EMA/HMA Annual Data Forum (9 December 2025) - Opportunities in Data, Policy, and Ethics

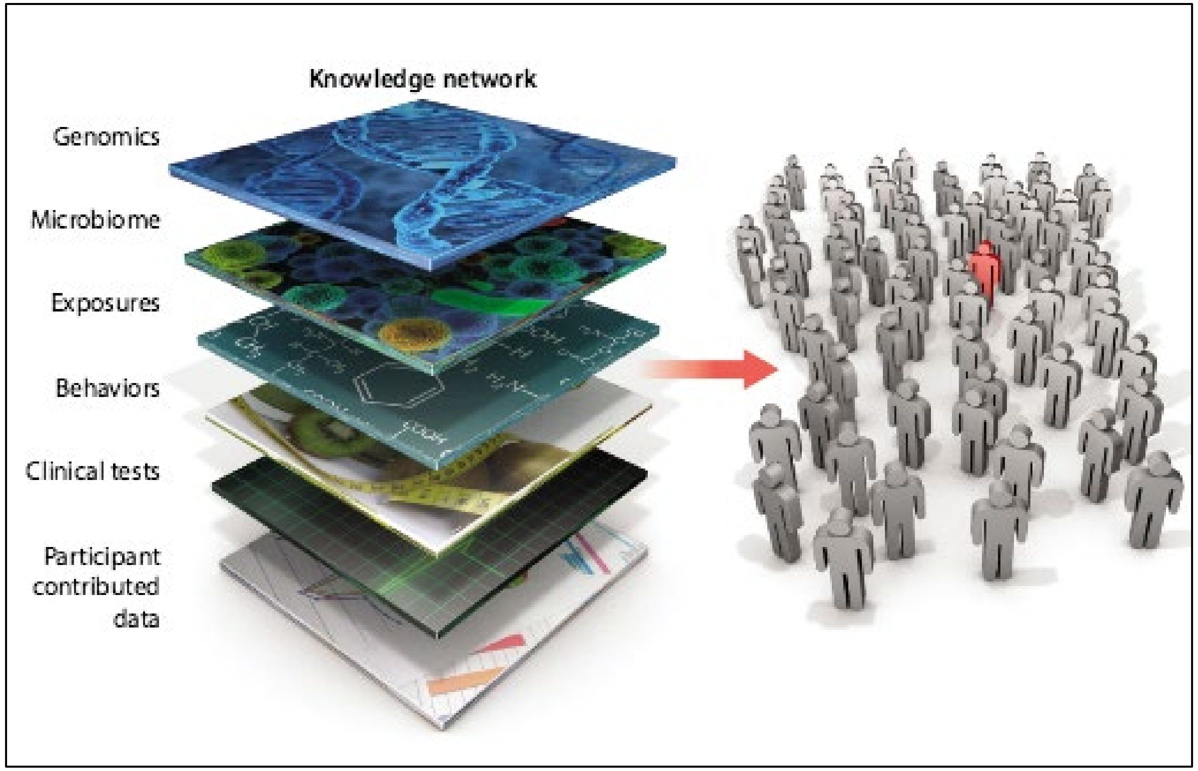
TRUST, TRUTH, AND TRANSFORMATION: THE ETHICS OF DATA IN A DIGITAL AGE

Alessandro Blasimme

DHEST
Department of Health Sciences and Technology

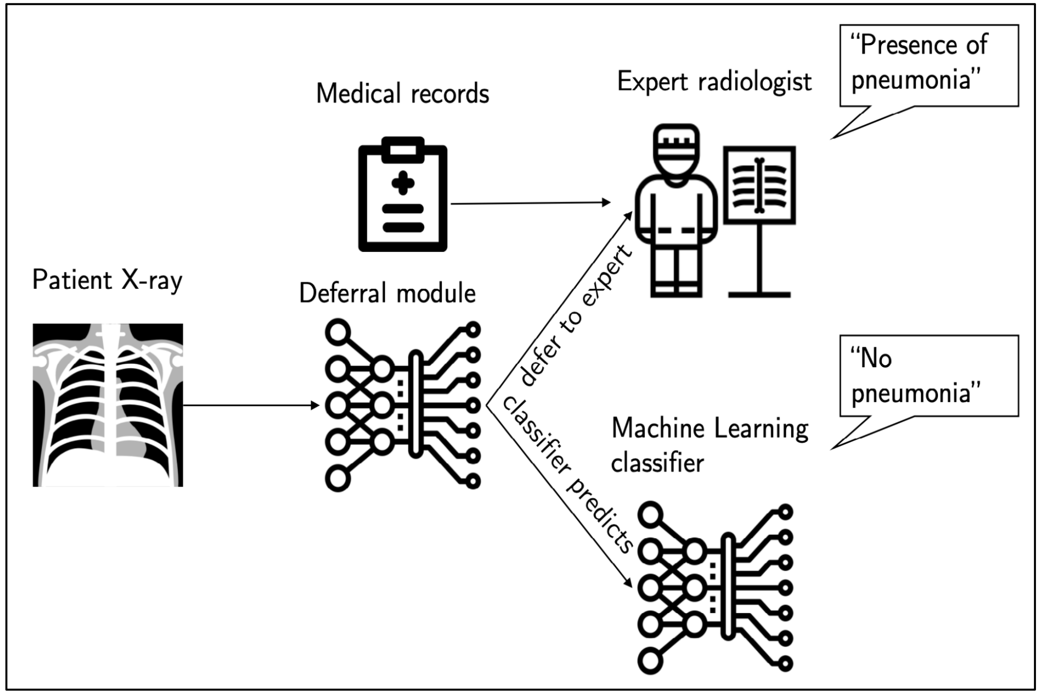
Two major trends in data ethics in the last 15 years

Understanding individual variation
(Precision Medicine)



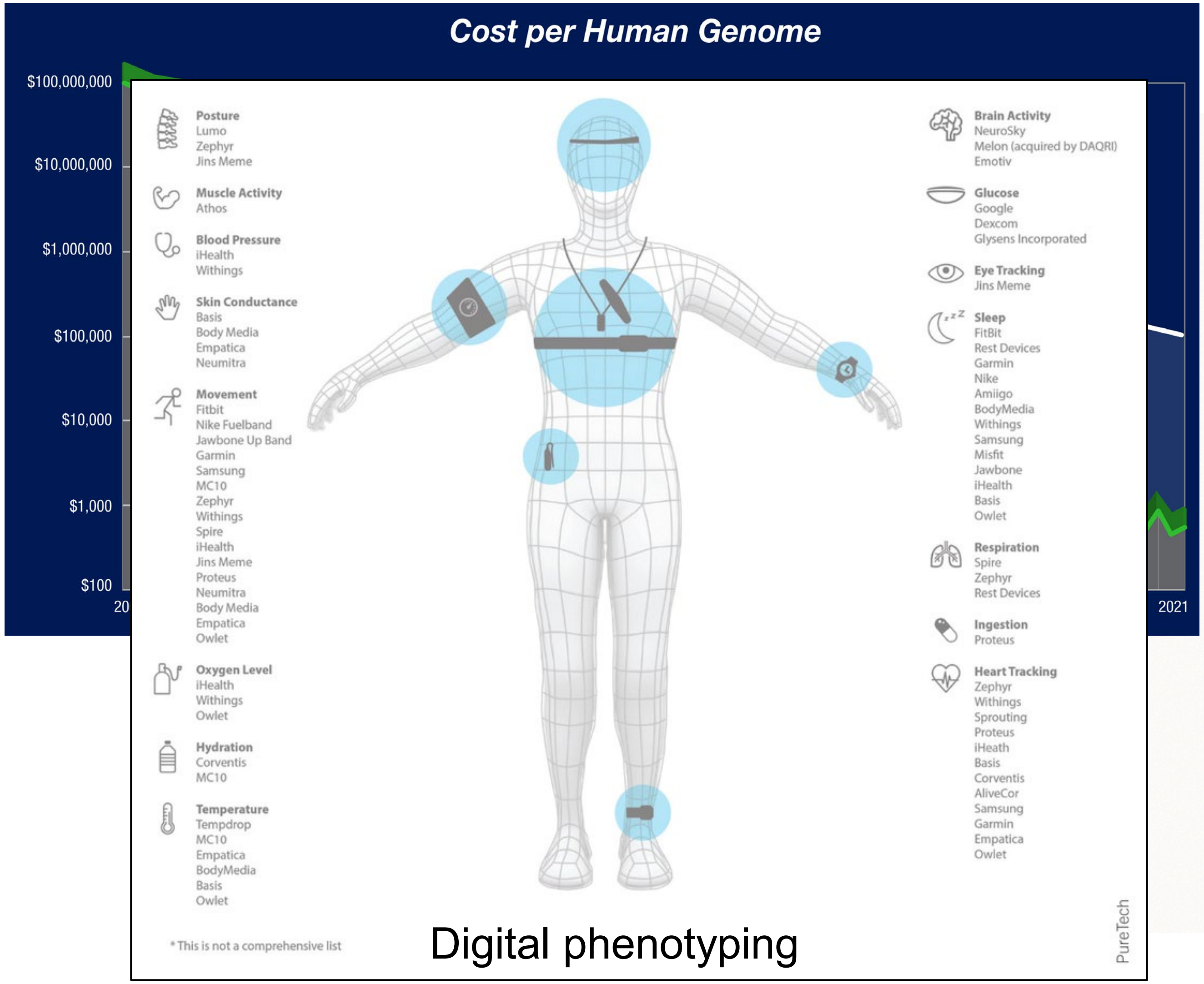
Hagwood et al. Science Translational Medicine 12 Aug 2015

Learning from common patterns
(Artificial Intelligence)



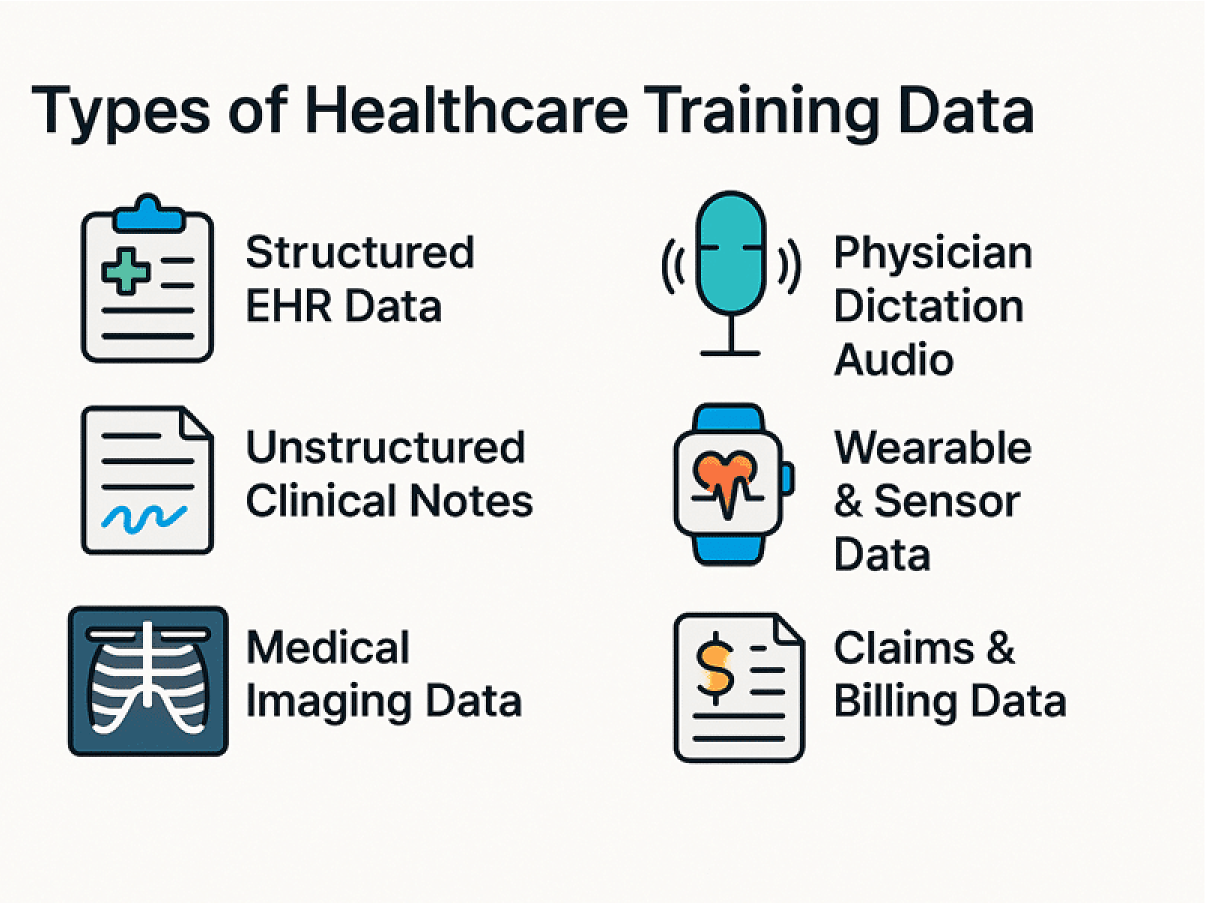
https://news.mit.edu/2020/machine-learning-health-care-system-understands-when-to-step-in-0731

Technological drivers



Digital phenotyping

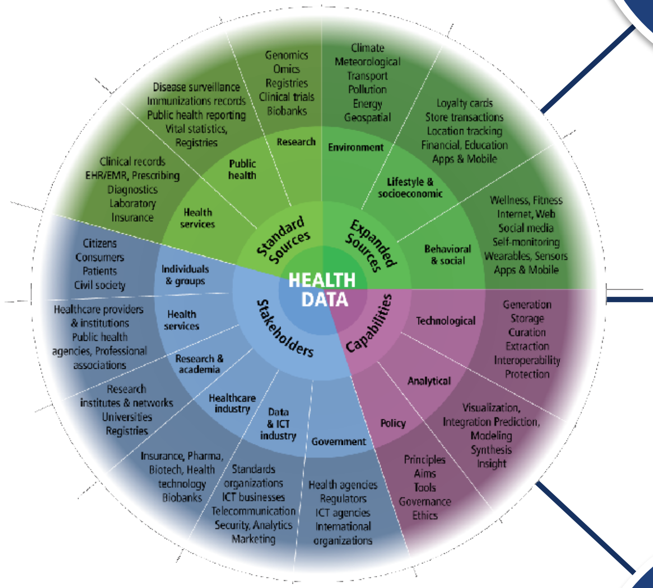
Elenko et al. 2015, Defininig Digital Medicine, *Nat Biotech.*



https://www.shaip.com/blog/what-is-healthcare-training-data-and-why-is-it-important/

An evolving health data ecosystem

Vayena et al. Bull World Health Organ. 2018 Jan 1;96(1):66-68.



Uncertainty

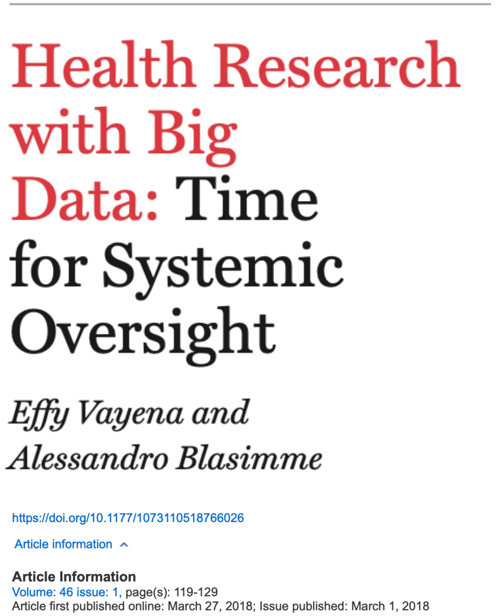
- What is the real purpose of data collection?
- Who is using my data and why?
- What uses of data are permitted/authorized?
- Who controls data? Who owns data?
- Who oversees data collection and use?

Expanded temporality

- For how long will data be stored/used?
- What happens to data after data subjects pass away?
- Will data access conditions evolve over time?
- How will new data analytics and technologies affect the use of collected data?

Relationality

- Big data reveal relations among data subjects: genes, environmental exposure, habits, group behaviours etc.
- Big data encourage classificatory practices: grouping people in categories.



Health Research with Big Data: Time for Systemic Oversight

Effy Vayena and Alessandro Blasimme

https://doi.org/10.1177/1073110518766026

Article information ^

Article Information
Volume: 46 issue: 1, page(s): 119-129
Article first published online: March 27, 2018; Issue published: March 1, 2018

The Ethics of AI in Biomedical Research, Patient Care, and Public Health

Alessandro Blasimme and Effy Vayena

The Oxford Handbook of Ethics of AI

Edited by Markus D. Dubber, Frank Pasquale, and Sunit Das

Print Publication Date: Jul 2020 Subject: Law, IT and Communications Law Online Publication Date: Jul 2020
DOI: 10.1093/oxfordhb/9780190067397.013.45

“Tailored-to-You”: Public Engagement and the Political Legitimation of Precision Medicine

Alessandro Blasimme, Effy Vayena

Perspectives in Biology and Medicine, Volume 59, Number 2, Spring 2016, pp. 172-188 (Article)

Published by Johns Hopkins University Press
DOI: https://doi.org/10.1353/pbm.2017.0002

Towards Adaptive Governance in Big Data Health Research: Implementing Regulatory Principles

Laurie Laurie G. (ed.). Cambridge Handbook of Health Research Regulation. Cambridge University Press, Cambridge, UK – Forthcoming.

24 Pages • Posted: 2 Jan 2020

Alessandro Blasimme
ETH Zurich

Effy Vayena
ETH Zurich

Date Written: October 2019

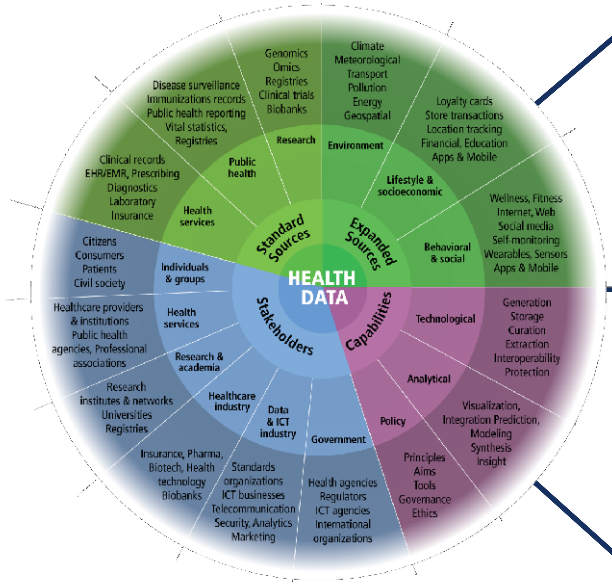
Abstract

In this chapter, we will first make the case for adaptive and principle-based governance of big data research. We will then outline six principles of adaptive governance (AFIRRM) for big data research and discuss key factors for their implementation into effective governance structures and processes.

Keywords: big data research, health research, research ethics, governance, adaptive governance, ethics

DHEST
Department of Health Sciences and Technology

Data governance: guiding principles



Uncertainy

- Adaptivity
- Flexibility

Expanded temporality

- Monitoring
- Responsiveness

Relationality

- Reflexivity
- Inclusiveness

Data governance: guiding principles

AFIRRM

- **A**daptivity
- **F**lexibility

- **M**onitoring
- **R**esponsiveness

- **R**eflexivity
- **I**nclusiveness

Data governance: guiding principles 1/3

- **Adaptivity**
  - Acknowledge the limits of anticipatory practices
  - Readiniess to devise specific oversight mechanisms for new data types and uses
- **Flexibility**
  - Treating data under different oversight rubrics depending on their actual use rather than on their source

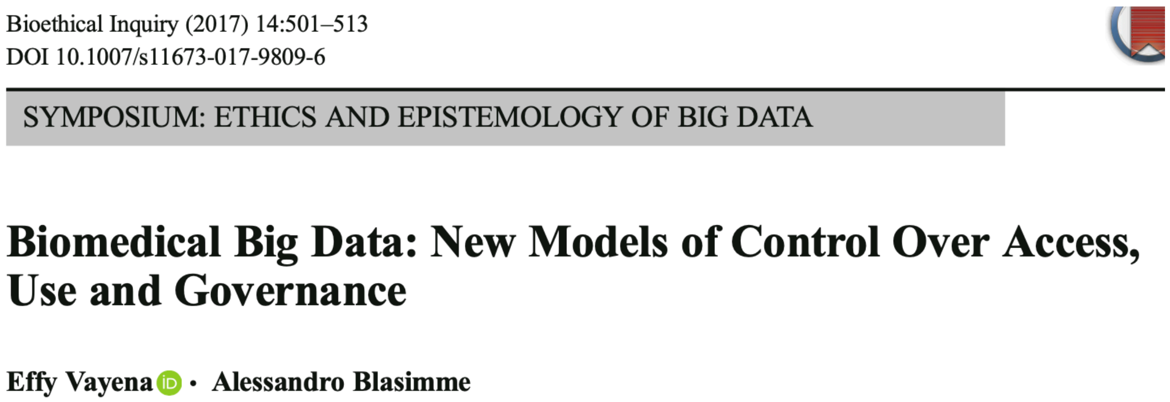
Data governance: guiding principles 2/3

- **Monitoring**
  - Continuous oversight activities to detect signals of new vulnerabilities for data subjects: potential harms, discrimination, privacy, surveillance, medicalization, etc.
- **Responsiveness**
  - Preventing vulnerabilities to result in actual harms
  - Ensure effective containment of harmful effects in case of failures: e.g. privacy breaches

Data governance: guiding principles 3/3

- **Reflexivity**
  - Monitor the effects of classificatory practices as they emerge from data-driven research
  - Scrutinize assumptions embedded in research practices and regulatory mechanisms (institutional reflexivity)
  - Cultivate awarness for how dominant interests and power structures affect rights and interests of individuals and communities
- **Inclusiveness**
  - Upstream engagement of relevant stakeholders – including lay publics – before technological path dependences become established
  - Reaching out to under-represented stakeholders

How to ensure data control?



Bioethical Inquiry (2017) 14:501–513
DOI 10.1007/s11673-017-9809-6

SYMPOSIUM: ETHICS AND EPISTEMOLOGY OF BIG DATA

Biomedical Big Data: New Models of Control Over Access, Use and Governance

Effy Vayena · Alessandro Blasimme

Bioethical Inquiry (2017) 14:501–513

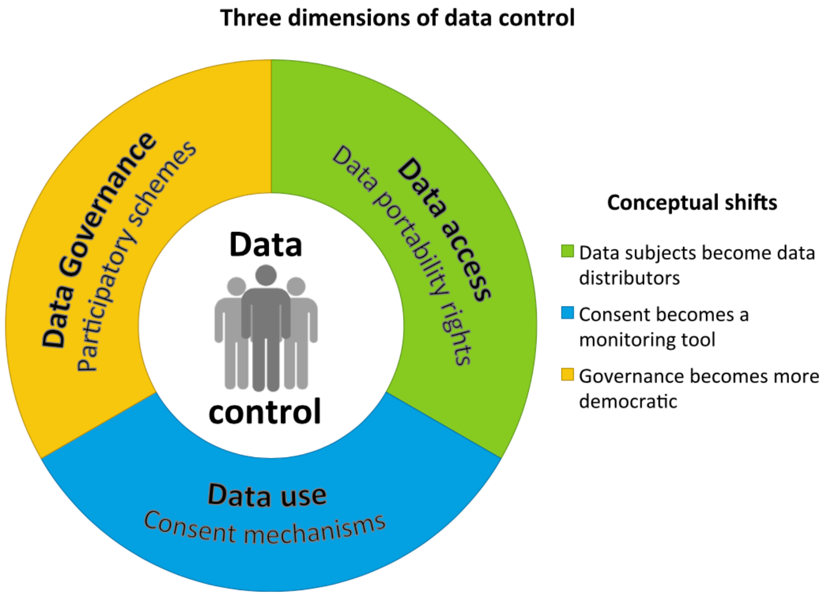


Fig. 1 Three dimensions of data control

How to ensure data control?

Informed consent for data use

Health Research with Big Data: Time for Systemic Oversight

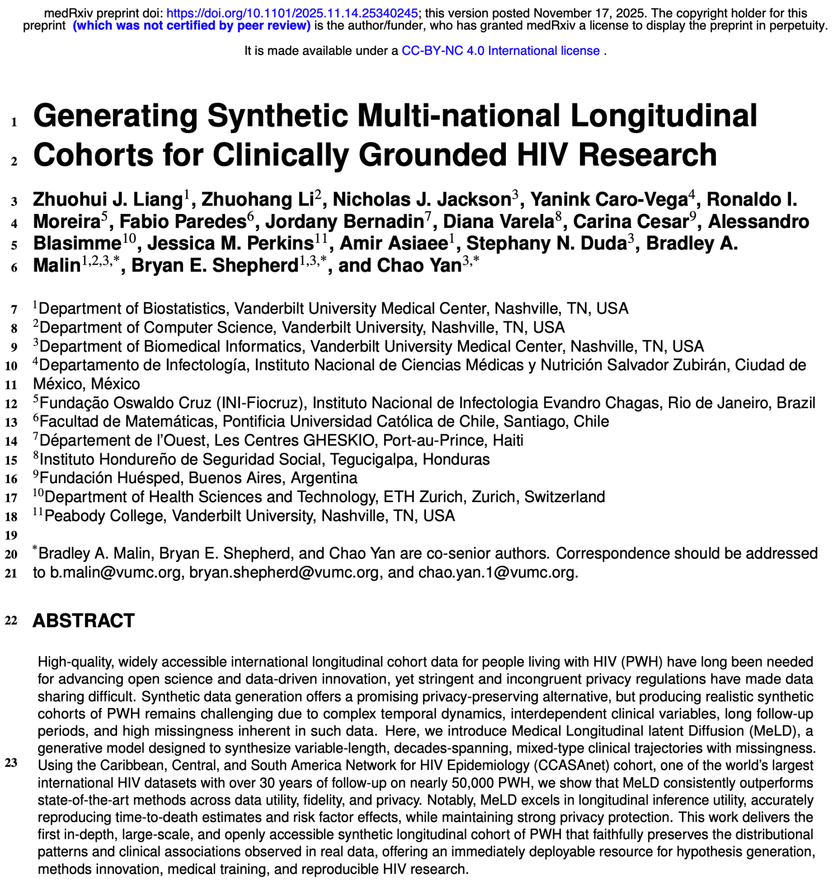
Effy Vayena and Alessandro Blasimme

- Use authorizations given for primary data use are unlikely to cover novel future uses.
- Given the expansion of the data ecosystem, it is unlikely that current oversight models will suffice.

The Journal of Law, Medicine & Ethics, 46 (2018): 119-129. © 2018 The Author(s)
DOI: 10.1177/1073110518766026
https://doi.org/10.1177/1073110518766026 Published online by Cambridge University Press

How to ensure data control?

Informed consent for data use



Generating Synthetic Multi-national Longitudinal Cohorts for Clinically Grounded HIV Research

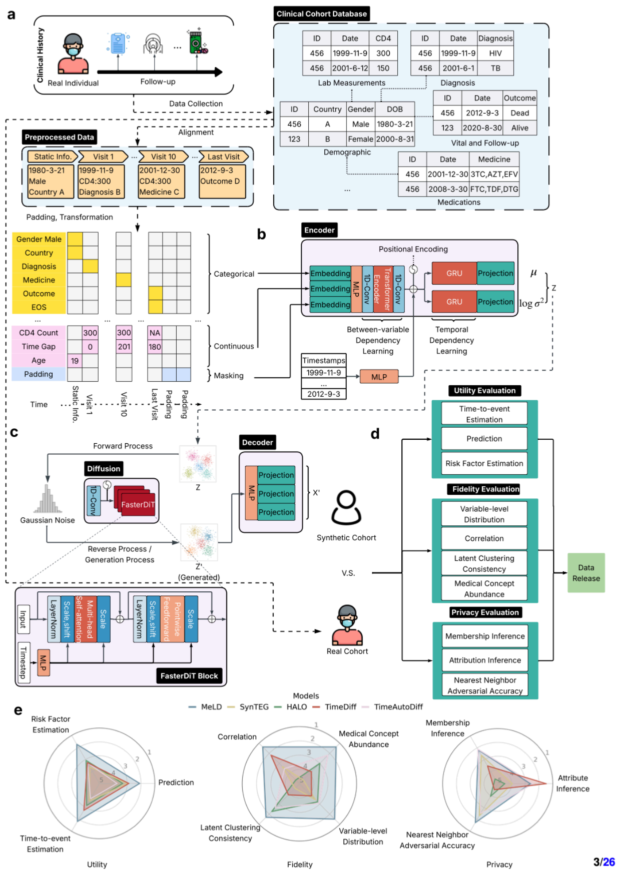
Zhuohui J. Liang[1], **Zhuohang Li**[2], **Nicholas J. Jackson**[3], **Yanink Caro-Vega**[4], **Ronaldo I. Moreira**[5], **Fabio Paredes**[6], **Jordany Bernadin**[7], **Diana Varela**[8], **Carina Cesar**[9], **Alessandro Blasimme**[10], **Jessica M. Perkins**[11], **Amir Asiaee**[1], **Stephany N. Duda**[3], **Bradley A. Malin**[1,2,3,*], **Bryan E. Shepherd**[1,3,*], **and Chao Yan**[3,*]

[1]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA
[2]Department of Computer Science, Vanderbilt University, Nashville, TN, USA
[3]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
[4]Departamento de Infectología, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Ciudad de México, México
[5]Fundação Oswaldo Cruz (INI-Fiocruz), Instituto Nacional de Infectologia Evandro Chagas, Rio de Janeiro, Brazil
[6]Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Santiago, Chile
[7]Département de l'Ouest, Les Centres GHESKIO, Port-au-Prince, Haiti
[8]Instituto Hondureño de Seguridad Social, Tegucigalpa, Honduras
[9]Fundación Huésped, Buenos Aires, Argentina
[10]Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland
[11]Peabody College, Vanderbilt University, Nashville, TN, USA

[*]Bradley A. Malin, Bryan E. Shepherd, and Chao Yan are co-senior authors. Correspondence should be addressed to b.malin@vumc.org, bryan.shepherd@vumc.org, and chao.yan.1@vumc.org.

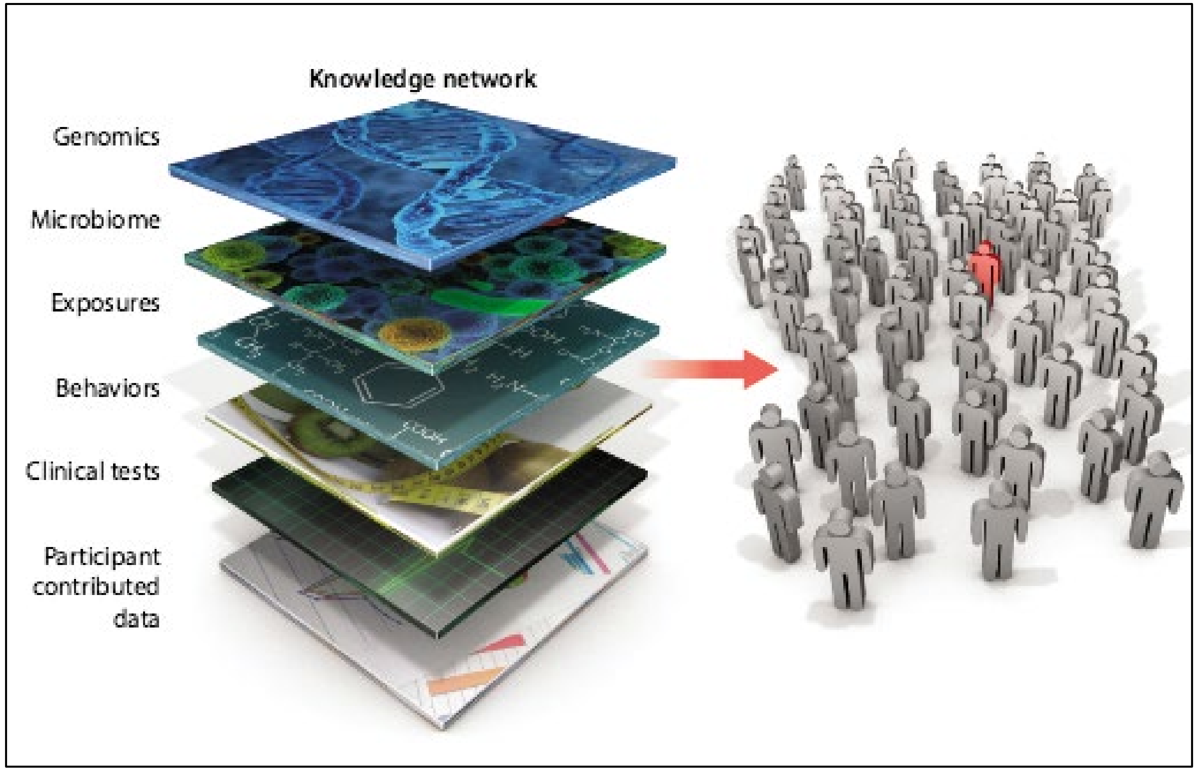
ABSTRACT

High-quality, widely accessible international longitudinal cohort data for people living with HIV (PWH) have long been needed for advancing open science and data-driven innovation, yet stringent and incongruent privacy regulations have made data sharing difficult. Synthetic data generation offers a promising privacy-preserving alternative, but producing realistic synthetic cohorts of PWH remains challenging due to complex temporal dynamics, interdependent clinical variables, long follow-up periods, and high missingness inherent in such data. Here, we introduce Medical Longitudinal latent Diffusion (MeLD), a generative model designed to synthesize variable-length, decades-spanning, mixed-type clinical trajectories with missingness. Using the Caribbean, Central, and South America Network for HIV Epidemiology (CCASAnet) cohort, one of the world's largest international HIV datasets with over 30 years of follow-up on nearly 50,000 PWH, we show that MeLD consistently outperforms state-of-the-art methods across data utility, fidelity, and privacy. Notably, MeLD excels in longitudinal inference quality, accurately reproducing time-to-death estimates and risk factor effects, while maintaining strong privacy protection. This work delivers the first in-depth, large-scale, and openly accessible synthetic longitudinal cohort of PWH that faithfully preserves the distributional patterns and clinical associations observed in real data, offering an immediately deployable resource for hypothesis generation, methods innovation, medical training, and reproducible HIV research.



Synthetic data have an uncertain ethical and legal status

How to ensure data control?
Participatory practices for data governance



Hagwood et al. Science Translational Medicine 12 Aug 2015

"Tailored-to-You": Public Engagement and the Political Legitimation of Precision Medicine

Alessandro Blasimme, Effy Vayena

Perspectives in Biology and Medicine, Volume 59, Number 2, Spring 2016, pp. 172-188 (Article)

Published by Johns Hopkins University Press
DOI: https://doi.org/10.1353/pbm.2017.0002

Data governance: participatory practices*

*genetic/genomic exceptionalism

Recommendation 4.7: The PMI-CP should ensure the ***responsible return of personal results and information*** *to individual participants and sharing of* ***aggregate findings*** *from its investigations with participants so all volunteers may* ***have opportunity to benefit from the science****.*

Recommendation 4.1: Research participants and their advocates should be ***central partners in the governance****, design, conduct, oversight, dissemination, and evaluation activities of the PMI-CP.*

NIH National Institutes of Health

All of Us RESEARCH PROGRAM

PMI-CP = Precision Medicine Initiative Cohort Program

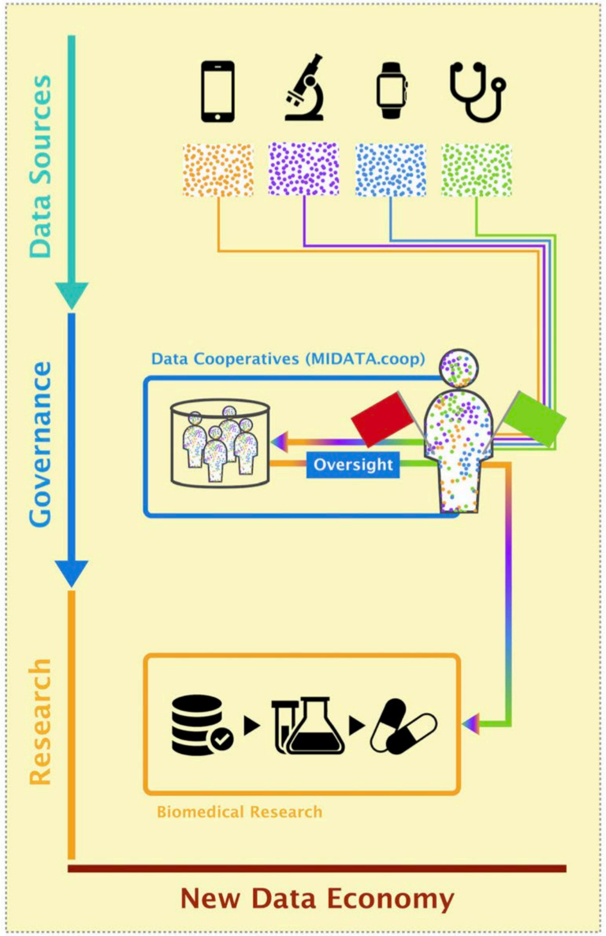
How to ensure data control?
Data portability for data access management

Philos. Technol. (2018) 31:473–479
https://doi.org/10.1007/s13347-018-0320-8

COMMENTARY

Democratizing Health Research Through Data Cooperatives

Alessandro Blasimme[1] · Effy Vayena[1] · Ernst Hafen[2]



graphic credit Manuel Schneider

Fig. 1 A new data economy for health research. Citizens aggregate data from different sources and make them available for research through data cooperatives. Cooperatives offer oversight mechanisms to filter data access requests and tools for the democratic governance of the data

The role of data infrastructures

www.nature.com/ejhg

ESHG

REVIEW ARTICLE OPEN

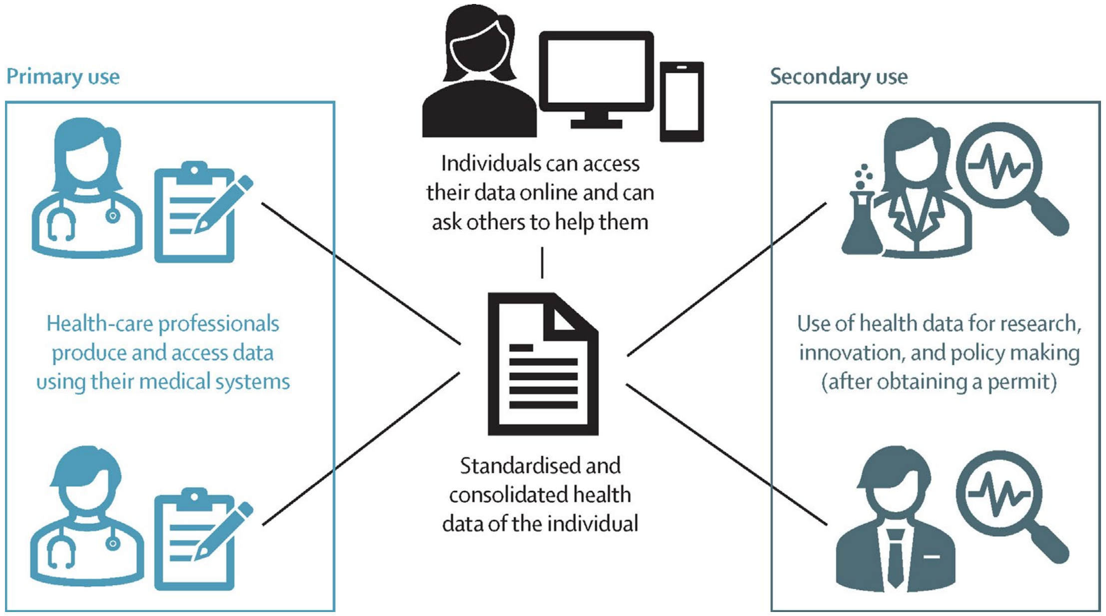
Check for updates

Ethical and social reflections on the proposed European Health Data Space

Ciara Staunton[1,2], Mahsa Shabani[3], Deborah Mascalzoni[1,4], Signe Mežinska[5] and Santa Slokenberga[6]

The COVID-19 pandemic demonstrated the benefits of international data sharing. Data sharing enabled the health care policy makers to make decisions based on real-time data, it enabled the tracking of the virus, and importantly it enabled the development of vaccines that were crucial to mitigating the impact of the virus. This data sharing is not the norm as data sharing needs to navigate complex ethical and legal rules, and in particular, the fragmented application of the General Data Protection Regulation (GDPR). The introduction of the draft regulation for a European Health Data Space (EHDS) in May 2022 seeks to address some of these legal issues. If passed, it will create an obligation to share electronic health data for certain secondary purposes. While there is a clear need to address the legal complexities involved with data sharing, it is critical that any proposed reforms are in line with ethical principles and the expectations of the data subjects. In this paper we offer a critique of the EHDS and offer some recommendations for this evolving regulatory space.

European Journal of Human Genetics (2024) 32:498–505; https://doi.org/10.1038/s41431-024-01543-9



Primary use

Health-care professionals produce and access data using their medical systems

Individuals can access their data online and can ask others to help them

Standardised and consolidated health data of the individual

Secondary use

Use of health data for research, innovation, and policy making (after obtaining a permit)

https://doi.org/10.1016/S2589-7500(23)00156-5

npj | digital medicine
Article
Published in partnership with Seoul National University Bundang Hospital
https://doi.org/10.1038/s41746-024-01105-9

Sync fast and solve things—best practices for responsible digital health

Check for updates

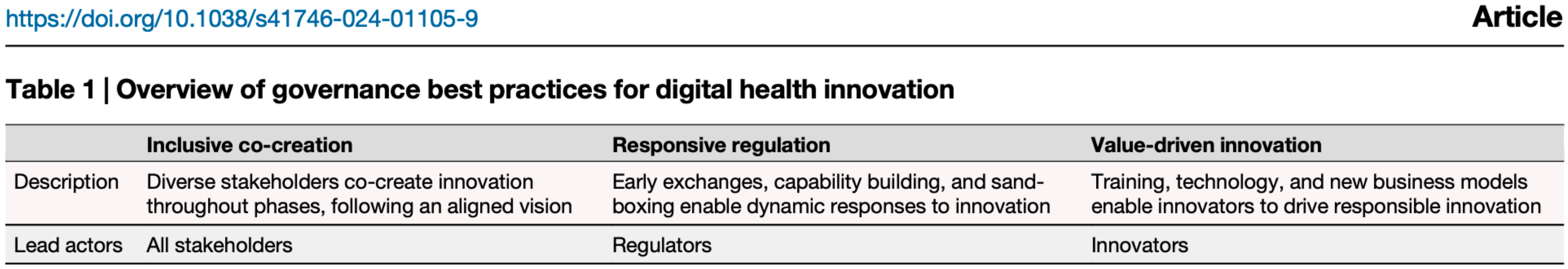
Constantin Landers, Alessandro Blasimme & Effy Vayena

npj | digital medicine
Comment
Published in partnership with Seoul National University Bundang Hospital
https://doi.org/10.1038/s41746-025-01724-w

Fostering inclusive co-creation in digital health

Alessandro Blasimme, Constantin Landers & Effy Vayena

Check for updates



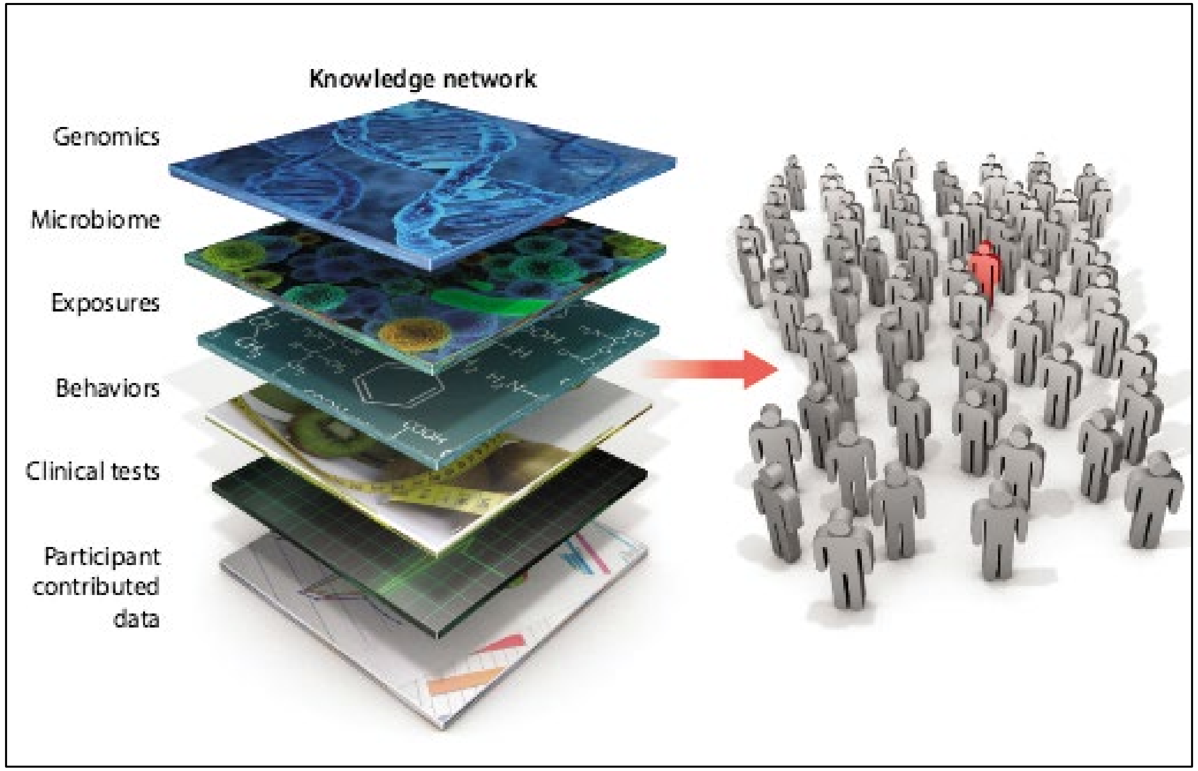
https://doi.org/10.1038/s41746-024-01105-9
Article

Table 1 | Overview of governance best practices for digital health innovation

	Inclusive co-creation	Responsive regulation	Value-driven innovation
Description	Diverse stakeholders co-create innovation throughout phases, following an aligned vision	Early exchanges, capability building, and sandboxing enable dynamic responses to innovation	Training, technology, and new business models enable innovators to drive responsible innovation
Lead actors	All stakeholders	Regulators	Innovators

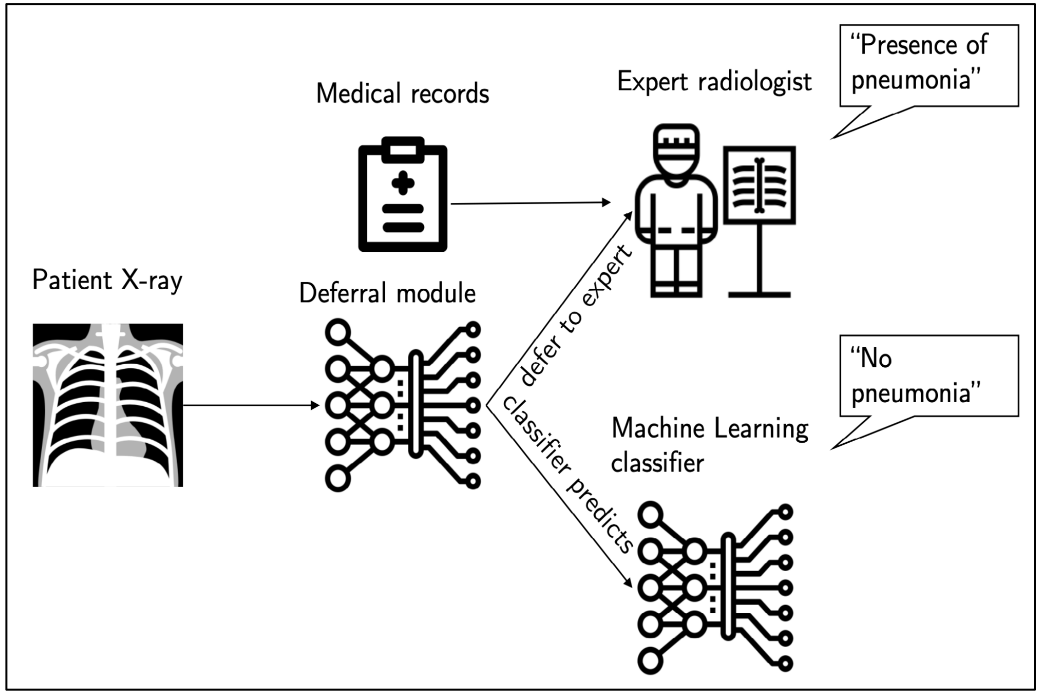
Lost in “transition”

Understanding individual variation
(Precision Medicine)



Hagwood et al. *Science Translational Medicine* 12 Aug 2015

Learning from common patterns
(Artificial Intelligence)



https://news.mit.edu/2020/machine-learning-health-care-system-understands-when-to-step-in-0731

Concluding considerations

- The conditions for data **use**, **governance** and **access** are evolving
- Regulation alone is insufficient to guarantee **data control:** participatory data governance and meaningful empowerment of data subjects
- Need to expand data infrastructure design to include data ethics best practices

THANK YOU FOR YOUR ATTENTION !

Acknowledgements

ETH zürich
Health Ethics & Policy Lab

Swiss National Science Foundation