Unlocking the Power of Health Data by Ensuring the Public Can Trust the EHDS

Dipak Kalra[a]

[a]**The European Institute for Innovation through Health Data, Ghent, Belgium**

DOI: 10.25929/wgxbmf28

ABSTRACT

It is vitally important to assure public trust during the European scaling up of health data reuse for research, public health and health systems strategy. The European Health Data Space (EHDS) proposals are catalysing the multi-stakeholder debate on this, and include some provisions that are aimed at giving that assurance, such as strict purposes for which data may be reused, and which may not be used. However, the public needs a way of trusting that the organisations that reuse their data are bound to by ethical and transparency principles. Furthermore, these organisations are to be held accountable for how they use data. This paper summarises the Big Data rationale for needing a European scale for the secondary use of health data, and the provisions in the proposal for a Regulation for the EHDS to enable and safeguard this secondary use. It proposes a complementary measure, a Societal Compact, that could strengthen public trust in the EHDS.

KEYWORDS

Health data, electronic health records, big data, information governance, transparency

1. The need for European scale real-world data

Health systems right across the developed world are challenged to meet current healthcare demands and public expectations of quality and timeliness. They are working within funding constraints following the depletion of public budgets due to the COVID-19 pandemic, in the face of an aging society and increasing multi-morbidity which increases treatment complexity and costs. [1] In parallel, there is a growing recognition of the importance of equity of access to health services and equity of outcomes. [2] Health systems need to make structural reforms and deliver better value, for which digital and data transformation is needed. [3]

Healthcare providers need to make better use of their patient-level data for monitoring the health outcomes they achieve, streamlining care pathways and interconnecting multi-professional care teams, delivering remote telehealth and more monitoring, and implementing more personalised care. Health systems need to transform on the basis of evidence, and public health systems need greater access to patient-level and population-level data for bio-surveillance and rapid outbreak management. Research increasingly needs large-scale data for the development of targeted therapies, medical devices and AI algorithms. [4,5]

2. The growth of European health data infrastructures

Over the past several years Europe has seen substantial investment in Big Data infrastructures, for example at national levels in Germany [6] and France, [7] and in particular health sub-sectors such as for medicines regulators, [8] life sciences research [9] and for rare diseases. [10] Some of these utilise a federated analysis architecture, in which a research computable query is cascaded across multiple repositories and only the analysis results are returned, which avoids centralising large volumes of fine-grained patient data. This provides a more safeguarded way of undertaking large-scale analysis, whilst conforming to the European General Data Protection Regulation (GDPR). [11]

The conduct of Big Health Data research is challenged when it comes to interpreting the GDPR, because different jurisdictions and sometimes even different data protection officers will arrive at different interpretations of the acceptability and/or required safeguards when conducting Big Data research. This includes the requirement for patient consent, which is nearly impossible to collect in a fully informed way when constructing a Big Data resource such as a research infrastructure, because the multitude of future research purposes cannot be explained accurately enough to comply with GDPR informed consent. Anonymisation of data is one method for conducting Big Data research, but is challenging to apply in a robust manner when it comes to rare disease groups where patient numbers are small, to genetic information, images, etc. Pseudonymised data, which are often necessary for longitudinally linked datasets, are still considered personal data under the GDPR and require a legal basis as well as stringent safeguards.

3. Proposals for a European Health Data Space (EHDS)

All eyes in Europe are now on an ambitious programme announced by the European Commission, with an allocated budget of many billions of Euro: the European Health Data Space (EHDS) [12] It is expected that the regulation to enact this pan-European channel for data sharing will be passed during 2024, and be implemented over the next several years across all 27 member states. The European Parliament and Council has debated this proposal over several months, and reached a final text that has strengthened many of its provisions, [13] which will now be finalised, passed and enacted.

For many people this is an exciting proposal. Firstly it will establish a patient-level communications ecosystem for identified patient data (containing a patient summary, recent prescriptions, test results and hospital discharge reports) to be accessible wherever a patient needs urgent healthcare across Europe. This will make their care safer and enable their original clinical team to know what treatments were issued whilst they were cared for abroad. Patients will also be able to access their health summary and recent correspondence, and it is likely that innovative apps will enable patients to understand and use their data for better self-care, for example for chronic diseases and prevention. A summary of the main provisions for patient-level data (primary use of data) is shown in Figure 1, reflecting the text agreed in March 2024, although it may still be subject to final changes.

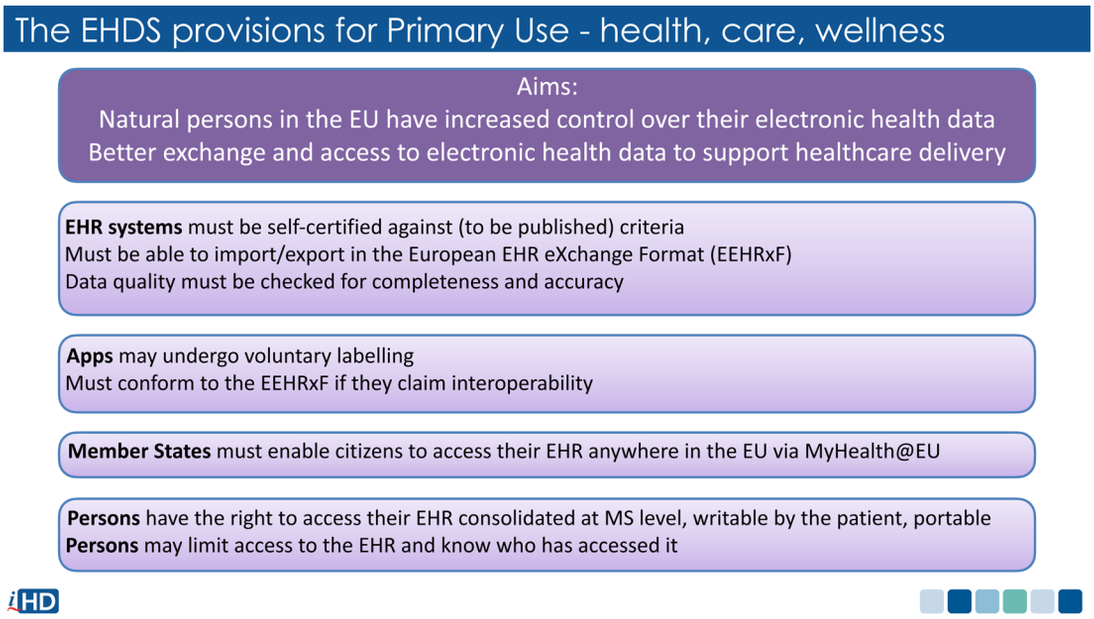

Figure 1: Summary of the provisions in the provisional political agreement on a European Health Data Space (March 2024) relating to the primary use of health data.

The provisions relating to the secondary use of data are quite distinct from those relating to primary use. It is generally considered that the prioritisation given to the health data space ahead of other sectoral data spaces is at least in part driven by a wish to strengthen Europe's capability to determine appropriate public policy and real-time evidence-based public health strategy in the case of another pandemic or a similar large-scale threat to the health of European citizens. The secondary use provisions cover the analysis of data sets for the purposes of public health, research, education, innovation, policy, regulation and the development of personalised medicine, but with a strong emphasis on enabling the access to data by public authorities. This is enabled through a governance mechanism to access data sets that will be curated and provided through national Health Data Access Bodies (HDABs) in all 27 member states. This part of the regulation places an obligation on health data holders (almost all kinds of public and private institutions) that have any of a wide range of categories of health data to document the datasets they hold in a transparent way, with standardised FAIR [14] metadata and optionally a data quality assessment (to be developed) through a centralised catalogue to be maintained by each HDAB. This catalogue will enable these data sets to be discovered, and to request from their HDAB a permit to access these data. The EHDS Regulation will include a list of permitted purposes and prohibited purposes that the data user must agree to adhere to.

HDABs will be the ultimate arbiters of whether to grant a data access permit or not. They will then require provision of the requested data from the data holders, and will seemingly be the arbiters of any safeguards or restrictions that are needed to protect intellectual property or other commercial sensitivities regarding the released data. HDABs will establish secure processing environments for using the data and require that the results of data use are openly published within 18 months.

A summary of these provisions is given in Figure 2.

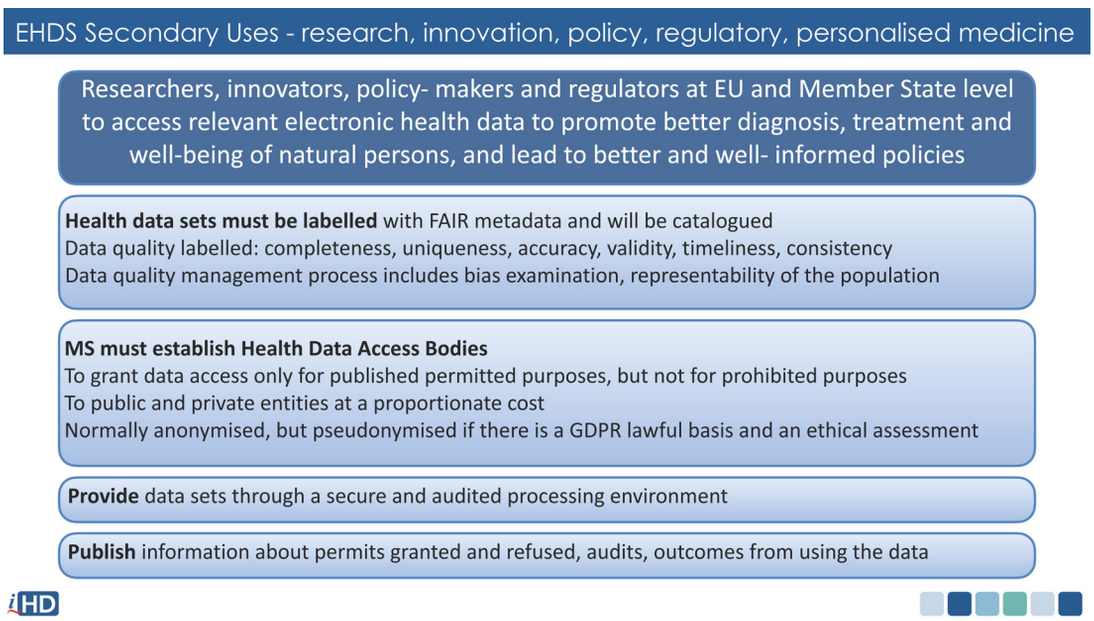

Figure 2: Summary of the provisions in the provisional political agreement on a European Health Data Space (March 2024) relating to the secondary use of health data.

There is debate amongst research organisations about the powers that will be bestowed on HDABs, and whether the requirements and mechanisms for data sharing will actually help or hinder innovation within Europe.

The regulation proposes that every member state puts in place a process for its citizens to opt out of secondary uses of data (other than by public authorities when responding to a public health crisis, which cannot be opted out of). It is not clear, and it may vary between countries, whether citizens might be enabled to opt out of the secondary use of certain categories of their data (such as genetic data), to opt out of certain purposes of secondary use (such as research) or to opt out of secondary used by certain types of organisation.

4. The challenge of winning public trust

Whilst there are many supporters of this right to citizen control and self-determination with regard to the use of personal information, there are concerns that public confidence in the secondary use of data may be vulnerable to misinformation campaigns and lead to an opt-out by particular sectors of society. This introduces a risk of important bias in the data which are used for research, such as for the development of AI, and may lead to solutions that are less accurate in those populations, placing that at a future health or safety risk.

What will encourage individuals to not opt out? The author proposes that one important factor will be a way for the public and data access decision-makers to determine who to trust to use their health data, and why. Reciprocally, bona fide research organisations need a way of demonstrating they are trustworthy data users.

5. Proposal for a societal compact

In late 2022 and early 2023, two not-for-profit organisations, which included the author's institute, convened multi-stakeholder round tables and further consultations to consider this issue, and to formulate a "proposal for a societal compact for the secondary use of health data". [15] It is offered as a voluntary agreement between a range of stakeholders who cooperate to achieve social benefits by granting access to and reusing health data. The compact aims to provide an assurance to all stakeholders in the health data ecosystem, especially to the public, that the organisations and individuals who sign it will reuse health data in legal, ethical and secure ways and that they will use the data in society's interests.

This proposal comprises a set of ethical and data usage undertakings that any data user should a priori commit to before being granted data access, and against which they should be held accountable and audited, and be sanctioned, if necessary. The ethical principles are summarised in Figure 3 below. The compact requires data users to only use data for a list of permitted purposes, and never for the listed prohibited purposes; these lists are aligned with the draft EHDS Regulation and will be updated when the Regulation is finalised. In addition to these principles, data users are required to adopt a number of stipulated measures to safeguard the data they are processing, including legal compliance and adopting data protection and information security practices, many of which align with the EU GDPR. It emphasises transparency, requiring data use of organisations to have a public location (such as a website) where they list and briefly summarise the accesses to data they have been granted and the purposes for which they are using it, later to be updated with a summary of the learning acquired or the product or service implemented. The compact proposal includes an operational and governance workflow, including the basis for applying sanctions. Some of these measures echo those in the recent Council draft of the EHDS Regulation, suggesting that this compact could fit well as a complementary measure.

Wider consultation is now taking place with stakeholders, globally, about this compact, and some bodies are considering if it might form the basis of self-regulation within their sector.

1. Reuse data to contribute and bring benefits to society in terms of improved opportunities for better health and care.
2. Never reuse data unethically, or by violating human rights, discriminating against individuals or groups of individuals, or to further individual or organisational interests exclusively without bringing benefits to society.
3. Reuse by always safeguarding the privacy of individuals whose data are being reused and apply the principle of data minimisation.
4. Reuse data respectfully toward data holders and adhere to data use terms agreed with the data holders.
5. Results should be published or shared in some way unless publishing such results violates Principle 2.
6. Organisations reusing data must make every effort to be transparent to the public about their use of health data and the outcomes of each data use, complying with the H2O data use commitments.

Figure 2: Summary of the ethical principles in the societal compact proposal.

6. Conflict of interest statement

The author declares that there is no conflict of interests regarding the publication of this paper.

References and notes

[1] The Organisation for Economic Co-operation and Development. *Fiscal Sustainability of Health Systems: How to Finance More Resilient Health Systems When Money Is Tight?*, OECD Publishing, Paris, 2024. https://doi.org/10.1787/880f3195-en. Accessed June 12, 2024..

[2] The European Commission. *Joint Report on Health Care and Long-Term Care Systems & Fiscal Sustainability.* The European Commission Institutional Paper 37, October 2016. https://economy-finance.ec.europa.eu/publications/joint-report-health-care-and-long-term-care-systems-fiscal-sustainability_en. Accessed June 12, 2024

[3] European Alliance for Value in Health. *Health systems after COVID-19 – Building resilience through a value-based approach.* August 2021. https://www.europeanallianceforvalueinhealth.eu/library/health-systems-after-covid-19-building-resilience-through-a-value-based-approach. Accessed June 12, 2024..

[4] Kalra D. Scaling up the big health data ecosystem: engaging all stakeholders. *International Journal of Telemedicine and Applications*, eHealth 2020;8:e16. https://doi.org/10.29086/JISfTeH.8.e16. Accessed June 12, 2024.

[5] Hughes N, Kalra D. Data Standards and Platform Interoperability. In: He, W., Fang, Y., Wang, H. (eds) *Real-World Evidence in Medical Product Development.* Springer, Cham, Switzerland. 2023. https://doi.org/10.1007/978-3-031-26328-6_6. Accessed June 12, 2024.

[6] Medical Informatics Initiative Germany. Strengthening research and advancing healthcare. https://www.medizininformatik-initiative.de/en/start. Accessed June 12, 2024.

[7] French Health Data Hub. https://www.health-data-hub.fr (in French). English FAQ at https://www.health-data-hub.fr/page/faq-english. Accessed June 12, 2024.

[8] European Medicines Agency. Data Analysis and Real World Interrogation Network (DARWIN EU®). https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-and-real-world-interrogation-network-darwin-eu. Accessed June 12, 2024.

[9] ELIXIR. https://elixir-europe.org. Accessed June 12, 2024.

[10] European Reference Networks. https://health.ec.europa.eu/european-reference-networks_en. Accessed June 12, 2024.

[11] The European General Data Protection Regulation. What is GDPR, the EU's new data protection law?, May 2018. https://gdpr.eu/what-is-gdpr. Accessed June 12, 2024.

[12] European Health Data Space. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en. Accessed June 12, 2024.

[13] European Parliament News. Announcement of the provisional political agreement on a European Health Data Space (EHDS). 15th March 2024. https://www.europarl.europa.eu/news/en/press-room/20240304IPR18765/eu-health-data-space-to-support-patients-and-research. Accessed June 12, 2024.

[14] Wilkinson M, Dumontier M, Aalbersberg I, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Nature Scientific Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18. Accessed June 12, 2024.

[15] The Digital Health Society and the European Institute for Innovation through Health Data. Proposal for a societal compact for the secondary use of health data. https://echalliance.com/news/round-table-6-proposal-for-a-societal-compact-for-the-secondary-use-of-health-data. Accessed June 12, 2024.