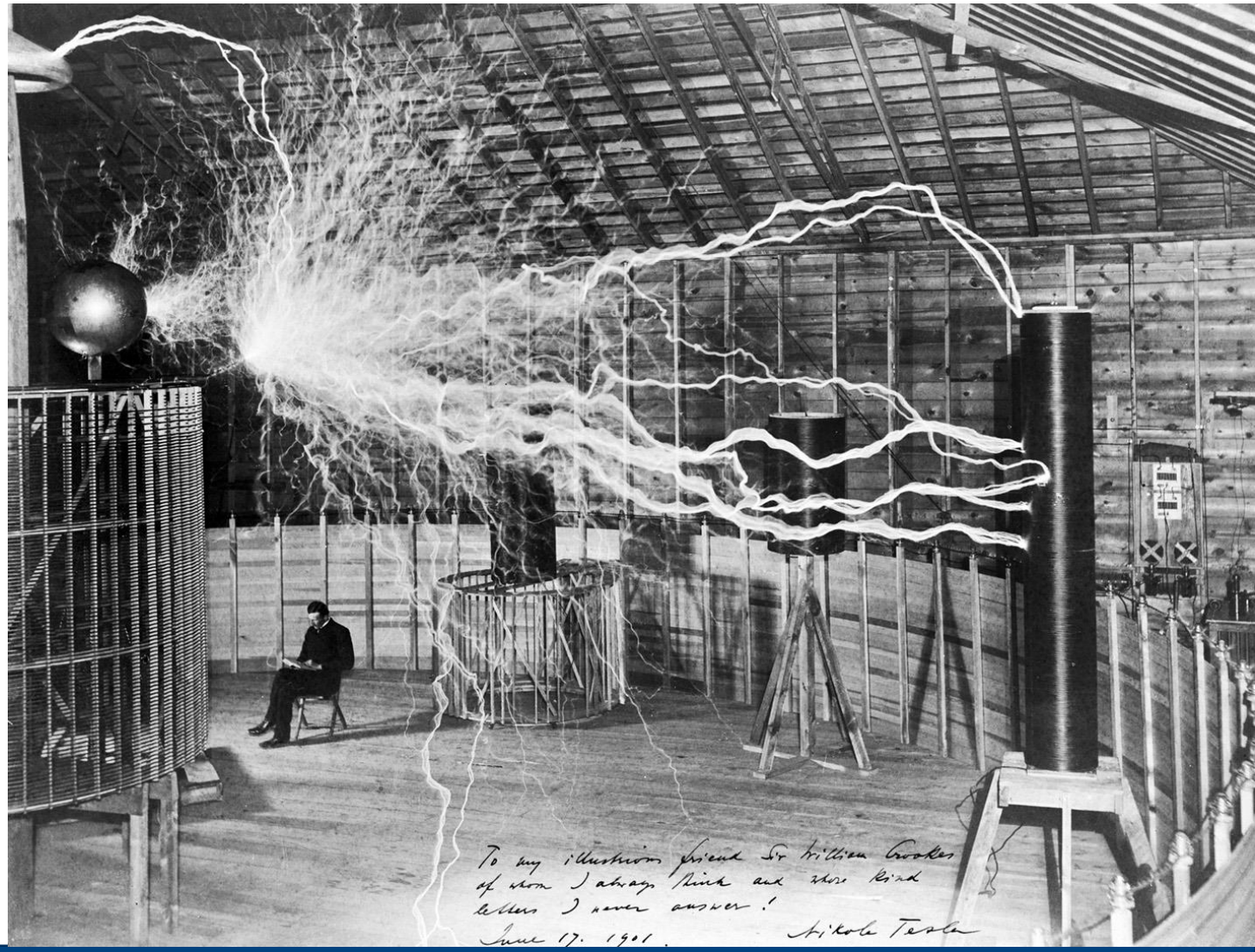
Jacob Leupold, Steam Engine, in *Theatri Machinarum Hydraulicarum II* (1720)

Duke | FORGE
HEALTH DATA IN ACTION

Four industrial revolutions

SECOND

Electric power creates mass production.

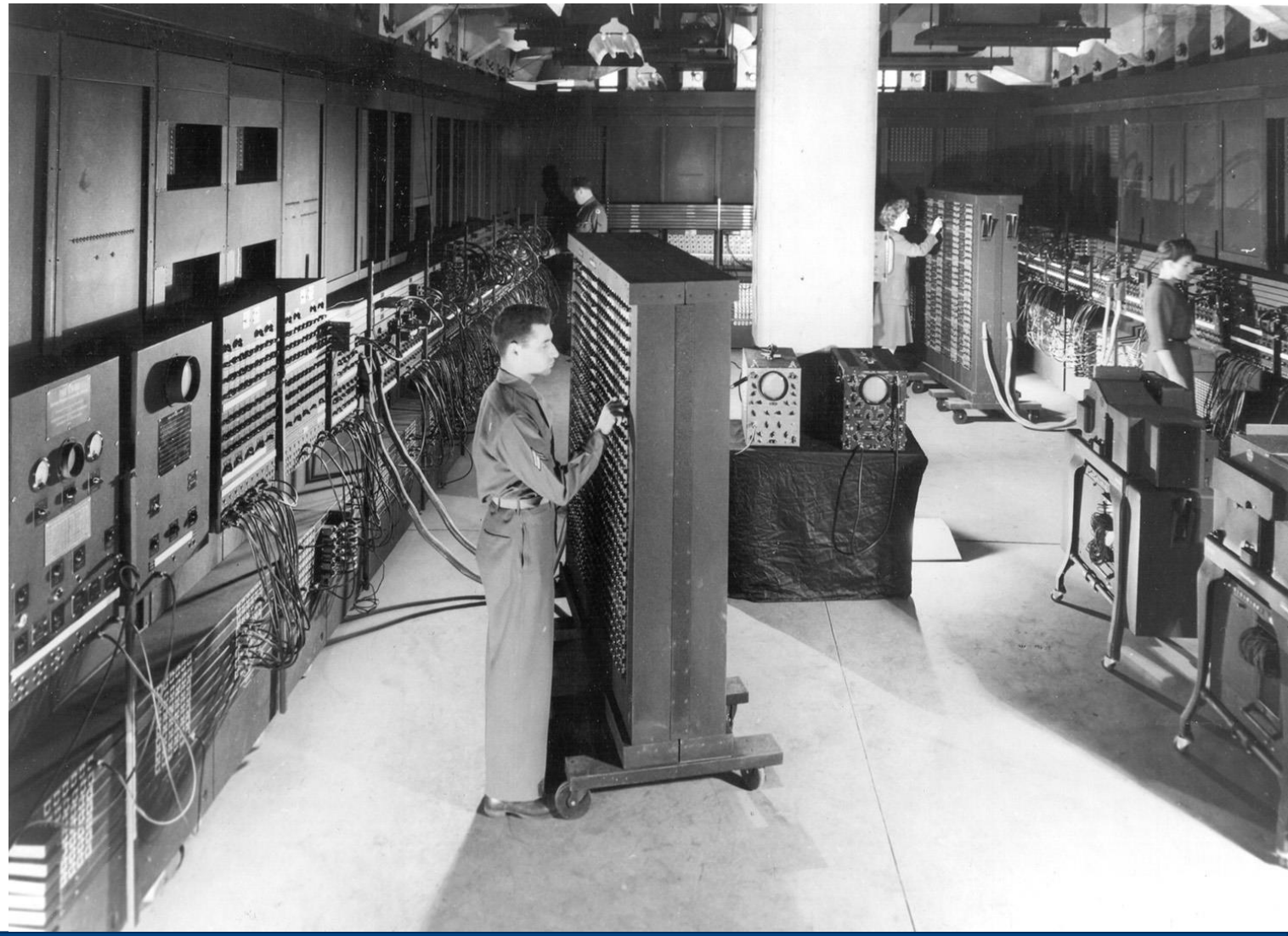


Duke | FORGE
HEALTH DATA IN ACTION

Four industrial revolutions

THIRD

Electronics and information technology automate production.



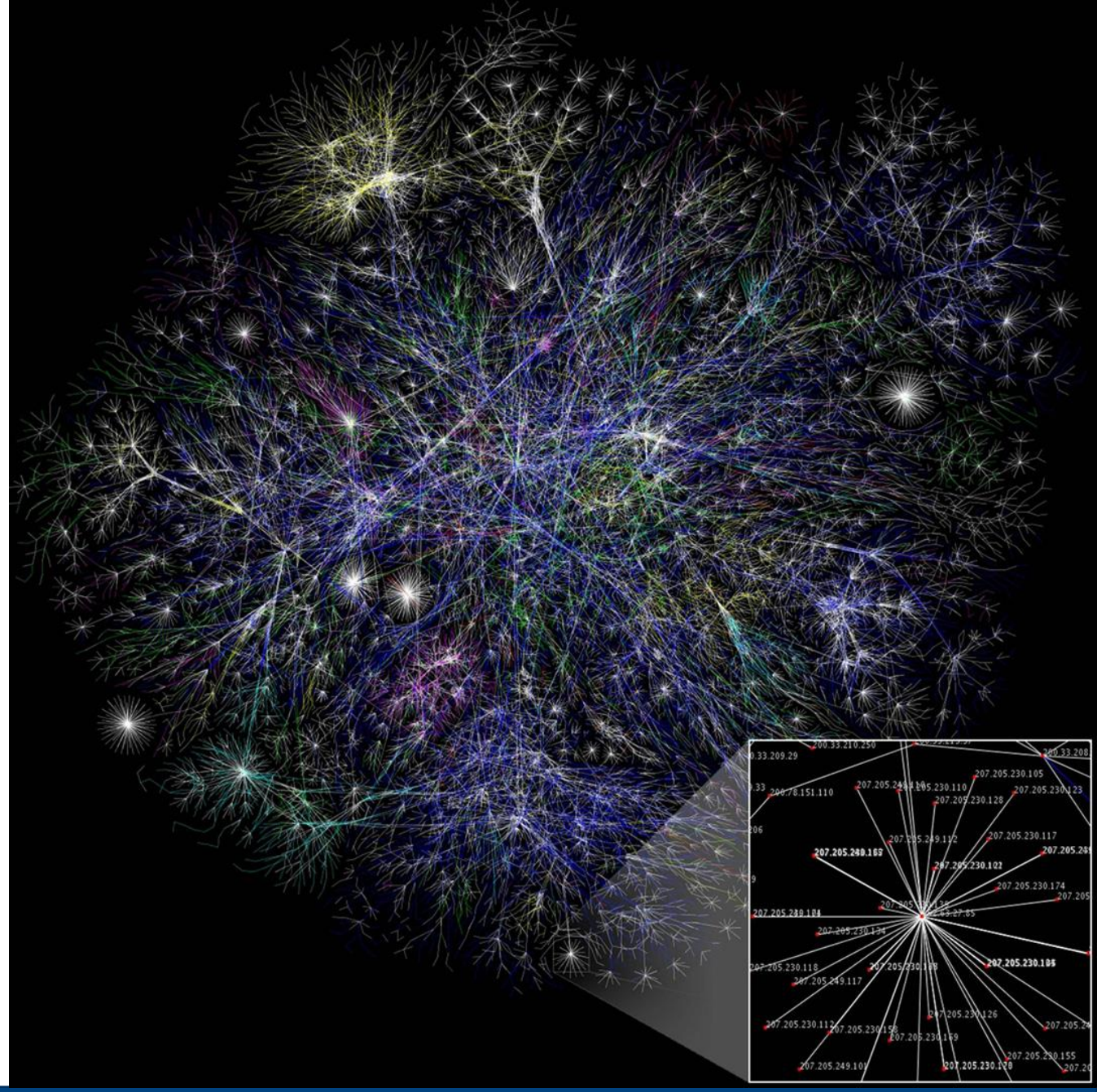
ENIAC digital computer. Unidentified U.S. Army photographer. Public Domain, https://commons.wikimedia.org/w/index.php?curid=978770

Duke | FORGE
HEALTH DATA IN ACTION

Four industrial revolutions

FOURTH

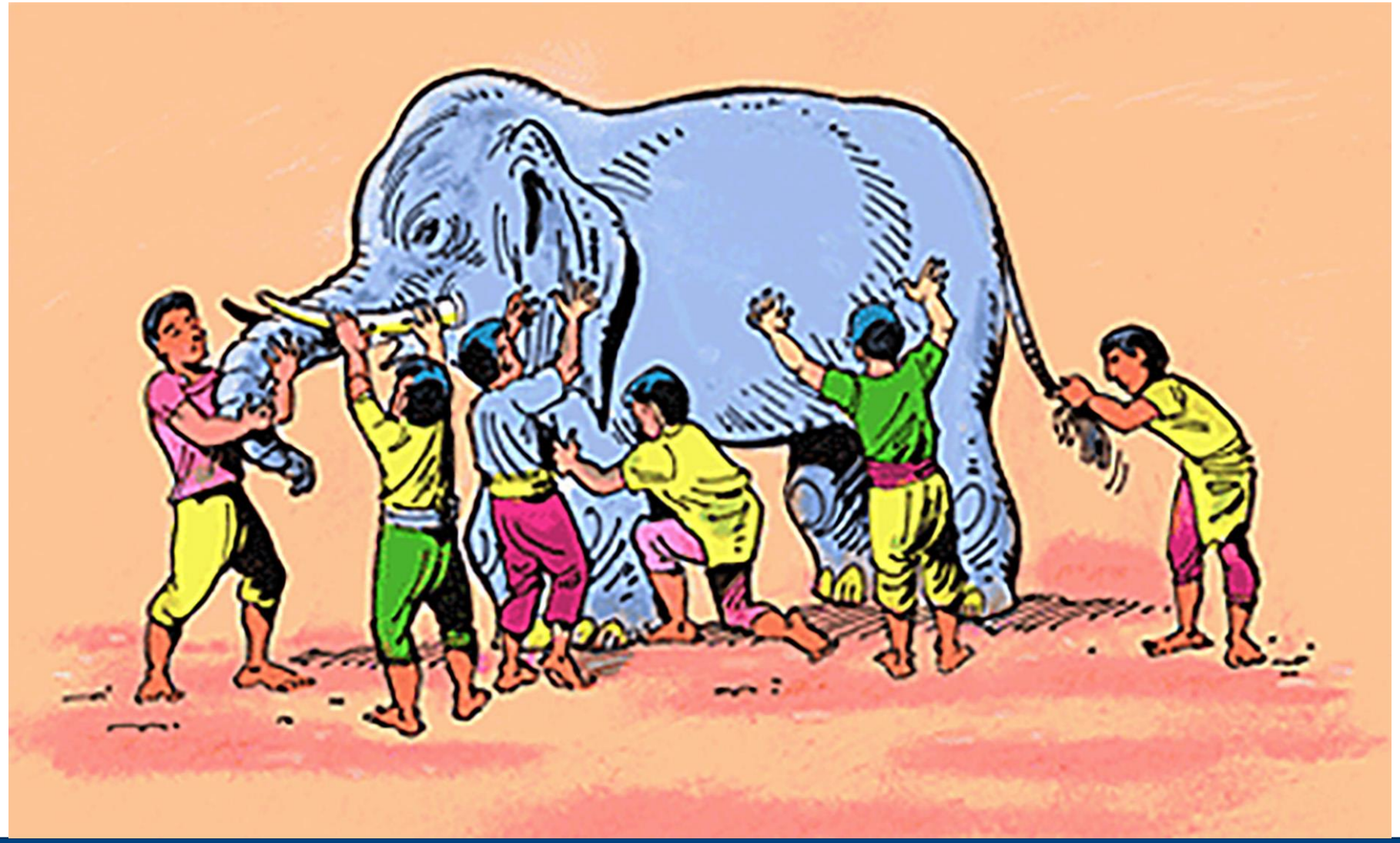
The digital revolution—characterized by a fusion of technologies—blurs the lines between physical, digital, and biological spheres.



Opte Project. Internet map. https://commons.wikimedia.org/wiki/File:Internet_map_1024.jpg

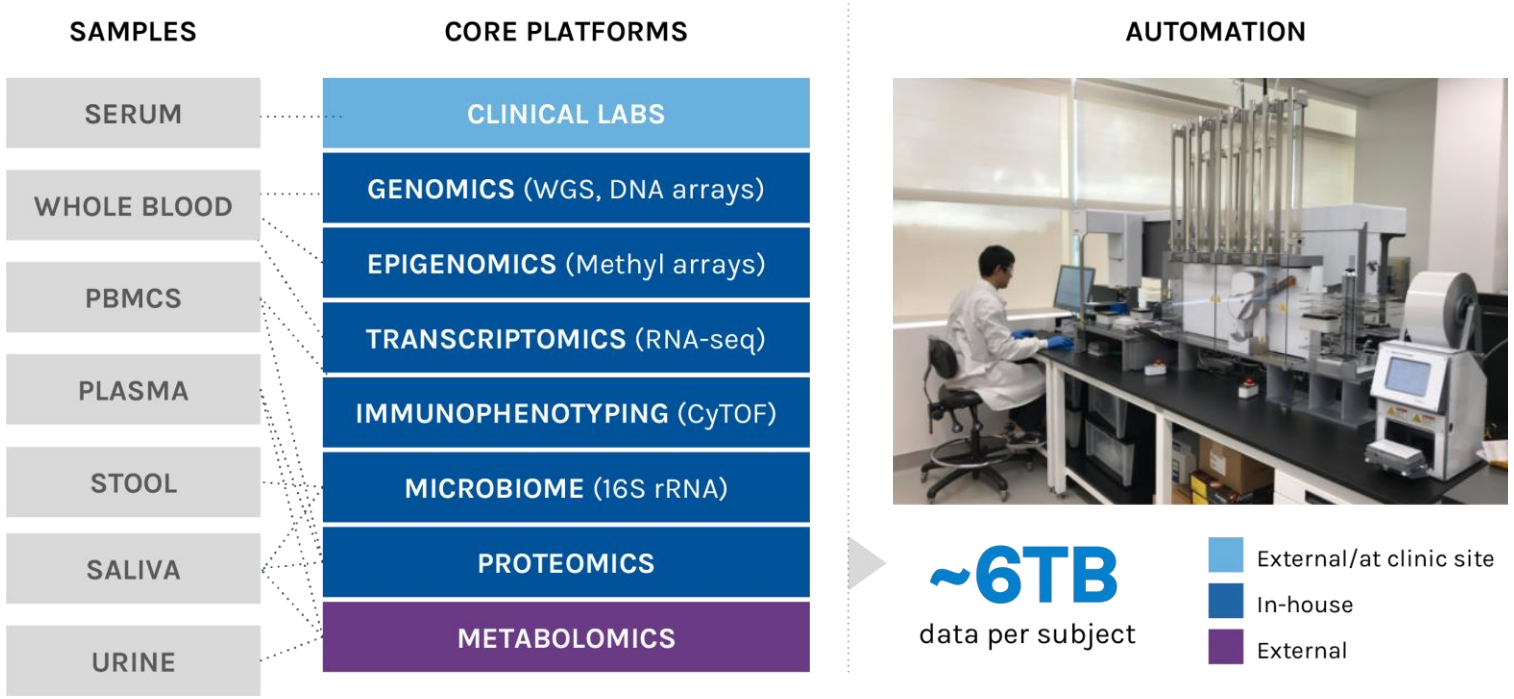
Duke | FORGE
HEALTH DATA IN ACTION

Duke | FORGE
HEALTH DATA IN ACTION



Duke | FORGE HEALTH DATA IN ACTION

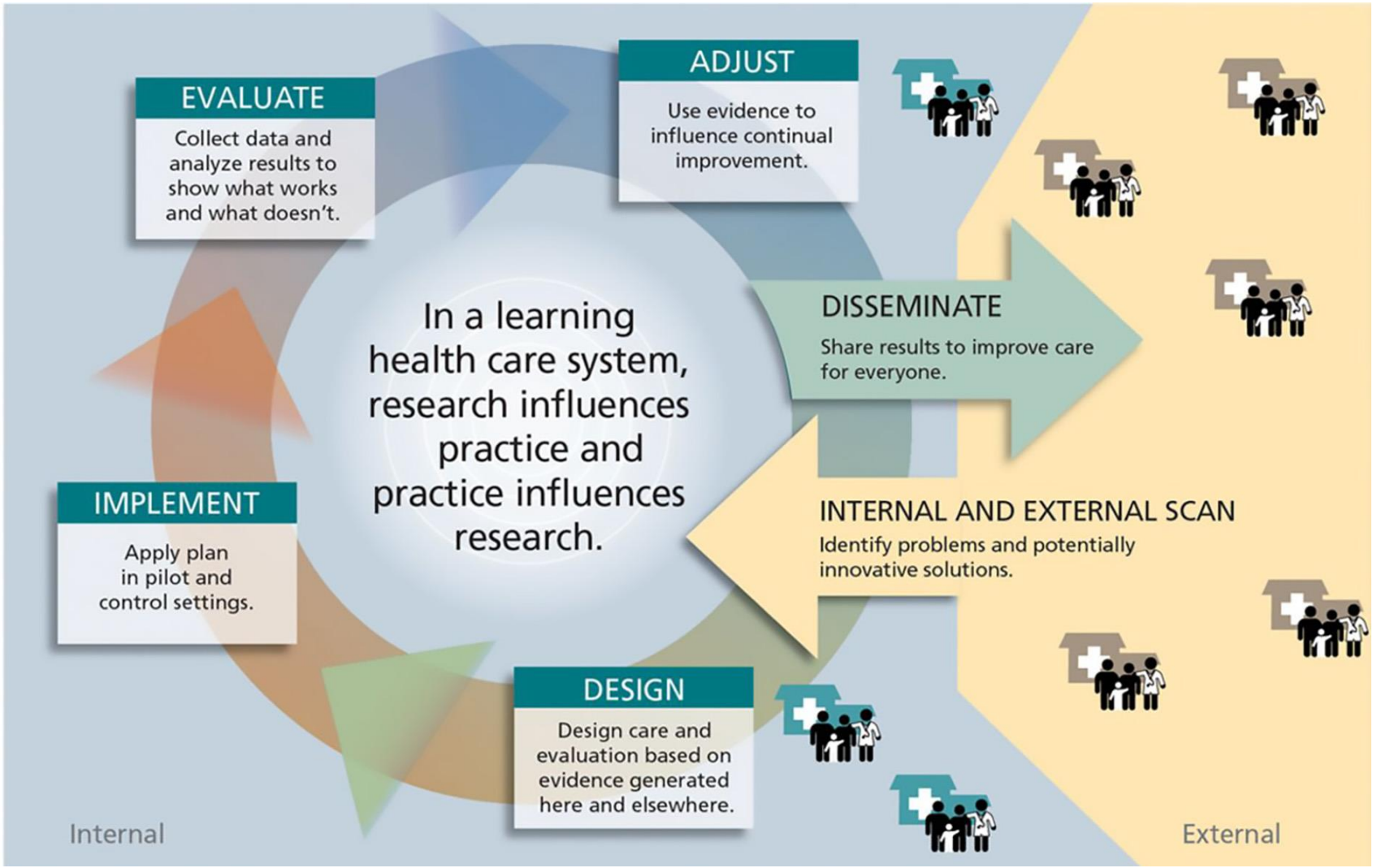
Deep molecular profiling



Project Baseline

Duke | FORGE
HEALTH DATA IN ACTION

Learning health care systems



www.fda.gov

Duke | FORGE HEALTH DATA IN ACTION

Continuous monitoring through passive sensors

Sleep sensor

Commercially available, placed under mattress to passively monitor multiple physiologic data parameters

Study watch

Investigational wrist-worn sensor for continuous recording of physiological and environmental data

Study hub

Safely sends device data to secure, encrypted Baseline database

App

Mobile interface for self-reported and passive data acquisitions

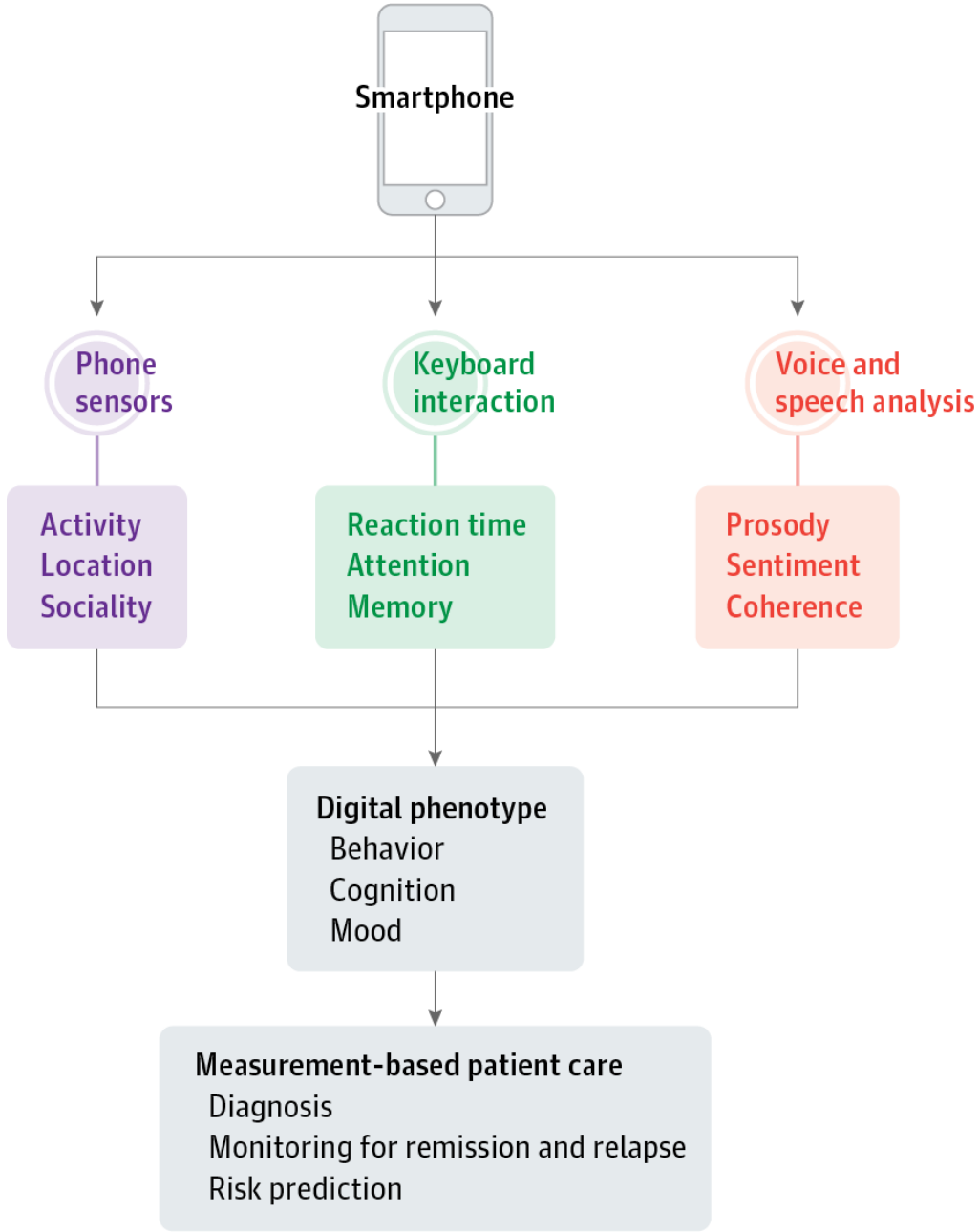
Project Baseline

Duke | FORGE HEALTH DATA IN ACTION

The process of digital phenotyping

Digital phenotyping involves collecting sensor, keyboard, and voice and speech data from smartphones to measure behavior, cognition, and mood.

The JAMA Network



Insel TR. Digital phenotyping technology for a new science of behavior. JAMA. 2017;318:1215-16. doi:10.1001/jama.2017.11295

Duke | FORGE
HEALTH DATA IN ACTION

Premise 1: Through a combination of clinical/epi expertise and quantitative methods, considerable effort needs to go into organizing and curating the data, especially the clinical, behavioral social and environmental data

- Even with standards as detailed by Dr. Keane, curating the data to place it in proper context and to assure a level of veracity and appropriate provenance should remain a high priority
- In addition the clinical, behavioral and social data are contextual

Duke | FORGE
HEALTH DATA IN ACTION

SECTIONS HOME SEARCH

The New York Times

SUBSCRIBE NOW LOG IN

Twitter's C.E.O., Dick Costolo, Is Set to Exit, Feeling Heat of Criticism

ROBOTICA EPISODE 5
Sex Dolls That Talk Back

STATE OF THE ART
For Twitter, Future Means Here and Now

Sidewalk Labs, a Start-Up Created by Google, Has Bold Aims to Improve City Living

Am
Bus
Eur
Reg

TECHNOLOGY

For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights

By STEVE LOHR AUG. 17, 2014

Duke | FORGE
HEALTH DATA IN ACTION

Print this Page for Your Records Close Window

Control/Tracking Number: 2018-SW-2357-ISPE
Activity: Symposium/Workshop
Current Date/Time: 2/14/2018 9:52:36 PM

Long Live the "Medical Data Janitors": International Data Quality Assurance Practices in Distributed Data Networks

Author Block: Judith C. Maro[1], Christian G. Reich[2], Keith Marsolo[3], Yoshiaki Uyama[4], Kristian B. Filion[5], Miriam C. J. M. Sturkenboom[6]. [1]Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; [2]IQVIA, Cambridge, MA; [3]Cincinnati Children's Hospital Medical Center, Cincinnatti, OH; [4]Pharmaceuticals and Medical Devices Agency, Tokyo, Japan; [5]McGill University, Montreal, QC, Canada; [6]University Medical Center Utrecht, Utrecht, Netherlands

Proposal / Abstract:
Background: Ensuring data quality for distributed data networks is challenging.
Objectives: We will examine international practices in five distributed data networks that house a mixture of administrative claims data and electronic health record data including: the U.S. Food and Drug Administration's (FDA's) Sentinel Initiative (Sentinel), the FDA's Biologics Effectiveness and Safety (BEST) Initiative, the U.S. National Patient Centered-Clinical Research Network (PCORnet), Japan's Medical Information Database Network (MID-NET), and the Canadian Network for Observational Drug Effect Studies (CNODES).

Duke | FORGE
HEALTH DATA IN ACTION

Data Provenance

- Prior to application of research to humans in a manner that directs diagnostic testing or recommendations for behavioral change or therapy either in clinical studies or in commercialization, findings must be reproducible (both internally and externally) based on additional documented data sets.
- Documentation of internal reproducibility should extend in a readily traceable manner back to the fundamental source of the original data, a concept known as *data provenance*.
- When possible, external reproducibility should be demonstrated in data external to those used to produce the original findings.
- Optimally, this could include a requirement that the data analysis be reproduced in an independent facility in addition to being replicated in second data sets.

Duke | FORGE
HEALTH DATA IN ACTION

Premise 2: The right teams of researchers are needed to enable the right data and best analysis for the specific question or purpose

- We will have access to immense amounts of data in multiple dimensions with tremendous possibilities of improving health outcomes and improving treatment
- The plethora of analyses and their complexity will lead to an increased risk of errors in analysis and interpretation
- How do we develop approaches to analyses that produce truthful and non-misleading findings?
- See U of Washington course: http://callingbullshit.org/

Duke | FORGE
HEALTH DATA IN ACTION

Data deluge

Illustration credit: Brett Ryder. https://www.economist.com/node/15579717

Duke | FORGE
HEALTH DATA IN ACTION

How it works: analytics and data science

Ask the right questions

Tap knowledge to specify the question that defines the challenge

Find the right data

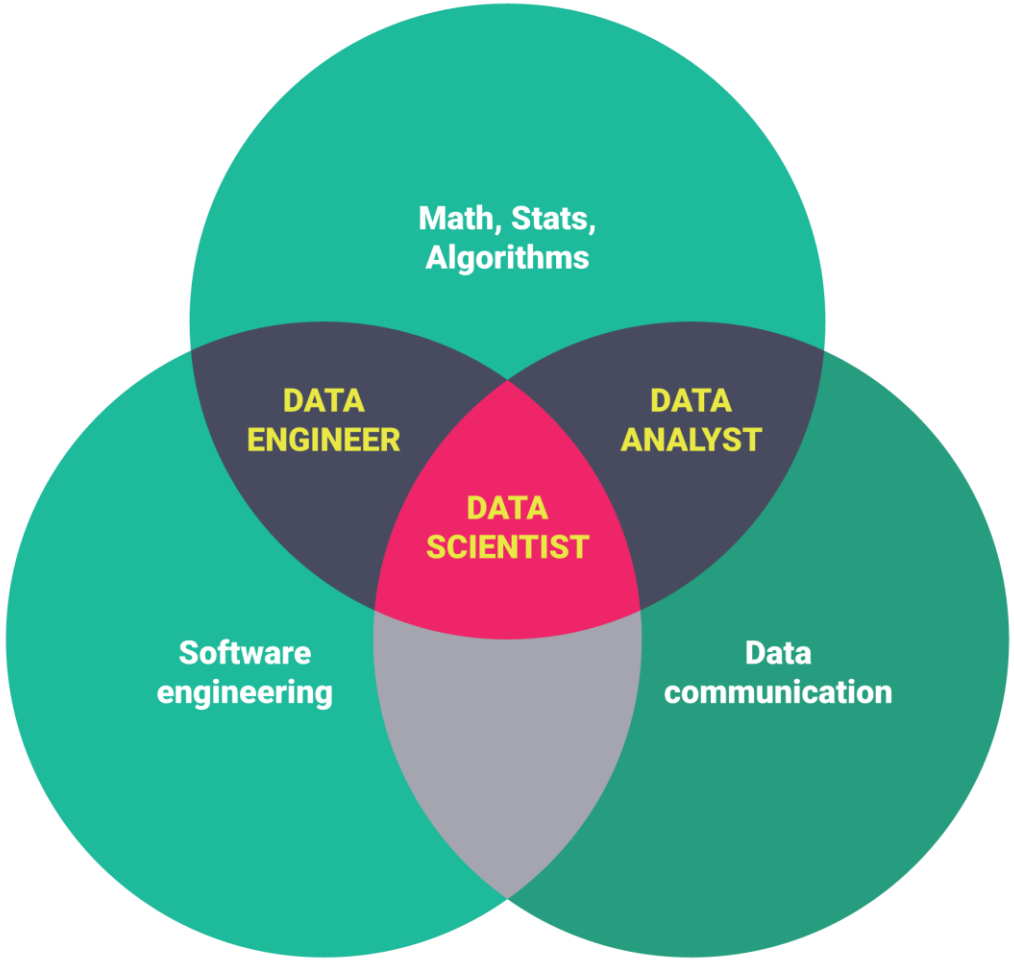
Examine all data sources (access, feasibility, content, quality)

Optimize methods

- Tailor methods to address the question
- Develop new tools to improve analytics

Decide wisely

- Improve health
- Improve care
- Reduce cost
- Improve clinical practice


Math, Stats,
Algorithms
DATA
ENGINEER
DATA
ANALYST
DATA
SCIENTIST
Software
engineering
Data
communication

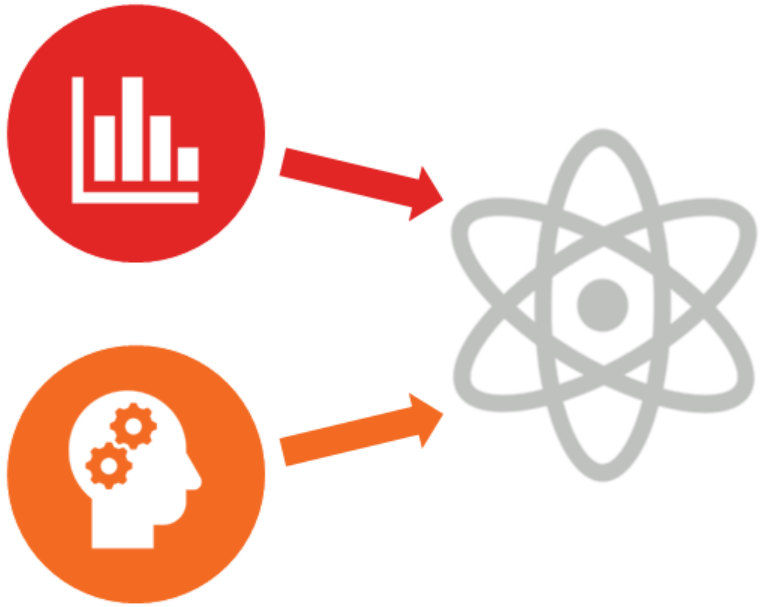
Duke | FORGE
HEALTH DATA IN ACTION

The grand fusion

Melding strengths across disciplines and between professionals

Fostering the **comprehensive toolbox across the spectrum** including frequentist statistics, Bayesian statistics, machine learning, and deep learning

Developing the **right framework for teams** including clinicians and quantitative expertise

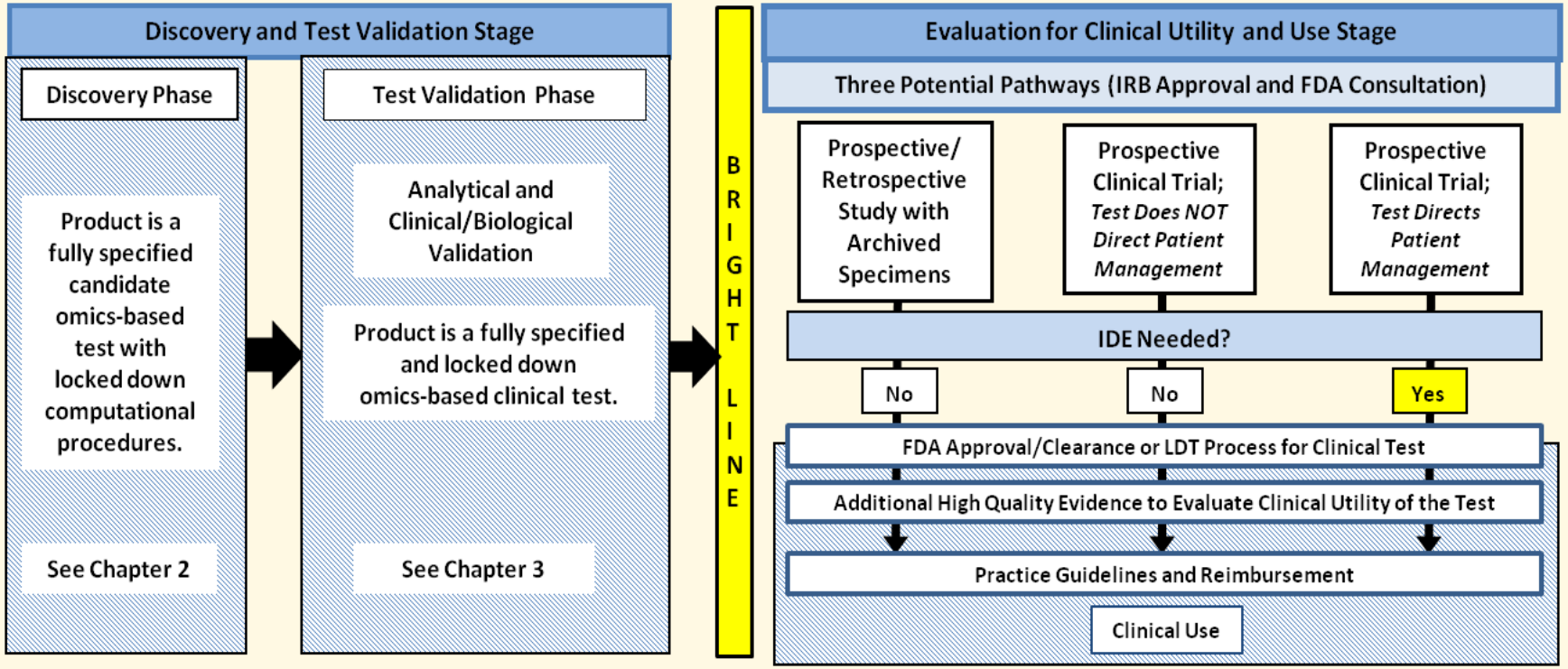


- Biostatistics and bioinformatics
- Population health
- Clinical research
- Research training and support
- Oversight and quality assurance

- Basic science departments
- Clinical departments
- Clinical research units

- Engineering
- Computer science
- Statistical science
- Big data analytics

Omics-Based Test Development Framework



Duke | FORGE
HEALTH DATA IN ACTION

Premise 3: The field must step up its approach to translating findings into truthful, understandable information

- Even with the ethical principles articulated by Laura, we have significant issues to gain/sustain public trust
- The public is losing faith in science and susceptible to more appealing versions of “the facts”
- For every question on health or biomedical data there is an answer: ”made up” answers are more appealing than scientific findings with all their uncertainty and caveats
- A brilliant analysis that isn’t understood has limited value

Duke | FORGE HEALTH DATA IN ACTION

Facebook Fiasco

The Economist | Topics | Current edition | More

What Zuckerberg should do

Facebook faces a reputational meltdown

This is how it, and the wider industry, should respond

Print edition | Leaders ›

Mar 22nd 2018

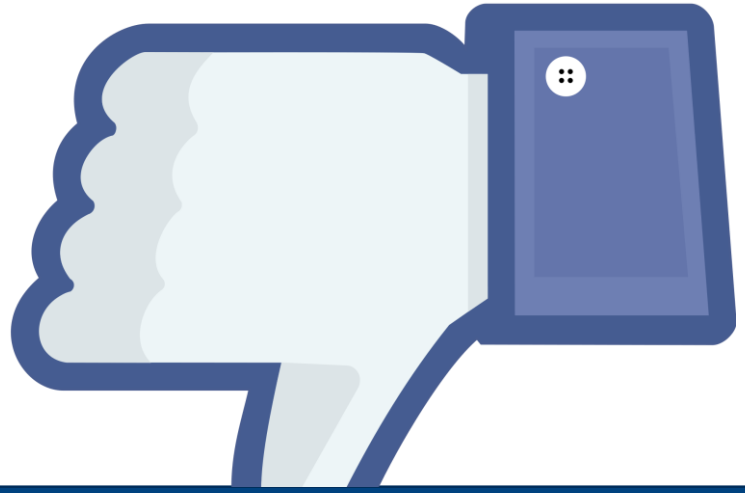
LAST year the idea took hold that Mark Zuckerberg might run for president in 2020 and seek to lead the world's most powerful country. Today, Facebook's founder is fighting to show that he is capable of leading the world's eighth-biggest listed company or that any of its 2.1bn users should trust it

https://www.economist.com/news/leaders/21739151-how-it-and-wider-industry-should-respond-facebook-faces-reputational-meltdown

The Guardian

Facebook: is it time we all deleted our accounts?

The Cambridge Analytica revelations may be the final nudge we need to turn away from the social network. And it's only the tip of the iceberg when it comes to big tech harvesting private information



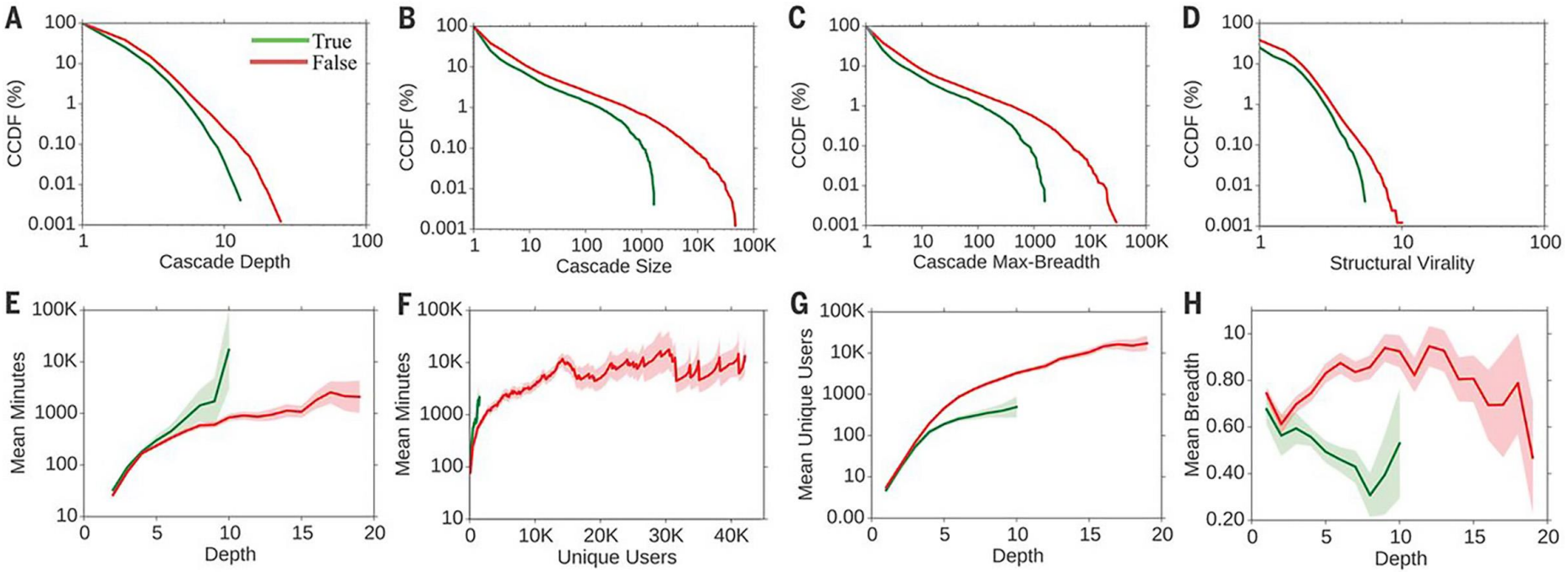
https://www.theguardian.com/technology/2018/mar/20/facebook-is-it-time-we-all-deleted-our-accounts

Duke | FORGE
HEALTH DATA IN ACTION

Krista Kennell / Stone / Catwalker / Shutterstock / The Atlantic
https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/

Duke | FORGE
HEALTH DATA IN ACTION

Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades

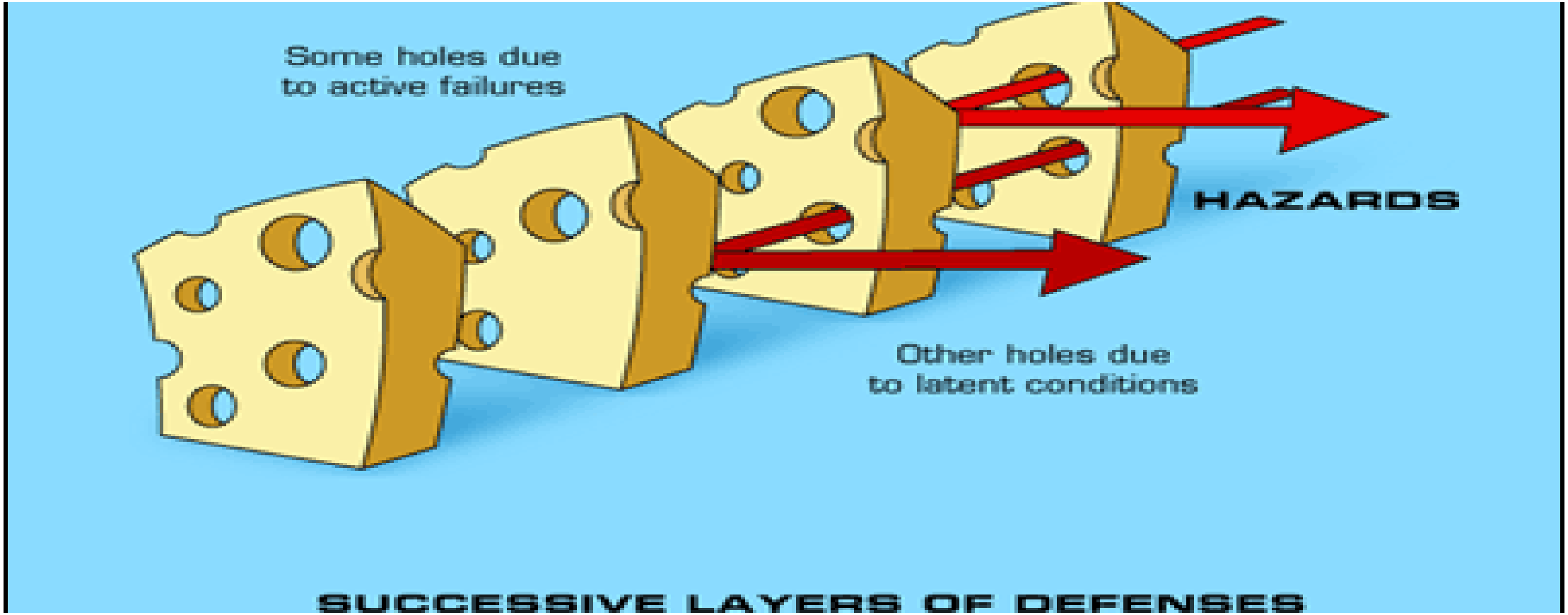


Science
AAAS

Vosoughi S, et al. Science. 2018;359:1146-51.

Duke | FORGE
HEALTH DATA IN ACTION

Swiss Cheese and Errors



Duke | FORGE
HEALTH DATA IN ACTION

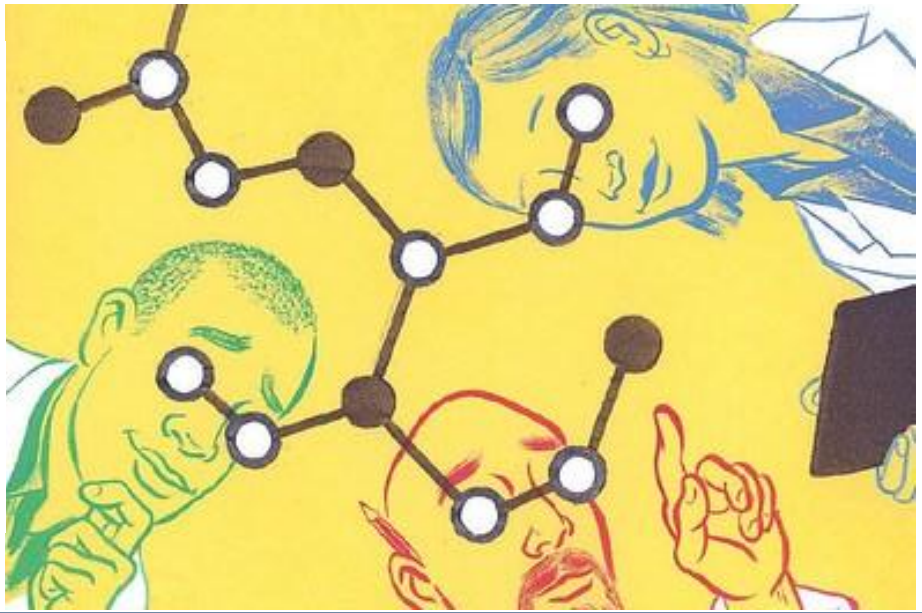
The New Einsteins Will Be Scientists Who Share

From cancer to cosmology, researchers could race ahead by working together—online and in the open

By MICHAEL NIELSEN

In January 2009, a mathematician at Cambridge University named Tim Gowers decided to use his blog to run an unusual social experiment. He picked out a difficult mathematical problem and tried to solve it completely in the open, using his blog to post ideas and partial progress. He issued an open invitation for others to contribute their own ideas, hoping that many minds would be more powerful than one. He dubbed the experiment the Polymath Project.

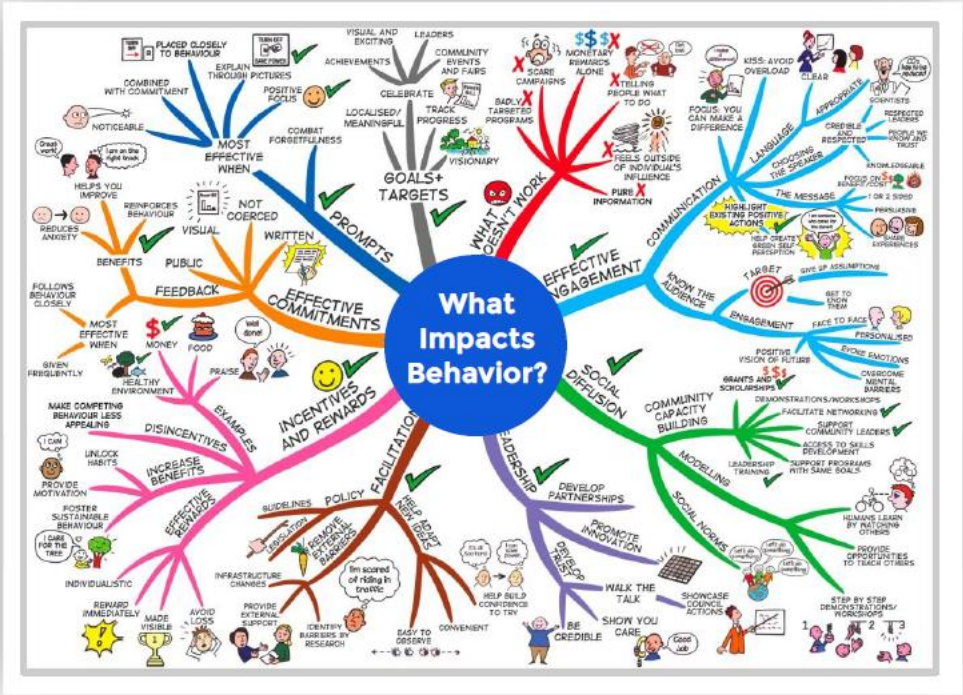
Several hours after Mr. Gowers opened up his blog for discussion, a Canadian-Hungarian mathematician posted a comment. Fifteen minutes later, an Arizona high-school math teacher chimed in. Three minutes after that, the UCLA mathematician Terence Tao commented. The discussion ignited, and in just six weeks, the mathematical problem had been solved.

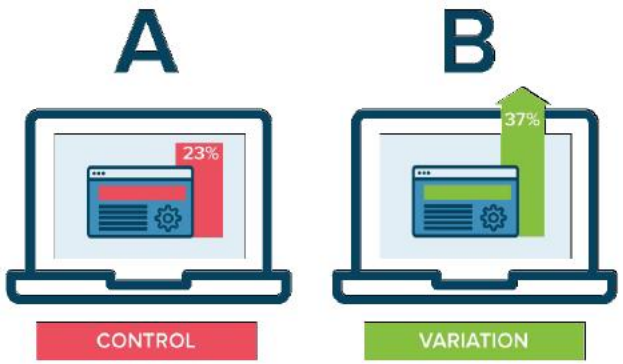


THE WALL STREET JOURNAL.

Duke | FORGE
HEALTH DATA IN ACTION

Data Activation and Testing Outcomes





Duke | FORGE
HEALTH DATA IN ACTION

Digital Transformation

- Data on premise, hard to access, analyze and use
- Productivity tools built for individual, local usage
- IT focusing on **where** it computes

- Data stored in cloud, simple to query
- Collaborative, cloud based productivity applications
- Machine learning drives deep, actionable insights
- IT changing **how** it computes

Duke | FORGE
HEALTH DATA IN ACTION

"Truthful and Not Misleading"

- Truthful
  - —the body of real things, events, and facts : actuality.
  - —accuracy, accurateness, actuality, authenticity, correctness, credibility, honesty, trustability, trustiness, trustworthiness, truthful ness, veracity dependability, reliability
- Not misleading
  - —to lead in a wrong direction or into a mistaken action or belief often by deliberate deceit
  - —to lead astray : give a wrong impression
  - —beguiling, deceitful, deceiving, deceptive, deluding, delusive, delusory, fallacious, false, specious

Duke | FORGE
HEALTH DATA IN ACTION

It takes a team

- Ethics is too important to leave to ethics
- Data science is too important to leave to data scientists