

SUPPORTING HEALTH INNOVATION WITH FAIR INFORMATION PRACTICE PRINCIPLES

Key issues emerging from the
OECD-Israel Workshop of
19-20 January 2021



ISRAEL
TECH POLICY INSTITUTE
An Affiliate of the Future of Privacy Forum

State of Israel
Ministry of Health
משרד הבריאות, ישראל

SUPPORTING HEALTH INNOVATION WITH FAIR INFORMATION PRACTICE PRINCIPLES

Key issues emerging from the OECD-Israel Workshop

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Acknowledgements

This report was drafted by Limor Shmerling Magazanik (Managing Director, Israel Tech Policy Institute (ITPI)), based on inputs from workshop experts, the OECD Working Party on Health Care Quality and Outcomes and the OECD Working Party on Data Governance. It summarizes insights from the discussions and presentations and includes post-workshop feedback received on possible next phases of OECD work.

Both this report and the workshop were organised and produced in collaboration with Jillian Oderkirk, Duniya Dedeyn and Ricardo Sanchez Torres from OECD DELSA and Elettra Ronchi and Lauren Bourke from OECD DSTI.

Particular thanks and recognition are extended to: Clayton Harris (WHO), Louisa Stüwe, International Policy Manager (France Health Data Hub), Thomas Duong, Data Protection Officer (France Health Data Hub), Antti Piirainen, Head of Communications THL (Finland Health Ministry), Dr. Jin Yong LEE, Chair of Health Insurance Review and Assessment Agency (HIRA Korea), Elisabeth Myers, ONC Deputy Director of Policy (HHS Office of the National Coordinator for Health IT, United States), Esti Shelly, Director of Digital Health (Israel Ministry of Health), Omer Tene, Vice President, Chief Knowledge Officer of IAPP (International Association of Privacy Professionals), Dr. Eyal Zimlichman, Deputy Director General, Chief Medical Officer and Chief Innovation Officer (Sheba Medical Center), David Hoffman, Duke Sanford Professor & Associate General Counsel, Senior Director of Data Policy Strategy (Intel Corporation), Jules Polonetsky, CEO (Future of Privacy Forum), Christine Runnegar, Senior Director ('Internet Trust' Internet Society, ISOC), Jean-Pierre Hubaux, EPFL (École Polytechnique Fédérale de Lausanne), Hadas Bitran, Head of Microsoft Health (Microsoft Israel R&D Center), Nigel Hughes, EHDEN Coordinator and Scientific Director, Epidemiology, Janssen, EFPIA (The European Federation of Pharmaceutical Industries and Associations), Sivan Tamir (Policy Counsel and Senior Researcher, ITPI), Richard Milne (Wellcome Connecting Science), Irene Schluender (BBMRI-ERIC), Kaisa Immonen, EFE (European Patients' Forum), Knut Mager (Novartis).

Abstract

This document reports on key issues emerging from the OECD Workshop: "Supporting Health Innovation with Fair Information Practice Principles". The workshop was held virtually on 19-20 January 2021, in collaboration with the Israel Ministry of Health (IMoH) and the Israel Tech Policy Institute (ITPI). The objective of the workshop was to further international dialogue on issues critical for the successful use of health data for the benefit of the public, focusing on the implementation of privacy protection principals and the challenges that arise in the process. This workshop was also an opportunity to contribute evidence-based policy insights to the evaluation of the 2017 OECD Recommendation on Health Data Governance, and discuss relevant follow-on actions and recommendations emerging from the recent review of the implementation of the 2013 OECD Privacy Guidelines.

Table of contents

Acknowledgements	3
Abstract	3
Introduction	6
Executive Summary	8
1 The right to inclusion in health research and the significance of public trust	10
Legal Frameworks offering Privacy Safeguards	12
2 Experiences developing health data and governance	14
France Health Data Hub	14
Finland Findata	17
Korea digital new deal	19
United States new rule within the 21 st Century Cures Act	21
Israel National and Organizational 'Data Lakes'	23
European Health Data & Evidence Network	26
3 Data governance policy issues	29
Insufficient clarity and harmonization of regulatory frameworks	29
Governance of cross-border data transfers	30
Patient perspectives and public trust	32
The Genomic Data Challenge	34
Legal basis for the secondary use of health data for research	37
Data quality and linkage	38
4 Tools to support adoption and scaling of research with health data	39
Privacy by Design & Privacy Enhancing Technologies	39
Synthetic data	41
Homomorphic Encryption Computation	42
The (Democratic) Federated Model	43

5 Priorities for future collaborative efforts	45
Annex A. Workshop agenda	46
Annex B. Speakers	49
Annex C. Participating delegations and organisations	58
Annex D. Summary of the report “Using health data for research: Evolving national policies”	63
Annex E. Summary of the OECD survey of national health data development and governance	68
Introduction	68
National health data availability, maturity and use	70
National health data governance	74
Next steps	79

Tables

Table 1. Key national health dataset availability, maturity and use	71
Table 2. Proportion of key national health care datasets with key data governance elements	79

Figures

Figure 1. Small group of countries score highly on both dataset availability, maturity and use and dataset governance	70
Figure 2. Percentage of key national health datasets available and regularly linked for monitoring and research	73

Boxes

Box 1. OECD Council Recommendation on Health Data Governance	69
--	----

Introduction

1. The interactions between data protection and scientific research are complex, with privacy and data protection enhancing individuals' trust and ensuring respect of fundamental rights and ethical standards, while at the same time, creating friction for data collection and sharing across organizations and borders. The COVID-19 pandemic has brought to the fore the crucial role that data collection, analysis, sharing, and dissemination play for governments, academic institutions, and private sector businesses racing to advance scientific research to help contain, mitigate and recover from the virus.
2. Data-driven health innovation raises the need for further development of fair information processing practices as well as national health data governance frameworks. Stakeholders should convene to determine the scope of health data available to researchers and identify the uses of data that are in the public interest.
3. The 2017 [OECD Recommendation on Health Data Governance](#) calls on countries to develop national frameworks that encourage the availability and secondary use of personal health data to serve public interest goals, while protecting privacy and data security and maintaining data subject rights. Since then, countries have made investments and implemented legal and organizational reforms to strengthen health data governance. While the OECD Recommendation provides principles for the development of national frameworks, challenges remain in implementing these principles in practice at the national level and allowing for international collaboration on health data for research.
4. At the time of the planning of this workshop the OECD was engaged in a comprehensive review of the implementation of the Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data [[OECD/LEGAL/0188](#)], which have been a cornerstone for data protection regulation in member countries and other countries around the world.
5. Health data are personal and sensitive. Their misuse could harm individuals by infringement of privacy; (individual and community) stigmatization; discrimination in areas such as insurance or employment; and individual and national security violations. Failure to provide policies and practices for secure and private data processing, access and sharing could undermine public trust not only in healthcare providers but also in the healthcare system itself.
6. At the same time, member countries must advance the public interest in improving health outcomes and providing high-quality healthcare solutions. This requires them to provide modern, patient-centered healthcare services, including contributing digital health data for scientific research.
7. This document reports on key issues emerging from the Workshop: "Supporting Health Innovation with Fair Information Practice Principles" held virtually on 19-20 January 2021. The workshop, jointly organized by the OECD, Israel Ministry of Health (IMoH) and the Israel Tech Policy Institute (ITPI), provided an opportunity for delegates of both the OECD Health Care Quality and Outcomes Working Party (HCQO) and the OECD Data Governance and Privacy Working Party in the Digital Economy group (DGP), together with experts in these fields, to discuss progress toward and challenges faced in implementing the 2017 OECD Recommendation on Health Data Governance, and to contribute to the ongoing review of the 2013 OECD Privacy Guidelines.

8. As background for this workshop, delegates and participants were invited to review a survey prepared by the ITPI and the Future of Privacy Forum (FPF), examining the legal frameworks for secondary use of healthcare data for research purposes in eight countries, including Australia, Finland, France, India, Ireland, Israel, England and the US (Annex D). The report demonstrates commonality across legal systems and regimes, essentially permitting secondary use of healthcare data for research purposes under certain conditions, including review by ethical boards, proper de-identification and additional administrative, technical and contractual safeguards. It also finds differences in the implementation of legal frameworks that may create barriers for cross border or cross sectors data sharing for collaborative research.

9. Delegates were also provided with two OECD reports. One, a summary of the findings of the [2019-20 OECD survey of countries' development of national health data, data uses and governance](#), which provides an overview of progress toward the implementation of the OECD Council Recommendation on Health Data Governance (OECD Health Data Survey). It highlights the considerable variability across countries and within countries regarding health data development, accessibility and data protection (Annex E). The second, a [report on data localisation trends and challenges](#) (OECD Data localisation report), surveys trends within, and attitudes towards, data localization amongst countries, consumers, industry and the expert community and highlights a complex situation in which data localization is both seen as useful and as a significant threat and obstacle.

Executive Summary

10. In order to learn about ongoing efforts and to identify priorities for future actions to support evidence-based health data governance policies, the OECD, Israel Ministry of Health and ITPI convened a two-day workshop in January 2021. The workshop sought to highlight country level insights and provide a platform for shared learning across countries, as well as to discuss priorities for future collaborative efforts.

11. The workshop agenda included the following topics:

- Significant national health data governance reforms implemented recently in countries, which included legal and operational reforms to strengthen health data governance. These examples were viewed in the context of the WHO Global Strategy on Digital Health and an Industry Consortium project.
- Safeguards for health data sharing to promote innovation while protecting people's privacy. These include: 1) ethical review board oversight; 2) de-identification; 3) administrative, technical and contractual safeguards; and 4) safeguards around cross border data flows.
- Privacy-by-Design and state-of-the-art solutions for safeguarding digital health data against unauthorized access and use.
- Individual & Community perspectives on the imperative to respect the individual's interest alongside those of the community and society; and on consent and alternative legal basis for the secondary use of patient data for research.

12. One key takeaway raised at the workshop is the recent conceptual evolution of a perception of research participation as a solidarity-based moral imperative or a moral obligation. This perception applies particularly to non-interventional, minimum-risk health data research and sample donation to (bio)-repositories. Seemingly, where there lies a promise for medical benefits for individuals and society at large, then the low health risk and minimal inconvenience entailed for research subjects in a data analysis invoke a stronger social and moral rationale for participation. At the same time, the autonomy, privacy and the right to self-determination of health data research participants must still be protected by the regulation.

13. Another key takeaway was that various countries are engaged in what may be dubbed the "3rd generation" of data protection regulations. Following the OECD Privacy Guidelines of 1980, and the amended OECD Guidelines of 2013, along with the EU GDPR, the 3rd generation of privacy and data protection principles is underway in health data research as well as in finance (Open Banking) and smart mobility. These would support research and innovation for the public good, including local and global data sharing, while protecting people's privacy rights and other liberties.

14. Key governance issues raised at the workshop were the need for clarity and harmonization of regulatory frameworks; challenges in cross-border data transfers; the importance of patient perspectives and public trust; genomic data challenges; views about the legal basis for the secondary use of health data for health research; and concerns about data quality and data linkages.

15. Focusing on the need for clarity and harmonization of regulatory frameworks, workshop participants discussed the legal basis for health data research. As seen in the ITPI/FPF report and the workshop presentations, countries usually decide on some equation where a compelling public interest

replaces consent for data use — and that interest is coupled with de-identification mechanisms and other safeguards. Moving away from consent for the use of health data for research to a public interest model is not without challenges. An interesting perspective raised is that reliance on (opt-in/opt-out) consent in the context of sharing health data might inadvertently introduce bias into the data, since those who are willing to engage in research, and are consequently represented therein, may not be fully representative of the population and lead to under-representation in research and its benefits.

16. An additional key takeaway raised is that countries use varying definitions for personal data and de-identified data, and that more knowledge is required about privacy enhancing solutions. Particularly, further work is needed around data de-identification to better assess the levels of privacy risk in the use of data with various levels of identifiability, and how they can be mitigated with additional safeguards.

17. The 2019-20 OECD Survey indicated that only half of countries would approve the sharing of de-identified national health datasets across borders with a university, non-profit or governmental body (Annex E). It is evident that international and multistakeholder collaboration in health research is of importance and requires interoperable regulation frameworks.

18. The following priority areas for future collaborative work emerged from the workshop discussions:

- Insufficient clarity and harmonization between national health data governance frameworks challenge their implementation within nations and between them.
- Data localisation laws and policies limit cross-border collaboration in health research.
- Conflicts about the appropriate legal basis for data processing: Consent vs. Public Interest – and the scope of “Public interest” – affect countries’ ability to conduct research and streamline data sharing project approvals.
- Insufficient awareness of and expertise in the use of Privacy Enhancing Solutions limit their wider adoption to strengthen data protection and limit their consideration when developing legal basis for data processing or approving data processing requests.
- Necessity of fostering public trust of individuals, communities and societies in scientific research that is in the public interest, through Inclusion, Transparency and Participation.
- Importance of recognizing and addressing inequalities in access to data for research, such as when researchers from diverse backgrounds are denied access to data due to payment or credit sharing requirements, and the risks of bias in research results emanating from this.
- Importance of recognizing and addressing the consequences of inequalities emanating from potential bias in the data explored, in and of itself, or due to insufficient inclusion of participants from diverse populations.
- Necessity of strong digital identification methods on a national level to support data linkage whilst preserving privacy and security.

1 The right to inclusion in health research and the significance of public trust

19. As was presented in the workshop, in some ways, the topic of this workshop is the result of two rather simultaneous evolutions: one technological and scientific – the development of data analytics, tools and their application to the abundance of accumulated health data – and the other a conceptual change regarding participation in research.

20. Over the past century,¹ research participation has gone from involuntary and uninformed inclusion of society's most disadvantaged members (perceived as a legitimate sacrifice for the greater good of society) – to a mere burden (subject to informed consent). It has then evolved to be perceived as a right – the right to be included in medical research – which is mainly asserted by underrepresented populations (what is known as *the ethics of inclusion*).

Research participation is increasingly perceived as a solidarity-based moral imperative or a moral obligation – particularly in the case of non-interventional, minimum-risk health data research and sample donation to (bio)-repositories.

21. The most recent conceptual evolution is toward the perception of research participation as a solidarity-based moral imperative or a moral obligation. This perception applies particularly to non-interventional, minimum-risk health data research and sample donation to (bio)-repositories. For example, an analysis of test results or medical patient files by a machine learning algorithm in search of patterns that will lead to early detection, prevention or cure of a medical condition, without conducting any physical contact with the patients or requiring any involvement from them. Seemingly, where there lies a promise for medical benefits for individuals and society at large, then the low health risk and minimal inconvenience entailed for research subjects invokes a stronger social and moral rationale for participation.

¹ A. Petersen and D. Lupton, *The New Public Health: Health and Self in the Age of Risk* (1996); S. Epstein, *Inclusion: the Politics of Difference in Medical Research* (2007); R. Tutton and B. Prainsack, "Enterprising or altruistic selves? Making up research subjects in genetics research," *Sociology of Health & Illness*, 33 (2011)

22. Such sentiments, social expectation or personal motivation for participation, typically do not exist in the context of interventional medical research, such as drug trials, which carry higher risks and more substantial burdens for participants.

23. And indeed, surveys of public attitudes to patient data use conducted in the UK, for example, have demonstrated that "a high proportion of people support sharing patient data for research, where there is public benefit". Even before the onset of the pandemic, public sentiment strongly supported healthcare data-based research. For example, a 2017 survey showed that 93% of Australians support using medical records for research, and more than 95% had moderate, high, or very high trust that researchers would use health data responsibly.²

24. The *EU General Data Protection Regulation (GDPR)* recognizes scientific research as an issue of prime interest to the public and includes dedicated articles about arrangements that govern data use for research in the public interest and for scientific research.

25. In addition to public benefit as a key condition for people accepting data use for research, public trust is also a crucial factor. Lack of transparency, data breaches, misuses of health data and societal skepticism about the benefits of research can all diminish trust.

26. Notwithstanding the evolution in public attitudes about participation in research, the autonomy, privacy and the right to self-determination of health data research participants must still be protected via appropriate jurisdiction-specific mechanisms, such as opt-in/opt-out consent or public interest legal bases, robust de-identification methods, privacy-enhancing technologies and where applicable – ethical approval or oversight.

27. Furthermore, analytical techniques, such as machine learning and artificial intelligence, bear potential benefits such as increased efficiency in the provision of health services; objectivity and neutrality in decision-making, and the consequent promotion of public trust in health services; the personalisation of healthcare services which may bring higher quality results; early detection of individuals at-risk and prevention of future harms by illness; more equal care and social justice; the alleviation of workload and more. Machine learning and artificial intelligence development and deployment also introduce new data requirements that raise concerns about privacy protection, as well as equality and fairness. These techniques require access to vast datasets and explore full data to discover patterns that are relevant to the research question. These techniques challenge concepts for privacy protection, such as data minimization, and project approval processes. Further, where data are biased, limited and unrepresentative, results of these techniques raise concerns about equality and fairness following the use of data. In general, use of real-world evidence casts a wider net for health research, far beyond the traditional data collected for medical care and the traditional approaches to scientific inquiry.

28. These factors and others raise the stakes for data research benefits at an individual and community level, as well as the risks at both.

² Office of the Australian Information commissioner, Australian Community Attitudes to Privacy Survey [2017]

Legal Frameworks offering Privacy Safeguards

29. For this workshop, the Israel Tech Policy Institute (ITPI) and the Future of Privacy Forum (FPF) issued a joint report, “Using Health Data for Research: Evolving National Policies”, which is a comparative study of the laws, regulations and organizational measures in eight countries. The report found large commonalities at the macro level, important differences in the micro level, and some remaining challenges in this area.

30. In clinical research settings, patients’ consent to the use of their health data to address specific research questions is typically required. But as demonstrated in the OECD Health Data Survey, the ITPI/FPF report and through country presentations in the workshop, consent is not the only legitimate basis for data use and is often impracticable, such as when the volume of research involving re-uses of the same data becomes high. In most countries, the use of patient data for research is authorized by law, foregoing individuals’ consent requirement, providing those certain conditions are met. Some countries provide patients with a mechanism to opt-out of research uses of their own data.

31. New technologies, such as secure patient portals can provide patients with greater engagement with transparency as well as control over their data. At the same time, the rapid advancement of data-based research increases the need to future-proof patient empowerment mechanisms. Alongside the interests of individuals, there are also communal and societal interests in improving health services and investing in preventive medicine.

32. Data Protection regulations have developed over the past 40 years, starting with the OECD Privacy Guidelines of 1980 and the Privacy Principles serving as the 1st generation approach to privacy protection. The amended OECD Guidelines of 2013, along with the EU GDPR, established a 2nd generation approach. We are now at the forefront of a 3rd generation in data protection approaches.

Three generations of data protection regulations:

1. *OECD Privacy Guidelines of 1980 and the EU 1995 Directive*
 2. *Amended OECD Guidelines of 2013 and the 2018 EU GDPR*
 3. *Privacy principles inserted into sector-specific laws and regulations to support Innovation for the public good*
-

33. The 3rd generation of privacy and data protection principles would support research and innovation for the public good, including local and global data sharing, while protecting people’s privacy rights and other liberties. For example, legislations and regulations are being introduced to support autonomous vehicles while accounting for data governance and cyber security issues and offering a sector-specific implementation of data protection principles.³ The same can be seen in the financial sector in Open Banking initiatives.⁴ Arguably, successful Smart City frameworks will also include these principals,

³ California Senate Bill [SB-346 In-vehicle cameras](#) [February 2021]; German Draft law amending the Road Traffic Act and the Compulsory Insurance Act - [Autonomous Driving Act](#) [February 2021]; NHTSA Advance notice of proposed rulemaking, [Framework for Automated Driving System Safety](#) [November 2020]

⁴ [DIRECTIVE \(EU\) 2015/2366](#) on payment services in the internal market; US Consumer Financial Protection Bureau, Consumer Protection Principles: Consumer-Authorized Financial Data Sharing and Aggregation [October

and this may also be the case when artificial intelligence regulation arrives at the next level.⁵ The 2017 OECD Council Recommendation on Health Data Governance supports this broader view by encouraging countries to develop national health data governance frameworks that encourage the development, re-use, linkage and accessibility of health data while protecting individual's privacy and data security.

34. Even with sector-specific legislation, implementation is not without challenges. At present, as this workshop and its background papers have shown, health data research provides a use case that can inform the development of privacy and data protection principles in other social and economic sectors.

35. The next chapter will present several health data sharing projects, some based on new regulations and some based on other mechanisms.

2017]; The Australian Competition and Consumer Commission, [Consumer Data Right in banking Rules](#) [September 2019]; [UK Open Banking Standard](#) [31 March 2021]; The Monetary Authority of Singapore (MAS), Financial Planning Digital Services (FPDS)

⁵ Theo Bass, Emma Sutherland, Tom Symons, [Reclaiming the Smart City](#): Personal Data, Trust and the new Commons, EU DECODE [July 2018]

2 Experiences developing health data and governance

36. This chapter presents new initiatives and regulations within five countries (France, Finland, Korea, United States and Israel) and at the European level that were discussed during the workshop. All of the projects aim to strengthen health data governance while supporting data development and accessibility. The projects had a number of shared goals in common including:

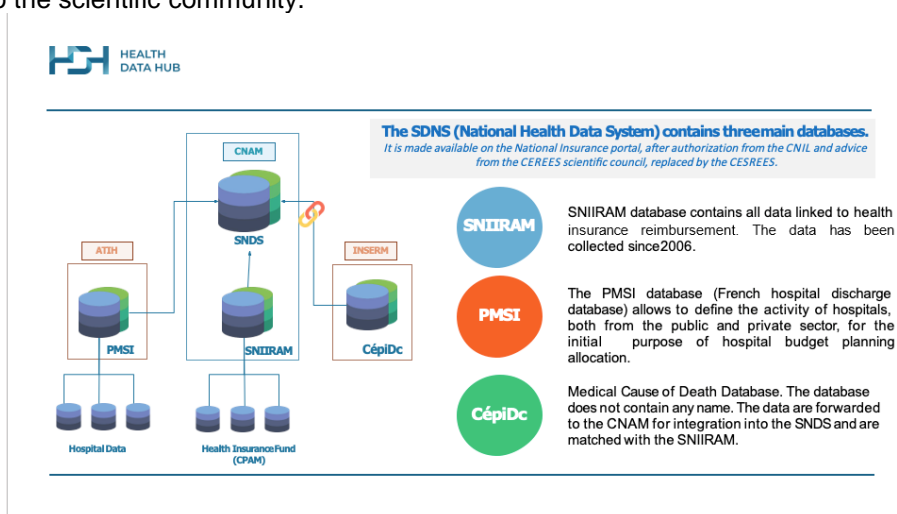
- Collection and linkage of data from multiple sources.
- Providing a secure and managed research platform.
- Allowing access from multiple sectors and regions.
- Streamlining approval processes.
- Safeguarding people's privacy.

37. These countries are among the top countries in the OECD for both the development, maturity and use of national health data and the development of national health data governance (see Annex E).

France Health Data Hub

38. The Health Data Hub (HDH) was launched in 2019 to support France in becoming a leader in Artificial Intelligence in health and to overcome barriers to the re-use of health data for research. Even though France is rich with data, these data were scattered among multiple databases and little was known or understood by researchers about which databases were available. Further, procedures to gain access to these databases were ambiguous and difficult to understand.

39. The HDH's primary goal is to support research and innovation in health and health care by providing a unique entry point for secure and privacy-protective data linkage services and access to health microdata for research projects that contribute to the public interest, while respecting patient rights and ensuring transparency with civil society. The second goal was to design a state-of-the-art platform at the highest level of security, offering data storage, computing, risk mitigation and analysis capabilities. Finally, the third goal was to create a documented data catalogue built in a progressive manner to make priority data known to the scientific community.



40. The HDH is a public interest group that was authorized by law and funded by the government to expand upon the existing national health data system (SNDS) to encompass all existing databases concerning publicly funded health activities (e.g. hospital electronic health records warehouses, cohorts, and registries). HDH was built on the infrastructure of the SNDS, the French administrative health care database, that covers around 99% of the population. The HDH catalogue unifies a collection of pseudonymised databases which the HDH is authorized to make available for research including data from clinical registries, research cohorts, hospital data warehouses, the nationwide emergency department supervision system (OSCOUR) and the [I-Share Cohort](#) consisting of health data of 30,000 university students spanning over 10 years. The catalogue includes mainly existing databases considered to be the most relevant for research and innovation.

41. The legal reform that launched the HDH aims to allow better visibility of common data assets for the entire ecosystem and to harmonize data access rules. This will be supported by the implementation of the HDH as a secure service for making the data catalogue available (the details of the national expert consultation on the design of the hub were published: [rapport de préfiguration](#)). Access to data is regulated and is carried out with respect for the rights of individuals. The catalogue is being built in a progressive and iterative manner, in partnership with data controllers, who are not obliged by law to provide HDH with the data. Citizen consent for the use of the data is not required, but each person is entitled to opt out of being included in the catalogue. There is no obligation to process health data in France within the technological platform of the HDH and it is still possible to conduct research in other partnerships. This initiative is aiming to make the availability of and the access to data for research easier and accessible to more researchers. HDH has so far launched 27 pilot projects, 9 of them COVID-19 related, after HDH received a specific mandate to accommodate COVID-19 related projects.

42. Permanent access to the HDH is granted to health authorities by decree of the French Ministry of Health. Other research requests for data are submitted to the “access team” that conducts a scientific and ethical assessment. If the request is found eligible, it is sent to the independent Scientific and Ethical Committee (CESREES). CESREES verifies that the purpose of the study is relevant and of public interest, that the data requested are in line with the study objective and that the proposed methodology is robust. If found positive, the project is submitted for authorization of the French Data Protection Authority (CNIL). So far, 8 projects were approved by the CNIL. The law also allows for the CNIL to publish a reference methodology for self-declaration by interested researcher groups that wish to self-declare and send their declaration to the CNIL in place of an approval request. This methodology was published as [MR-004](#).

A unique gateway to accessing health data

The data will be accessible to all project coordinators contributing to public interest after a positive opinion from an independent committee and an authorization by the Data Protection Authority (CNIL).

Step 1. Data access request. Project coordinators who wish to access the data submit an access request to the Health Data Hub. They can get support in their procedures.

Step 2. Scientific and ethics assessment If eligible, the request is sent to the independent Scientific and Ethical Committee (CESREES). It verifies that the subject of the study is relevant and of public interest, that the data requested is in line with the study objective and that the proposed methodology is robust.

Step 3. Data Protection Authority. On the basis of these elements and after positive opinion, the CNIL is contacted to give its authorisation to launch the project.

Step 4. Project Space. Once the CNIL authorisation has been obtained, the Health Data Hub team consolidates the required data and prepares a secure “project space” on its technological platform, which contains only the necessary data.

Step 5. Project in progress. Users of the technological platform have remote access to their “project spaces” and process the data on the platform without being able to retrieve it. The Health Data Hub can charge for access to its services.

Step 6. Publication of results. The project results are made public on the Health Data Hub website, with due respect for academic and industrial competitiveness.

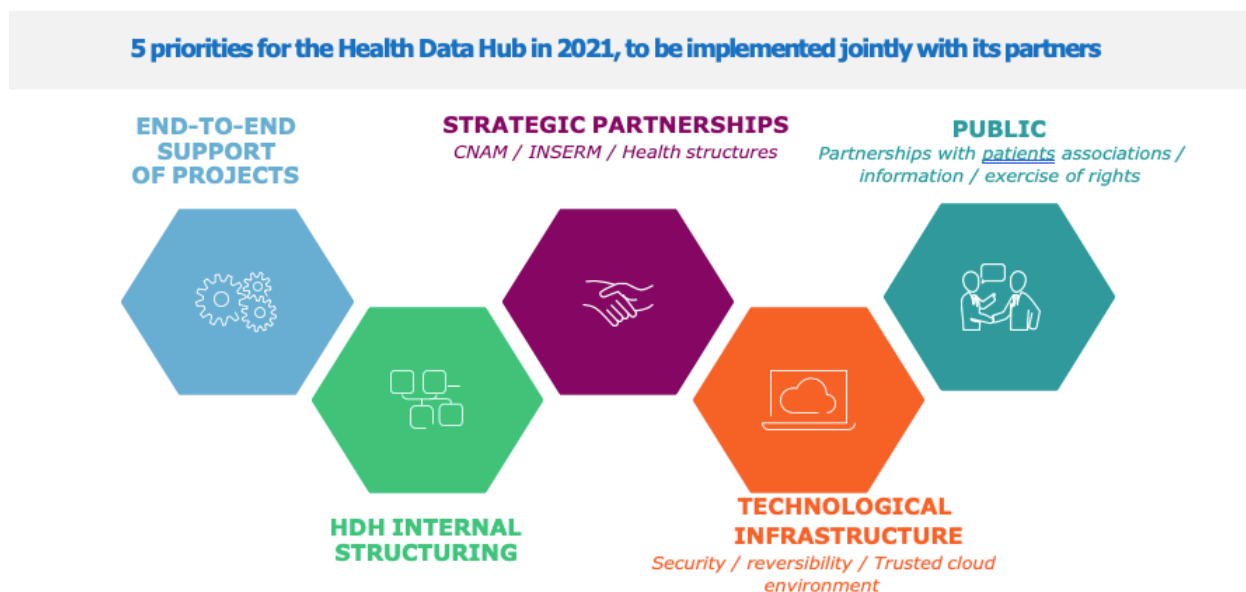


Health DataHub

43. A question was raised in the workshop about whether researchers can undertake 'blue sky mining' of the data (to explore unexplored data without a specific research question, as is typically done through artificial intelligence models). The reply of the HDH was that such broad access to data would be problematic because the CESREES committee is required to verify that the purpose of the study is relevant and of public interest, that the data requested are in line with the study objectives and that the proposed methodology is robust, in strict compliance with the GDPR principle of data minimization.

44. HDH makes health data available to health researchers in Europe, not just France. Furthermore, the HDH is the competent authority in France for the Joint Action "Towards a European Health Data Space" (TEHDaS). A [law](#) on the European Data Spaces, including the Health Data Space, is currently being drafted. The European Health Data Space has the potential to act as a powerful federator between national data hubs, promoting interoperability standards, best practices for data sharing across the European Union and setting a coherent governance framework.

45. As the HDH moves forward, it is seeking to foster strategic partnerships to connect with the general public, in order to communicate the benefits of using health data for the public good. HDH consults with civil society by carrying out studies and consultations on the relationship that citizens have with health data and on their perceptions, needs and expectations. This knowledge is necessary to orient and adapt public communications, and to evaluate them and ensure they are clear. HDH also contributes to the implementation of a "health data culture" by providing educational tools to enable citizens to understand the data and to learn how to use them and how to carry out projects with them.



*Model: Central repository of health databases
(built on an existing one - SNDS) with a Data
Catalogue, Data Linkage Services and Secure
Research Infrastructure*

Management: Public Authority

Goal: Health research in the public interest

*Users: Public agencies and private sector
entities from Europe*

*Project approval process: Ethics Committee and
Data Protection Authority*

Finland Findata

46. Findata is authorized by law to support secondary uses of health and social data in Finland for projects that contribute to the public interest. Findata is the only authority that can issue permits for the secondary use of health and social data when the data is compiled from more than one data custodian. Findata provides for the secure linkage and research access to publicly funded datasets and registries including the data holdings of the Finnish Institute for Health and Welfare (THL), the Social Insurance Institution of Finland (Kela), the Population Register Centre, the Finnish Centre for Pensions and Statistics Finland. From 2021, Findata will expand to include data within the national EHR system (Kanta).

47. Findata is a centralized system issuing permits and a one-stop shop for the secondary use of health and social care data in Finland. It grants data use permits when data are requested from multiple registries or from the private sector; collects, links and prepares the data; provides the data in a secure IT-environment for data users; offers electronic tools for data permit applications; offers a help desk for data users; and works in collaboration with the controllers of the data.

FINDATA

What?

A centralised system issuing permits: one-stop shop for the secondary use of health & social care data

- **Grant data permits** when data are requested from multiple registries
- **Collect and process** the data
- **Provides the data** in a secure IT-environment for data users
- **Electronic tools** for data permit applications, information requests and access to data.
- **Offers a Help Desk** for data users
- **Works in collaboration** with the controllers of the data.





@FindataFi | #Findata

48. Findata is not a permanent data repository, but a hub in which the data flows. It exists to streamline and secure the secondary use of health and social care data for four main purposes: 1) enabling effective and safe processing and access to data; 2) enhancing data protection and security; 3) eliminating overlapping administrative burden; and 4) improving data quality.

49. The *Act on the Secondary Use of Health and Social Data* (enacted in May 2019) gives Findata the authority to grant secondary use for research within Finland. It is noteworthy that this is made possible due to Finland's personal identification code that remains unchanged throughout an individual's life and is the key to linking personal information from various registries.

FINDATA

Why? To streamline and secure the secondary use of health and social care data for different purposes

-  **Enable effective and safe processing and access to data**
-  **Enhance data protection and security**
-  **Eliminate overlapping administrative burden**
-  **Improve register data quality**

@FindataFi | #Findata

50. The permits are awarded for a specific time period and for specific research purposes, following the data minimization principle (the scope of data to be given access to is limited to the required data for the specific purpose of the research). Findata can grant permits, conduct dataset linkages and provide access to large amounts of data if the research need is justified. Broad access for 'fishing for signals' or 'Blue Sky Mining' is challenging due to the data minimization principle, that allows disclosure of only the data that are essential for the research question.

51. The permit holder has access to analyze the data on the Findata platform. The applicant needs to order a "remote research environment" which is permit specific. It is possible to apply for an extension of the permit or to add researchers to the work environment by an amendment application. As a rule, the data are always disclosed to Findata's secure operating environment. However, the Act empowers Findata to make the data available in another environment as well, if it is necessary for the research purpose. From 1.5.2021 onward these other environments need to be audited against the regulation.

FINDATA

How?

Act on the Secondary Use of Health and Social Data gives Findata the authority to grant secondary use..

... Permits on:

- All health and social care data collected in connection of primary use in Finland
- National registers (health and social data)
- Population studies

For purposes:

- Scientific research
- Statistics
- Development & innovations
- Education
- Knowledge based management
- Authorities' steering, supervision, planning and reporting duties

Personal identification code

- Remains unchanged throughout life
- A key to linking personal information from various registries

@FindataFi | #Findata

Model: Central Permits Authority collects databases for approved projects, provides data linkage services and provides access to data on a secure virtual research infrastructure

Management: Public Authority

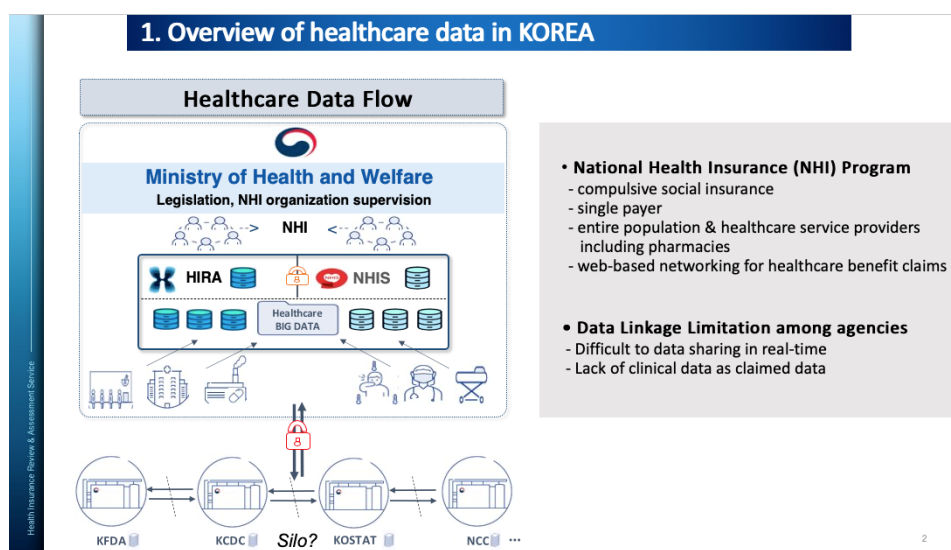
Goal: Scientific research, innovation, education, policy decisions

Users: Public agencies and private sector entities from Finland or outside it

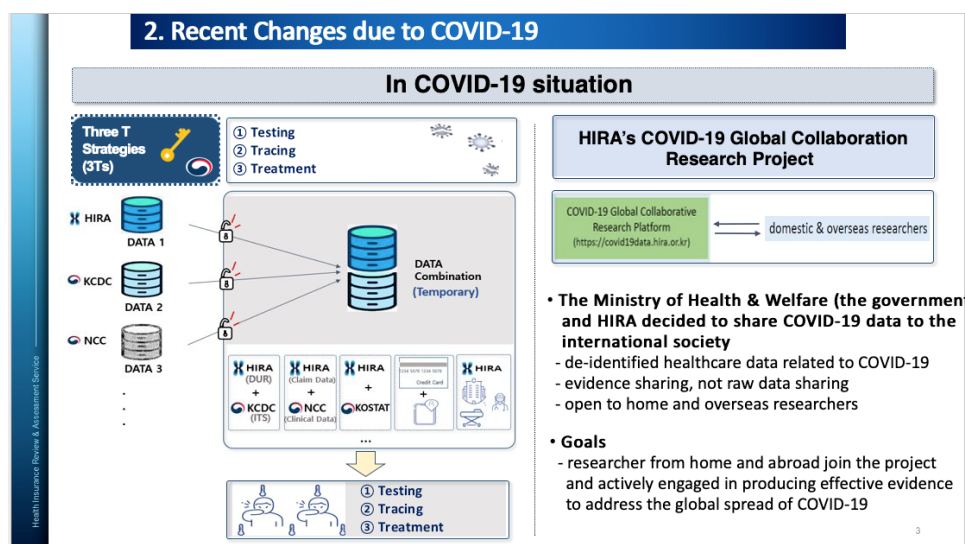
Project approval process: Permit authority assessment

Korea digital new deal

52. In Korea, the National Health Insurance (NHI) Database was established many years ago and organizations have been required to submit data to the NHI Program to obtain reimbursement. Korea already has real-time data at the national level across the continuum of health care services from insurance claims and these data are routinely linked for statistics and research.



53. The Covid-19 pandemic, however, has inspired an expansion of health data under a “Digital New Deal” that is being developed now. Under the New Deal, real-time insurance claims data can be linked with real-time clinical data. For example, Korea is developing the capability to monitor adverse events from the Covid-19 vaccination in real time. The Ministry of Health and Welfare and Health Insurance and Review Assessment Agency (HIRA) have been authorized to share COVID-19 data with the international community to find an effective response to COVID-19.



54. In order to further promote health data sharing for research, Korea has prepared legislation to establish a national data lake that will allow public bodies and private companies to have access to health data. Korea aims to link additional repositories to this national initiative. Under this new law (anticipated in 2021), Korea plans to maintain the data lake after the pandemic ends so it may continue to support international researchers' access to updated COVID-19 patient data.

55. De-identification techniques such as pseudonymisation are being used as a safeguard, and qualified organizations will perform data preparation. Engagement with the data lake is by application to qualified agencies.

Model: Expanding an existing National Health Insurance Database to support research and innovation

Management: Public Authority

Goal: Covid-19 response

Users: Local & International community

Project approval process: Application to qualified agencies (HIRA, NHIS, Samsung SDS)

United States new rule within the 21st Century Cures Act

56. In the United States, each state manages their own public health reporting programs, and these practices are regulated by state law. Each individual hospital system may have their own network — which can include thousands of payor systems. This fragmentation impedes patients' access to their complete records, as well as the availability of health data for research. To address this, the Department of Health and Human Services (HHS) proposed a new rule within the 21st Century Cures Act to support the seamless and secure exchange and use of electronic health records. The rule asks the health care industry to utilize Application Programming Interfaces (APIs) and to adopt the HL7 Fast Healthcare Interoperability Resources (FHIR) standard for health data exchange. Further, a Trusted Exchange and Common Agreement (TEFCA) sets out principles, terms and conditions to enable the nationwide exchange of electronic health information across disparate health information networks.

2

Application Programming Interfaces (APIs)



- Established a new application programming interface (API) certification criterion that requires health IT developers to support standardized APIs for single patient and population services.
- Certification criterion is limited to API-enabled “read” services using the HL7® Fast Healthcare Interoperability Resources (FHIR) Release 4 standard.
- The use of the FHIR standard and a set of implementation specifications provides known technical requirements against which third-party apps can be developed.

Supports two types of API-enabled services:

- » Services for which a **single patient's data** is the focus
- » Services for which **multiple patients' data** are the focus




57. Standardization of the data sources is required for health data to be exchanged across all networks, not just the major networks like Medicare. The Office of the National Coordinator of Health IT (ONC) plans to introduce a United States Core Data for Interoperability Standard that will be the content and vocabulary baseline for health data, beginning 24 months after the publication of the final rule. This standard includes new data classes and data elements, such as provenance, clinical notes, pediatric vital signs, addresses, email addresses and phone numbers. These data pieces were not universally exchanged before — but are essential for patient matching and identifying risk factors. Leveraging this data allows better demographic information to be available to health care providers so that they can evaluate patients' risks and needs.

3

United States Core Data for Interoperability Standard

The Office of the National Coordinator for Health Information Technology

The United States Core Data for Interoperability (USCDI) standard will be the content and vocabulary standards baseline beginning 24 months after publication of the final rule.




USCDI includes the following new required data classes and data elements:

- Provenance
- Clinical Notes
- Pediatric Vital Signs
- Address, Email & Phone Number

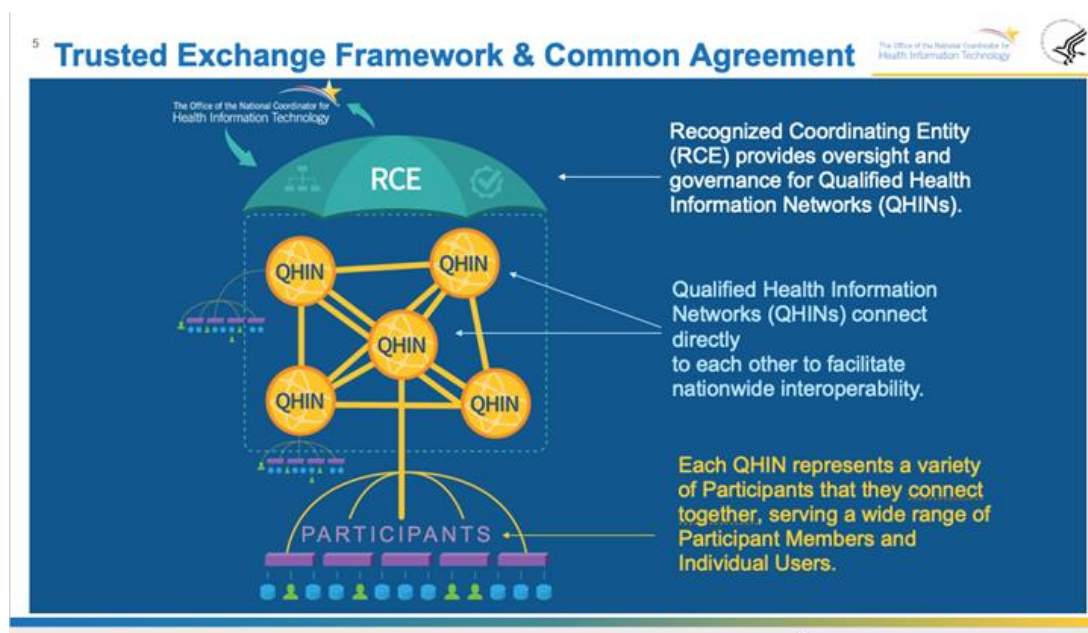
USCDI Standard Annual Update Schedule

ONC will establish and follow a predictable, transparent, and collaborative process to expand the USCDI, including providing stakeholders with the opportunity to comment on the USCDI's expansion.



58. ONC has several pathways for public engagement and input into these data interoperability standards including a federal advisory committee made up of representatives from health care, health IT, and patient advocacy organizations. It publishes proposals for public comment and conducts targeted listening sessions with different groups. Finally, on the technical aspects, it works closely with the standards organizations which include public input and consensus- based balloting processes.

59. Generally, there isn't financial support to all stakeholders to invest in this, but there is some support for states to implement these capabilities in their networks. For health care providers, there was previously a program that provided incentive payments for adoption of an electronic health record system, but there has not been new funding approved by Congress to continue support. However, there are requirements for hospital systems that are paid under the Medicare (National) program to adopt and use technology that is certified to certain standards and functionalities. ONC has added these new requirements to the existing program requirements. There is also a program that requires the payers (the plans that administer Medicare and Medicaid) to build APIs, as well to allow the data they hold to also be accessible. And finally, ONC requires technology developers, through a certification program, to make this technology available to their customers. At present, there is no requirement to participate in the program. ONC is exploring with other federal, state, tribal, etc. partners about if and when such a requirement would be feasible or desirable.



Model: Application Programming Interfaces (APIs), Fast Healthcare Interoperability Resources (FHIR) and data terminology standards

Management: Public Authority

Goal: National exchange of standardized health data enabling patients to access their own health records and providing a large volume of standardized health data that could support research and innovation

Users: Patients, health care providers, health care industry and the research community

Project approval process: Legislation

Israel National and Organizational 'Data Lakes'

60. On a national level, Israel has a rich and well computerized health data ecosystem consisting of 30 years of central public healthcare. Israel's government has been working on designing a policy framework for secondary use of health data for research to enable collaborative data research initiatives. This framework (not yet finalized) was surveyed in the ITPI/FPF report (Annex D).

61. In addition, the Ministry of Health is also working on an initiative to form a 'Data Lake' that will include Israel's digital health data from hospitals as well as HMO's and the Ministry of Health itself. In the current COVID-19 context, relevant data for related research is an opportunity, since 95% of patients in Israel are currently being treated by HMOs. There is value in bringing all of this data together. The 'Data Lake' policy framework consists of IRB certificate mechanisms, transparency, de-identification mechanisms, secure environment, user controls, opt-out mechanisms, and data use agreements.

62. In the process of creating the Israel Covid-19 Data Lake, there were a number of challenging discussions. First, there was a debate regarding whether to have an opt-in or an opt-out mechanism for individuals whose data would be part of the lake. It was decided that the public interest in making the data available for research allows for an opt-out mechanism. Second, there was a debate regarding how the opt-out option could be communicated so that it becomes public knowledge and the information reaches all communities. The decision was to send a text message to everyone to be able to reach the maximum number of people in Israel. A third concern was a belief that people would not want to opt-out due to fear that they would not receive proper care in hospitals. Fourth, there was a debate regarding whether to collaborate with industry. There are positive effects from collaboration, including innovation and cost sharing; however, it was difficult to generate a standard agreement for an unknown future collaboration; so, it was determined that the National Data Lake would only be available for pure research and would not have any industry collaboration. Fifth, an area of concern was the effects that COVID-19 would have on public trust, particularly given public trust in government has been weakening in Israel, as has been the case in other countries. Possible solutions to strengthening public communication to build trust in the national data lake are under consideration.

63. In order to streamline the application process to the Data Lake, Israel is preparing one formal agreement for researchers that want to access the data, since this data is needed in a timely manner due to COVID-19. Further, Israel is considering new technologies for privacy enhancement that support researchers' ability to access complete records (raw data).

64. Strengthening the argument supporting the decision to offer an opt-out mechanism were previous decisions regarding the National Patient File (summary health record). The National Patient File requires all providers in Israel to use the same central system for data management, so that they can easily communicate with each other. There were discussions in the Ministry of Health to determine if this system should have an opt-in or opt-out structure. An opt-out structure was chosen because there was strong evidence that having all of the data available for patient care provides for more accurate findings and better healthcare services; and allows for more effective decisions to be made, which in turn allows costs to decrease and is in the public's best interest.

65. A question was raised in the workshop about whether the current Covid-19 research data lake will be needed after a vaccination is offered on a global scale? The Israel answer was that research regarding Covid-19 will remain relevant for some time. It is also a valuable case study that could be scaled. The Covid-19 Data Lake is the first stage in a national effort to allow research and new innovative healthcare and digital health development via secondary use of health data.

66. Another question was whether the Covid-19 Data Lake should or would be joined with other research data lakes? Israel's answer was that cross-country research is challenging, but not impossible. The experience of Covid-19 has demonstrated that trans-border cooperation is crucial. There are multiple challenges regarding this, both privacy issues and data standardization challenges, but it is most definitely the way forward. The HL7 FHIR standard (already mentioned in the previous session) is an important tool to allow such local, organizational, national and international collaboration.

Creating multi-stakeholder partnerships with a large hospital's data registry

67. On an organizational level, Sheba Medical Center, the largest hospital in Israel by license of the Ministry of Health, and with its approval, launched an innovation program, the New ARC Initiative (ARC). Arc is a global ecosystem that aims to redesign healthcare in the next decade. Various institutions are part of this network, with 70 startup companies, corporate strategic partners such as Microsoft, Boston Scientific, Allscripts, and other international institutions. ARC focuses on digital health aiming to impact how healthcare will be done in the next decade, with big data and AI as the main focus.

68. ARC COVID Data Challenge is a registry ('Data Lake') that was established in March 2019. It allows access to COVID-19 data on the cloud for timely and life-saving research. ARC works on a cloud infrastructure (on a hybrid model) daily, uploading data from clinical systems. Access to the data was granted to 40 data groups after they signed a confidentiality agreement committing them to not share the data with other parties and to use the data only for research. The participating groups were from the defense cluster in Israel, academics, and corporations. At a competition in June 2020, a group working at Intel Israel won by developing an algorithm that predicts which patients will require ventilation earlier. The algorithm was implemented into Sheba's electronic systems and assisted in saving lives.

69. Technology and algorithms are available in containers that give customers a choice between running analysis on-premises or in the cloud. An advantage of the cloud is that it is always up to date with the most recent security patches and constantly scanned and validated for security breaches.



Model: National Data Lake & Organizational data registries

Management: Public Ministry & Licensed Health Organizations

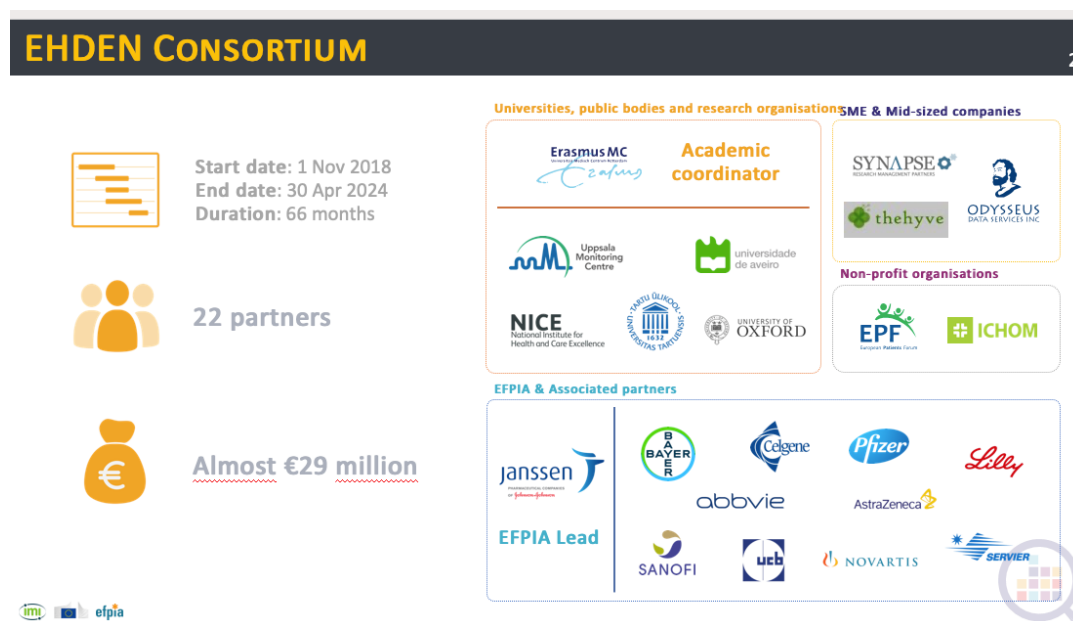
Goal: Research & Innovation, Evidence based policy decisions

Users: Public sector & private sector

Project approval process: Review committees at the national & organizational level

European Health Data & Evidence Network

70. The European Health Data & Evidence Network (EHDEN) is a public-private data partnership that aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care. It is developing a federated network that manages and shares research methodologies, standardizes a wide range of health data to a common data model, and expands education in open science and collaboration. The EHDEN Consortium runs from Nov 1, 2018 to Apr 30, 2024, has 22 Partners, and a budget of 29 million (euros).

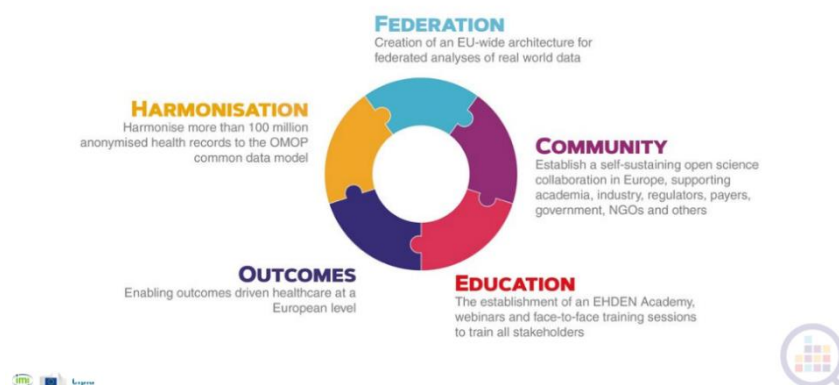


71. Generally, EHDEN's mission can be divided into five categories. The first category is federation: the creation of an EU-wide architecture for federated analyses of real-world data. The second is community: where the goal is to establish a self-sustaining open science collaboration in Europe, supporting academic, industry, regulators, payers, government, NGOs and others. Third is education: where the EHDEN Academy, webinars, and face to face training sessions offer training for all stakeholders. The fourth goal relates to outcomes: specifically, enabling outcomes driven healthcare at a European level. And finally, harmonization: where EHDEN has set the goal to harmonize more than 327 million anonymized health records to the OMOP common data model.⁶ This was achieved in collaboration with 60 partners from 20 countries.

⁶ As defined by the Observational Health Data Sciences and Informatics (OHDSI) the OMOP Common Data Model allows for the systematic analysis of disparate observational databases. The concept behind this approach is to transform data contained within those databases into a common format (data model) as well as a common representation (terminologies, vocabularies, coding schemes), and then perform systematic analyses using a library of standard analytic routines that have been written based on the common format.

EHDEN IS ABOUT ...

1

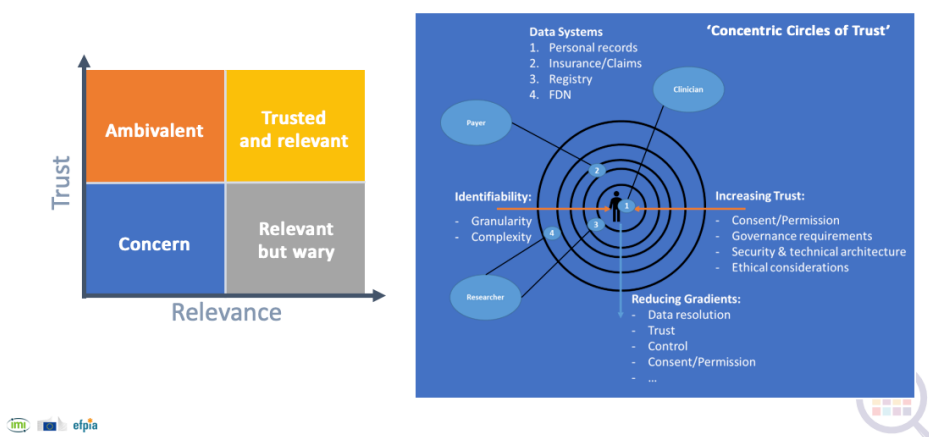


72. Regarding research quid pro quo for sustainability, the EHDEN project is building circles of trust that radiate outward from the trust of individuals whose data are included within the federation and include all stakeholders in the development and use of the data. Building these circles of trust is part of the development of the EHDEN data systems, which include personal clinical records; insurance claims; registries; and a federated data network. Elements of increasing trust include consent/permission; governance requirements; and security and technical architecture. Another component is gradients in data accessibility, where data resolution; trust; control; and consent/permission serving as the elements. And finally, identifiability, where granularity and complexity serve as the two elements of this component.

73. The EHDEN project fosters trust with a distributed analytics framework, where software and statistical analysis programs travel to where data are located, rather than data flowing to a central data lake for analysis. With this method, data collectors retain control of their data at all times and data privacy and security remain protected by local legal requirements and operational practices. Data collectors also pseudonymise their own data before they are analyzed. Further, because there is no transfer of record-level data, associated security risks are avoided. Key to the success of this method is the coding of data within the Federation to a common data model (OMOP) which permits the analytical programs or software code to run smoothly at each node of the network.

A RESEARCH QUID PRO QUO FOR SUSTAINABILITY

4



74. In the workshop a question was asked about the sustainability of the project, given past EU projects have developed consortium and infrastructure that were not taken forward by the European Commission. EHDEN replied that they are working on sustainability and creating a legal entity now and are fortunate that their timing aligns with development of the European Health Data Space and EU DARWIN (Horizon 2020 funded project focused on improving responses to expected and unexpected crises affecting critical societal structures during natural disasters, e.g., flooding, earthquakes, and man-made disasters, e.g., cyber-attacks). EHDEN also collaborates closely with the European Commission and the European Medicines Agency (EMA) on the integration of all these EU initiatives. EHDEN is also linked to a global network and will be launching a not-for profit legal entity in 2021.

Model: Industry initiative for federated data partnership with academic and public partners

Management: Steering Committee

Goal: Coding health data to a common data model and a distributed analytics approach supporting multi-country health research

Users: Public sector & private sector across Europe

Project approval process: N/A

3 Data governance policy issues

75. This chapter summarises discussions at the workshop and following engagements regarding issues encountered when developing approaches to health data governance. The key governance issues raised at the workshop were the need for clarity and harmonization of regulatory frameworks; challenges in cross-border data transfers; the importance of patient perspectives and public trust; genomic data challenges; views about the legal basis for the secondary use of health data for health research; and concerns about data quality and data linkages.

Policy issues when governing health data for research:

- *Alternative legal basis to consent for research uses of data*
 - *Distinction between data research and human interventional research*
 - *The scope of “Public Interest”*
 - *Definitions of de-identification levels and the data protection safeguards still required after data has undergone them*
 - *Types of review boards participating in the research approval process*
 - *National and International Data transfer conditions*
 - *Data Linkage abilities*
-

Insufficient clarity and harmonization of regulatory frameworks

76. As presented in the workshop and the ITPI/FPF report, legal frameworks for governing the use of healthcare data for research purposes typically involve three layers of laws and regulations: First, are the general data protection laws which regard healthcare data as sensitive information, subject to strict protections. Second, are patient rights and patient privacy laws, for example, laws around genetic data. Third, are public health regulations. To make matters more complex, some countries like Australia⁷ and

⁷ The Federal Government of Australia has several initiatives aimed at facilitating sharing of data across jurisdictions and the Federal government. These include the 2020–25 Addendum to the National Health Reform Agreement, signed by all jurisdictions, which includes an action to scale up a national approach to data governance arrangements. The Intergovernmental Agreement on Data Sharing which commits all jurisdictions to share public sector data as a default position aiming to support better public services to citizens and the Data Availability and Transparency Bill 2020, which

the United States also have a federal and state system, which impacts the flow of healthcare information. These layers create issues around determining the best definition of research, as there are many types of research, including corporate research, and questions arise as to whether research exemptions should apply to various kinds of corporate research. Additional requirements may be applicable stemming from conventions on ethics in health research such as the Helsinki Declaration and the Council of Europe Convention.

Regulations governing health data for research:

- *General data protection laws*
 - *Patient rights and patient privacy laws*
 - *Public health regulations*
 - *Federal, State & County statutes*
 - *Conventions on Ethics in health research*
-

77. As was further presented in the workshop, at a general level, the eight countries studied in the ITPI/FPF report (Annex D) permit secondary use of healthcare data for research purposes under certain safeguarding conditions (reviews by ethical boards over data use purpose, and data collection, sharing, and analysis; de-identification and privacy by design techniques, and additional administrative and technical organizational requirements like data sharing agreements, etc.). Each country also has the ability to impose restrictions on cross-border data flows. Notably, another common thread between these countries is the role of consent, whether an opt-in consent structure is essential, or whether an opt-out system better serves the purpose of making data available for research. Consent is an important pillar in data protection law generally, and specifically in the health context. Countries usually decide on some equation where a compelling public interest replaces consent — and that interest is coupled with de-identification mechanisms and other safeguards.

78. Another issue is that countries use different definitions for personal data and define de-identified data differently. With respect to de-identification, a major issue that has been shown is that some countries allow a more risk-based approach, seeing de-identification or identifiability on a risk-based spectrum, whereas others see this as a more absolute term. Further, ethical review boards also vary across countries, as some have national standing IRBs, while others have organizational ones.

79. In summary, workshop participants expressed that there is still ambiguity around health data governance legislation and political issues sidelining efforts of linking data repositories and effectively analyzing and acting upon quality data in real time. Although in a significant part of research, identity is not required, it was recommended that we explore these ambiguities and implementations to allow for research that will support public health management, prevention and care. Particularly, further work is needed around definitions of data de-identification and risk mitigations and processes to authorize use of data with various levels of identifiability.

Governance of cross-border data transfers

80. The workshop revealed some examples of data transfer restrictions that are required in some countries with respect to health data for research. In Australia, the law restricts health data from being

is currently before the Australian Parliament, that seeks to establish a scheme for the sharing of public sector data by agencies to accredited users.

taken out of the country's borders. This is referred to as data localisation. The Indian draft Data Protection Bill has some data localisation requirements for sensitive information, and the Israeli Draft Health Data Research Regulations also has some form of data localisation as well. The 2019-20 OECD Survey indicated that only half of countries would approve the sharing of de-identified national health datasets across borders with a university, non-profit or governmental body (Annex E).

81. The issue of data localisation connects to the general discourse on this challenge across economic sectors and to the OECD work in that area. A new OECD report⁸ on trends and challenges of data localization was shared with workshop participants. It defines 'data localisation' as a mandatory legal or administrative requirement directly or indirectly stipulating that data be stored or processed, exclusively or non-exclusively, within a specified jurisdiction. Such data localisation requirements have emerged as a major issue in transborder data flows and are of relevance to the topic of sharing health data for research as well.

82. The report asserts that within most data protection regulations, data transfer conditions do not necessarily amount to data localisation measures. It argues that the proportionality test, as articulated by the OECD Privacy Guidelines, should be considered a key mechanism for the evaluation of data localisation measures, including a holistic and context-dependent assessment of whether localisation is supportive of or obstructive to the public interest.

83. In other words, the OECD Proportionality principal should guide a risk-based assessment of a data governance framework that considers or requires data localisation for reasons of data protection. Sometimes, localisation may not be the right mechanism to support better data protection, for example when the level of security at the local location is lower or similar to the security offered at the cross-border location. In these circumstances the goal of better data protection is not actually supported while the benefit from collaborative data access for the purpose of a public interest goal is prevented.

84. Furthermore, since today's economy and society are reliant on global sharing of data, the OECD Proportionality test should include consideration of the consequences of a country's data localisation measures to the international community.

85. Svantesson's report suggests that where a legal or administrative requirement is found to constitute a data localisation measure, and it amounts to a restriction to transborder flows of personal data under paragraph 18 of the OECD Privacy Guidelines, the assessment of whether it is proportionate (under that same paragraph) to the risks presented, ought to take into account multiple factors, such as:

- (a) the sensitivity of the data;
- (b) the purpose and context of the processing;
- (c) the extent to which it is demonstrated that the data localisation measure effectively achieves the goals for which it was introduced;
- (d) whether there are any less restrictive measure that could be enacted;
- (e) the direct and indirect, domestic and international, implications of the measures;
- (f) evidence of intent where it is possible to establish; and
- (g) the implications likely to arise if also other countries adopt the same measure ('scalability' as a consideration in the assessment of proportionality).

⁸ Svantesson, D. (2020), "Data localisation trends and challenges: Considerations for the review of the Privacy Guidelines", OECD Digital Economy Papers, No. 301, OECD Publishing, Paris, <https://doi.org/10.1787/7fbaed62-en>

86. In the context of health data research, projects often involve stakeholders from different nations, entity types and organizations. Therefore, it is fundamental to make progress toward interoperable data governance frameworks that support data sharing between different types of entities and across borders.

87. Reasons why countries pursue data localisation policies include national security concerns, legitimate law enforcement functions and economic competitiveness, such as the desire to use and implement local AI-based cloud solutions. Of course, data localisation can also have a negative impact upon innovation and economic development, by limiting national engagement in international research projects. Solutions to data location requirements regarding sensitive health data include technologies that avoid the need for data to be transferred while still allowing access to the data. These include remote access to data on secure monitored platforms. These approaches were taken by Finland, France, and Israel, as presented in the previous chapter. Another approach is the distributed analytics method of the EHDEN project that was also discussed earlier. However, these solutions will not fit all situations. For example, not all research methods can be applied using these solutions; and, for startups and tech companies, being obliged to use unique methods (e.g. a proprietary algorithm) on a platform controlled by the data holder may constitute an IP protection risk that will undermine their willingness to take part in the research.

Patient perspectives and public trust

88. As was discussed earlier, research participation is increasingly perceived as a solidarity-based moral imperative or a moral obligation – particularly in the case of non-interventional, minimum-risk health data research and sample donation to (bio)-repositories.

89. For example, an analysis of test results or medical patient files, by a machine learning algorithm, in search of patterns that will lead to early detection, prevention or cure of medical conditions, without conducting any physical contact with the patients or requiring any involvement from them. Seemingly, where there lies a promise for medical benefits for individuals and society at large then the low risk and minimal inconvenience entailed for research subjects invokes a stronger social and moral rationale for participation.

90. Such sentiments, social expectation or personal motivation for participation, typically do not exist in the context of interventional medical research, such as drug trials, which carry higher risks and more substantial burdens for participants.

91. And indeed, surveys of public attitudes to patient data use conducted in the UK⁹ for example, have demonstrated that "a high proportion of people support sharing patient data for research, where there is public benefit". Even before the onset of the pandemic, public sentiment strongly supported healthcare data-based research. For example, a 2017 survey showed that 93% of Australians support using medical records for research, and more than 95% had moderate, high, or very high trust that researchers would use health data responsibly.¹⁰

92. The EU General Data Protection Regulation (GDPR), which has become the dominant data protection standard globally, recognizes scientific research as serving the public interest. However, while public benefit is a condition for people to accept patient data sharing; public trust is a crucial factor for public support. Lack of transparency, concerns about misuses of health data in general, and skepticism about data uses for research can erode public trust.

⁹ Understanding Patient Data, Public attitudes to patient data use: A summary of existing research [July 2018]

¹⁰ Office of the Australian Information commissioner, Australian Community Attitudes to Privacy Survey [2017]

93. Furthermore, inequalities in access to data for research and inequalities that result from research with incomplete or biased data will erode public trust.

94. Inequalities have become exacerbated and are being transposed onto data. Without appropriate data governance that takes into account equality in access to data by diverse groups of researchers and equality in representation of groups in the data, we are going to keep seeing groups marginalized from the benefits of health data research and subsequently from equal health care. As a workshop participant stated: “It is clear that we can't afford to have people who can't afford to have public health”.

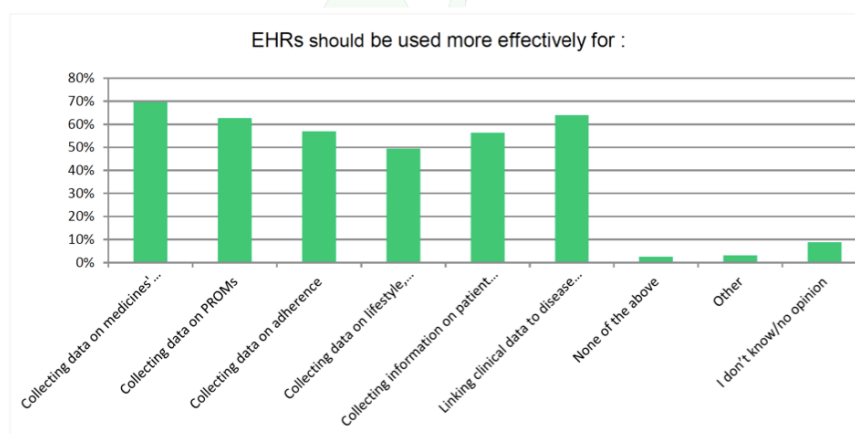
95. In a recent survey about the use of electronic health record (eHR) data for research conducted by the European Patients Forum (EPF) and presented in the workshop, most respondents saw benefits in better use of eHR data for research purposes and were happy to share their health data.

Use of EHR data for research



Most respondents saw benefits in better use of EHR data for research

- Assess medicine safety and efficacy after authorisation
- Collect patient-reported outcome measures (PROMs)
- Improve knowledge on adherence to treatments
- Lifestyle and demographic information
- Gain insights on the patient journey → develop patient experience measures (PREMs)
- Link to disease registries



“ A STRONG PATIENTS' VOICE TO DRIVE BETTER HEALTH IN EUROPE ”

96. On the issue of the type of informed consent required, respondents were evenly split with a slight advantage for the position favoring explicit consent.

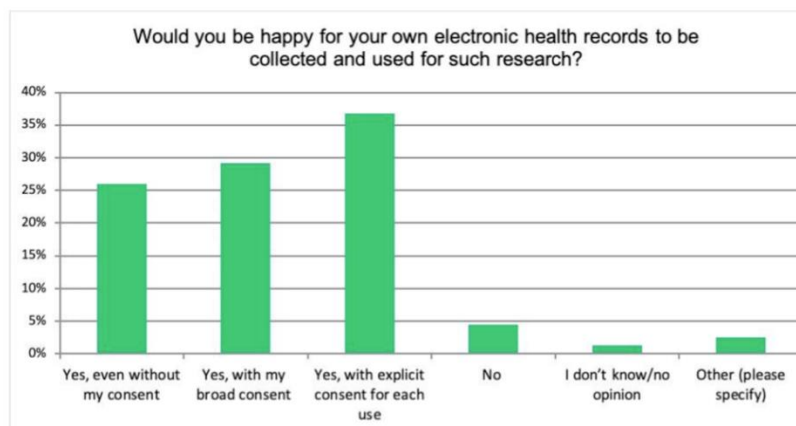
Consent or not? What kind of consent?



Most respondents happy to share their data

Opinions split between consent options

Slight majority favouring explicit consent



“ A STRONG PATIENTS' VOICE TO DRIVE BETTER HEALTH IN EUROPE ”

97. The main worry expressed by respondents was that data would be inappropriately shared with insurance or non-medical companies, where respondents made a distinction between for-profit and non-profit entities.

The Genomic Data Challenge

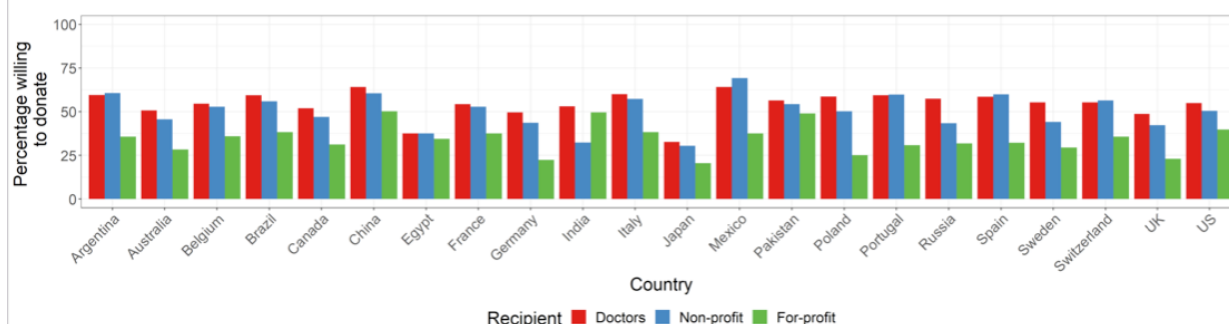
98. Where genomic data is concerned, public trust is even more critical. These data are particularly sensitive because they may disclose a broad spectrum of genetic conditions and personal hereditary traits, are predictive in nature, can reveal information not just about a data subject but also about their family members (blood relatives), and carry relevance for reproductive decision-making. As a result, some countries taking a *genetic exceptionalism* approach and have excluded genomic data from their frameworks supporting the secondary use of health data for research.

99. A study presented¹¹ in the workshop indicated that across 22 countries studied, the public are more comfortable with their genomic data being used by doctors and by non-profit researchers, such as universities, than by for-profit companies. While there is variation in who is trusted with data, the most trusted is a person's own doctor.

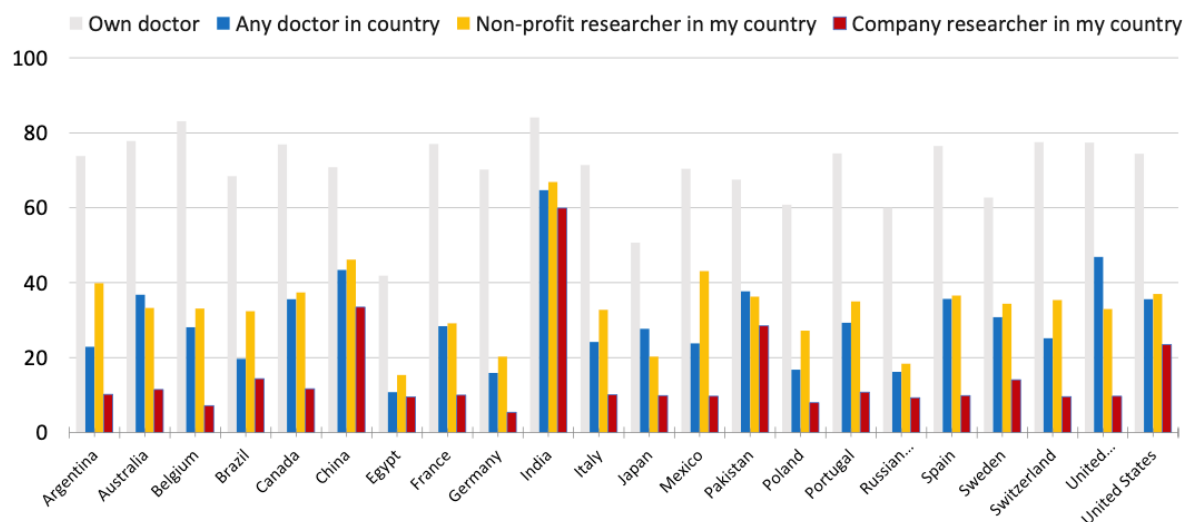
¹¹ Dr. Richard Milne, Society and Ethics Research, Wellcome Connecting Science, Trust and the donation of genomic and health information.

Overall willingness to donate is low

Across 22 countries, global publics are more comfortable with their data being used by doctors than for-profit companies

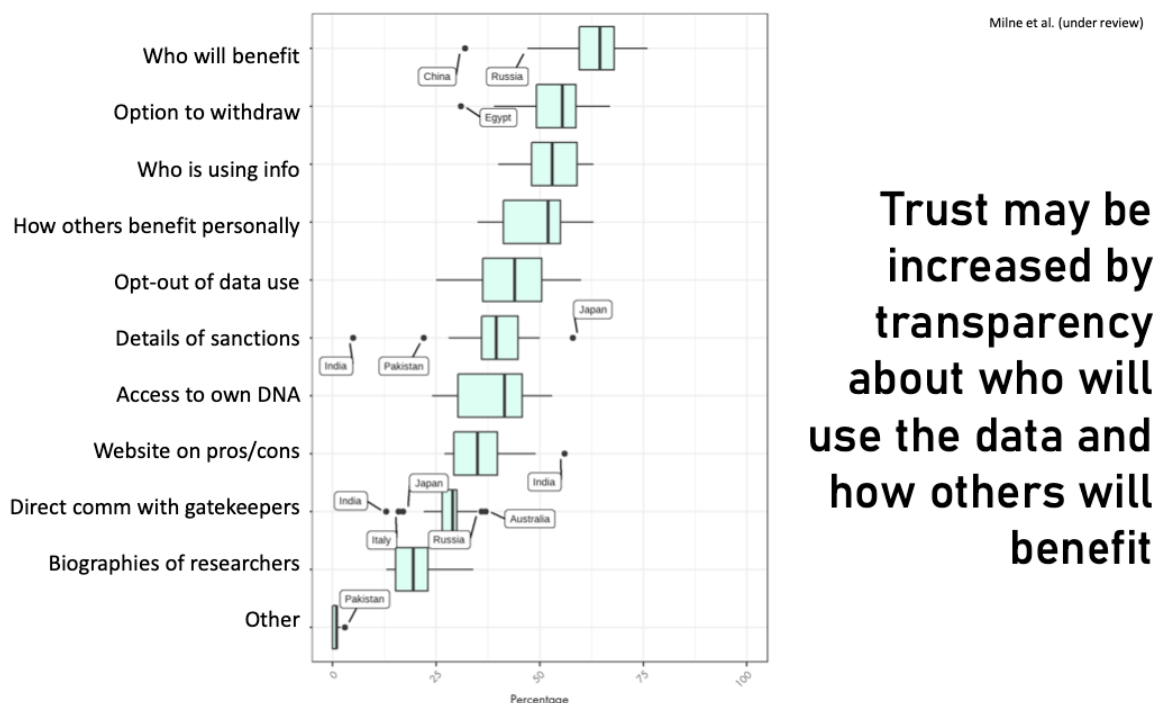


There's variation in who is trusted with data but the most trusted is a person's own doctor



100. While the role of trust in the donation and sharing of genomic data is important, it is not the only factor. Reciprocity and solidarity are also important, and these core values often rest on an individual's familiarity with the use of genetic data. With greater familiarity, individuals are more likely to share their data, and are also more aware of potential risks.

101. Trust, and willingness to donate data, can be developed with transparency about who will use the data and information about how others stand to benefit from the data. The option to withdraw from data research is also supportive of trust.



Trust may be increased by transparency about who will use the data and how others will benefit

102. While all population groups should have the opportunity to be included and represented in health research, research being beneficial for all does not exclude the need for consent or an alternative robust legal basis, and the right of patients to have a say about how their data is used. Health and data literacy are useful tools for achieving these goals and should be co-designed with patients to increase participation. Inclusivity, participation and transparency are tools for building trust and are critical for data governance. Central authorities for approval of health data research applications should include public and patients' representatives, as well as mechanisms for public involvement. Authorities need to communicate in an agile, clear, honest and relevant way to minimize the impact of misinformation.

Trust building tools:

- *Inclusion and diverse representation in research*
 - *Transparency about how and by whom data are used*
 - *Participation through public engagement and promotion of health and data literacy*
-

Legal basis for the secondary use of health data for research

103. In the workshop, a discussion evolved around ‘Consent’ as a legal basis for the processing of health data for research. Under the GDPR, Articles 6 and 9 refer to consent as a legal basis for data processing. However, consent is not the only lawful basis. Some suggest that if we rely on consent provided during clinical care or research, it should be broad (albeit not a “blanket consent”). That is, consent for future research projects where the concrete aims are not specified. Admittedly, industry is often limited when it comes to research, because companies are normally perceived as collecting data for their own marketing purposes, and as research cited previously shows, people are less inclined to trust in private entities in giving them future consent for data use for purposes which are not previously identified. Advocating for, and requiring consent, is a professional requirement—and as real-world data becomes increasingly accessible and actionable, we need to keep in mind that obtaining consent in new contexts may not be feasible, even if it is aspired to.

104. Another challenge is that relying on consent often means relying on the choices of individuals who do not have the time, or the ability, to give fully informed consent. We should provide practical opportunities to give consent in appropriate circumstances, while acknowledging the suitable use of other legal basis, when there is benefit in using data, rather than inferring consent in an artificial manner.

105. Furthermore, efforts to incentivize consent and rely solely on it for big health data research may lead to less attention to implementation of privacy-by-design measures, thereby increasing the risk of data breaches or data mishandling.

106. It was voiced in the workshop, that when it comes to big data, consent can be difficult to collect since the quantities of unique individuals in the dataset are vast. Additionally, when conducting longitudinal data studies, some of the people may be out of reach or deceased. This is another reason to rely on alternative legal basis, with requirements for ethical approvals that measure risk; and privacy-enhancing technologies that perform risk analysis and mitigate re-identification.

107. Other positions consider the reliance on (opt-in/opt-out) consent in the context of sharing health data as being somewhat misguided. There is a need for unbiased information at a system level. Requiring consent might inadvertently introduce bias, since those who are willing to engage in research, and are consequently represented therein, shape the consent framework. We need to ask ourselves what promises, or representations are being made to patients and individuals, and how to fulfill the expectations of those individuals in a trustworthy manner. We ought to ensure that individuals are well informed about the purpose of use of their data, to avoid consequential harm. And we need to seek legislative solutions, rather than always rely on consent as a basis for acquiring information.

108. Overall, while all participants in the workshop valued health research; there were different viewpoints on the use of consent or alternative legal basis expressed. If we are to move forward with harmonized approaches to data governance, it will be important to bridge these gaps.

109. Shortly after this workshop took place, the EDPB (European Data Protection Board) published a response document to resolve inconsistencies between the prominent autonomy-respecting bioethics principle of consent required for participation in medical research, and Articles 6 and 9 of the GDPR. The EDPB's response stated as follows:

"Ethics standards cannot be interpreted in such a way that only explicit consent of data subjects can be used to legitimize the processing of health data for scientific research purposes. Article 6 and Article 9 GDPR contain other options for a legal basis and an exemption, that can be relied on for processing health data for scientific research purposes. The requirement of informed consent for participation in a scientific research project can and must be distinguished from explicit consent as a possibility to legitimize the processing of personal data for scientific research purposes."

110. The EDPB position essentially distinguishes between the bioethical requirement of informed consent to participate in medical research projects, and 'consent as a legal basis for processing of personal data', embedded in Article 6(1)(a) of the GDPR. Articles 6(1) and 9(2) of the GDPR recognize, respectively, legal basis other than consent, as well as exemptions other than explicit consent, as alternative legal grounds to be relied upon for the processing of health data for scientific research purposes or for public interest when it is legislated in union or member state law. The EDPB's position is that these are not incompatible with bioethical standards for research participation and consent.

111. It is noteworthy that the EDPB states that the bioethical requirement of informed consent for participation in medical research projects still needs to be met when relying on Article 6 or on one of the exemptions in Article 9(2) of the GDPR. The important distinction is that this requirement pertains to interventional medical studies involving the human person and body. The data research referred to in the workshop is non-interventional in nature, that is retrospective studies typically conducted on databases using data analytics tools and algorithms, that do not entail personal human interactions, substance administration or physical touch.

Data quality and linkage

112. The OECD Health Data Survey circulated as background to the workshop revealed findings on the use and governance of health data among surveyed countries (Annex E). On a high-level, the conclusion is that there is considerable variability across countries and within countries regarding health data development, accessibility and data protection.

113. Among countries with lower scores for dataset availability, maturity and use, the challenge lies in making data available for research and statistical purposes and there is work to be done to develop collaborative policies and practices among government authorities in custody of key health data and considerable work and investments required to improve data quality, linking to other data sets and sharing with researchers so that data can serve the health-related public interest. Among countries with lower scores for dataset governance, there are gaps to address in data privacy and security protections for key health datasets such as having a data protection officer and providing staff training, access controls, managing re-identification risks, and protecting data when they are linked and accessed.

114. One of the elements surveyed and found to be fundamental in the report was also highlighted by workshop participants: data linkage between repositories. Finland stressed that it finds the key to success of its health data for research project is the citizen ID that each citizen has from birth. Unique identification is also essential for data linkages within the French Health Data Hub and the Korean data lake. It was stated in the workshop that a national ID is instrumental to maintain data quality through data linkage, especially with respect to the health sector. Data linkages enable the information value within one dataset to grow, connecting treatments to outcomes and providing essential context to assess outcomes. As a result, it is fundamental for health data governance frameworks to consider the impact upon data linkage capabilities when developing rules regarding data pseudonymisation so that data linkages for future approved projects can be possible; and on whether the legal basis for processing health data, including dataset linkages, will be an alternative authorization to consent. Seeking consent to link older and large datasets may be impracticable and result in biased data that invalidates the research.

4 Tools to support adoption and scaling of research with health data

115. This chapter presents key approaches and technologies that were discussed during the workshop that can support the advancement and scaling of research involving the secondary use of health data. These include privacy by design and privacy enhancing technologies; synthetic data; homomorphic encryption computation; and federated (democratic) models of data development and use.

Privacy by Design & Privacy Enhancing Technologies

116. To enable innovation, access to health data is key. A relatively new topic just a few years ago, privacy-by-design is now a recognized profession of privacy engineering and risk management. It calls for the whole information technology eco-system, from platform through applications to data, from initiation to procedures, design and development, to incorporate privacy enhancing tools.

117. Privacy-by-design best practices involve a combination of privacy enhancing technologies with data access processes and tools. The latest example of this would be the protection of privacy, while processing COVID-19 patient and population level data. Privacy-by-design encourages consideration of data protection at all stages of data development, use, sharing and access; and in the design of IT systems for data management. These principles serve as a risk mitigation tool. An example of a privacy-by-design practice raised during the workshop is the use of ‘containers’ for storing data both on an organisation’s premises and on the cloud service because data are encrypted when they move between containers.

118. A workshop participant presented that though implementing privacy-by-design incurs costs, these can be mitigated by privacy-enhancing technologies. Some traditional options include three typical examples of privacy enhancing tech: HTTPS; global privacy control; and infrastructure for privacy respecting metrics (a relatively new initiative by W3C). Opting for a privacy-enhancing technology largely depends on: (1) the intended use-case; (2) what data are actually needed; and (3) trust boundaries. Barriers to adoption of such technologies include: the common belief that privacy comes at the expense of something else; the belief that businesses will lose access to personal data and the revenue that comes from them; too little emphasis on first parties and too much focus on third parties; policies and laws that threaten end-to-end encryption; and tensions between privacy and competition objectives.

119. Furthermore, researchers and entrepreneurs have brought to market various state-of-the-art solutions in support of privacy-by-design, for safeguarding digital personal data against unauthorized access and use. Emerging concepts and mechanisms specific to privacy enhancing technologies include differential privacy, federated learning, de-identification, homomorphic encryption, synthetic data, multi-party computation, distributed analytics or real-time remote data access. Each of these mechanisms, however, is context-dependent and presents unique benefits and limitations depending on the use case.

120. De-identification or Anonymisation are perhaps the most desired outcomes to make the claim that data are not personalised. It is a challenge, however, to know to what extent data are truly anonymous after de-identification. It has been demonstrated by researchers that fully anonymising a dataset so that it

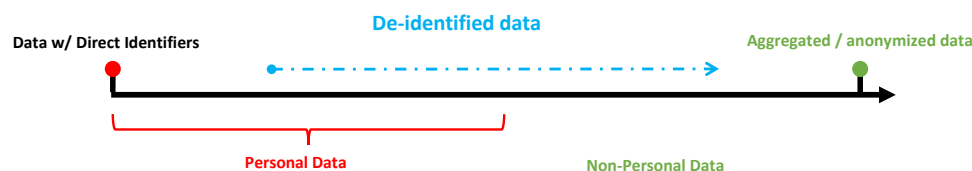
will be absolutely impossible to re-identify a person from the data is nearly impossible, making it necessary to move forward with caution when considering the level of privacy safeguards employed to personal data.

¹² The GDPR language has taken this into its definitions by terming the procedure 'Pseudonymisation'.

121. Depending on the measures invested in re-identification, identifiability can be anywhere on a range of probabilities (see image below) of assigning a real identity to a record in a data set, varying from zero to one. Where it is impossible to assign an identity to a record, this is when probability equals zero. Any personal dataset can have a probability of identification along the line except zero. Along this line, it is possible to determine a threshold value that distinguishes between personal data and non-personal data, depending on the policy adopted.

122. Privacy-enhancing technologies move the risk of re-identification along the line to different levels or different degrees. Data processors should exercise due diligence by analysing and measuring the potential risk and mitigating it accordingly.

123. Zero risk is unlikely to be fully achieved and is very difficult to attain in practice. Therefore, some researchers believe it should not be the goal, and that de-identification in practice should be considered as a risk-based method for measuring the probability of re-identification as a function of both data transformations and the controls to further reduce the risk of identifiability. These may include contractual agreements disallowing attempts to re-identify the data, controlled, secure and managed research platforms, researcher vetting, type and scope of data involved and rigorous oversight.¹³



124. Appropriately anonymizing health data is ever more challenging to achieve, considering the utility required for successful research. When health data are processed to the point they may be considered (nearly)-anonymous (partially masked or highly aggregated) then they may be un-linkable and not granular enough or accurate enough for medical research or any research for that matter.

125. This is especially true for longitudinal research that mandates the ability to update each subject's file over time. Sometimes there is also an incentive or even a moral obligation to maintain the ability to re-identify the participants, so that it remains possible for researchers to provide critical information to research participants. This presents a challenge to the level of protection that can be practically achieved.

126. Additionally, de-identifying data is inertly associated with utility for some machine learning model development. For example, de-identifying demographic data may reduce the utility of studying the impact of environmental factors.

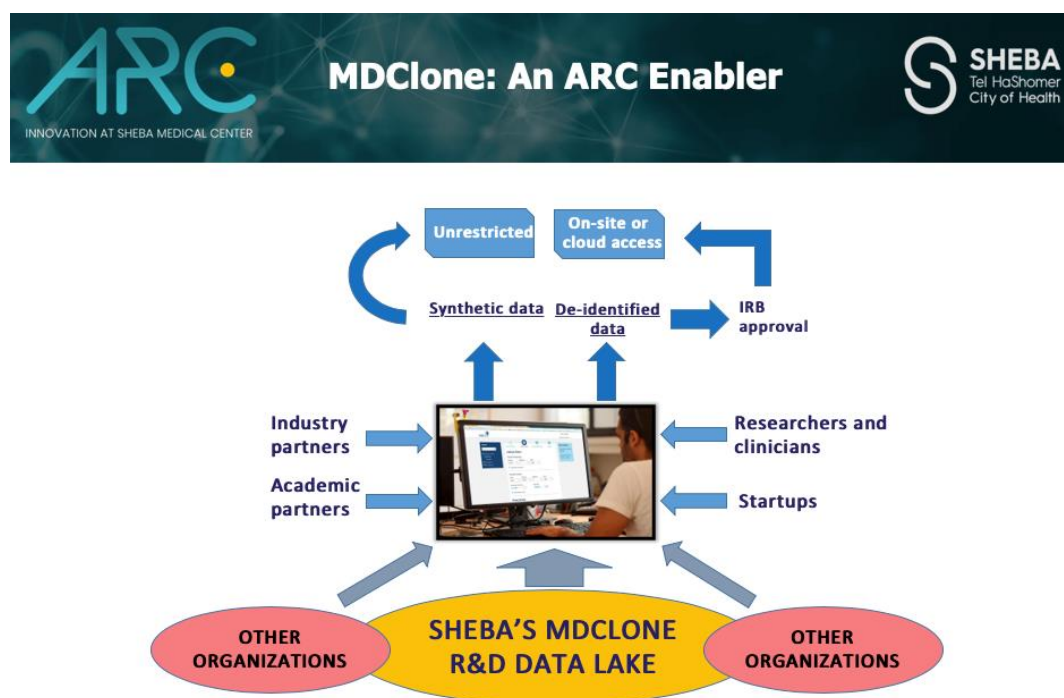
127. It should be noted that some privacy enhancing technologies (such as federated learning and homomorphic encryption, presented hereinafter) offer a solution that employs encryption of the data rather than de-identification, thereby aiming to provide a non-relative masking of the data, along with full utility. These models require further study.

¹² Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCLA Law Review 1701; Rubinstein, Ira, [Big Data: The End of Privacy or a New Beginning?](#) (October 5, 2012). International Data Privacy Law (2013 Forthcoming), NYU School of Law, Public Law Research Paper No. 12-56,

¹³ El Emam K, Alvarez C., 'A critical appraisal of the Article 29 Working Party opinion 05/2014 on data anonymization techniques', Oxford University Press, International Data Privacy Law, 2015, Vol. 5, No. 1; El Emam K, Malin BA. Appendix B: Concepts and methods for de-identifying clinical trial data. In: Committee on Strategies for Responsible Sharing of Clinical Trial Data; Board on Health Sciences Policy; Institute of Medicine, editor. Sharing clinical trial data: Maximizing benefits, minimizing risk. Washington (DC): National Academies Press (US); 2015

Synthetic data

128. In the workshop, the Israeli ARC participant shared experience with a solution by an Israeli company that generates synthetic data. The synthetic data batches (fake data) are created by a query built on top of real data, while maintaining data integrity. Generating the synthetic data file gives a report on how closely it matches the real data. Since the data are not real, IRB approval is not mandated, and data protection regulations do not apply to them. Synthetic data can be very easily shared between corporations and institutions, which, in turn, accelerates research. The user interface allows any user to query the data and it takes only 4-6 hours of training on the system. This allows clinicians to create their own datasets without needing data preparation by programmers or data intermediaries. Synthetic data can be accessed from anywhere, whereas real data can usually be accessed only on secure environments. A caveat with synthetic data however is that they are usually not suitable for the generation of publishable research results. A final step where statistical programs or algorithms are run on the 'real' data may be needed.



129. Synthetic data is a new concept, but there are already scientific publications validating synthetic data and its uses as compared to real data. Two-hundred Sheba hospital doctors have gone through the training and use the system. At first, synthetic data was a prelude to working on the real data. However, if researchers do not plan on publishing their findings, but instead simply hope to learn something useful from the data, it has been shown in practice that these researchers solely use the synthetic data and do not apply for access to the real data; therefore, circumventing a lot of time and bureaucracy.

130. Synthetic data are also used to get an initial impression from the data. For example, when pharmaceutical companies consider a new drug, synthetic data are as beneficial as real data, and more easily available to a private sector company. However, there are limitations to synthetic data, since they are created from real health data. There is also a tradeoff between the level of synthesis, i.e. divergence from the real data that provides more privacy protection when it is higher, but carries a price in data utility. Furthermore, inference of the real subjects from synthetic data may be possible in some cases, so that

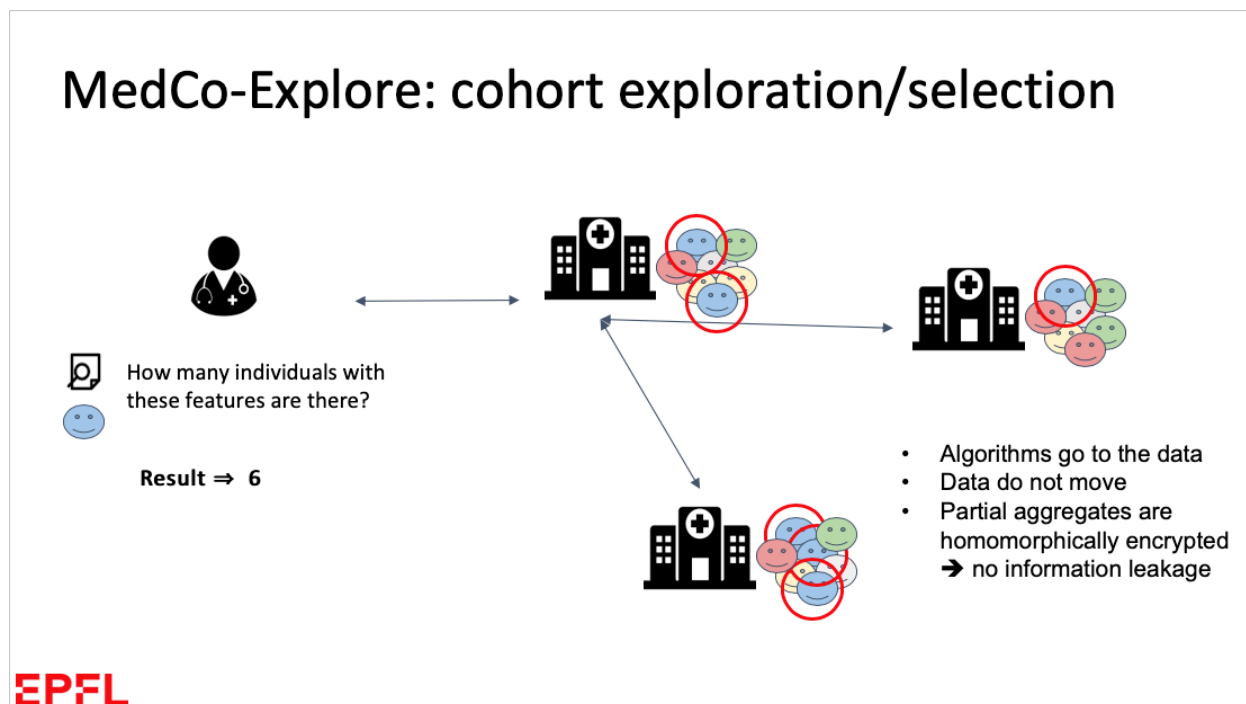
avoiding re-identification may require a higher level of synthesis, resulting in it being unrealistic for machine learning purposes.¹⁴ In the context of health research there may also be a challenge since there is a need to add data to a person's file from various sources over time, therefore requiring a level of identifiability and data integrity over time so that synthetic data records are realistic representations of the population.

Homomorphic Encryption Computation

131. As presented in the workshop, multi-site studies benefit from a higher volume of available data with an improvement in the statistical significance of findings as the main benefit. However, there are four challenges that multi-site studies face: (1) the need to interact with several ethics' committees and IT services; (2) the need to adhere to diverse regulations; (3) heterogeneity of data semantics and data quality; and (4) a reluctance to share data control.

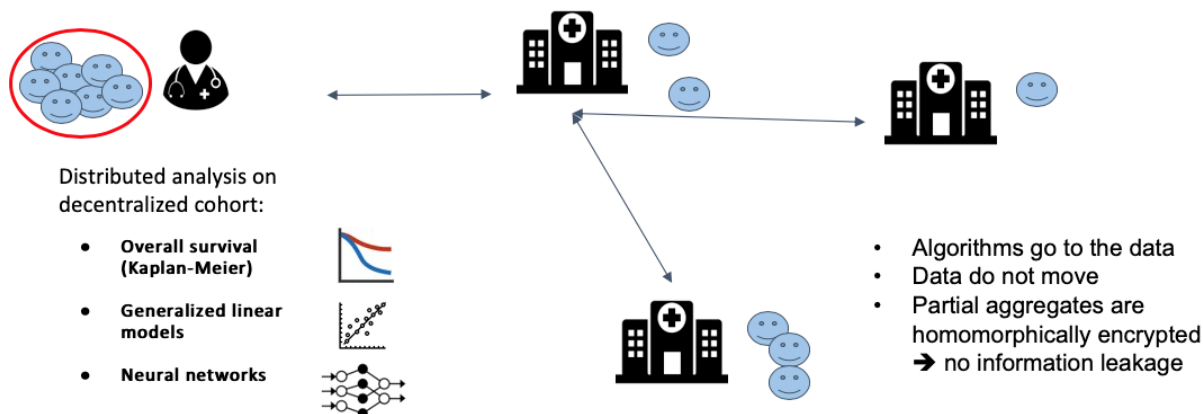
132. An available solution is to extract info without transferring data, as shown in the presentation of the MedCo toolbox, intended for both data research and clinical research.

133. In sum, it is possible to extract information out of data sets without transferring them and while keeping them encrypted, thereby avoiding loss of control over the data; driving higher motivation to share data; and a higher willingness to harmonize data fronts. This can be achieved with techniques like homomorphic encryption and secure multi-party computation. These solutions are being deploying at Swiss hospitals and beyond: yielding more efficient healthcare and better preparation for the next pandemic. (More info: SecureCovidResearch.org).



¹⁴ ICO website, "[How should we assess security and data minimization in AI?](#)", What privacy-enhancing methods should we consider?, synthetic data

MedCo-Analysis: distributed analytics



EPFL

134. Under the umbrella of homomorphic encryption, there are different operations that are determined by what one is able to conform to. Since the data are being investigated while under encryption, by several actors who participate in the computation, but are not granted access to the raw data, this technology can be seen as anonymization to the level of the GDPR Framework. The relationship between GDPR and advanced cryptography tools such as homomorphic encryption and secure multiparty computation is explained [here](#).

135. For analytics that require data not be encrypted, for example AI, often it is required to have access at a fairly granular level. There are techniques to mitigate re-identification attacks, but there is no silver bullet, at least for the moment, so with enough effort re-identification may occur. Beyond a certain level of granularity, the researcher may be required to conform to additional guarantees and the data controller should employ compensating safeguards on the data.

136. The distributed approach using micro-aggregate data has been applied in several EU projects. It was also used by the OECD hospital performance benchmarking project (for further reading on the application of this approach from a GDPR perspective [see here](#)).

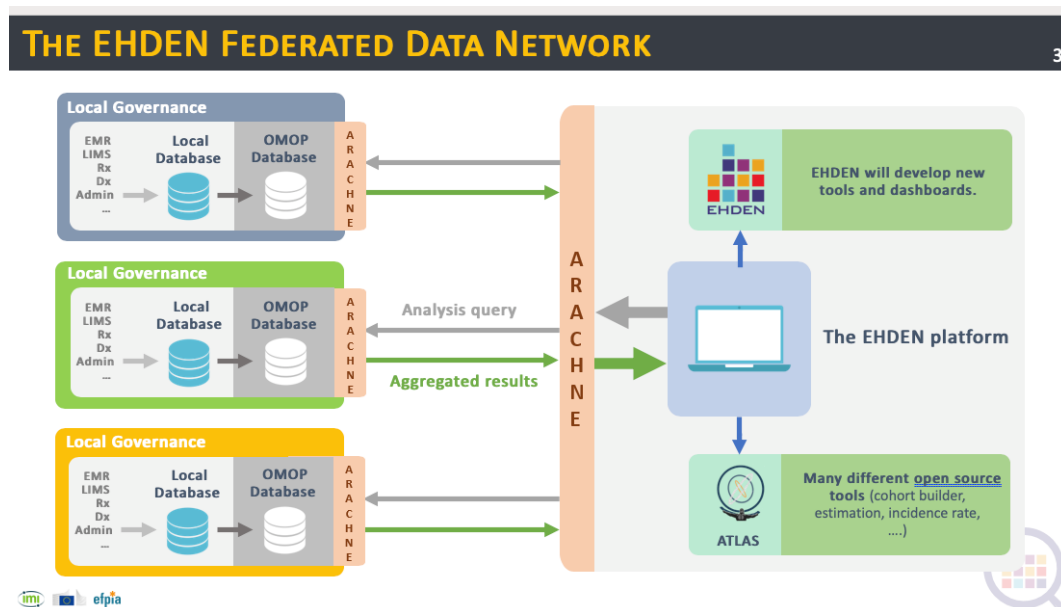
137. Homomorphic encryption and secure multiparty computation would work on defined data sources and requires appropriate data formats to be established beforehand.

The (Democratic) Federated Model

138. Some argued at the workshop that the future is, or should be, based on a federated model of citizens holding their own data. For example, patient-reported outcome measures (PROMs), personal health records and data collected via sensors etc. Responses included a view that this is still aspirational, though working with patient bodies, focusing on ePROs and patient generated data and also linking with other projects, such as [IMI2 and H2O](#) is a path worth walking.

139. There was also mention of a debate among epidemiologists that federated data with machine learning would introduce bias compared to the analysis of pooled individual data. For example, the global CONCORD 3 study included individual records for 37.5 million patients diagnosed with cancer during a 15-year period from 2000–14 that were provided to the study team from cancer registries in 71 countries. The scientific argument for pooling such a vast volume of health microdata includes that it permits data verification and improves the estimates of cancer survival and the adjustment of the data for underlying differences in population characteristics.¹⁵

140. A question was asked about whether you can run a Cox PH Model using federated data until you replicate the entire time series? The reply was there are numerous publications and methodological studies outlining this risk, yet the participant argues it is not as challenging as some say and there are recent publications of survival analysis applied to federated data via distributed analytics.¹⁶ Indeed, there are certain caveats and Federated Data Networks cannot provide all answers for all studies. There is a need for an ecology of approaches. In terms of the time series aspect, there is considerable focus on this within The Observational Health Data Sciences and Informatics (OHDSI) in terms of the common data model and the methodological tools right now, which is at the core of the EH DEN project.



¹⁵ Allemani C, Matsuda T et al (2018), Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries, *The Lancet* 391(10125): 1023-1075, [https://doi.org/10.1016/S0140-6736\(17\)33326-3](https://doi.org/10.1016/S0140-6736(17)33326-3).

¹⁶ You SC, Rho Y, Bikdeli B, et al. Association of Ticagrelor vs Clopidogrel With Net Adverse Clinical Events in Patients With Acute Coronary Syndrome Undergoing Percutaneous Coronary Intervention. *JAMA*. 2020;324(16):1640–1650. doi:10.1001/jama.2020.16167.

5

Priorities for future collaborative efforts

141. The following key themes emerged from the workshop presentations, discussions and background materials that identify priority areas for future collaborative efforts.

- Insufficient clarity and harmonization between national health data governance frameworks challenge their implementation within nations and between them.
- Data localisation laws and policies limit cross-border collaboration in health research.
- Conflicts about the appropriate legal basis for data processing: Consent vs. Public Interest – and the scope of “Public interest” – affect countries’ ability to conduct research and streamline data sharing project approvals.
- Insufficient awareness of, and expertise in the use of, Privacy Enhancing Solutions. This gap limits their wider adoption to strengthen data protection and limit their use when developing a legal basis for data processing or approving data processing requests.
- Necessity of fostering public trust of individuals, communities and societies in scientific research that is in the public interest, through Inclusion, Transparency and Participation.
- Importance of recognising and addressing inequalities in access to data, such as when researchers from diverse backgrounds are denied access to data due to payment or credit sharing requirements, and the risks of bias in research results emanating from this.
- Importance of recognizing and addressing the consequences of inequalities emanating from potential bias in the data explored, in and of itself, or due to insufficient inclusion of participants.
- Necessity of strong digital identification methods on a national level to support data linkage.

142. The OECD will consider the results of this workshop as we monitor progress toward the implementation of the OECD Council Recommendation on Health Data Governance. It will further inform follow-on work from the review of the OECD Privacy Guidelines and the next phase of the OECD Going Digital horizontal project which focusses on Data Governance in health and other sectors of the economy.

Annex A. Workshop agenda

DAY ONE - 19 January 2021	
13:00	–Opening and welcome remarks
13:10	<ul style="list-style-type: none"> • OECD, Francesca Colombo, Head of Health Division, Directorate for Employment, Labour and Social Affairs • Israel, Asher Salmon, Head of the Department of International Relations, Ministry of Health
13:10	–Session 1: National health data governance reforms
14:10	<p>While the benefits of strong and well-governed health data systems are clear, enabling progress in medicine, delivery of patient-centered services and access to information, health care quality and performance improvements, protecting patient safety, and more; the project of implementing national data governance frameworks is challenging. This panel session will discuss the importance of national health data governance reforms in the context of the WHO Global Strategy on Digital Health and asks 4 countries who have implemented recent significant legal and operational reforms to strengthen health data governance the following key questions.</p> <ul style="list-style-type: none"> • What were your main aims and objectives when you launched legal and operational reforms? • What are the most significant reforms you have launched and what are the major changes that your reforms have brought about? • What obstacles did you face in implementing reforms and what challenges are you still facing? <p>Moderator: Jillian Oderkirk, OECD</p> <p>Discussants:</p> <ul style="list-style-type: none"> • WHO Europe, Clayton Hamilton, Coordinator, Digital Health Division of Country Health Policies and Systems • France, Louisa Stuwe, International Policy Manager, Health data Hub • Finland, Antti Piirainen, Head of Communications THL (Health Ministry) • Korea, Dr. Jin Yong LEE, the Chair of Health Insurance Review and Assessment Agency (HIRA) • United States, Elisabeth Myers, ONC Deputy Director of Policy, HHS Office of the National Coordinator for Health IT
14:10	–Open discussion
14:30	
14:30	–Session 2: Safeguards for health data sharing
15h40	<p>Countries around the world are developing data governance policies and mechanisms to promote innovation around the use of digital health information while protecting patient privacy expectations. These may include: 1) ethical review board oversight over data use purpose, and data collection, sharing, and analysis; 2) de-identification of datasets; 3) administrative, technical and contractual safeguards; 4) safeguards around cross border data flows. In this session,</p>

presenters will discuss these policies and mechanisms in light of the latest evidence about how they support large-scale health data digitization.

The session will focus on the following questions:

- Under what circumstances should ethical review boards scrutinize and approve the use of health data for research?
- Who can access health data for research and under what conditions?
- Under what conditions can data be shared with private for-profit organisations for research within the health-related public interest?
- What is the standard for anonymising or de-identifying health data? How is this standard articulated in various privacy laws?
- What are the implications of data localisation requirements for research data? What are the specific challenges in the context of fighting the pandemic? Under what conditions can health data be shared cross border for research within the health-related public interest?
- What are possible solutions to the misalignment between data protection frameworks, policies and standards around various countries?

Moderator: Limor Shmerling Magazanik, ITPI (Israel Tech Policy Institute)

Discussants:

- **Israel**, Esti Shelly, Director of Digital Health at Israel Ministry of Health
- **IAPP** (International Association of Privacy Professionals), Prof. Omer Tene, Vice President, Chief Knowledge Officer
- **Intel Corporation**, David Hoffman, Associate General Counsel, Senior Director of Data Policy Strategy
- **Sheba Medical Center**, Dr. Eyal Zimlichman, Deputy Director General, Chief Medical Officer and Chief Innovation Officer.

15:40 - **Open discussion**

16:00

DAY TWO - 20 January 2021

13:00 – **Opening and welcome remarks**

13:10

- **OECD** – Audrey Plonk, Head of Division - Digital Economy Policy, Directorate for Science, Technology and Innovation
- **Israel**, Dr. Shlomit Wagman-Ratner, Head of the Israeli Privacy Protection Authority, Ministry of Justice.

Session 3: Privacy by Design

13:10 – Researchers and entrepreneurs have brought to market various state-of-the-art solutions in support of privacy by design for safeguarding digital health data against unauthorised access and use. Emerging mechanisms include de-identification, differential privacy, homomorphic encryption, synthetic data, multi-party computation, distributed analytics and real-time remote data access. Each of these mechanisms, however, is context-dependent and presents unique benefits and limitations.

14:10 This session will focus on the following questions:

- What are the costs and benefits of various privacy enhancing tools?
- What are best practices or standards for deploying various privacy enhancing tools?

- What are barriers to wider adoption?

Moderator: Jules Polonetsky, CEO, Future of Privacy Forum

Discussants:

- **Microsoft**, Hadas Bitran, Head of Microsoft Health, Microsoft Israel R&D Center
- **EFPIA** (The European Federation of Pharmaceutical Industries and Associations), Nigel Hughes, Coordinator and Scientific Director, JCI Patient Data for Research, Janssen
- **Internet Society** (ISOC), Christine Runnegar, Senior Director, Internet Trust
- **EPFL** (École Polytechnique Fédérale de Lausanne), Prof. Jean-Pierre Hubaux

14:10- **Open discussion**

14:20

14:20 – **Session 4: Individual & Community perspectives on using health data for research**

15:25 In clinical research settings, patients' consent to the use of their health data to address specific research questions is typically required. But consent is not the only legitimate basis for data use and is impracticable when the volume of research uses of the data become high. In most countries, the use of patient data for research is authorized by law, foregoing individuals' consent requirement, providing that certain conditions are met. Some countries provide patients with a mechanism to opt-out of research uses of their own data. New technologies, such as secure patient portals, provide patients with greater engagement with, transparency about and control over their data. At the same time, the rapid advancement of data-based research increases the need to ensure future-proof patient empowerment mechanisms. Alongside the interests of individuals, there are also the communal and societal interests in improving health services to growing populations and investing in preventive medicine.

The session will focus on the following questions:

- What type of research should be regarded as a public good?
- How to prevent bias and discrimination in health data research?
- What are the effects of limiting data use in a machine learning environment?
- What are some of the challenges around reliance on patient consent to enable health data research? How can these challenges be mitigated or resolved?
- What policy or program mechanisms can foster and ensure the essential public trust and confidence in the protection of personal health data used for scientific research?

Moderator: Dr. Sivan Tamir, ITPI (Israel Tech Policy Institute)

Discussants:

- **Wellcome Genome Campus**, Dr. Richard Milne, Society and Ethics Research
- **BBMRI-ERIC**, Adv. Irene Schlünder, ELSI expert
- **EPF** (European Patients' Forum), Kaisa Immonen, Director of Policy, Belgium
- **Novartis International**, Knut Mager, Head of Global Data Privacy

15:25 – **Key take aways and discussion**

15:50 **Moderators:** Elettra Ronchi and Jillian Oderkirk

15:50 – **Closing Remarks: OECD**

16:00

Annex B. Speakers

Day 1 - 19 January 2021

Opening remarks

Mrs. Francesca Colombo, M.Sc., is Head of the Health Division at the Organization for Economic Co-operation and Development.

She oversees OECD work on health, which aims at providing internationally comparable data on health systems and applying economic analysis to health policies, advising policy makers, stakeholders and citizens on how to respond to demands for more and better health care and make health systems more people centered. Major activities of the OECD Health Division cover the response to the COVID-19 crisis and solutions to make health systems more resilient; trends in health spending; measuring of health care outcomes, activities and inputs; health care quality policies; assessing health system efficiency and value for money; health workforce; long-term care systems and ageing; the economics of public health; pharmaceutical policies, new technologies, big data and Artificial Intelligence in health. Mrs. Colombo has over 20 years of experience leading international activities on health and health systems. Over her career, she travelled extensively in Europe, South America and Asia, advising governments on health system policies and reforms.





Dr. Asher Salmon, Department of International Relations, Ministry of Health, Israel.

Previous positions: Deputy Director, Hadassah University Hospital, Ein-Kerem. Deputy C.E.O & Medical Director, Barzilai Medical Center. Head, Oncology Inpatient Ward, Sharett Institute of Oncology, Hadassah University Hospital. Head, Medical Audit Department, Israel Ministry of Health. Head, Outpatient Clinic, Sharrett Institute of Oncology, Hadassah University Medical Center. Education: Masters (MA) in health administration, Ben-Gurion University, Israel Summa cum laude Clinical research fellowship in Clinical Oncology & Cancer Genetics, The Royal Marsden NHS Trust & Institute of Cancer Research, London, UK; Residency in Oncology, Hadassah University Hospital, Jerusalem Ph.D. studies in molecular biology, Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University of Jerusalem. Research fellowship in Nuclear Medicine joint program, Harvard University Medical School, Boston, MA, USA. M.D studies, Faculty of Medicine, The Technion, Israel Institute of Technology, B.Sc. Medical Sciences.



Session 1

<p>Jillian Oderkirk is a Senior Health Economist in the Health Division of the Organisation for Economic Co-operation and Development (OECD) in Paris, France. Jillian studies OECD countries' progress in advancing national health information infrastructure including the development of electronic health record systems, adoption of digital technologies and health data governance. Prior to joining the OECD in 2011, Jillian was Director, Health Analysis Division at Statistics Canada in Ottawa, Canada. Jillian has a master's degree in economics from McMaster University, Hamilton, Canada.</p>	
<p>Clayton Hamilton leads the digital health flagship program within the WHO Regional Office for Europe's Division for Country Health Policies and Systems, providing support and guidance on all aspects of digital health and innovation to support health systems strengthening and reform in the Region's 53 Member States. With a background in ICT development and business management and over 22 years of UN experience working in the fields of technology, digital health, and public health, Mr. Hamilton provides strategic guidance to assist countries in their development of national digital health programs within the framework of delivering on WHO's triple billion targets and in working towards the achievement of the UN Agenda for Sustainable Development. In working to accelerate the uptake of safe and inclusive digital health services for all, Mr. Hamilton leads the development of dialogue and regional guidance on digital health and innovation, establishes partnerships with major international stakeholders working to reform health systems and public health, and provides thought leadership on the adoption of frontier digital technologies to improve access to healthcare and empower individuals in their choices for better health and well-being.</p>	
<p>Louisa Stüwe is international policy manager at the French Health Data Hub where she is in charge of coordinating the French contribution to the European Health Data Space including the TEHDaS Joint Action. She fosters international partnerships with institutional, private and academic actors to promote cross-border research collaborations and the dissemination and exchange of best practices. Before joining the Hub, Louisa managed health, research and innovation programs and activities, funding, policy, projects and networks at the Pan American Health Organization/WHO, the French Ministry of Health, the World Bank Group, GlaxoSmithKline and as a consultant. Louisa studied Public Administration and International Politics at Sciences Po Paris and Georgetown University as well as Public Health at the French School of Public Health (EHESP).</p>	

<p>Antti Piirainen is the Head of Communications and part of the management group at Findata, the Finnish Health and Social Data Permit Authority. He has previously worked in the field of communications at the Finnish Tax Administration and at an independent think tank Demos Helsinki, before jumping in to Findata in 2019 as one of its first employees. Antti has a Master of Social Sciences (M. Soc.Sci) degree in Communication and Media Studies from University of Helsinki.</p>	
<p>Dr. Jin Yong Lee is Chair of HIRA Research Institute, Health Service Review and Assessment Service. He also serves as a professor in the Public Healthcare Center at Seoul National University Hospital and in the department of Health Policy and Management in Seoul National University College of Medicine. His research interests are Korean healthcare systems, quality of life, and the management of public hospital. In particular, he is trying to find the reasons of malfunction of primary care, excessive competition among healthcare providers, and inefficiency in public hospital. He received MD degree from Chung-Buk National University and PhD from Seoul National University and MHA from the University of North Carolina at Chapel Hill.</p>	
<p>Elisabeth Myers, Deputy Director (Acting Executive Director) Office of Policy, ONC. Elisabeth Myers has worked on health IT policy at HHS since 2012 working on CMS quality programs, the CMS eHealth Initiative, and the EHR Incentive Programs before moving to the Office of the National Coordinator for Health IT. Prior to her work at HHS, Elisabeth worked on health care initiatives in the non-profit and private sector, and at the state level in the Governor's Office of Health Care Reform in Pennsylvania. In her role at ONC, Elisabeth is helping to lead the team implementing the 21st Century Cures Act which addresses a wide range of health IT provisions from interoperable standards development to health IT for specialty settings and sites of service including pediatric care. Elisabeth also leads ONC policy efforts related to the health IT provisions within the Support for Patients and Families Act which was signed into law in December of 2018 to drive policy initiatives in support of OUD prevention and treatment.</p>	
<h2>Session 2</h2>	
<p>Limor Shmerling Magazanik is the Managing Director of the Israel Tech Policy Institute. In this role, Limor provides leadership on day-to-day operational matters of ITPI, including directing ITPI's policy agenda; engaging policymakers, regulators, academics, and business leaders; convening multi stakeholder groups for discussion; and overseeing communications with the public and the advisory board. Limor comes to ITPI after a decade with the Privacy Protection Authority, serving most recently as Director of Strategic Alliances and previously as Director of Licensing & Inspection. She led policy initiatives and regulation in technology driven sectors and promoted compliance with data protection, privacy, cybersecurity and digital identity regulation. She was an adjunct lecturer at the Hebrew University Faculty of Law and the IDC Herzliya School of Law, has LL.B., MA and LL.M. degrees from Tel Aviv University and is a CIPP/E, CIPP/US, CIPM.</p>	

<p>Esti Shelly is the Director of Digital Health Division in the Israeli Ministry of Health. Esti is leading the digital health initiative in the Israeli healthcare system. In her current position, Esti is responsible to drive digital transformation and lead policy, implementation and change, with a diverse team with a range of skills and expertise. In her previous position, Esti served as the deputy director of the Medical Technology, Health Information and Research Directorate, in the Israeli Ministry of Health. Esti led national projects such as the National Program for increasing the accessibility and availability to MR exams, policy making in various fields especially medical exposure to ionizing radiation, as well as international collaboration with the IAEA on the field of medical technologies. Esti holds a MPA in Public Administration from Harvard Kennedy School, a M.Sc. in Technology Management from the Hebrew University and B.Sc. Med in Medical Science from the Hebrew University. Esti is alum of Wexner Israel Fellowship (class 26 (and Atidim – Cadets for Public Service (Class 1), both are leadership programs of the Israeli Public Sector.</p>	
<p>Omer Tene is Vice President of Research and Education at the International Association of Privacy Professionals. He is an Affiliate Scholar at the Stanford Center for Internet and Society and Senior Fellow at the Future of Privacy Forum. He is an Associate Professor at the College of Management School of Law (on leave of absence). Together with Jules Polonetsky, he co-founded the Israel Tech Policy Institute.</p>	
<p>David Hoffman is the Steed Family Professor of the Practice of Cybersecurity Policy at the Sanford School of Public Policy at Duke University. He is also the Associate General Counsel and Senior Director of Data Policy Strategy for Intel Corporation. Hoffman currently chairs the Civil Liberties and Privacy Panel for the Director's Advisory Board for the US National Security Agency. He also chairs the board of the Center for Cybersecurity Policy and Law and serves on the Advisory Boards for the Future of Privacy Forum and Mine. Hoffman also founded and chairs the board for the Triangle Privacy Research Hub, which highlights and fosters cybersecurity and privacy academic research done in the North Carolina Research Triangle. Hoffman previously served on the Department of Homeland Security's Data Privacy and Integrity Advisory Committee and the Board of Directors of the National Cyber Security Alliance. He has also served in the U.S. Federal Trade Commission's Online Access and Security Committee, the Center for Strategic and International Studies Cyber Security Commission, the Steering Committee for BBB Online, the TRUSTe Board of Directors and the Board of the International Association of Privacy Professionals. He is the author of many papers and articles on cybersecurity and privacy and has testified to Congress on these topics. Hoffman has a JD from Duke Law School, where he was a member of the Duke Law Journal. He received an AB from Hamilton College.</p>	

Dr. Eyal Zimlichman serves as Deputy Director General; Chief Medical Officer; and Chief Innovation and Commercialization Officer at Sheba Medical Center.

He is founder and director of Sheba's ARC Innovation Program, an open global innovation ecosystem which aims to redesign healthcare through digital health solutions by 2030. He is board certified in internal medicine, a graduate of the Harvard School of Public Health Executive Health Care Management Master of Science program and earned his MD at the Technion Israel Institute of Technology in Haifa, Israel. In his position as Deputy Director General and Chief Medical Officer at Sheba Medical Center, Dr. Zimlichman is in charge of clinical affairs, quality, patient safety and risk management. He is also head of clinical innovation at Sheba, leading innovation efforts both internally and through establishment of partnerships with industry and leading academic medical centers worldwide. Prior to this Dr. Zimlichman held the position of Lead Researcher at Partners Health Care Clinical Affairs Department in Boston where he was involved in the efforts to bring about a strategic care redesign initiative. In that capacity, he established a program for Partners Healthcare, for collecting and reporting patient reported outcomes across the continuum of care, a program that won international appraisal. Dr. Zimlichman is a founding member of the International Academy for Quality and Patient Safety and currently holds an appointment at Brigham and Women's Hospital and Harvard Medical School affiliated Center for Patient Safety Research and Practice, where he is conducting research on implementing technology to improve health care quality and patient safety. He served as an advisor to the Office of the National Coordinator for Health Care Information Technology in the U.S. Department of Health and Human Services. Dr. Zimlichman currently serves as co-chairman of a policy steering committee dealing with strategic uses of healthcare data at the Israeli Ministry of Health. In 2013 Dr. Zimlichman was appointed as an International Expert at the International Society of Quality in Healthcare (ISQua). He is board certified in internal medicine, completed a Master of Science in Health Care Management at the Harvard T.H. Chan School of Public Health, and has earned his MD at the Technion Israel Institute of Technology.






Day 2 - 20 January 2021
Opening remarks

Ms. Audrey Plonk is Head of the Digital Economy Policy (DEP) Division of the Directorate for Science, Technology and Innovation (STI) at the OECD.

Ms. Plonk is responsible the Committee on Digital Economy Policy (CDEP) and the Committee for Consumer Policy (CCP). She contributes to the development of evidence-based policies through multi-stakeholder processes to i) stimulate the growth of an accessible, innovative, open, inclusive and trustworthy digital economy for sustained prosperity and well-being, and ii) provide policymakers with the tools needed to develop a forward-looking, whole-of-government policy response that leverages the potential of digitalization for growth and well-being. Prior to joining the OECD, Ms. Plonk was Senior Director, Global Security Policy at Intel Corporation where she was also a Senior Director for Public Policy based in Santa Clara, California.



<p>Dr. Shlomit Wagman is the Acting Director of the Israel Privacy Protection Authority since March 2019. She took part in the establishment of the IPPA in 2006. Dr. Wagman holds J.S.D. and LL.M. degrees from Yale Law School and a joint LL.B. and B.A. degree in law and business management from the Hebrew University of Jerusalem (magna cum laude). She clerked for the Chief Justice of the Israeli Supreme Court, the Honorable Professor Aharon Barak. She served as adjunct lecturer at Tel Aviv University, Bar-Ilan University and the Interdisciplinary Center (2007-2012), was a fellow of the Information Society Project at Yale Law School (2003-2007) and co-edited the book "Cybercrime: Digital Cops in a Networked Environment" (NYU Press, 2007), with Prof. Jack Balkin et al. During the past two years, which have indeed been challenging both due to the political situation and as a result of the enormous changes caused by Covid-19, Dr. Wagman has been proactive and influential, especially by initiating and promoting the right to privacy and protection of personal information in digital databases. She is also the Director General of Israel's Money Laundering and Terror Financing Prohibition Authority (IMPA), the Israeli Financial Intelligence Unit, since 2016.</p>	
<h3>Session 3</h3>	
<p>Jules Polonetsky serves as CEO of the Future of Privacy Forum, a Washington, D.C.-based non-profit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies. FPF is supported by the chief privacy officers of more than 175 leading companies, several foundations including the US National Science Foundation, as well as by an advisory board comprised of the country's leading academics and advocates. FPF's current projects focus on AI, Adtech, Ethics and Data Driven Research, Health, Smart City, Connected Cars and Student Privacy.</p>	
<p>Hadas Bitran is Head of Microsoft Health Israel, at the Microsoft Israel R&D Center, within Microsoft AI & Research. Hadas and her organization are focused on building technologies for Health & Life Sciences. She is the founder and manager of a multi-disciplinary group within Microsoft Health Next that drives Artificial Intelligence for healthcare. During her career in Microsoft, Hadas also developed products for User Privacy. Before Microsoft, Hadas held senior leadership positions managing software R&D and Product groups. In addition to her role at Microsoft, Hadas volunteers as Management Board Director at SNOMED International, non-for-profit organization that promotes global standards for clinical terminology. Hadas holds B.Sc. Computer Science from Tel Aviv University, and MBA from Kellogg School of Management, Northwestern University in Chicago.</p>	

<p>Nigel Hughes has a thirty-five-year career spanning the NHS in the UK (16 years), NGOs and patient organizations (10 years) and within the pharmaceutical industry (17 years). He has worked clinically in HIV and viral hepatitis, liver disease, and in sales & marketing, medical affairs, market access and health economics, R&D, precision medicine, advanced diagnostics, health IT and Real-World Data/Real World Medicine. His experience covers clinical, education, as an advisor, consulting, communications and lobbying over the years. He is currently the Project Lead for the IMI2 European Health Data & Evidence Network (EHDEN) and was Platform Co-Lead for the IMI1 European Medical Information Framework (EMIF), as well as consulting on numerous projects and programs in the domain of RWD/RW.</p>	
<p>Christine Runnegar is Senior Director, Internet Trust at the Internet Society, which advocates for policies that support an open, globally connected, secure and trustworthy Internet. With a background in regulatory litigation and an international outlook, Christine complements her privacy and security expertise with a deep appreciation of technology. Christine rolls up her sleeves to work with governments and other stakeholders to write guiding policies on cybersecurity and privacy and is not afraid of hackathons. Christine co-chairs the W3C Privacy Interest Group (PING), has served as a member of the ENISA Permanent Stakeholders Group, and currently serves on the Board of the Internet Security Research Group (ISRG), the home of the Let's Encrypt certificate authority.</p>	
<p>Jean-Pierre Hubaux is a full professor at EPFL and head of the Laboratory for Data Security. Through his research, he contributes to laying the foundations and developing the tools for protecting privacy in today's hyper-connected world. He has pioneered the areas of privacy and security in mobile/wireless networks and in personalized health. He leads the Data Protection in Personalized Health (DPPH) project funded by the Swiss ETH Council and is a co-chair of the Data Security Work Stream of the Global Alliance for Genomics and Health (GA4GH). He is a Fellow of both IEEE (2008) and ACM (2010). He is among the most cited computer scientists in privacy protection and in information security.</p>	
<h2>Session 4</h2>	
<p>Dr. Sivan Tamir is a Senior Researcher and Policy Counsel at the Israel Tech Policy Institute. Dr. Tamir is a researcher in the field of law and bioethics, specializing in Genethics. She was the Coordinator of the National Helsinki Committee for Human Medical Research (Ministry of Health) and has served as a researcher in the Genetic Policy & Bioethics Unit, at the Gertner Institute for Epidemiology & Health Policy Research. She also served as the Director of Foreign Relations for the Ministry of Justice. Dr. Tamir is a teaching fellow at the Faculty of Law, University of Haifa, and a member of the Advisory Committee on Bioethics of the Israel Academy of Sciences and Humanities. She holds a Ph.D. in law from the Hebrew</p>	

<p>University of Jerusalem (specializing in bioethics). Dr. Tamir's principal research interests lie in the fields of medical law and bioethics, particularly 'genethics', ethics of emerging technologies and health data, reproductive ethics, neuroethics, clinical research ethics, and health policy.</p>	
<p>Dr. Richard Milne is Senior Social Scientist in the Society and Ethics Research Group at the Wellcome Genome Campus and Senior Visiting Research Fellow in the Department of Public Health and Primary Care at the University of Cambridge. His research focusses on social and ethical questions associated with the development of medical technologies, particularly related to genomics and biodata. He leads work on Ethical, Social and Legal Implications within the Cambridge Public Health research initiative and is currently working on a study of how experts and members of the public address ethical questions associated with the development of data-driven tools for the detection of cognitive decline and the diagnosis of dementia.</p>	
<p>Adv. Irene Schlünder is a lawyer and expert in EU data protection law and database governance as well as IP rights. She works for TMF e.V., a member based German umbrella organization for academic biomedical research in Berlin. She has been involved in the IMI projects EHR4CR (co-author of the IMI Code of Conduct) and DO-IT (developing an Informed Consent Template for clinical trials including secondary use of the data for future research) as well as in H2020 projects like BioMedBridges, CORBEL and EOSC Life and HealthyCloud. She is a member of the Common Service ELSI of BBMRI-Eric and belongs to the core drafting group of the BBMRI lead initiative for a "Code of Conduct for health research" under the GDPR and has recently been appointed legal adviser to EJP-RD.</p>	
<p>Kaisa Immonen is Director of Policy at the European Patients' Forum (EPF). She has a master's degree in International Relations (UK and Finland). She is responsible for the overall lead of EPF's policy and advocacy work at EU level, strategic planning, policy analysis, and building positive relationships with EU institutions, international organizations such as the OECD, WHO and stakeholders. Her personal interests include patient empowerment and involvement in healthcare improvement, patient safety and quality of care, and health literacy. She is currently co-chair of the European Medicines Agency's Patients and Consumers Working Party since 2016. From October 2020 she is a patient representative on the EMA COVID-19 task force. She also represents EPF on the OECD's health care quality and outcomes working party, an advisory group on PRIMIS, the International patient advisory group for the OECD Paris initiative, and has been involved in CIOMS working group XI on patient involvement. In her personal capacity she is a member of the BMJ Patient Panel and was involved in the ICHOM working group on general adult health. Before joining EPF in 2010 she worked in EU health policy, advocacy and external communications roles in both the private and non-profit sectors.</p>	

<p>Knut Mager is Novartis' Head Global Data Privacy and Group Data Protection Officer. As a frequent speaker and member of advisory boards, Knut contributes actively to the development of principles and standards in Data Privacy. From 2015 - 2019, Knut chaired the Data Protection working group of the European Federation of Pharmaceutical Industries and Associations (EFPIA). Knut is a German lawyer; he has been working in the pharmaceuticals industry in various roles since 1990. In 2003, he joined Novartis as General Counsel, Head of Legal of the Sandoz. Since 2007, he held various leadership positions with Novartis Group Legal.</p>	
<h3>Closing remarks</h3>	
<p>Elettra Ronchi, PhD, MPP, is Senior Policy Analyst at the OECD. Since 2015 she has been Head of Unit in the Division for Digital Economy Policy, where she coordinates work on privacy, risk management and data governance. She is currently leading the review of the implementation of the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. A related stream of work under her team involves the development of principles for enhancing access and sharing of data. Elettra Ronchi has more than 20 years of experience as policy analyst, evaluating the instruments available to governments to improve the public benefits from investments in health, science and technology. Since 2006 she has also contributed to the organization's work on digital health, including the development of the 2016 OECD Council Recommendation on Health Data Governance. Elettra started her policy career in 1993 as consultant for the United Nations Development Program, before which she held academic research and teaching positions in the U.S. and France.</p>	

Annex C. Participating delegations and organisations

Country/Region Name	Organization
Australia	Commission on Excellence and Innovation in Health
Australia	SA NT DataLink
Australia	University of Melbourne
Australia	University of Western Australia (PHRN)
Belgium	INTEL
Belgium	Ministry public health
Belgium	FPS Health, Food Chain Safety and Environment
Belgium	Intel
Belgium	European Commission
Canada	Public Health Agency of Canada
Canada	Health Canada
Canada	Office of the Privacy Commissioner of Canada
Canada	Innovation, Science and Economic Development Canada
Canada	Public Safety Canada
Chile	Ministerio de Salud
Chile	Consejo para la Transparencia
Colombia	Ministerio de Salud y Protección Social
Colombia	Ministry of Information, Communications Technologies
Colombia	ANDI
Costa Rica	CCSS
Costa Rica	Caja Costarricense de Seguro Social
Costa Rica	Ministerio de Salud
Czechia	Institute of Health Information and Statistics of the Czech Republic
Czechia	MF
Finland	Ministry of Economic Affairs and Employment
Finland	VTT Technical Research Centre of Finland
Finland	The Finnish innovation fund Sitra
France	OECD
France	Business at OECD (BIAC)
France	Health Data Hub

France	DNS, Directeur de projets Expert Recherche & Innovation
France	DNS, Directeur de projets Expert Europe & International
France	ANS, Directrice Adjointe Responsable des Projets Européens et Internationaux
France	ANS Agence du Numérique en Santé
France	ANS, Chargée de Projets Européens
France	Trade Union Advisory Committee to the OECD (TUAC)
France	Public Services International
France	Délégation ministérielle du numérique en santé
France	Permanent Delegation of Japan to OECD
France	Direction de la recherche, des études, de l'évaluation et des statistiques
France	INTESA SANPAOLO
Germany	Federal Ministry of Health
Germany	BMW i
Germany	TMF e. V.
Germany	Caresyntax
Germany	Federal Commissioner for Data Protection and Freedom of Information
Greece	NATIONAL DOCUMENTATION CENTER (EKT)
Hungary	National Directorate General for Hospitals
Hungary	Permanent Delegation of Hungary to OECD
Ireland	Science Foundation of Ireland
Israel	Ministry of Health
Israel	privacy protection authority
Israel	Yale
Israel	Surgical AI
Israel	Start-up Nation Central
Israel	University of Haifa
Israel	HIT
Israel	Novartis
Israel	Geneyx
Israel	Government ICT Authority
Israel	Clalit health services
Israel	Ministry of Justice
Israel	Naschitz Brandes Amir
Israel	Haifa Center for Law and Technology, University of Haifa
Israel	Privacy Protection Council
Israel	Or-Hof Law; Tel Aviv University
Israel	GSK
Israel	Tel Aviv University
Israel	Hebrew University, Jerusalem

Israel	Google
Israel	Ministry of Economy & Industry
Israel	Digital Israel Bureau
Israel	Foreign Trade Administration
Israel	Head of Israel's Economic and Trade Mission to Romania and Ukraine
Israel	Knesset RIC
Italy	Università del Salento
Italy	Italian DPA
Italy	Israel Trade Mission - Embassy of Israel
Italy	Pharma & Biotech Advisors srl
Italy	GVM Care and Research
Italy	Origgi
Italy	ISS
Italy	PIEMONTE AGENCY
Italy	Ospedale Bambino Gesù
Italy	TIM
Japan	MHLW
Japan	Privacy by Design Lab
Japan	Japan National Institute of Public Health
Japan	田辺三菱製薬株式会社
Japan	世界経済フォーラムC4IRJ
Kazakhstan	Ministry of Healthcare of Kazakhstan
Kenya	I-DAIR
Korea, Republic of	OECD
Korea, Republic of	Korea UNIV.
Korea, Republic of	HEALTH INSURANCE REVIEW & ASSESSMENT SERVICE(HIRA)
Korea, Republic of	Korea Internet & Security Agency
Korea, Republic of	Korea National Information Agency
Latvia	Riga Stradins University
Latvia	Centre for Disease Prevention and Control
Latvia	Latvian Hospital association
Latvia	Ministry of Health
Latvia	Riga Stradiņš University
Latvia	Ministry of Health
Lithuania	The Institute of Hygiene
Lithuania	Ministry of Justice of the Republic of Lithuania
Lithuania	State Data Protection Inspectorate
Lithuania	Ministry of Economy and Innovation of the Republic of Lithuania
Mexico	MX-INAI
Mexico	Secretaría de Economía

Mexico	Permanent Delegation of Mexico to the OECD
Mexico	Asociacion Mexicana de Industrias de Investigacion Farmaceutica
Mexico	AMIIF
Netherlands	OECD
Netherlands	Ministry of Health, Welfare and Sport
Netherlands	Philips
Peru	Ministerio de Salud
Poland	Ministry of Health
Portugal	Direção Geral da Saúde
Portugal	ACSS - Administração Central do Sistema de Saúde, IP
Portugal	SPMS
Romania	ANMCS - Romanian National Authority on HealthCare Quality Management
Russian Federation	Federal Research Institute for Health Organization and Informatics of Ministry of Health of the Russian Federation
Slovenia	Ministry of Health
Slovenia	National Institute of Public Health (NIJZ)
Slovenia	Ministry of Health
Spain	Ministerio de Sanidad
Switzerland	WHO
Switzerland	Foraus
Switzerland	EDÖB
Switzerland	Digital Law Center - Faculty of Law - University of Geneva
Switzerland	University of Geneva, QoL Lab
Taiwan	National Academy of Science
Turkey	Turkish Data Protection Authority (KVKK)
Turkey	TCCB Digital Transformation Office
United Kingdom	Federal government Belgium
United Kingdom	Dept. of Health NI
United Kingdom	Department for Digital, Culture, Media and Sport
United Kingdom	ETH Zürich
United Kingdom	UHG
United Kingdom	AGENS
USA	OECD
USA	Alfred B. Sloan Foundation
USA	Intel
USA	TrustArc
USA	Duke University Sanford School of Public Policy
USA	U.S. Chamber of Commerce
USA	Loeb & Loeb
USA	DLA Piper

USA	Pfizer
USA	Hogan Lovells
USA	The Brookings Institution
USA	LiveRamp
USA	IMPACT Institute for the Digital Economy
USA	US Chamber of Commerce / US-Israel Business Council
USA	Israel Trade & Innovation Office, San Francisco
USA	Wonk Policy & Communications, LLC
USA	IQVIA
USA	Mastercard
USA	BigID
USA	Kaleidoscope Health Ventures
USA	BARDA DRIVE
USA	Center for AI and Digital Policy
USA	GlaxoSmithKline
USA	George Washington University
USA	INAI
USA	Microsoft
USA	National Science Foundation
USA	NSF
USA	Johnson & Johnson
USA	FTC
USA	DISCUS
USA	Google
USA	US DHHS
USA	Lumen
USA	DoD
USA	Biotechnology Innovation Organization
USA	US Patent and Trademark Office
USA	Walgreens

Annex D. Summary of the report “Using health data for research: Evolving national policies”

143. The COVID-19 pandemic has brought to the fore the crucial role that data collection, analysis, sharing, and dissemination play for governments, academic institutions, and private sector businesses racing to advance scientific research to help combat the virus. It also illustrates that data protection safeguards are essential to build public trust for the swift adoption of data-based solutions, such as the myriad efforts related to pandemic related research. The interactions between data protection and scientific research are complex, with privacy and data protection enhancing individuals’ trust and ensuring respect of fundamental rights and ethical standards, while at the same time setting parameters and boundaries for data collection and sharing across organizations and borders. Nowhere is this balancing of interests and rights clearer than in the context of secondary use of healthcare data for scientific research. Countries around the world are charting new paths to seek the insights that health data can reveal while at the same time respecting individual rights.

144. Even before the pandemic, public sentiment strongly supported healthcare data-based research. For example, a 2017 survey showed that 93% of Australians support using medical records for research, and more than 95% had moderate, high, or very high trust that researchers would use health data responsibly. The EU General Data Protection Regulation (GDPR) recognizes scientific research as an issue of prime interest to the public. Recital 159 of the GDPR reads, “For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research.” Explicit allowances are made for secondary use of personal data for scientific research in Article 89 of the GDPR as well as Article 5(1)(b) of the regulation, which sets forth the purpose limitation principle, while at the same time providing, “further processing for scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes.”

145. However, concerns about misuses of health data and skepticism about uses for research are also widely documented. Machine learning techniques that are a priority for many researchers require access to vast datasets and raise fears of bias and discrimination. Use of real-world evidence casts a wider net for health research far beyond the traditional data collected for medical care. These factors and others raise the stakes for data research benefits and the risks at an individual and community level.

146. In this survey, the Israel Tech Policy Institute and the Future of Privacy Forum examined the legal frameworks for secondary use of healthcare data, including demographics, diagnoses, symptoms, prescriptions, immunizations, tests and other medical conditions, for research purposes in eight countries, including Australia, England, Finland, France, India, Ireland, Israel and the US.¹⁷ The research demonstrates large commonality across legal systems and regimes, permitting secondary use of healthcare data for research purposes under certain conditions, including review by ethical boards, proper de-identification and additional administrative, technical and contractual safeguards.

17 See the full report here: <https://techpolicy.org.il/wp-content/uploads/2021/02/Using-Health-Data-for-Research-Evolving-National-Policies-FV-.pdf>

147. Legal frameworks typically extend the definition of secondary research to academic, government, and pharmaceutical research, including pharmacovigilance, but excluding uses of data for clinical purposes or for the operation of the healthcare system. These purposes, in turn, are based on other legal mechanisms. The Health Insurance Portability and Accountability Act (HIPAA), for example, defines a sphere of healthcare data uses commonly referred to as Treatment, Payment, and Healthcare Operations (TPO). Under HIPAA, organizations may use patient data for TPO, as opposed to research purposes, even without specific patient consent. Irish law excludes from the obligations that apply to research purposes activities such as service evaluations, clinical audits, and “usual practice” (e.g., investigations of the health of a population and the causes of disease).

148. The legal frameworks governing secondary use of healthcare data for research purposes are complex. They typically include at least three layers of laws and regulations: general privacy and data protection laws, which regard healthcare data as sensitive information subject to strict protections; patient privacy laws; and public health regulation. In certain countries, such as the US and Australia, a federal system adds complexity with state and provincial laws providing additional layers of regulation. Consider the recent California Consumer Privacy Act of 2018 (CCPA), which sets new obligations for health data that is exempt from HIPAA, when conducted by commercial entities. In England, for example, secondary use of healthcare data is governed by a complex legal framework, including common law, statutory law, and government decisions and recommendations from health, ethics, and data protection authorities. And in this context, GDPR provides latitude for national derogations that have led to extensive member state legislation.

149. Several of the examined countries have created or are in the process of creating centralized national hubs for healthcare data. In 2019, Finland, for example, enacted a law explicitly intended to clarify the legal basis for broader secondary use of healthcare data. The law aims to encourage use of healthcare data for socially valuable purposes by enabling data users to have a “one-stop-shop” for access requests to such data. The law also created a new government agency, specifically tasked with collecting and managing healthcare data from numerous disparate sources, securing the data, and efficiently processing access requests from researchers. One key benefit of such a centralized system is the simplification of bureaucratic hurdles. Where researchers once had to wait two years or more to access information, they can now clear the process within three months.

150. France, for example, established another model, whereby the national data protection authority (CNIL) published reference methodologies for those wishing to access the national health data system. Researchers that do not certify to these standards must undergo a CNIL review process before accessing the data.

Consent

151. All legal frameworks explored in this project generally empower patients to authorize healthcare data research with explicit consent. This, however, is a basis that may often be infeasible, especially where data is used for purposes unexpected at the time of collection or for population level research. All countries surveyed therefore also permit data research on other grounds and with additional safeguards, such as ethical reviews, pseudonymisation and security measures. In some countries, to be approved without consent, research must also be deemed to be in the public interest. In others, such as the US, research must contribute to generalizable knowledge. The manner in which these legal grounds and safeguards are structured or prioritized and the extent to which consent is privileged creates variations across countries. And these divergences, in turn, pose limitations for research and scientific collaboration.

152. Under Irish law, for example, a data controller seeking an alternative to consent must apply to a special central committee to assess whether the public interest in conducting health data research “significantly outweighs” the interest in obtaining patient consent. In the US, HIPAA allows several pathways for healthcare data to be used or disclosed for research, including pursuant to a waiver of

consent by a local Institutional Review Board or Privacy Board; for limited purposes “preparatory to research”; as part of a “Limited Data Set” subject to a Data Use Agreement prohibiting re-identification; and more. English law authorizes secondary use of healthcare data if a researcher shows that obtaining patient consent is not practical and that anonymized information cannot be used; the activity has a medical purpose, such as medical research that has received ethics approval from a Research Ethics Committee; and that the activity is in the public interest or in the interests of improving patient care. In Australia, research is permitted where obtaining consent is impracticable; the information is “necessary” for research; the research outcome is relevant to public health or public safety; a Human Research Ethics Committee has approved the project; and sufficient de-identification is deployed.

153. Some countries provide patients with a right to opt out from having their health data analyzed for scientific research. Critics argue that in some cases general opt out rights could unintentionally skew research results injecting bias into public health decision making or into machine-learning algorithms. In England, for example, government operations, including NHS resource allocation and treatment protocols, rely on analyses of patient data made available under data research mandates. The risk is that if opt outs concentrate in certain groups or communities, such analyses may become less accurate and useful, leading to inappropriate treatment and deployment of government services or critical adverse impacts on certain populations.

De-identification

154. Under most data protection frameworks, anonymized data is not subject to legal protections. For data to be considered anonymized, some countries require that re-identification be practically impossible; others accept anonymization if re-identification is unlikely, or risks are remote. France and Finland are guided by stricter de-identification standards, according to the European Data Protection Board (EDPB) guidance under GDPR. In Finland, the law focused on secondary use of health and social data creates a dedicated agency to centralize and manage access. Under Finnish law, only aggregate data is freely available to the public, whereas all other forms of personal data, including pseudonymised data, are available only in a secure user environment and subject to contractual restrictions. The UK has historically taken the position that de-identification risk can be remote, but maybe aligning more closely with EDPB guidance. Under US law, data that is “de-identified” under the HIPAA privacy rule is no longer considered Protected Health Information and is thus not subject to any restrictions. This is the case where health information does not identify – and there is no reasonable basis to believe that it can be used to reidentify – an individual patient. HIPAA provides two methods for de-identification, a Safe Harbor method, whereby an organization removes 18 direct and indirect identifiers and does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual; and an Expert Determination method, whereby an expert with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods of de-identification determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify a data subject.

Ethics

155. Ethical review is generally required for all human subject research, independently from data protection law. Review panels are tasked with protecting the rights, safety, dignity and well-being of research participants. Ethics committee guidance, however, may also set parameters for consent to participate in research or may provide a waiver of consent requirements. If a research project involves further sharing of data for research, the ethics committee may require or waive consent with regard to such data sharing. In the US, the criteria for waivers of consent under HIPAA are that the use or disclosure of healthcare data involves no more than minimal risk to the privacy of individuals and that the research could not practicably be conducted with patient consent or without access to and use of the data. In Australia, a

committee approving the waiver of consent for a research project must be satisfied that several preconditions are met, including that involvement in the research carries no more than low risk to participants; the benefits from the research justify any risks of harm associated with not seeking consent; it is impracticable to obtain consent; there is no known or likely reason for thinking that participants would not have consented if they had been asked; and there is sufficient protection of patient privacy and data confidentiality.

However, it is important to note that the EDPB has advised that the informed consent that ethics committees often require for research data sharing may not always align with the legal basis under data protection law, due to concerns that in a clinical trial relationship data protection consent may not be freely given. Thus, another legal basis other than consent is required under data protection law for access to the same research data.

Additional safeguards

156. Legal frameworks ensure additional safeguards for healthcare data research. In England and Finland, researchers must execute a data-sharing agreement and comply with strict information governance protocols. In the US, in contrast to de-identified data, which is outside the remit of the law, data in a Limited Data Set remains Protected Health Information subject to HIPAA restrictions. While it may be used and disclosed for research without consent, such use or disclosure requires entry into a Data Use Agreement containing specific re-identification obligations and other prohibitions. In Finland, the central data hub allocates access rights in line with the issued permits; monitors access to devices, systems, and office sites; deploys security controls to prevent unauthorized access; and monitors and restricts data communications. It also deploys extensive logging systems that record all data-processing events. In Ireland, processes and procedures relating to the management and conduct of health data research projects include: an assessment of the data protection implications of the research; where an initial assessment indicates a high risk, a data protection impact assessment; measures that demonstrate compliance with GDPR data minimization requirements; controls to limit access to health data to prevent unauthorized consultation, alteration, disclosure or erasure; controls to log whether and by whom health data have been accessed; measures to protect data security; arrangements to anonymize, archive or destroy health data once the research project has been completed; and other technical and organizational measures designed to ensure that the processing is carried out in accordance with the GDPR, together with processes to test and evaluate the effectiveness of such measures.

Cross-border data transfers

157. Data protection frameworks apply generally applicable restrictions to the transfer of health data for research purposes abroad, seeking to continue to guarantee a high level of protection. Because remote access to personal data is considered a transfer under the GDPR, the regulation's restrictions on cross-border data transfers apply to the consultation of national health repositories. Researchers thus face a range of obstacles to participation in cross border projects and may be further limited by the consequences of the Court of Justice of the European Union decision in Schrems II. Some countries apply specific additional transfer limitations to health data. Under draft regulations, Israel, which typically applies liberal data transfer rules, requires localization of health data when used for secondary purposes, including research. India too has introduced draft legislation which would require data localization, in the form of retaining a copy of any dataset in the country if it is transferred abroad. Australia also requires localization of health records made available for research. The US generally does not restrict international transfers of data, but extra-territorial risks to compliance must be considered.

Conclusion

158. The fight against COVID-19 has highlighted the urgency of facilitating the utility of health data for scientific research while protecting individual rights to ensure trust and safety. This report examines the legal frameworks governing secondary use of clinical health data in a number of countries. The legal and ethical frameworks are complex, comprising several layers of privacy, patient rights and public health regulations, at the national and state level. But the commonalities between frameworks outweigh the differences. All countries allow data research for compelling public interests such as public health. This includes repurposing data without patients' consent, subject to protections that include review by ethical committees, de-identification, contractual arrangements, and elevated security obligations as well as general data transfer restrictions and in certain jurisdictions, localization requirements. Significant variations in the manner in which these legal and ethical safeguards are structured and interpreted pose difficulties for research and collaboration. These differences may be bridged by a deeper mutual understanding of the commonalities.

Annex E. Summary of the OECD survey of national health data development and governance

Introduction

159. Health data provide the means to respond to public health challenges, to manage and improve health system performance and health care quality, and to advance scientific discoveries that improve medical treatments and save lives. Health data can fuel a transformation of the health sector toward 21st century treatments and care; but they will not do so without effective health data governance that enables secure and privacy-protective data uses.

160. While the improvement of health information infrastructure and governance has been a policy priority of the OECD for the past decade, the global COVID-19 pandemic brought into sharp focus the need for and value of high quality and timely data to identify high risk patients for testing; to track and trace individuals infected with the SARS-Cov2 virus; to allocate and manage human and physical resources in an emergency; to develop and deploy tools to support diagnosis and treatment; and to contribute to research and evaluation of treatments and vaccines.

161. This annex provides a summary of findings from the OECD Health Care Quality and Outcomes Working Party Survey of Health Data Use and Governance that was administered in 2019/20. Twenty-three countries participated in the survey: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Germany, Estonia, Finland, France, Ireland, Israel, Japan, Korea, Latvia, Luxembourg, the Netherlands, Norway, Singapore, Slovenia, Sweden, United Kingdom (Scotland) and the United States.

162. It is the first survey of health data governance since the implementation of the *OECD Council Recommendation on Health Data Governance in 2017* and it provides insights into strengths and weaknesses in data development, use and governance across countries (Box 1). Please consult the working paper for the full survey results¹⁸.

¹⁸ Oderkirk, J. (2021). Survey Results: National Health Data Infrastructure and Governance. Paris: OECD Publishing. Retrieved from <https://doi.org/10.1787/55d24b5d-en>.

Box 1. OECD Council Recommendation on Health Data Governance

The *OECD Council Recommendation on Health Data Governance* was adopted by the OECD Council on 13 December 2016 and was welcomed by OECD Health Ministers at their meeting in Paris on 17 January 2017. The Recommendation provides policy guidance to:

- Encourage the availability and use of personal health information, to the extent that this enables significant improvements in health, health care quality and performance and, thereby, the development of healthy societies while, at the same time, continuing to promote and protect the fundamental values of privacy and individual liberties;
- Promote the use of personal health data for public policy objectives, while maintaining public trust and confidence that any risks to privacy and security are minimized and appropriately managed; and
- Support greater harmonisation among the health data governance frameworks of Adherents so that more countries can benefit from statistical and research uses of data in which there is a public interest, and so that more countries can participate in multi-country statistical and research projects, while protecting privacy and data security.

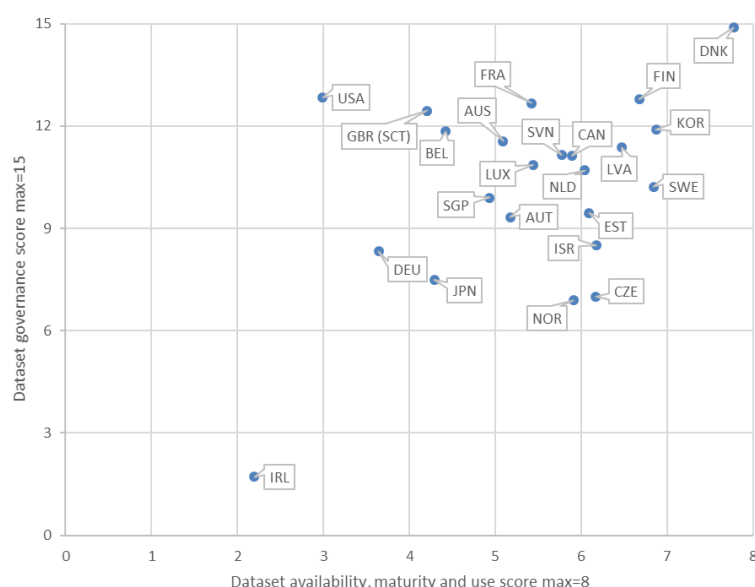
Governments adhering to the *OECD Council Recommendation on Health Data Governance* will establish and implement a national health data governance framework to encourage the availability and use of personal health data to serve health-related public interest purposes while promoting the protection of privacy, personal health data and data security.

The Recommendation sets out twelve key principles for the development and implementation of national health data governance frameworks. The principles encourage greater cross-country harmonisation of data governance frameworks so that more countries can use health data for research, statistics and health care quality improvement.

Source: OECD (2019), Recommendation of the Council on Health Data Governance, OECD Legal Instruments, Paris, <http://legalinstruments.oecd.org>.

163. Overall, the survey identified a small cluster of countries with policies, regulations and practices that foster the development, use, accessibility and sharing of key national health datasets for research and statistical purposes while also having a high degree of recommended health data governance policies and practices in place. Countries reporting the strongest national health data availability, maturity and use and health dataset governance policies and practices were Denmark, Finland, and Korea. These countries were followed by Australia, Canada, France, Latvia, the Netherlands, Slovenia, Sweden and United Kingdom (Scotland) (Figure 1).

Figure 1. Small group of countries score highly on both dataset availability, maturity and use and dataset governance



Note: Dataset governance score is the sum or the proportion of health care datasets meeting 15 dataset governance elements and the dataset availability, maturity and use score is the sum of the proportion of health datasets meeting 7 elements of dataset availability maturity and use. See Tables 1 and 2.

Source: Oderkirk, J. (2021). Survey Results: National Health Data Infrastructure and Governance. Paris: OECD Publishing.

National health data availability, maturity and use

164. Key national health datasets are widely available across the countries surveyed and significant investments are made in health and health care monitoring and research in all countries. Overall, the countries with the strongest indicators of dataset availability, maturity and use in 2019/20 are Denmark, Korea, Sweden, Finland, and Latvia (Table 1).

165. Dataset availability, maturity and use includes eight elements: dataset availability, coverage, automation, timeliness, unique identification, coding, data linkage and regular reporting of indicators of health care quality and system performance. These elements are discussed in this section.

166. The top half of countries tended to report progress in dataset availability, maturity and use since 2013; while the lower half of countries tended to report a drop in capability, with the exception of Japan, which is making clear progress.

167. The OECD has put a priority on supporting countries in measuring quality in health care, strengthening health data governance, developing knowledge-based health systems, and advancing health statistics. Nonetheless, cross-country variability remains significant in 2019/20 and points to challenges not yet overcome (Table 1).

168. The results presented in this report reflect the health data systems in OECD countries just before the onset of the COVID-19 pandemic in March 2020. The pandemic has since heightened governments' attention upon the long-standing gaps in health data and health information systems that we describe here.

169. Eleven countries reported having all or virtually all of the 13 key national health datasets included in this study: Australia, Austria, Denmark, Estonia, France, Korea, the Netherlands, Norway, Singapore, Sweden and the United Kingdom (Scotland). Only two national datasets, however, were available in all

countries: hospital in-patient data and population health survey data. The least available national dataset was a cardiovascular disease registry dataset, which is available in ten countries.

Table 1. Key national health dataset availability, maturity and use

Country	% of key national health datasets available ¹	% of available health care datasets with coverage of 80% or more of the population	% of available health care datasets where data extracted automatically from electronic clinical or administrative records	% of available health datasets where the time between record creation and inclusion in the dataset is one week or less	% of available health datasets sharing the same unique patient ID	% of available health care datasets where standard codes are used for clinical terminology	% of available health datasets used to regularly report on health care quality or health system performance (published indicators)	% of available health datasets regularly linked for research, statistics and/or monitoring (indicators)	Sum
Australia	92%	100%	56%	17%	17%	78%	83%	67%	5.09
Austria	92%	100%	78%	0%	33%	89%	75%	42%	5.17
Belgium	69%	71%	86%	11%	22%	71%	78%	33%	4.42
Canada	85%	75%	75%	0%	64%	100%	91%	100%	5.89
Czech Republic	77%	100%	100%	0%	90%	100%	90%	60%	6.17
Denmark	100%	100%	100%	77%	100%	100%	100%	100%	7.77
Estonia	92%	89%	78%	50%	83%	100%	92%	25%	6.09
Finland	85%	100%	56%	36%	100%	100%	91%	100%	6.67
France	92%	78%	56%	8%	58%	100%	83%	67%	5.42
Germany	31%	100%	33%	0%	0%	100%	100%	0%	3.64
Ireland	77%	86%	29%	0%	0%	29%	0%	0%	2.20
Israel	85%	88%	100%	18%	64%	100%	100%	64%	6.18
Japan	85%	100%	75%	0%	45%	88%	27%	9%	4.29
Korea	92%	89%	89%	58%	100%	100%	92%	67%	6.87
Latvia	77%	88%	63%	80%	80%	100%	90%	70%	6.47
Luxembourg	77%	100%	71%	10%	70%	86%	100%	30%	5.44
Netherlands	92%	70%	100%	0%	75%	100%	83%	83%	6.04
Norway	100%	80%	90%	0%	77%	90%	85%	69%	5.91
Singapore	100%	80%	100%	0%	62%	90%	31%	31%	4.93
Slovenia	77%	100%	100%	0%	70%	100%	70%	60%	5.77
Sweden	92%	100%	100%	8%	92%	100%	100%	92%	6.84
UK (Scotland)	92%	100%	67%	0%	0%	78%	67%	17%	4.20
United States	54%	33%	17%	0%	14%	67%	57%	57%	2.99

Note: The sum column is the sum of the preceding columns and the maximum is 8.

1. Thirteen national datasets including 10 health care datasets (hospital in-patient, mental hospital in-patient, emergency health care, primary care, prescription medicines, cancer, diabetes, cardiovascular disease, mortality and formal long-term care); patient experiences survey, population health survey and population census/registry.

Source: Oderkirk, J. (2021). Survey Results: National Health Data Infrastructure and Governance. Paris: OECD Publishing.

Quality of national health care datasets

170. Focusing on ten national datasets that are directly related to health care¹⁹, this study probed elements of data quality including population coverage, coding of clinical terminology, extraction of data from electronic clinical records, and timeliness.

171. Most countries' datasets cover 100% of the target population; however, there are important gaps in some cases. Data gaps must be closed to have a full understanding of health care provision and outcomes. The most common reason why national datasets are not covering the full population is because they are missing records for care provided by private sector providers and institutions or that are covered by private insurance.

172. Only seven countries (Czech Republic, Denmark, Israel, the Netherlands, Singapore, Slovenia and Sweden) reported that all datasets rely to some extent on data extracted automatically from electronic clinical data and/or electronic insurance claims or billing data. In most countries, available key national health care datasets have some mixture of data entry from paper records and data extracted automatically from electronic records. The benefits of automatic data extraction include improvements in timeliness of data capture, avoidance of costs associated with paper data capture, and minimisation of errors that occur from transcription of information.

173. Thirteen countries reported that for all of their key health care datasets clinical terminology is coded by assigning standard codes using a classification system, such as coding diagnosis to an International Classification of Diseases (ICD) code or coding prescription medicines to an Anatomical Therapeutic Chemical Classification System (ATC) code. Fifteen countries reported that the majority of their available key health care datasets included data that had been coded by a clinician, such as a physician or nurse. Thirteen countries reported that most health care datasets were coded by a health care coding professional. In most countries, both coding professionals and clinicians are doing the work of data coding within national datasets.

174. Reliance on electronic data, as well as upon clinical professionals for coding, have a positive impact upon the timeliness of data within key national datasets. Data that are available in real time or near real time open the possibility of monitoring health care quality and performance in time to detect and address issues as they are emerging, including a rapid detection and response to adverse events. Denmark, Estonia, Korea and Latvia stand out for having a very short time lapse, of one week or less, between when a data record is first created and when it is included in the national dataset used for analysis for all or most key national datasets.

175. For the most part, however, countries are not reporting the use of national health datasets for either clinical or managerial decision-making in real time. The exceptions are Canada, where the inter-RAI tool within long-term care data has algorithms applied to it to alert clinicians in real time to areas of intervention, such as the risk of falls; the Netherlands, where national diabetes registry data are included in a dashboard available to clinicians for clinical decision making; and Sweden, where regional components of the diabetes registry are used for clinical decision-making.

Capacity for national dataset linkages

176. Record linkages enable the information value of individual datasets to grow, permitting connections between health care provided and the outcomes of that care over time; and permitting data within one dataset to be put into context with data from other sources. This study examined a set of

¹⁹ The ten national health care datasets included in this study were hospital in-patient data, mental hospital in-patient data, emergency care data, primary care data, prescription medicines data, cancer registry data, diabetes registry data, CVD registry data, mortality data and long-term care data.

technical factors related to capacity for record linkages including availability of a unique ID that could be used to link datasets, identifying variables that might facilitate linkages, consistency of the use of unique IDs, and the regularity of conducting dataset linkages.

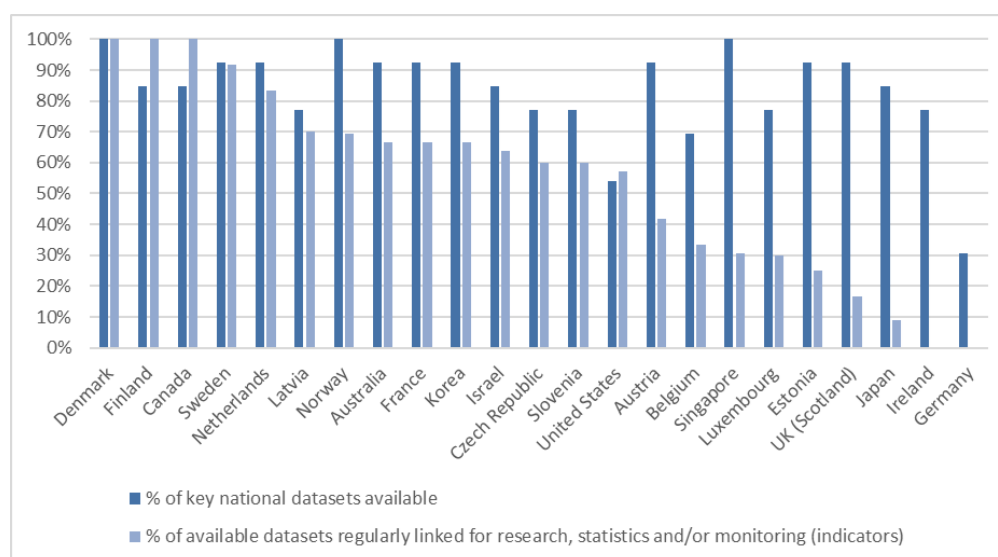
177. Only for four key national health datasets (hospital inpatient data, mental hospital inpatient data, mortality data and cancer registry data) do the majority of countries report that the dataset contains a unique patient ID that could be used for record linkage and that the data are regularly linked for research, statistics or monitoring (indicators). Opportunities to gain additional information value from other key national datasets through record linkages appear to be not pursued in many countries (Figure 2).

178. Seven countries (Czech Republic, Finland, Israel, Korea, Norway, Singapore, and Sweden) report having a unique patient/person identifying number that could be used for record linkage that is included within 90% or more of their national health datasets. Fourteen countries report having the same unique ID number within 60% or more of their national health datasets.

179. Probabilistic data linkages involving matching records on other identifying variables (such as name, sex, birth date, address) could be used for the linkage of the majority of national health datasets in sixteen countries. In only Australia and United States, however, it was possible to link the majority of datasets via these other identifying variables, but not via a unique patient/person ID number.

180. Over half of countries report that dataset linkages are conducted on a regular basis with most of their national health datasets (Australia, Canada, Czech Republic, Denmark, Finland, France, Israel, Korea, Latvia, Netherlands, Norway, Slovenia, Sweden and United States). In contrast, dataset linkages are conducted on a regular basis with a minority of national datasets in Austria, Belgium, Estonia, Japan, Luxembourg, Singapore and United Kingdom (Scotland) and with no national datasets in Germany and Ireland.

Figure 2. Percentage of key national health datasets available and regularly linked for monitoring and research



Source: Oderkirk, J. (2021). Survey Results: National Health Data Infrastructure and Governance. Paris: OECD Publishing.

181. There are indications within this study that different unique ID's are used among national health datasets in some countries. Unless there are ways to match across different ID's, then these differences will prevent the use of these ID's for dataset linkages. Countries challenged with 50% or fewer of national

health datasets sharing a common unique patient/person ID number include Australia, Austria, Belgium, Germany, Ireland, Japan, United Kingdom (Scotland) and United States.

182. A contributing factor to whether or not dataset linkages are conducted regularly is the number of custodians of key national health data sets. Most countries have 3 to 5 different organisations in custody of the thirteen key health datasets studied. However, in Ireland and the Netherlands there are nine different organisations in custody of key national datasets and in France there are seven different organisations. These countries would have considerably higher challenges integrating and linking data across the pathway of care than in other countries, as laws and policies governing health data accessibility and sharing would need to be considered and applied across multiple organisations.

183. Countries regularly linking data also shared insights into the purpose of these linkages that include: improving the quality of national information, such as validating data and filling in information gaps; providing new information about health care quality, outcomes, performance, accessibility and equity; and advancing epidemiological and health services research.

184. Eighteen countries reported regularly linking datasets to monitor health care quality and/or health system performance. Examples of the types of indicators and analysis they undertake on a regular basis with linked data to monitor health care quality and/or health system performance include indicators of mortality at intervals after procedures, treatments or health care episodes; indicators of readmission to hospital; indicators of rates of prescribing medicines; and indicators of survival after diagnosis or treatment.

Challenges developing and analysing national health datasets

185. Eighteen countries reported experiencing challenges developing one or more of their key national health datasets. These included limitations to the quality of the data used to build the datasets, such as limits to the content and quality of insurance claims; gaps in population coverage; lack of unique patient identifiers; exclusion of non-reimbursed services; data coding problems; lack of digitised records; lack of timeliness of data; decreasing survey response rates; and breaks in time series.

186. There were also challenges related to policies and data governance, such as restrictions on analysis; high costs of data access; lack of resources for dataset development; and barriers to dataset linkages, to extraction of data from e-HRs, and to data sharing within a country.

National health data governance

187. The 2019/20 survey measured elements of national health data governance. Overall, 17 of 23 countries reported that a national health data governance framework is established or is being established. Six countries reported that the national health data governance framework is set out in law (Austria, Czech Republic, Denmark, Finland, France, Germany).

188. Through open and public dialogue about potential benefits, risks and risk mitigations it is possible to promote a balanced approach to the governance of personal health data within society. In 2019/20, 14 of 23 countries reported that a public consultation had taken place or was planned about the elements of a national health data governance framework.

189. Most countries reported health data falling under a national health data privacy legislation; other data used in health studies falling under a national privacy legislation; and certain health datasets or health data programmes falling under other legislations governing ministries, data collections or registries. Some countries have legislations at different levels of government. Overall, 21 of 23 countries reported that a national law or regulation exists that speaks to the protection of health information privacy and/or to the protection and use of electronic clinical records.

190. Seventeen countries reported that a central authority for the approval of requests to process personal health data is established or planned.

Challenges experienced in developing national health data governance

191. The survey asked about challenges and difficulties countries were experiencing in the development of health data governance. Virtually all countries reported experiencing one or more data governance or technical challenges at the national level. The most commonly experienced challenges to developing health data infrastructure were legal or policy barriers to public authorities undertaking data linkages (16 countries); concerns with the quality of data that limit their usefulness (15 countries); and legal or policy barriers to sharing data among public authorities (13 countries). A group of countries reported experiencing a high number of health data governance and technical challenges: Belgium, Australia, Ireland, Luxembourg, Germany and Netherlands.

Governance of national health care datasets

192. Considering 15 key dataset governance elements measured in this survey and presented in this section, the countries with the strongest national health care dataset governance were Denmark, United States, Finland, France, UK (Scotland), Korea and Belgium.

193. In most countries all, or nearly all, of the 10 key national health care datasets included in this study are authorised by law. Thirteen countries reported that 100% of their key national health care datasets were authorised by law and four countries reported that 85% or more of these datasets were legally authorised (Table 2). Patient consent is rarely the legal authorisation for national dataset creation in health care.

194. Most countries have a data protection or privacy official within the data custodian's offices for all of their national health care datasets. In Germany and the United Kingdom (Scotland) there was an officer reported for the majority of national health care datasets.

195. The survey asked if the organisations responsible for 10 key national health care datasets provide regular training to staff regarding their responsibilities to protect privacy and data security. Sixteen countries reported that regular training was provided to staff across all of the organisations responsible for key health care datasets. The United Kingdom (Scotland) reported that this was the case for the majority of key health care datasets. Six countries did not confirm that regular staff training is provided for most key national health care datasets.

196. Identities of staff accessing all key national health care datasets are controlled and tracked in 14 countries and for most national health care datasets in 4 countries.

197. In this survey, de-identification alters data so that the resulting data does not identify individuals directly. De-identification requires the removal of name, exact address and details that might support identification. Seventeen countries reported that all of their key health care datasets are de-identified prior to analysis. This is also the case for most health care datasets in four countries. Norway and Ireland did not report that data are de-identified prior to analysis.

198. Fifteen countries reported that pseudonyms are created from direct identifiers on all of their health care datasets and four countries reported this was the case for most of their health care datasets. Reversible pseudonyms facilitate re-identification to conduct future approved data linkages and analysis or to inform an individual of a specific condition or research outcome, where appropriate.

199. Six countries reported that testing is conducted to ensure that realistic re-identification attacks will have a very small probability of success for all or the majority of their national health care datasets. In five countries, such testing is conducted for a minority of datasets. In contrast, in twelve countries, such testing is not reported.

200. Thirteen countries reported that there are practices for the treatment of variables that pose a re-identification risk (such as rare diseases, exact dates, locations, or ethnic origins) for all of their key health care datasets.

Governing health care dataset sharing and access

201. Eight countries indicated that all of their key health care data sets are shared with other public data custodians or government agencies and nine more countries indicated this sharing with the majority of key health care datasets. Some countries reported sharing was possible with one dataset only, which is the cancer registry in Slovenia, long-term care data in Singapore and hospital in-patient data in Germany. In Japan, Ireland and Czech Republic, key national health data sets were not reported as ever shared with other data custodians or public entities.

202. This survey asked countries about whether key national health care datasets, after data de-identification, can be shared for research, statistics or other uses in the public interest and with whom.

203. Analysts employed by a university or non-profit research centre may apply for and be approved access to all de-identified personal health care datasets in 11 countries. In another nine countries, this was true for the majority of health care datasets. In Austria this was true for hospital in-patient, cancer registry, long-term care and mortality data. In the Czech Republic and Ireland, no sharing with university or non-profit researchers was reported.

204. Analysts employed by a health care provider can apply for and be approved access to all de-identified personal health care datasets in 8 countries. In another ten countries this is possible for the majority of health care datasets and in 2 countries for a few datasets. Such sharing was reported in Sweden for the diabetes registry. In Czech Republic and Ireland such sharing was not reported.

205. Analysts employed by a for-profit business could apply for and be approved access to all key de-identified health care datasets in Denmark, Finland, Norway, Singapore and United States. Access to the majority of key de-identified health care datasets for this group was possible in United Kingdom (Scotland), France, Estonia, Canada and Australia. In Austria, de-identified hospital in-patient, cancer registry and mortality data can be shared with analysts employed by a for-profit business. In the Netherlands, de-identified prescription medicines and cardio-vascular disease registry data can be shared with analysts in this group. In Japan and Germany, cancer registry data could be shared with this group. In contrast, 9 countries have no sharing of de-identified personal health care datasets with analysts employed by a for-profit business, even if the data use is within the health-related public interest and is subject to safeguards.

206. Countries were asked under what authority data could be shared with other government entities or external researchers. In Slovenia, Norway, Korea, Germany, Finland and Denmark, the sharing of health care datasets is authorised by law for all key health care datasets. In the United States, the Netherlands, Luxembourg, Japan, France, Canada, Austria and Australia, legislation authorises the data sharing for most key health care datasets that are shared.

207. Far less common is authorising the sharing of health care datasets within government or with external researchers by patient consent. Only in Latvia (8 datasets), United States (5 datasets) and Canada (5 datasets) is patient consent an authorisation for the sharing of a majority of health care datasets. Three key health care datasets in the Netherlands, two in Korea and Estonia, and one in Australia, Austria, Germany, Norway and United Kingdom (Scotland) were authorised to be shared by patient consent.

208. Data subjects were rarely given an opportunity to opt-out of the sharing of key health care datasets. Only France, the Netherlands and United States reported that one-half or more of key health care datasets had an opt-out of data sharing with government entities or external researchers. Opt-out to data sharing with government entities or external researchers was reported for the sharing of diabetes registry and/or

cancer registry data in Sweden, Norway, Luxembourg and Australia. Norway also offers an opt-out to the sharing of long-term care data.

Governing cross-border data exchange and collaboration

209. In this survey, many countries reported recent policy relevant projects involving multiple countries in the linkage of their datasets or in the extraction of data from clinical record systems. These projects included parallel studies, where researchers in each country follow a common study protocol, and studies where data were shared across borders. Examples included studies of prescription drug use and harms between Australia and Canada; between United States and Canada; among Denmark, Finland, Norway and Sweden; and among Australia, United States, Denmark, Finland, Iceland, Norway and Sweden. There are examples of indicator development and research to improve health system performance including projects between Latvia and Slovenia; among Japan, Singapore, China and Hong Kong; and among Finland, Hungary, Italy, Netherlands, Norway, UK (Scotland) and Sweden. There are multiple examples of global and European projects examining cancer incidence and survival; and multiple examples of European projects involving indicator development and research.

210. In some OECD countries, data localisation regimes either explicitly forbid health data processors from approving the sharing of data with an organisation located outside of their country or create obstacles such as a lack of clarity about how data sharing outside of the border might be approved. Existing regimes can also result in processes to obtain approval that would be prohibitive in terms of time and resources. In federated countries, laws and policies within states, provinces or regions may entrench data localisation at a national level.

211. In this survey, countries were asked if de-identified data from key national health datasets may be shared with approved researchers working in foreign academic or non-profit research organisations. Seven countries, Australia, Belgium, Denmark, Finland, Norway, Singapore and Slovenia reported that de-identified data from all health care datasets could be shared for approved research to take place outside of their country. Another six countries reported sharing data outside of their country was possible with the majority of health care datasets. A minority of de-identified health care datasets could be shared cross-border in Austria and Latvia. Eight countries reported that sharing de-identified health care data outside of the country does not occur for any health care dataset.

Data sharing agreements

212. Eighteen countries reported that they have a standard data sharing agreement for disclosing data from all or the majority of their health care datasets. A standard data sharing agreement was reported for one dataset in Ireland. The use of standard data sharing agreements was not reported in Austria, Czech Republic, Estonia, and Norway.

213. Standard data sharing agreements include requirements for certain data privacy and security practices at the data recipient's site for all or the majority of health care datasets shared by sixteen countries (Australia, Belgium, Canada, Denmark, Finland, France, Israel, Japan, Korea, Luxembourg, Netherlands, Singapore, Slovenia, Sweden, United Kingdom (Scotland) and United States).

214. Seven countries reported providing training to data recipients regarding data privacy and security practices when all or the majority of health care datasets are shared (Austria, Denmark, Finland, France, Korea, United Kingdom (Scotland) and United States).

215. The survey asked whether data sharing agreements include penalties that would occur if the required data privacy and security practices were not respected. Penalties were reported as included in standard data sharing agreements in Australia, Canada, Estonia, Finland, Germany, Japan, Korea, Latvia, Luxembourg, Netherlands, Singapore, Sweden, United Kingdom (Scotland) and United States.

Data transfers to approved applicants and alternatives – remote data access and research data centres

216. Data transfers are a particular area of risk for health data custodians as data can be lost or stolen in the transfer process. Countries were asked if they transfer data to approved applicants, such as by sending a copy of a dataset. This method is used for transferring all health care datasets to approved applicants in eleven countries and for transferring some key health care datasets in another nine countries.

217. Fourteen countries described secure portals/file transfer protocols to transfer data over the internet (Australia, Belgium, Canada, Denmark, Finland, Germany, Ireland, Israel, Korea, Luxembourg, Netherlands, Singapore, United Kingdom (Scotland) and United States).

218. Overall, 10 countries out of 23 provided secure access to all or most de-identified national health datasets via remote data access, a research data centre or both (Austria, Denmark, France, Korea, Luxembourg, Netherlands, Singapore, Slovenia, United Kingdom (Scotland) and United States).

219. Remote data access is a service providing access to data stored on a computer or network from a remote distance. Remote data access services are often secured to ensure that users can only access data to which they have been approved and that users cannot alter or withdraw/copy the data from the system without permission.

220. Six countries provide access to all or most key health care datasets to external approved applicants via a remote data access facility: United Kingdom (Scotland), Luxembourg, Korea, France, Denmark and Austria. Remote data access to a minority of datasets was reported in Sweden, Belgium, Germany, the Netherlands, Finland and Australia.

221. A research data centre is a secure physical setting, such as a secure room, where access is provided to data. Research data centres may have physical security, such as supervision and locked doors, as well as computer and data security, such as computer systems that ensure users can only access data to which they have been approved and that users cannot alter or withdraw/copy data from the system without permission.

222. A research data centre is provided for all or most health care datasets in seven countries: Denmark, Israel, Korea, Singapore, Slovenia, United Kingdom (Scotland) and United States. A minority of health care datasets are available via a research data centre in nine countries: Australia, Austria, the Netherlands, Canada, Sweden, Germany, Finland, Belgium and Luxembourg.

Transparency through public information

223. Twenty countries reported that for all or most key health care datasets there is a publicly available description of the dataset purpose and content and most provided a web-link to this public information. Singapore reported that a public description was available for two datasets; and Ireland reported this for one dataset.

224. Seventeen countries reported that the description of all or most health care datasets includes the health-related public interests served by the data and seventeen countries reported that the description for all or most datasets includes the legal basis for the processing. The procedure to request access to the data and the criteria used to approve access to the data are publicly available for all or most health care datasets in seventeen countries.

225. Fourteen countries reported that the procedure to request a record linkage or other further processing of all or most health care datasets and the criteria used to approve these requests are publicly available. When asked if there is a summary of approval decisions for the record linkage or further processing of the datasets that is publicly available, 10 countries answered yes for all or most key health care datasets. When asked whether the summary describes or identifies the data recipient of an approved

record linkage or further processing of the datasets, only Denmark, France, Israel, Japan, Korea, Latvia, Sweden, and United Kingdom (Scotland) said yes for all or most health care datasets.

Next steps

226. In 2022, the OECD will report on countries progress toward the adoption of the *OECD Recommendation on Health Data Governance*. The results of this survey will provide an important input toward monitoring progress; as will findings from a 2021 OECD survey investigating developments toward and barriers to standardised and interoperable electronic clinical records and the use of data within records for statistics and research. Further input to the report to the OECD Council in 2022 will be progress and recommendations toward strengthening data privacy and security protections for personal data undertaken by the Working Party on Data Security and Privacy of the OECD Committee on Digital Economy Policy.

Table 2. Proportion of key national health care datasets with key data governance elements

Country	Legislation authorises datasets	Data privacy/data protection officer	Staff are trained in data protection	Staff data access controls	Data de-identified prior to analysis	Testing re-identification attack risk	Data shared with public sector	Data shared with academic/non-profit sector	Data shared with for-profit sector	Data shared cross-border	Standard data sharing agreement	Either remote data access service or research data centre	Public description of dataset	Description includes legal basis for the dataset	Procedure to request and approval criteria for data linkage	sum
Australia	67%	100%	100%	100%	100%	0%	78%	89%	89%	100%	78%	33%	100%	22%	100%	11.56
Austria	100%	100%	100%	100%	100%	0%	89%	44%	33%	33%	0%	78%	78%	78%	0%	9.33
Belgium	100%	100%	100%	100%	100%	43%	57%	100%	0%	100%	29%	100%	100%	100%	57%	11.86
Canada	25%	100%	100%	100%	100%	0%	88%	88%	75%	75%	88%	25%	88%	75%	88%	11.13
Czech Republic	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	100%	100%	0%	7.00
Denmark	100%	100%	100%	100%	100%	90%	100%	100%	100%	100%	100%	100%	100%	100%	100%	14.90
Estonia	100%	100%	33%	100%	100%	0%	100%	89%	89%	89%	0%	0%	100%	33%	11%	9.44
Finland	100%	100%	100%	67%	100%	0%	100%	100%	100%	100%	100%	11%	100%	100%	100%	12.78
France	78%	100%	100%	100%	100%	78%	100%	100%	78%	78%	67%	67%	89%	67%	67%	12.67
Germany	67%	67%	33%	33%	67%	33%	33%	67%	33%	67%	67%	33%	100%	100%	33%	8.00

any				%			%		%	%						33
Ireland	100%	14%	14%	0%	0%	0%	0%	0%	0%	0%	14%	0%	14%	0%	14%	1.71
Israel	88%	100%	0%	0%	100%	0%	88%	88%	0%	0%	100%	63%	63%	63%	100%	8.50
Japan	100%	100%	13%	75%	88%	0%	0%	88%	13%	0%	75%	0%	100%	13%	88%	7.50
Korea	100%	100%	100%	100%	100%	89%	89%	89%	0%	0%	78%	78%	89%	89%	89%	11.89
Latvia	100%	100%	100%	100%	100%	0%	100%	100%	0%	38%	100%	0%	100%	100%	100%	11.38
Luxembourg	100%	100%	100%	100%	100%	14%	100%	100%	0%	57%	100%	86%	57%	57%	14%	10.86
Netherlands	80%	100%	100%	60%	100%	20%	70%	80%	20%	60%	100%	50%	100%	70%	60%	10.70
Norway	90%	0%	0%	0%	0%	0%	100%	100%	100%	100%	0%	0%	100%	100%	0%	6.90
Singapore	40%	100%	100%	100%	70%	80%	100%	100%	100%	100%	50%	100%	20%	20%	0%	9.90
Slovenia	100%	100%	100%	100%	100%	0%	14%	100%	0%	100%	100%	100%	100%	100%	0%	11.14
Sweden	89%	100%	100%	11%	100%	11%	100%	100%	0%	0%	100%	11%	100%	100%	100%	10.22
UK (Scotland)	89%	89%	89%	89%	89%	89%	89%	89%	89%	0%	89%	89%	89%	89%	89%	12.44
United States	100%	100%	100%	100%	100%	100%	83%	100%	100%	0%	100%	100%	100%	17%	83%	12.83

Note: The sum is the addition of the preceding columns and the maximum sum is 15.

Source: Oderkirk, J. (2021). Survey Results: National Health Data Infrastructure and Governance. Paris: OECD Publishing.