

### Challenges and enablers for implementation

The quality of the data that flows into the EHDS will determine the value of the insights generated and health solutions developed through its secondary use. Use of incomplete or unrepresentative datasets for research, innovation and policymaking alike carries the risk of introducing bias and leading to discriminatory outcomes for EU citizens. To aggregate data volumes on the scale made possible by the EHDS, however, the data needs to be interoperable not just from the technical perspective of its transferability across different systems, but also in terms of the language and coding systems used by different organisations and health systems to document patient information. While international data standards exist, their application varies drastically across the Member States and between different healthcare settings such as hospitals, community medicine or nursing care. A common data quality framework will need to be implemented consistently in each country to ensure that all Member States can participate and benefit equally from the secondary use of their data going forward. Extensive work has already been done at EU level in this area to inform the implementation process.

Still, many questions remain, starting with whether and how to approach the standardisation of vast amounts legacy electronic data held by the different health systems. While this data relates predominantly to older citizens who would stand to benefit significantly from its secondary use, the current reality is that health institutions have particularly low capacity and budgets for the data quality improvement measures this would require. In addition, indiscriminately integrating all data in the relevant categories would not necessarily

be useful or desirable: many experts involved in the roundtable series highlighted the need to define much more precisely the data requirements and to tie these to concrete use purposes to ensure an efficient, goal-oriented approach to data-sharing. This applies equally to data from medical devices, the utility of which for different research applications has yet to be determined. Its inclusion in the EHDS also poses its own challenges as regulatory data quality requirements differ across different categories of devices and applications—an issue which is even more acute for the data that could come from wellness apps, which are not currently subject to any standard evaluation.

Considering the large cost associated with obtaining high-quality data for research, standardisation and improvement measures on the side of data holders, and potentially patients themselves, will need to be met with efforts on the side of secondary users to manage inevitable variations in data quality. Work in this area will span from educating and supporting researchers in the appropriate use of different types of datasets, to developing standard methods to scientifically validate the datasets themselves as well as the algorithms and digital health innovation they are fed into.

**Disclaimer:** This text reflects only the point of view of the experts and stakeholders involved in the Think Tank roundtable series and consultation processes held in 2023 based on the European Commission's first legislative proposal

## Solutions should aim to

► Foster a common understanding of data quality throughout the EU

► Enforce an inclusive system of quality control throughout the EU

► Focus on the reuse of primary data

► Support appropriate use of different datasets

## Key actors, findings and solutions for implementation

### At European level

Establish a common language, a common (logical) data model, a common semantic and a unified way of collecting metadata to improve system and data interoperability in consultation with relevant stakeholders.

Further define EU-wide standards and concepts of data quality and data utility in the EHDS legislation delegated acts, building on the data equity and FAIR principles.

Agree on common European health record standards in line with EU and international best practices, in consultation with industry and research actors.

Establish guidelines for robust data traceability across primary and secondary use systems, as well as data robustness and access across all society.

Define an efficient EHR certification system that companies of all sizes can cope with, including user-centricity and the standardisation of terms and coding systems (vocabularies) as evaluation criteria for health information systems' data quality

### At national and regional levels

Adopt and implement a common data quality framework (like the one created under the Joint Action TEHDAS) at national level to ensure a harmonised European approach to data quality.

Adopt a coordinated strategy to educate clinicians on data standards to optimise the research potential of the data they record. Similar strategies should be adopted for researchers on data collection and analysis applied to healthcare and policymaking.

Establish legal obligations and financial support for the adoption and maintenance of EU and international data collection standards.

### Healthcare providers

Ensure the consistent capture of metadata in the primary use setting.

Contribute to developing standard approaches to improving primary data quality that are compatible with routine work processes.

Leverage new technologies such as natural language processing to expedite the standardisation of legacy free text data in EHRs.

Ensure pseudonymisation (by anonymisation of all personal identification information) of data at the data collection point to allow context-specific pseudonymisation and controlled use of privacy-preserving record linkage services when legal and ethical conditions are met

### Health data access bodies

Establish data traceability across primary and secondary use platforms, especially with regard to where, how and by whom data was collected to help data users better understand and manage variability in data quality within the EHDS and ensure appropriate use of different datasets.

### Higher education providers

Educate secondary users of health data on the suitability of EHRs and other primary data sources for different types of research.

Develop and validate data quality testing methods and algorithms, as applied in health research.

### ► EXAMPLE

In France, the University Hospital of Nantes, Nantes University, INSERM and École Centrale Nantes have developed the Avatar method to completely anonymise data and therefore avoid the risk of re-identification. To ensure privacy, Avatar meets three criteria of the European Data Protection Committee. The first is individualisation. Avatars datasets are individualizable, so you cannot isolate an individual. The second criterion is correlation: Avatar prevents two datasets about the same group of individuals from being linked. The last point is the inference. This process blocks the ability to infer sensitive information about an individual. This data can then be used to train AI models, contribute to research and statistical studies in medicine and many other fields.