AI Needs High-Quality Health Data at Scale - Will the EHDS Deliver?*

Dipak Kalra,[a] Eva Sabajova,[b] Birgit Bauer,[c] Dmitry Etin,[d] Henrique Martins[e]

[a]The European Institute for Innovation through Health Data, Ghent, Belgium
[b]Digital Health Slovakia
[c]european digital health academy edha gGmbH, Abensberg, Germany
[d]Deggendorf Institute of Technology (DIT)/European Campus Rottal-Inn (ECRI), Pfarrkirchen, Germany
[e]ISCTE-IUL, Lisbon, Portugal

DOI: 10.25929/d5p2-np32

ABSTRACT

A panel discussion was held during the DigiHealthDayS-2024 Scientific Congress on 15th November 2025. It explored the potential for the European Health Data Space (EHDS) to become a high-value resource of good quality data for the development of AI innovations to support better and safer healthcare. Panel members discussed whether the measures presently intended for the provision of secondary used data sets within the EHDS and AI Act will be suitable for AI development. At present, a standardized data quality label will be defined, but its use will be optional, interoperability standards for data sets sharing for research are not mandated. However, AI innovators need access to large-scale, high-fidelity data sets that have well-documented provenance and quality, and which are accurately representative of the populations on whom the innovators wish to target their solutions. The EHDS has the potential to accelerate the availability of high-quality data sets, but the adoption of the data quality and utility label to assess the data sets must be strongly promoted and accompanied by measures and incentives for health systems to actively improve the quality of the data they routinely collect within EHR systems.

KEYWORDS

Artificial intelligence, data bias, health data quality, European Health Data Space

1. Introduction

A panel discussion was held during DigiHealthDayS-2024 Scientific Congress on 15th November 2024, comprising the authors of this paper. It explored the potential for the European Health Data Space (EHDS) to become a high-value resource of good quality data for the development of responsible AI innovations to support better and safer healthcare. However, its provisions for the secondary use of data, in particular the limited obligations that it places on the data holders of data sets regarding the structural and semantic representation of the data (the use of interoperability standards) and the optional nature of the data quality and utility label may lead to the proliferation of data sets that are difficult to use and that

---

INTERACTIVE DISCUSSION – AI Needs High-Quality Health Data at Scale: Will the EHDS deliver? Hosts: Eva Sabajova (Digital Health Slovakia) & Prof. Dipak Kalra (i-HD, UK). Speakers: Birgit Bauer (DSL Germany), Dmitry Etin (Austria), Prof. Henrique Martins (ISCTE, Portugal).

are of low quality. There is a risk that the EHDS will not contribute successfully to the development of safe and trustworthy AI, also in conjunction with the newly established AI Act.

2. Key Considerations

AI innovators need access to large-scale, high-fidelity data sets that have well-documented provenance and quality and which are accurately representative of the populations on whom the innovators wish to target their solutions.

AI developers also need to assure their end users, potential purchasers and national assessment bodies that they have utilized high-quality data that is traceable, as free from bias as possible, and that they have followed robust European ethical principles.

The EHDS is the European Union's flagship program for scaling up the availability of health data, for supporting continuity of patient care across borders and for providing the wide public and private stakeholder access to the vast array of data sets held across Europe. These data sets exist in healthcare provider organisations, research organizations, registries, public health agencies and other sources, which will be catalogued and made accessible for permitted purposes which include the development of AI algorithms for healthcare.

The regulation endorses the FAIR principles (Findable, Accessible, Interoperable, and Reusable). This should mean that data sets are discoverable through Member State and European catalogues, and that the use of interoperability standards (if any) is declared. However, none are mandated and whilst providing transparency about standards used this does not foster consistency and might not have any impact on the data harmonization workload and risk of error when data sets need to be combined to achieve a sufficient scale for AI learning. Similarly, the EU AI Act is regulating relevant quality management, data governance requirements, which could be fulfilled using EHDS mechanisms, if implemented well.

The panel members felt that a data provenance, data quality label is important for data sets that are published via catalogues maintained by Health Data Access Bodies (HDABs) – the designated agencies to regulate EHDS implementation in EU Member States – to provide transparency on the composition, origin and quality for potential data users such as AI developers. However, this transparency should lead to minimum standards and the promotion of better-quality data at its origin, which is usually healthcare provider EHR systems, but increasingly patient-generated data from wearables and wellness apps increasing the need to use EHDS provisions to also promote interoperability from these sources. Common EU efforts in semantics and data catalogues need to be reinforced and may benefit from new governance structures under the EHDS regulation and should be seen in light of primary use, secondary use of healthcare data including for AI applications.

The EHDS has the potential to drive progress also in Member States with less developed health data systems by mandating minimum standards for data quality, interoperability, and governance. It motivates these countries to adopt international standards and invest in modern digital health infrastructure.

Additionally, the panel agreed that especially patients and also citizens must be informed and educated in the use of their health data and also about the EHDS. It is essential that participation is a key element to create meaningful and ethical guidelines and frameworks for the use of health data in general and also especially for the secondary use of health data.

3. Future Directions

Panel members felt that a more proactive approach including mandating a limited subset of interoperability, common data model standards, provenance (with only limited grounds for exceptions) and the mandatory adoption of the data quality label should be enforced through forthcoming implementing acts.

Healthcare providers need support to better encourage and incentivize a culture of high-quality data capture by (admittedly busy) clinicians, and that the quality of the EHR systems (especially user interface design) needs to facilitate high-quality data entry and prompt users to correct clear-cut data entry errors in real time.

Citizens must be informed transparently and be involved in the development of tools, standards and frameworks.

3. Conclusion

Panel members were all enthusiastic about the EHDS and its importance for European research including AI development and adoption of the EU AI Act. However, more needs to be done to ensure that AI innovation in Europe can benefit from it.