Canada Health Infoway

Enterprise Imaging Community Webinar Series:

Privacy and Access to Healthcare Data for (DI) Machine Learning

Jeff Curtis

DBA, MBA, MSc, CISSP, CDPSE, CGEIT, CRISC, CISM

Chief Privacy Officer

Sunnybrook Health Sciences Centre

February 18, 2022

Canada Health Infoway

Sunnybrook

when it matters MOST

Privacy and Access to Healthcare Data

Agenda

Three considerations for privacy assurance within AI studies:

1. **Premise: What data? Whose access? For what purpose?**
2. **Challenges: Governance and de-identification is not enough**
3. **Emerging Technical solutions for multi-site analytics using PHI**

Privacy and Access to Healthcare Data

The Challenge of Access to PHI at Increasing Scale and Scope

For what purpose? What data? Whose access?

"The advent of Artificial Intelligence (AI) and Machine-Learning (ML) is showing that AI in medical imaging is requiring the "availability of high-quality image data for testing and validation. ***Open access data sharing*** *is critical to enable adequate sampling of the human population and the wide variety of disease states. It is equally critical to* ***protect patient privacy and to comply with national and international regulations*** *governing protected or sensitive health information and personally identifiable information"*

Privacy and Access to Healthcare Data

Open Data Management for PHI

The *Open Data Handbook* provides the following definition of **Open Data**:

"Open data is data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and share-alike."

Open data exists in many forms such as ***datasets, survey results, and metadata.*** *Data should exist in a form that can be used to duplicate and verify research findings. Open data is structured data and machine-readable. Therefore Open data policies often permit data to be accessed by machines for extraction, modification, and analysis.*

Open data does not include personal data about individuals."

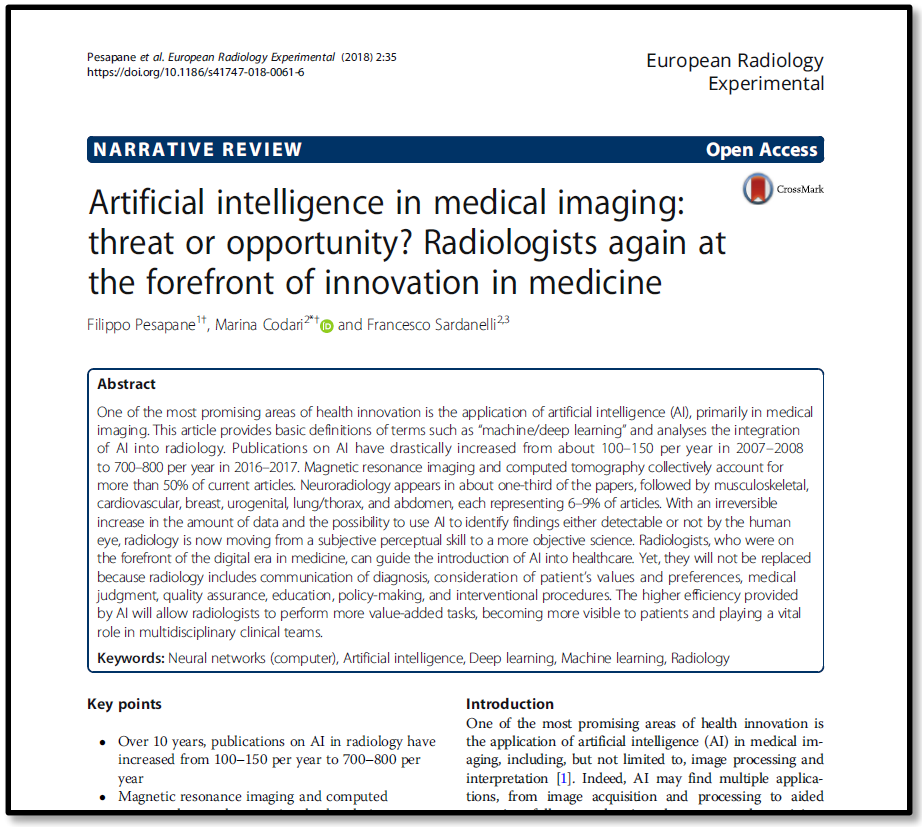
➔ OK, but in a healthcare context, we need to share lots of data that starts or may need to persist as PHI…what to do?

Open Access and Open data Sharing Committee – York University

https://www.library.yorku.ca/web/open/overview/open-data/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

No end of demand for PHI in sight…



Pesapane *et al. European Radiology Experimental* (2018) 2:35
https://doi.org/10.1186/s41747-018-0061-6

European Radiology Experimental

NARRATIVE REVIEW — Open Access

Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine

Filippo Pesapane[1†], Marina Codari[2*†] and Francesco Sardanelli[2,3]

Abstract

One of the most promising areas of health innovation is the application of artificial intelligence (AI), primarily in medical imaging. This article provides basic definitions of terms such as "machine/deep learning" and analyses the integration of AI into radiology. Publications on AI have drastically increased from about 100–150 per year in 2007–2008 to 700–800 per year in 2016–2017. Magnetic resonance imaging and computed tomography collectively account for more than 50% of current articles. Neuroradiology appears in about one-third of the papers, followed by musculoskeletal, cardiovascular, breast, urogenital, lung/thorax, and abdomen, each representing 6–9% of articles. With an irreversible increase in the amount of data and the possibility to use AI to identify findings either detectable or not by the human eye, radiology is now moving from a subjective perceptual skill to a more objective science. Radiologists, who were on the forefront of the digital era in medicine, can guide the introduction of AI into healthcare. Yet, they will not be replaced because radiology includes communication of diagnosis, consideration of patient's values and preferences, medical judgment, quality assurance, education, policy-making, and interventional procedures. The higher efficiency provided by AI will allow radiologists to perform more value-added tasks, becoming more visible to patients and playing a vital role in multidisciplinary clinical teams.

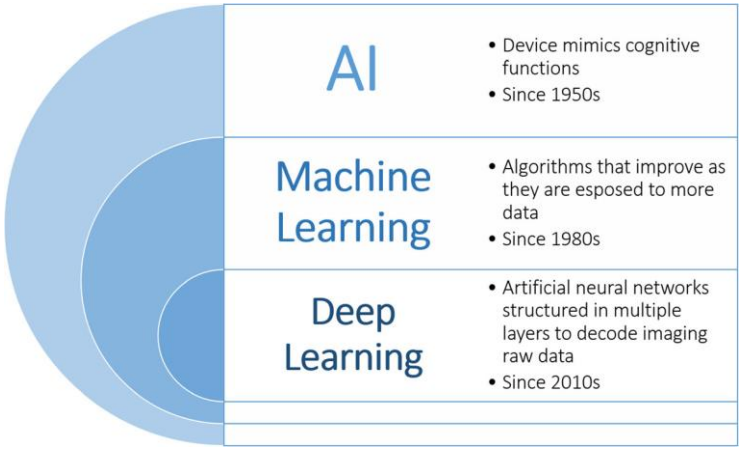
Keywords: Neural networks (computer), Artificial intelligence, Deep learning, Machine learning, Radiology

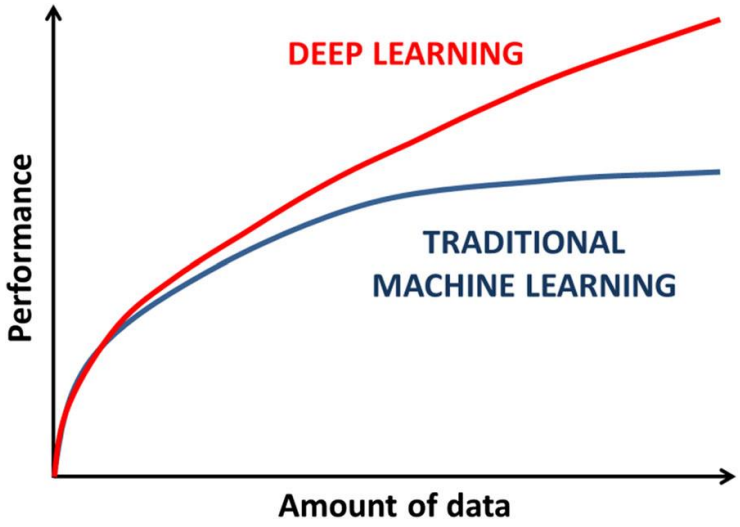
Key points

- Over 10 years, publications on AI in radiology have increased from 100–150 per year to 700–800 per year
- Magnetic resonance imaging and computed

Introduction

One of the most promising areas of health innovation is the application of artificial intelligence (AI) in medical imaging, including, but not limited to, image processing and interpretation [1]. Indeed, AI may find multiple applications, from image acquisition and processing to aided



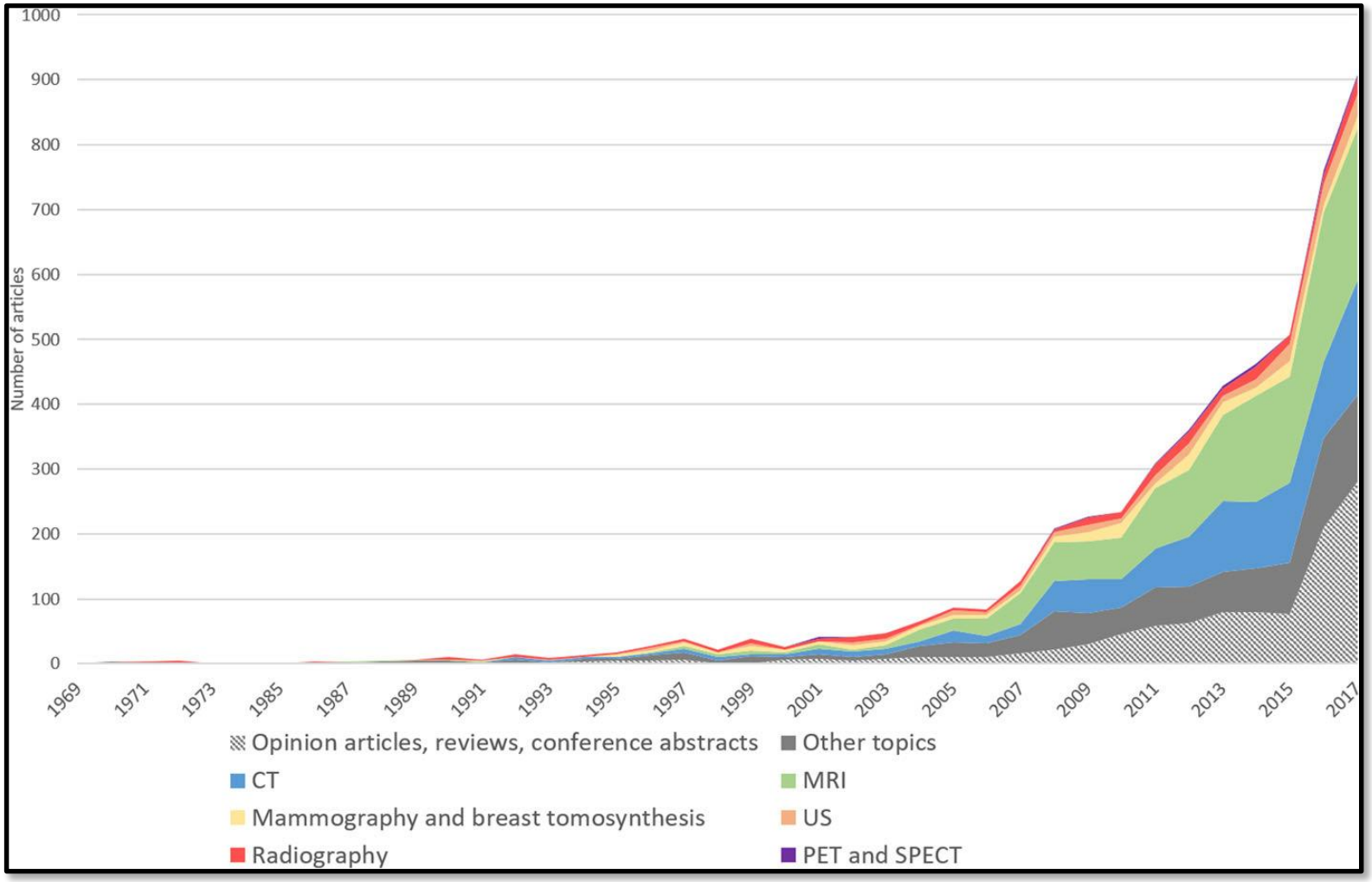


Pesapane, F., Codari, M. & Sardanelli, F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur Radiol Exp* **2,** 35 (2018).

https://eurradiolexp.springeropen.com/articles/10.1186/s41747-018-0061-6
(checked Feb 12_22)

Privacy and Access to Healthcare Data

No end of demand for PHI in sight…

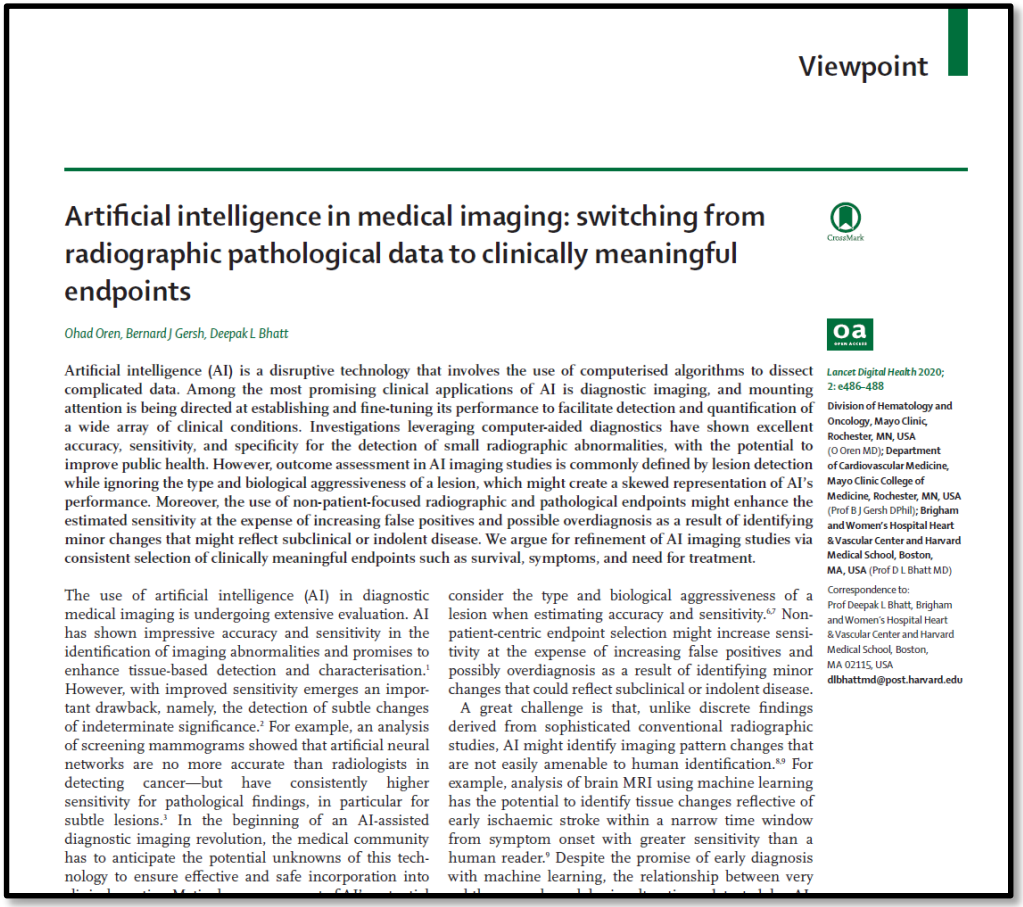


Pesapane, F., Codari, M. & Sardanelli, F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur Radiol Exp* **2,** 35 (2018).

https://eurradiolexp.springeropen.com/articles/10.1186/s41747-018-0061-6 (checked Feb 12_22)

Privacy and Access to Healthcare Data

No end of demand for PHI in sight…



Viewpoint

Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints

Ohad Oren, Bernard J Gersh, Deepak L Bhatt

Artificial intelligence (AI) is a disruptive technology that involves the use of computerised algorithms to dissect complicated data. Among the most promising clinical applications of AI is diagnostic imaging, and mounting attention is being directed at establishing and fine-tuning its performance to facilitate detection and quantification of a wide array of clinical conditions. Investigations leveraging computer-aided diagnostics have shown excellent accuracy, sensitivity, and specificity for the detection of small radiographic abnormalities, with the potential to improve public health. However, outcome assessment in AI imaging studies is commonly defined by lesion detection while ignoring the type and biological aggressiveness of a lesion, which might create a skewed representation of AI's performance. Moreover, the use of non-patient-focused radiographic and pathological endpoints might enhance the estimated sensitivity at the expense of increasing false positives and possible overdiagnosis as a result of identifying minor changes that might reflect subclinical or indolent disease. We argue for refinement of AI imaging studies via consistent selection of clinically meaningful endpoints such as survival, symptoms, and need for treatment.

The use of artificial intelligence (AI) in diagnostic medical imaging is undergoing extensive evaluation. AI has shown impressive accuracy and sensitivity in the identification of imaging abnormalities and promises to enhance tissue-based detection and characterisation.[1] However, with improved sensitivity emerges an important drawback, namely, the detection of subtle changes of indeterminate significance.[2] For example, an analysis of screening mammograms showed that artificial neural networks are no more accurate than radiologists in detecting cancer—but have consistently higher sensitivity for pathological findings, in particular for subtle lesions.[3] In the beginning of an AI-assisted diagnostic imaging revolution, the medical community has to anticipate the potential unknowns of this technology to ensure effective and safe incorporation into

consider the type and biological aggressiveness of a lesion when estimating accuracy and sensitivity.[6,7] Non-patient-centric endpoint selection might increase sensitivity at the expense of increasing false positives and possibly overdiagnosis as a result of identifying minor changes that could reflect subclinical or indolent disease.

A great challenge is that, unlike discrete findings derived from sophisticated conventional radiographic studies, AI might identify imaging pattern changes that are not easily amenable to human identification.[8,9] For example, analysis of brain MRI using machine learning has the potential to identify tissue changes reflective of early ischaemic stroke within a narrow time window from symptom onset with greater sensitivity than a human reader.[9] Despite the promise of early diagnosis with machine learning, the relationship between very

Lancet Digital Health 2020; 2: e486–488

Division of Hematology and Oncology, Mayo Clinic, Rochester, MN, USA (O Oren MD); Department of Cardiovascular Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA (Prof B J Gersh DPhil); Brigham and Women's Hospital Heart & Vascular Center and Harvard Medical School, Boston, MA, USA (Prof D L Bhatt MD)

Correspondence to: Prof Deepak L Bhatt, Brigham and Women's Hospital Heart & Vascular Center and Harvard Medical School, Boston, MA 02115, USA
dlbhattmd@post.harvard.edu

The need for even <u>more</u> personal data?

"Investigations leveraging computer-aided diagnostics have shown **excellent accuracy, sensitivity, and specificity** for the detection of small radiographic abnormalities, with the potential to improve public health.

However, outcome assessment in AI imaging studies is **commonly defined by lesion detection while ignoring the type and biological aggressiveness of a lesion**, which might create a skewed representation of AI's performance.

Moreover, the use of non-patient-focused radiographic and pathological endpoints might **enhance the estimated sensitivity at the expense of increasing false positives and possible over-diagnosis** as a result of identifying minor changes that might reflect subclinical or indolent disease.

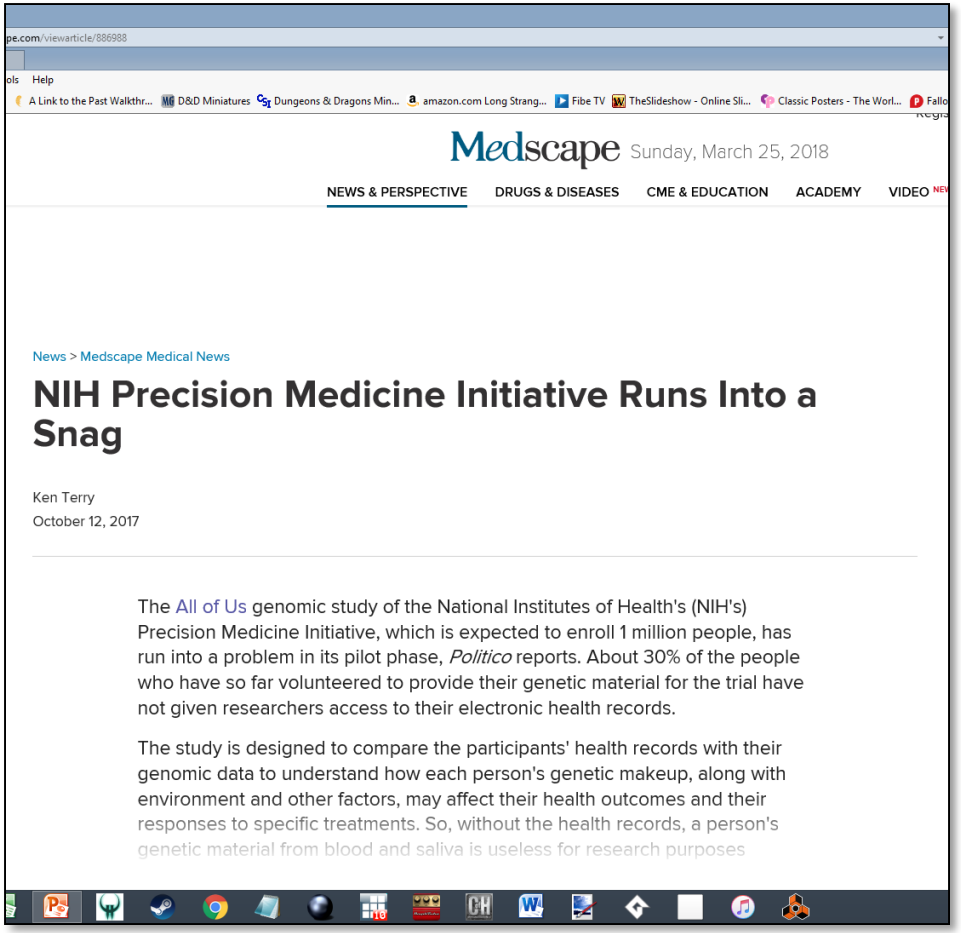
We argue for refinement of AI imaging studies via consistent selection of clinically meaningful endpoints such as survival, symptoms, and need for treatment."

Ohad, O., Gersh, Bernard J., Bhatt, Deepak L. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *The Lancet Digital Health, Pages e441-e492 (September 2020)*

https://www.sciencedirect.com/science/article/pii/S2589750020301606
(checked Feb 12_22)

Privacy and Access to Healthcare Data

Bias begins with the data…more data is the solution, and the problem



pe.com/viewarticle/886988

ols Help

A Link to the Past Walkthr... D&D Miniatures Dungeons & Dragons Min... amazon.com Long Strang... Fibe TV TheSlideshow - Online Sli... Classic Posters - The Worl... Fallo

Medscape Sunday, March 25, 2018

NEWS & PERSPECTIVE DRUGS & DISEASES CME & EDUCATION ACADEMY VIDEO NEW

News > Medscape Medical News

NIH Precision Medicine Initiative Runs Into a Snag

Ken Terry
October 12, 2017

The All of Us genomic study of the National Institutes of Health's (NIH's) Precision Medicine Initiative, which is expected to enroll 1 million people, has run into a problem in its pilot phase, *Politico* reports. About 30% of the people who have so far volunteered to provide their genetic material for the trial have not given researchers access to their electronic health records.

The study is designed to compare the participants' health records with their genomic data to understand how each person's genetic makeup, along with environment and other factors, may affect their health outcomes and their responses to specific treatments. So, without the health records, a person's genetic material from blood and saliva is useless for research purposes

"The All of Us genomic study of the National Institutes of Health's (NIH's) Precision Medicine Initiative, which is expected to enroll 1 million people, has run into a problem in its pilot phase, *Politico* reports. **About 30% of the people who have so far volunteered to provide their genetic material for the trial have not given researchers access to their electronic health records.**

Another big challenge for All of Us will be to enroll participants who represent the genetic pool of the entire US population. **In 2009, only 4% of 1.7 million participants in genome-wide association studies were people not of European descent, a *Nature* study found.** A subsequent study in 2016 showed that the non-European percentage had climbed to 20%, but that still does not reflect the diversity of Americans."

https://www.medscape.com/viewarticle/886988

Privacy and Access to Healthcare Data

'Inherently identifying information' and re-identification risk

Science

REPORTS

Cite as: Y. Erlich *et al.*, *Science* 10.1126/science.aau4832 (2018).

Identity inference of genomic data using long-range familial searches

Yaniv Erlich[1,2,3,4]*, Tal Shor[1], Itsik Pe'er[2,3], Shai Carmi[5]

[1]MyHeritage, Or Yehuda 6037606, Israel. [2]Department of Computer Science, Fu Foundation School of Engineering, Columbia University, New York, NY, USA. [3]Center for Computational Biology and Bioinformatics (C2B2), Department of Systems Biology, Columbia University, New York, NY, USA. [4]New York Genome Center, New York, NY, USA. [5]Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel.

*Corresponding author. Email: erlichya@gmail.com

"Consumer genomics databases have reached the scale of millions of individuals. Recently, law enforcement authorities have exploited some of these databases to **identify suspects via distant familial relatives.** Using genomic data of 1.28 million individuals tested with consumer genomics, we investigated the power of this technique. **We project that about 60% of the searches for individuals of European-descent will result in a third cousin or closer match, which can allow their identification using demographic identifiers. Moreover, the technique could implicate nearly any US-individual of European-descent in the near future. We demonstrate that the technique can also identify research participants of a public sequencing project.** Based on these results, we propose a potential mitigation strategy and policy implications to human subject research."

Y. Erlich *et al.* Science 10.1126/science.aau4832 (2018)
http://science.sciencemag.org/content/362/6415/690

Privacy and Access to Healthcare Data

Governance, risk and compliance controls for PHI: it's more than legal

'Necessary': Somewhat preventative, but generally recuperative (i.e. after the damage is done…)

- **Ethical principles:** (TCPS2, professional codes of practice…)
- **Legal requirements:** (PHIPA in Ontario - varies by jurisdiction; generally behind schedule and backward looking; lowest common denominator hard to establish or comply with…)
- **Contractual Requirements:** cost and complexity scales with the size of the project or study

Technical Methods/Controls: Largely preventative, permits 'real world' manipulation of PHI (multi-site, rich data sets) in a confidential manner…requires specialist/technical capabilities not yet mainstream:

- **Distributed / centralized computing methods**
- **Homomorphic encryption methods**
- **Differential privacy**
- **Secure hardware**

Privacy and Access to Healthcare Data

REB/ DAC Governance is not enough…

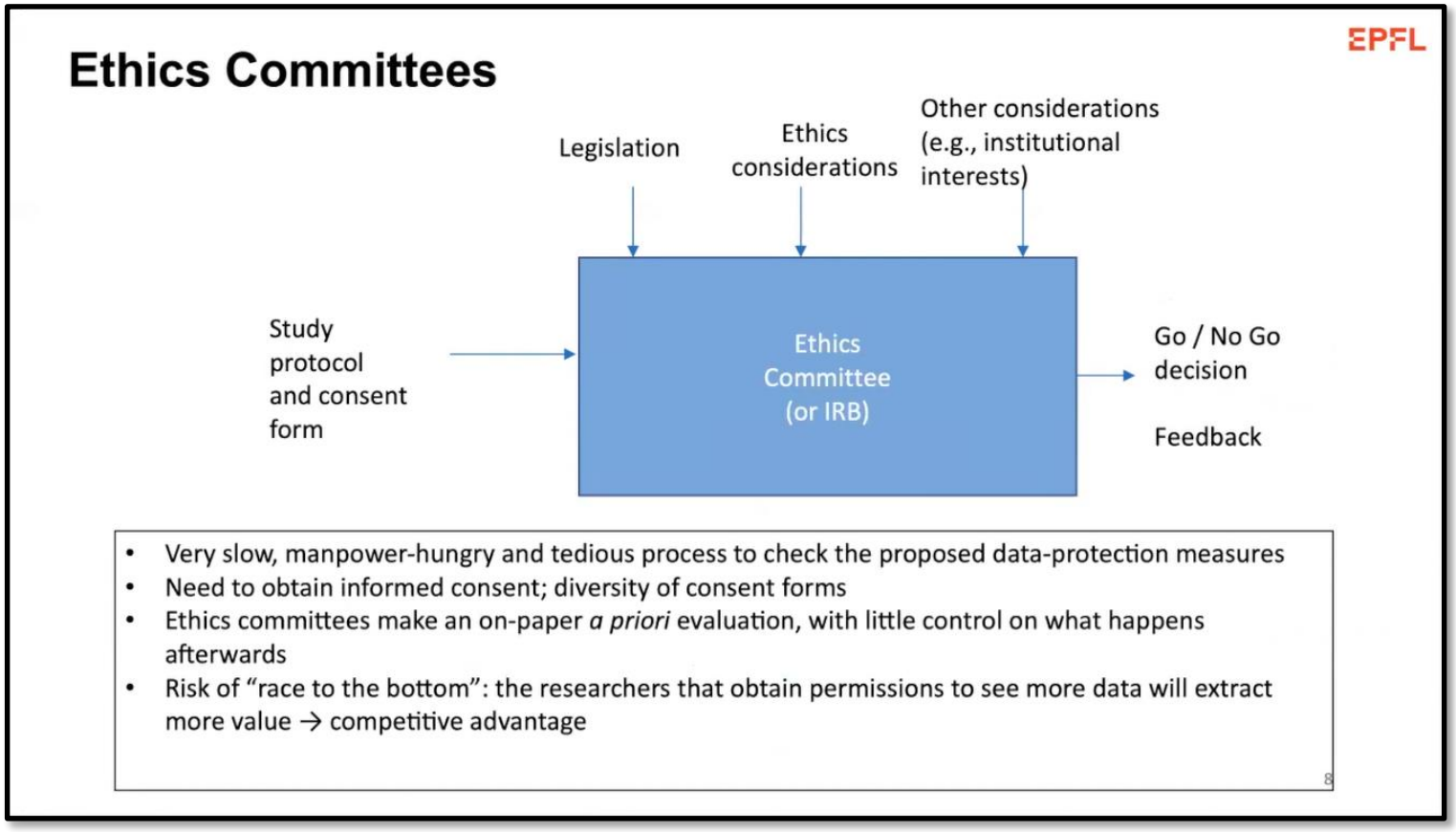
Table 5.3: Types of Access. RAS=Resource Access Subcommittee					
	Open Access (no login required)	Anyone with a login (no application necessary)	After approval of brief project review	With RAS and IRB review	With RAS and IRB review and additional participant consent
Newsletters, ongoing PMI studies and general updates	X				
Aggregate counts of individuals	X				
Graphical query to assess study feasibility using counts		X			
Query interface exact counts of rare events			X		
Access to de-identified individual-level data				X	
Access to identified data				X	
Recontact individuals				X	X
Clinical trials					X

'Real world' privacy problems start here:

https://allofus.nih.gov/

Privacy and Access to Healthcare Data

Governance is not enough…



Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

Privacy and Security Considerations – the emerging fine print…

On the one hand:

Recommendation 7.8: To safeguard against unintended release of the information, NIH should seek to establish an exemption under the Freedom of Information Act (FOIA) for release of genomic and other data held by the federal government.

On the other hand:

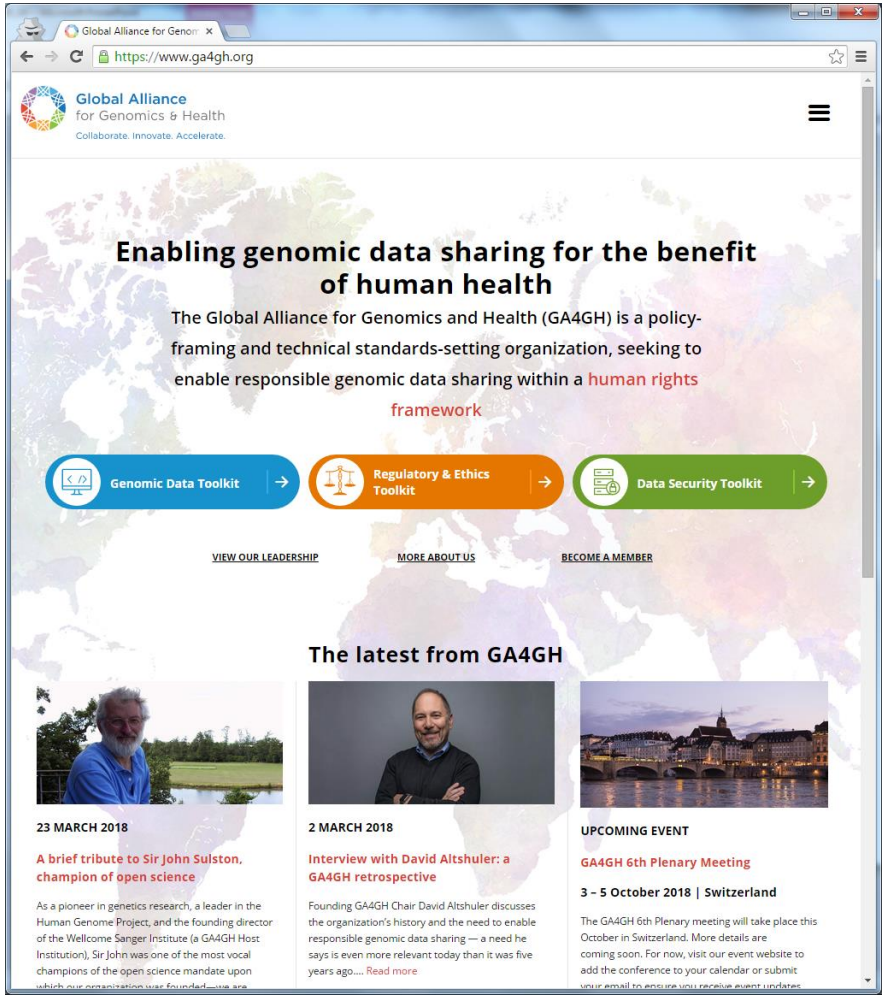
Recommendation 7.9: Unauthorized re-identification or re-contacting of participants should be expressly prohibited in agreements for the use of specimens and data, and NIH should pursue legislation penalizing such actions.

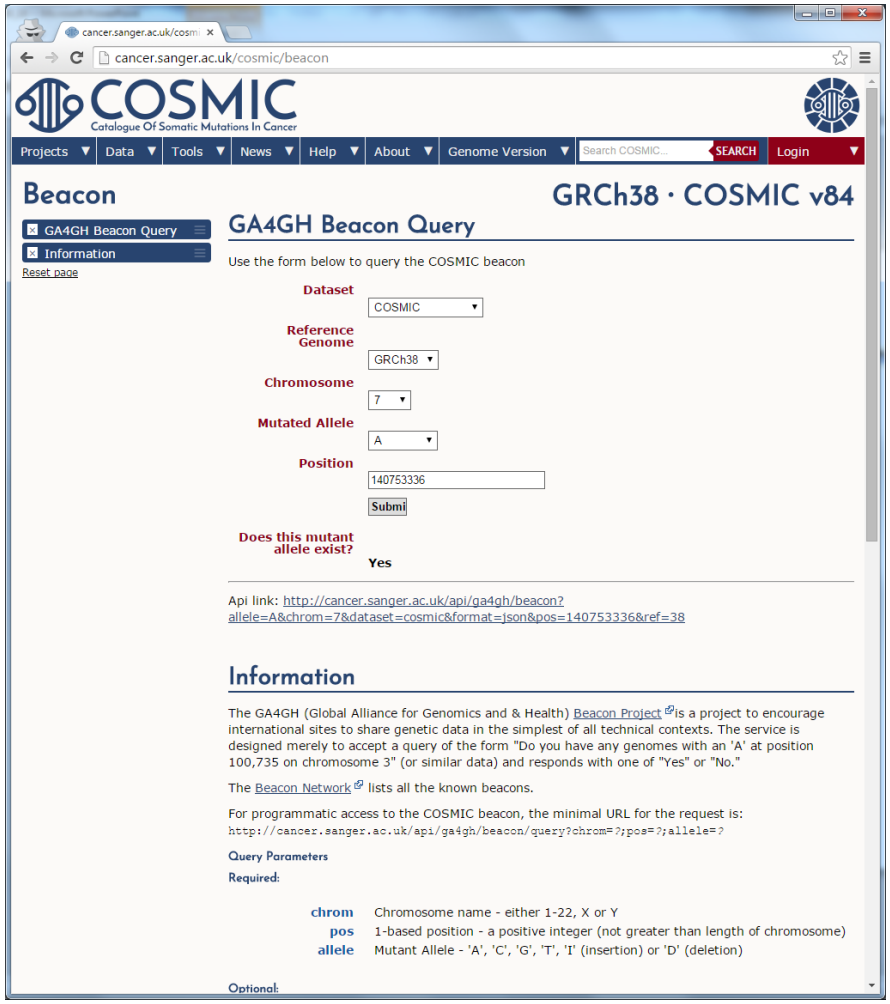
➔ Goodwill is not enough to protect privacy…

https://allofus.nih.gov/

Privacy and Access to Healthcare Data

Community Tools



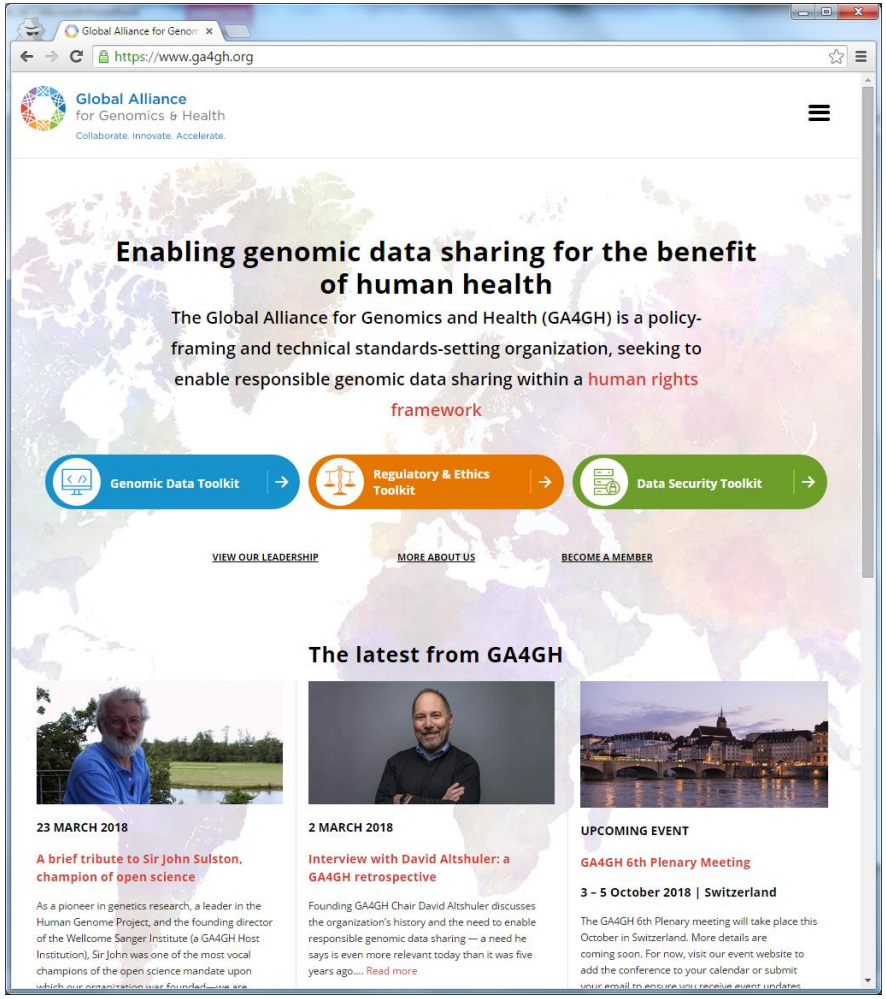


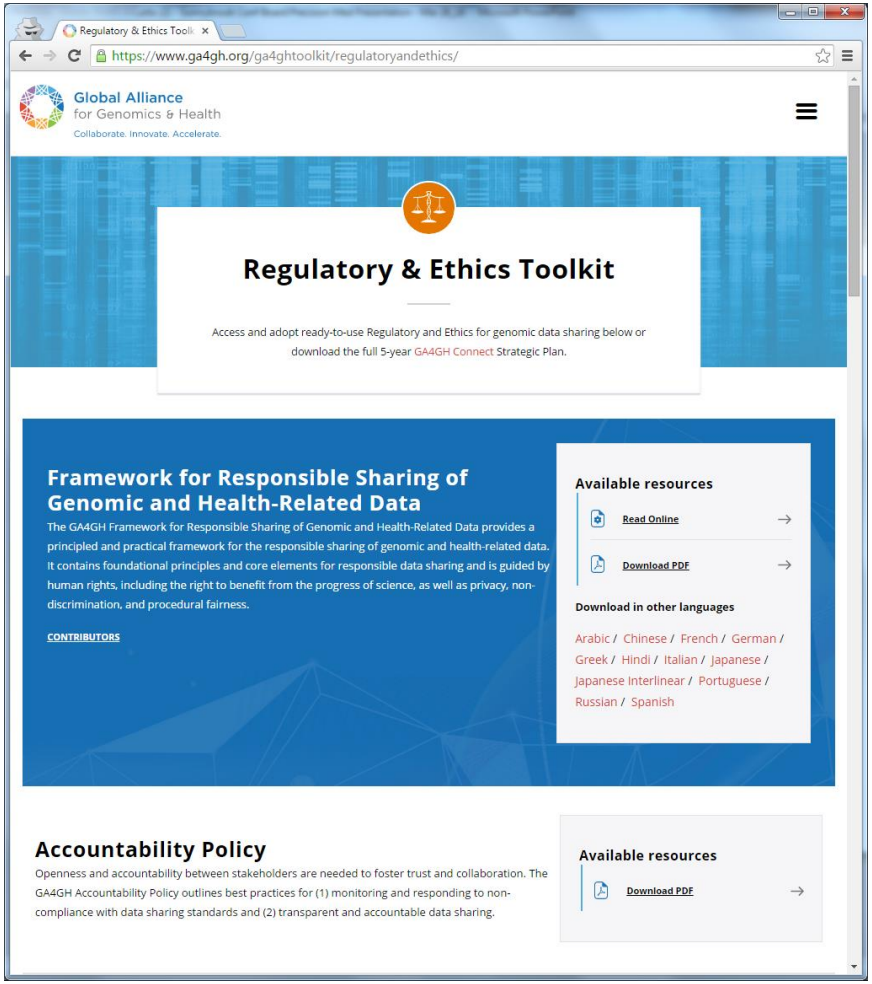
https://www.ga4gh.org/

http://cancer.sanger.ac.uk/cosmic/beacon

Privacy and Access to Healthcare Data

Community Tools





https://www.ga4gh.org/

http://cancer.sanger.ac.uk/cosmic/beacon

Privacy and Access to Healthcare Data

Current next-generation (technical) methods for privacy preservation

nature machine intelligence

PERSPECTIVE

https://doi.org/10.1038/s42256-020-0186-1

Check for updates

Secure, privacy-preserving and federated machine learning in medical imaging

Georgios A. Kaissis[1,2,3], Marcus R. Makowski[1], Daniel Rückert[2] and Rickmer F. Braren[1]

The broad application of artificial intelligence techniques in medicine is currently hindered by limited dataset availability for algorithm training and validation, due to the absence of standardized electronic medical records, and strict legal and ethical requirements to protect patient privacy. In medical imaging, harmonized data exchange formats such as Digital Imaging and Communication in Medicine and electronic data storage are the standard, partially addressing the first issue, but the requirements for privacy preservation are equally strict. To prevent patient privacy compromise while promoting scientific research on large datasets that aims to improve patient care, the implementation of technical solutions to simultaneously address the demands for data protection and utilization is mandatory. Here we present an overview of current and next-generation methods for federated, secure and privacy-preserving artificial intelligence with a focus on medical imaging applications, alongside potential attack vectors and future prospects in medical imaging and beyond.

Artificial intelligence (AI) methods have the potential to revolutionize the domain of medicine, as witnessed, for example, in medical imaging, where the application of computer vision techniques, traditional machine learning[1,2] and—more recently—deep neural networks have achieved remarkable successes. This progress can be ascribed to the release of large, curated corpora of images (ImageNet[3] perhaps being the best known), giving rise to performant pre-trained algorithms that facilitate transfer learning and led to increasing publications both in oncology—with applications in tumour detection[4,5], genomic characterization[6,7], tumour subtyping[8,9], grading prediction[10], outcome risk assessment[11] or risk of relapse quantification[12]—and non-oncologic applications, such as chest X-ray analysis[13] and retinal fundus imaging[14].

To allow medical imaging AI applications to offer clinical decision support suitable for precision medicine implementations, even larger amounts of imaging and clinical data will be required. Large cross-sectional population studies based solely on volunteer participation, such as the UK Biobank[15], cannot fill this gap. Even the largest current imaging studies in the field[4,5], demonstrating better-than-human performance in their respective tasks, include considerably less data than, for example, ImageNet[3], or the amount of data used to train algorithmic agents in the games of Go or StarCraft[16,17], or autonomous vehicles[18]. Furthermore, such datasets often stem from relatively few institutions, geographic regions or

and Communications in Medicine (DICOM)[21] is the universally adopted imaging data format, and electronic file storage is the near-global standard of care. Even where non-digital formats are still in use, the archival nature of, for instance, film radiography allows post hoc digitization, seen, for example, in the CBIS-DDSM dataset[22], consisting of digitized film breast radiographs. Digital imaging data, easily shareable, permanently storable and remotely accessible in the cloud has driven the aforementioned successes of medical imaging AI.

The second issue representing a stark deterrent from multi-institutional/multi-national AI trials[23] is the rigorous regulation of patient data and the requirements for its protection. Both the United States Health Insurance Portability and Accountability Act (HIPAA)[24] and the European General Data Protection Regulation (GDPR)[25] mandate strict rules regarding the storage and exchange of personally identifiable data and data concerning health, requiring authentication, authorization, accountability and—with GDPR—AI interpretability, sparking considerations on data handling, ownership and AI governance[26,27]. Ethical, moral and scientific guidelines (soft law[28]) also prescribe respect towards privacy—that is, the ability to retain full control and secrecy about one's personal information. The term privacy is used in this article to encapsulate both the intention to keep data protected from unintended leakage and from deliberate disclosure attempts (that is, synonymous with

The DI AI privacy challenge…and emerging solutions

"The broad application of artificial intelligence techniques in medicine is **currently hindered by limited dataset availability for algorithm training and validation, due to the absence of standardized electronic medical records, and strict legal and ethical requirements to protect patient privacy.**

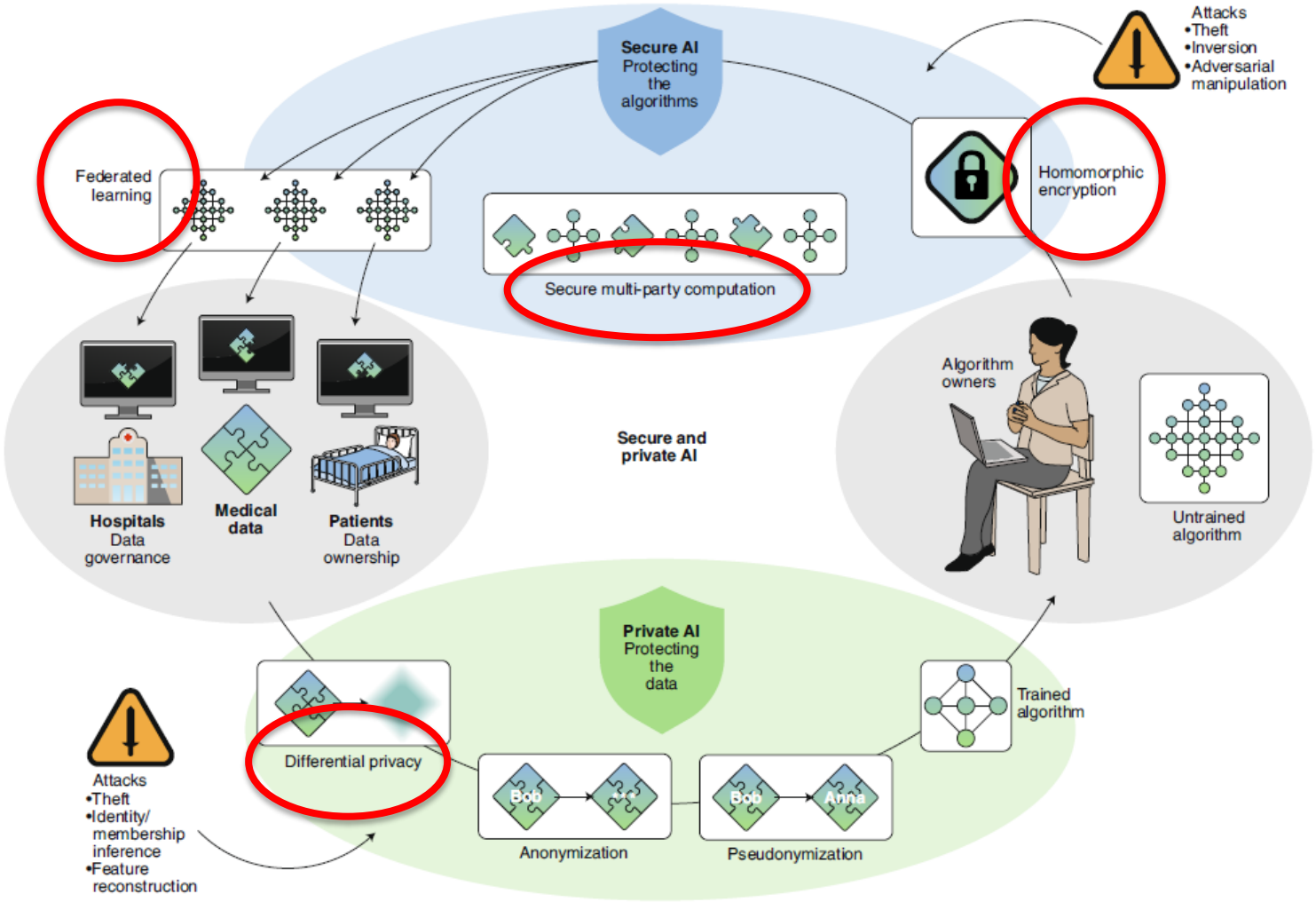
… To prevent patient privacy compromise while promoting scientific research on large datasets that aims to improve patient care, **the implementation of technical solutions to simultaneously address the demands for data protection and utilization is mandatory**. Here we present an overview of current and **next-generation methods for federated, secure and privacy-preserving artificial intelligence** with a focus on medical imaging applications, alongside potential attack vectors and future prospects in medical imaging and beyond."

Kaissis, G.A., Makowski, M.R., Rückert, D. *et al.* Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell* **2,** 305–311 (2020).

https://www.nature.com/articles/s42256-020-0186-1 (checked Feb 12_22)

Privacy and Access to Healthcare Data

Current next-generation (technical) methods for privacy preservation



Kaissis, G.A., Makowski, M.R., Rückert, D. *et al.* Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell* **2,** 305–311 (2020).

https://www.nature.com/articles/s42256-020-0186-1 (checked Feb 12_22)

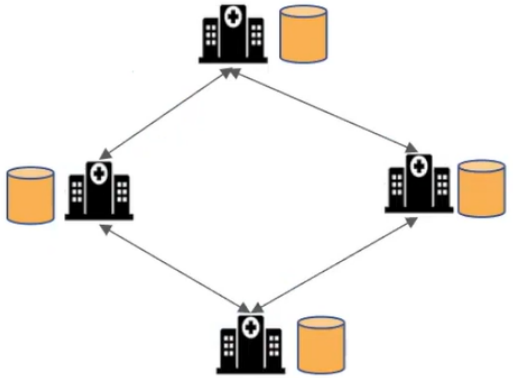
Privacy and Access to Healthcare Data

De-identification is the norm…but it's not enough

EPFL

Multi-site studies

- Benefit: increase the amount of available data and thus the statistical significance of findings
- Challenges
  - Need to interface **several** ethics committees and IT services
  - **Diverse** legislations (especially if international)
  - **Heterogeneity** of data semantics and of data quality (partially due to reluctance to data transfer)
  - **Reluctance** to share data (control, publication advantage)
  - **Privacy protection/regulations**
- **Widespread practice:** de-identification of the data



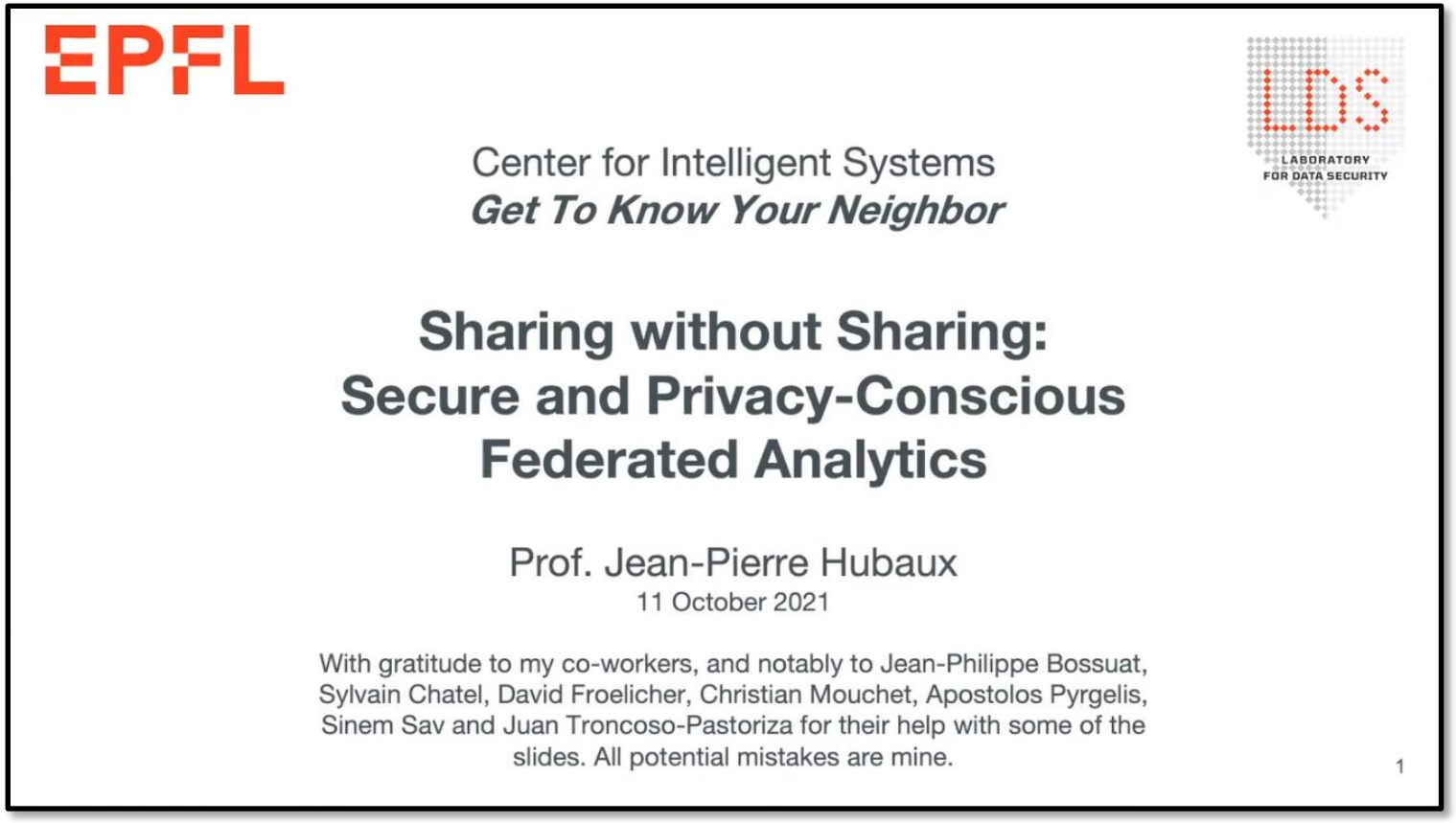
- Overall, very little awareness of cryptography and differential privacy in hospitals

6

Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

"Federated Analytics" example



EPFL

Center for Intelligent Systems

Get To Know Your Neighbor

LABORATORY FOR DATA SECURITY

Sharing without Sharing: Secure and Privacy-Conscious Federated Analytics

Prof. Jean-Pierre Hubaux

11 October 2021

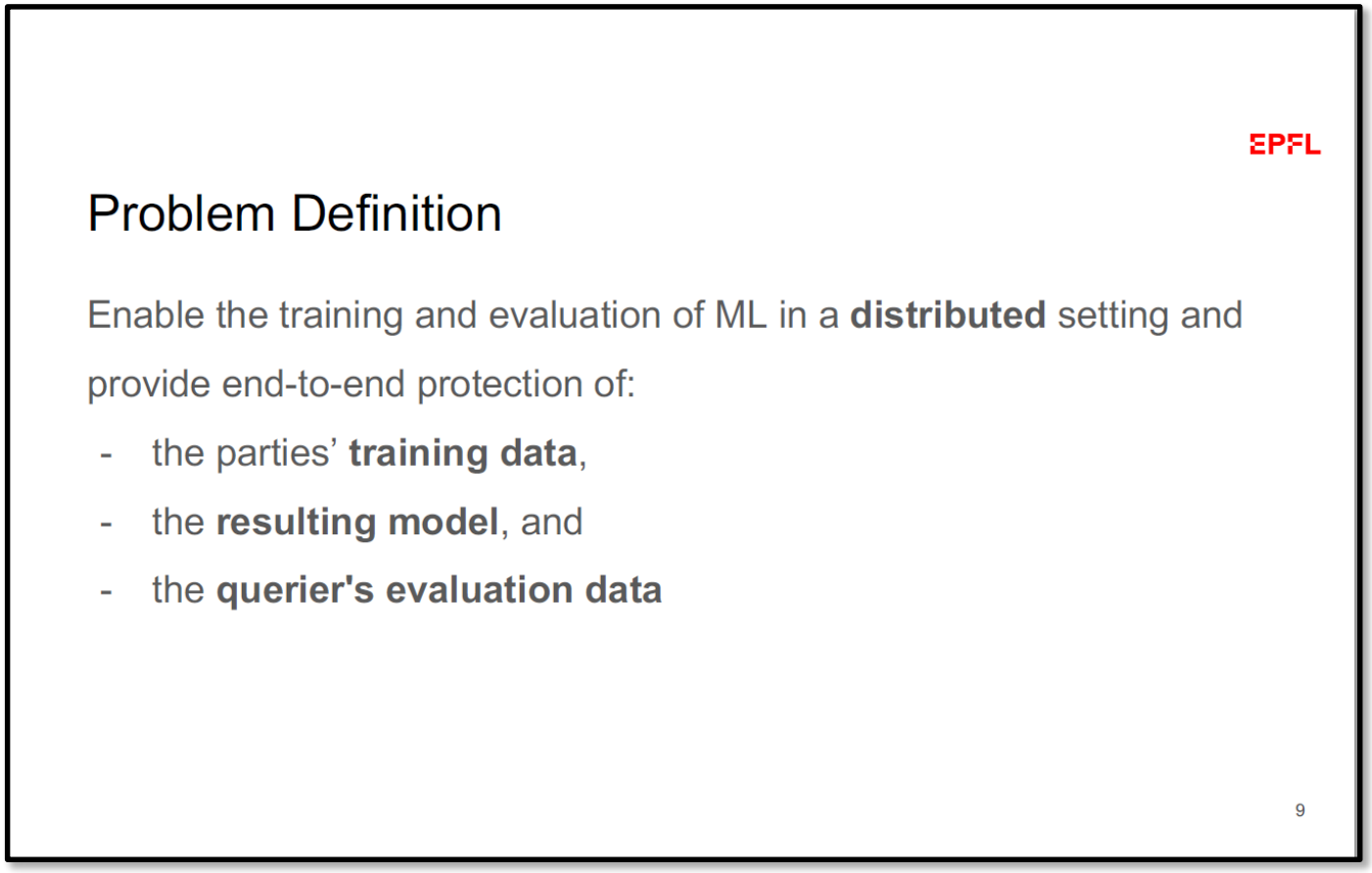
With gratitude to my co-workers, and notably to Jean-Philippe Bossuat, Sylvain Chatel, David Froelicher, Christian Mouchet, Apostolos Pyrgelis, Sinem Sav and Juan Troncoso-Pastoriza for their help with some of the slides. All potential mistakes are mine.

1

Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

”Federated Analytics” example



Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

"Federated Analytics" example

Motivation for Federated Analytics

MIT Technology Review

Why is it so hard to review the Johnson & Johnson vaccine? Data.

The clock is ticking for regulators looking into covid vaccine side effects. But their task is made harder by America's fragmented data systems.

- More than 1 billion people worldwide are fully vaccinated against COVID-19
- Severe (life threatening) reactions are extremely rare and dispersed around the globe
- Studying these cases requires the international sharing of dispersed sensitive patients' data

However, sensitive/personal data are difficult to share because of:

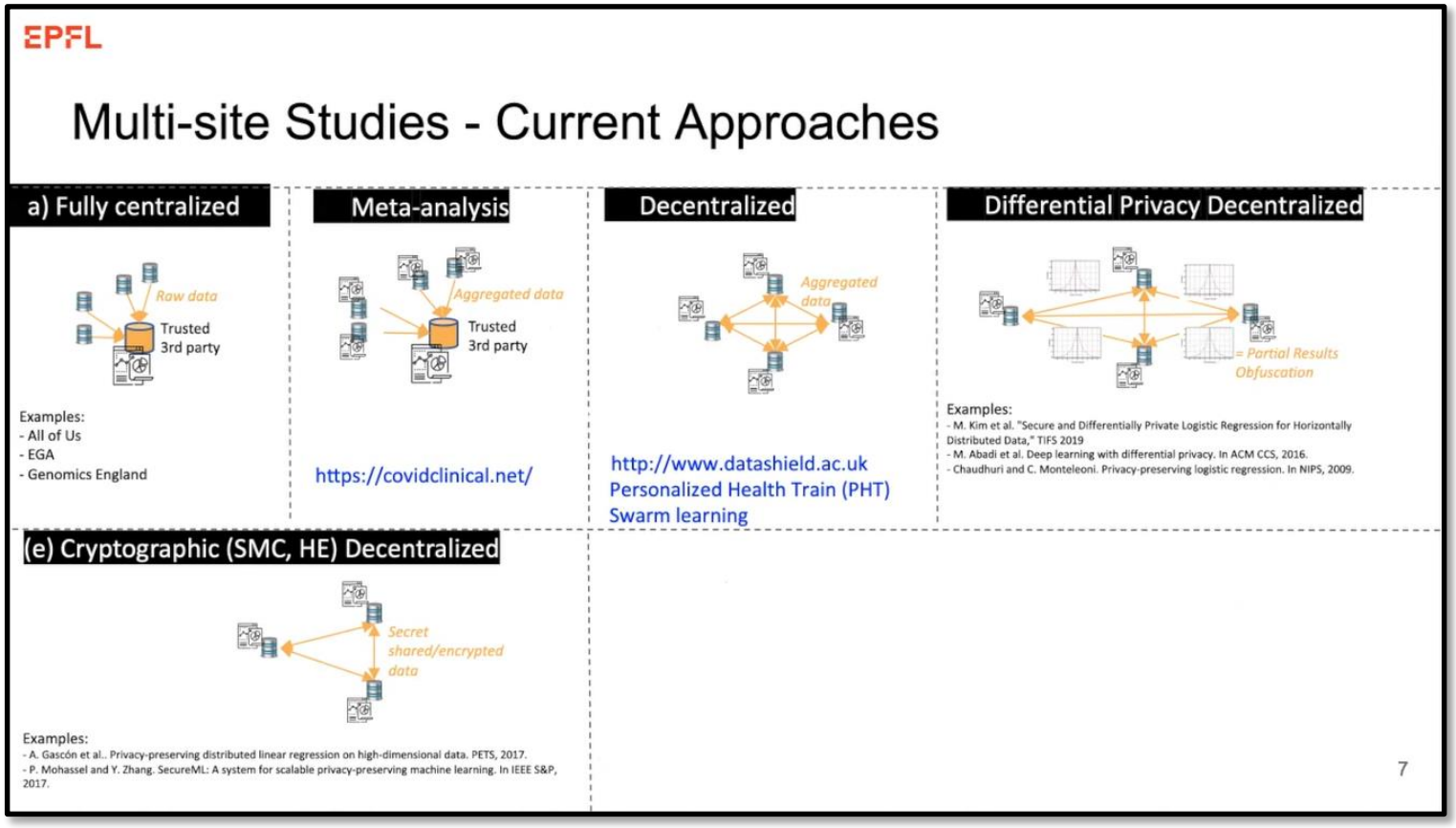
- Stringent regulations, e.g., GDPR.
- Complex/costly data-access agreements
- High repercussions in case of data leakage
- Competition among stakeholders

→ **Sensitive data are often siloed**

Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

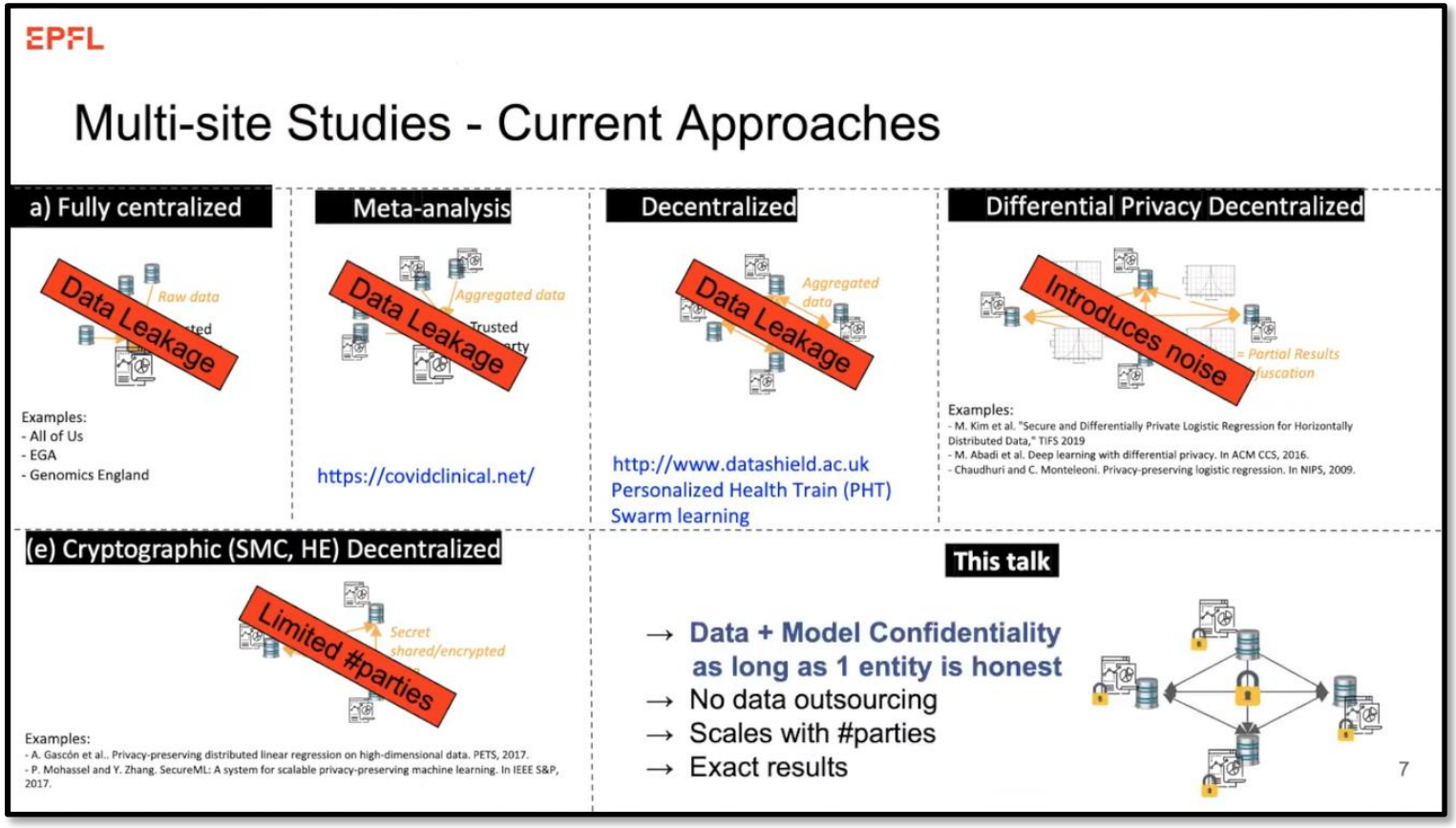
The Challenge of Multi-site studies



Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

The Challenge of Multi-site studies



Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

"Federated Analytics" example – PIA Summary

Data Protection Impact Assessment (DPIA) for multisite medical data analysis (June 2021)

Centralized approach with standard pseudonymization

Threat	Threat likelihood	Threat impact	Risk	Risk level
Unlawful access to the system	Unlikely	High	Loss of data confidentiality	Moderate
Malicious use of the system	Possible	High	Loss of data confidentiality	High
Loss of data	Unlikely	Minor	Loss of data integrity, data unavailability	Minor
Data leak of host/cloud	Possible	High	Loss of data confidentiality	High
Collusion of host/cloud	Possible	High	Loss of data confidentiality	High
Corrupted or malicious host/cloud	Possible	High	Data unavailability, loss of data integrity, loss of data confidentiality, loss of data correctness	High
Unavailability of host/cloud	Possible	Minor	Data unavailability, loss of data correctness	Moderate
Re-identification/attribute inference	Possible	High	Loss of data confidentiality	High

Federated approach enhanced **with MedCo**

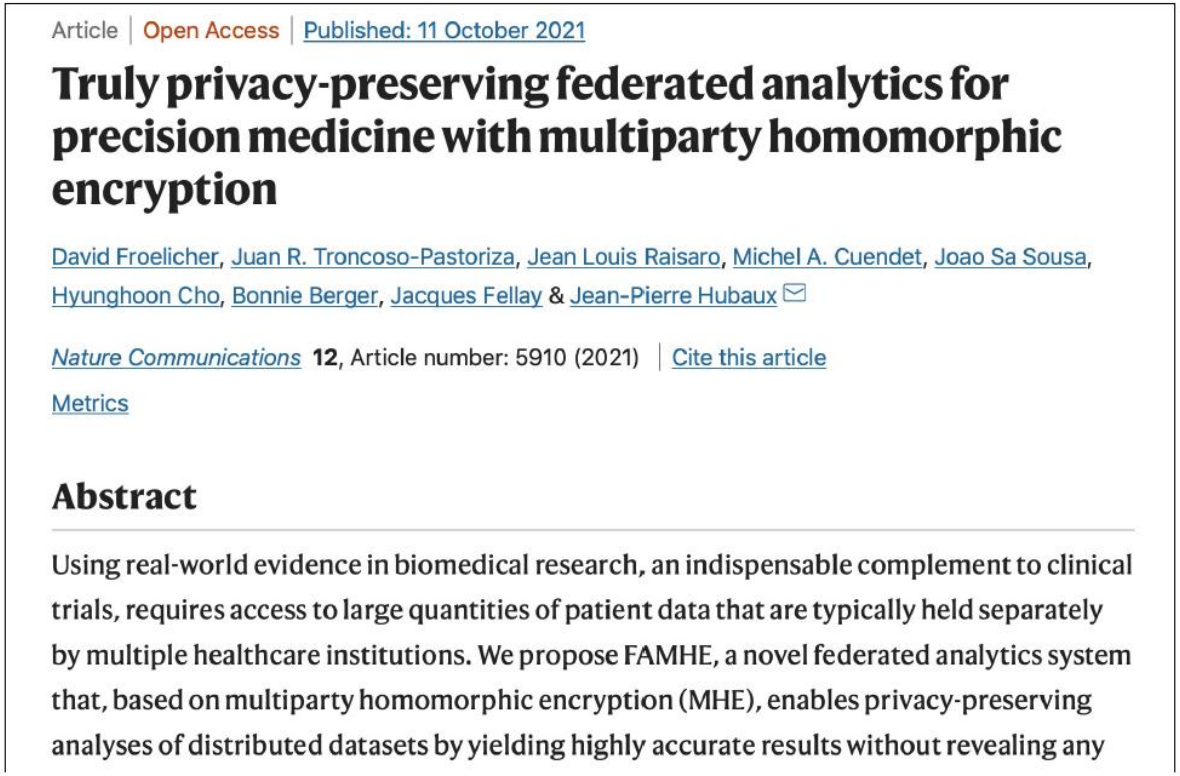
Threat	Measure introduced with MedCo	Threat likelihood	Threat Impact	Risk	Risk level
Unlawful access to the system	1	Unlikely	Minor	Loss of data confidentiality	Low
Malicious use of the system	1, 2, 4, 10	Possible	Minor	Loss of data confidentiality	Low
Loss of data	3, 5	Unlikely	Minor	Loss of data integrity, data unavailability	Low
Data leak	4, 5, 8, 9, 10	Unlikely	Minor	Loss of data confidentiality	Low
Collusion between nodes	4, 9	Unlikely	Moderate	Loss of data confidentiality	Moderate
Corrupted or malicious nodes	2, 5, 6, 7, 8, 9	Unlikely	Moderate	Data unavailability, loss of data integrity, loss of data confidentiality, loss of data correctness	Moderate
Unavailability of of nodes	6, 7	Possible	Minor	Data unavailability, loss of data correctness	Moderate
Re-identification or attribute inference	1, 2, 4, 9, 10	Unlikely	Minor	Loss of data confidentiality	Low

Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

"Federated Analytics" example

The "Holy Grail" for secure federated analytics of health data: FAMHE



Article | Open Access | Published: 11 October 2021

Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption

David Froelicher, Juan R. Troncoso-Pastoriza, Jean Louis Raisaro, Michel A. Cuendet, Joao Sa Sousa, Hyunghoon Cho, Bonnie Berger, Jacques Fellay & Jean-Pierre Hubaux

Nature Communications **12**, Article number: 5910 (2021) | Cite this article

Metrics

Abstract

Using real-world evidence in biomedical research, an indispensable complement to clinical trials, requires access to large quantities of patient data that are typically held separately by multiple healthcare institutions. We propose FAMHE, a novel federated analytics system that, based on multiparty homomorphic encryption (MHE), enables privacy-preserving analyses of distributed datasets by yielding highly accurate results without revealing any

Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

"Federated Analytics" example: Data and Model Confidentiality

Objectives

Data Confidentiality: During training and prediction, no party (including the querier) should learn more information about the input data of any other honest party, other than what can be deduced from its own input data.

Model Confidentiality: During training and prediction, no party (including the querier) should gain more information about the trained model weights, other than what can be deduced from its own input data.

41

Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Privacy and Access to Healthcare Data

EPFL Publications

Publications

Analytics Platform

NDSS 2021
POSEIDON: Privacy-Preserving Federated Neural Network Learning

PETS 2021
Spindle: Scalable Privacy-Preserving Federated Neural Network Learning

Cryptographic core

Eurocrypt 2021
Efficient Bootstrapping for Approximate Homomorphic Encryption with Non-Sparse Keys

Medical Application

Nature Communications 2021
FAMHE: Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multi-party Homomorphic Encryption

Nature Comp. Sci. 2021
Citizen-Centered, Auditable, and Privacy-Preserving Population Genomics

Security Framework

PETS 2021
Multiparty Homomorphic Encryption from Ring-Learning with Errors

60

Sharing without Sharing – Secure and Privacy-Conscious Federated Analytics
Prof. Jean-Pierre Hubaux, Head of the Laboratory for Data Security (LDS), EPFL
https://www.epfl.ch/research/domains/cis/center-for-intelligent-systems-cis/events/gtkyn/prof-jean-pierre-hubaux/ (checked Feb 12_22)

Questions?

Thank You!

jeff.curtis@sunnybrook.ca

Canada Health Infoway

Sunnybrook
when it matters MOST

Questions?

Thank You!

jeff.curtis@sunnybrook.ca

Canada Health Infoway