Proposal to supplement and improve the EDPB guidelines on the anonymisation of health data for clinical research purposes

Introduction

In light of developments in data processing technologies and the growing importance of clinical research based on large volumes of health data, it is considered appropriate to propose an update to the European Data Protection Board's (EDPB) guidelines on anonymisation, with particular reference to the needs of biomedical and health research.

The aim is to strengthen the protection of data subjects' rights and freedoms, whilst at the same time promoting the use of health data in a manner that is scientifically sound, proportionate and consistent with the European regulatory framework on the protection of personal data.

Health data have specific characteristics that make assessing anonymity particularly complex. The presence of rare information, the possibility of linking different databases, the growing availability of genomic and longitudinal data, and the use of advanced artificial intelligence techniques may make it possible – even in the absence of direct identifiers – to re-identify the individuals to whom the data relate.

In this context, it seems appropriate to supplement the general principles of the EDPB with operational guidelines specifically tailored to health research, capable of supporting research organisations, healthcare facilities, clinical trial sponsors and data protection officers in the assessment and management of risks.

Proposals for improvement

1. Development of specific criteria for the anonymisation of health data

It is proposed that a technical annex dedicated to health data be drawn up, setting out differentiated assessment criteria based on the type of information processed, the purpose of the research and the context of use.

This annex should consider, in particular, longitudinal clinical data, information relating to rare diseases, genetic and genomic data, diagnostic images, data from wearable devices and data derived from the integration of electronic health records.

For each category, it would be appropriate to identify the main risk factors, the applicable transformation techniques and the methods for verifying the effective reduction in the risk of identification.

2. Strengthening re-identification tests

It is proposed to clarify how to apply, in the healthcare context, the three assessment criteria referred to by the EDPB: singling out, linkability and inference.

In particular, the assessment should include documented tests designed to verify:

- the possibility of identifying an individual through combinations of clinical characteristics;
- the possibility of linking the dataset to other accessible information sources, including public or scientific databases;
- the possibility of inferring sensitive information not directly present in the dataset;
- the ability of artificial intelligence techniques to identify correlations or reconstruct identifying information.

The tests should be proportionate to the risk, updated periodically and documented in such a way as to allow for independent verification of the conclusions reached.

3. Joint assessment of privacy risk and scientific utility

A further area for action concerns the need to systematically consider the impact of anonymisation techniques on the quality and reliability of scientific results.

The removal or transformation of certain variables may, in fact, compromise the analysis of rare diseases, the assessment of infrequent clinical events, the identification of patient subgroups and the reconstruction of a disease's progression over time.

It is therefore proposed to introduce a documented procedure that jointly assesses the residual risk of identification and the loss of scientifically relevant information. This procedure should make it possible to compare different technical solutions, giving priority to those that ensure adequate protection without unduly compromising the purposes of the research.

This assessment should not be interpreted as a balancing exercise intended to justify an excessive risk of identification, but rather as a tool for identifying, amongst the alternatives that are adequate from a data protection perspective, the one most compatible with scientific requirements.

4. Guidance on the re-use of data for further research purposes

It is proposed to introduce specific guidance for the re-evaluation of datasets intended for new studies or for scientific purposes beyond those initially envisaged.

The assessment should take into account the new purposes, any additional information sources available, developments in analytical technologies and any changes to the conditions of access to the data.

Anonymity, in fact, should not be regarded as an absolute and immutable property of the dataset, independent of the context in which it is used. The possibility of linking information from different archives or applying new analytical techniques may alter the previously assessed risk.

It is therefore recommended that criteria be established to determine when it is necessary to repeat re-identification tests and update the technical documentation.

5. Promoting secure environments for health research

Where anonymisation results in an excessive loss of scientifically necessary information, it is proposed to promote the use of secure data processing environments, in which access is controlled and operations are subject to appropriate technical and organisational measures.

Such environments may include federated research infrastructures, controlled processing systems and solutions that allow data to be analysed without directly transferring individual information to researchers.

It should be noted that the use of such tools does not, in itself, result in the anonymisation of the data processed. Where data remain personal, the relevant provisions of Regulation (EU) 2016/679 (GDPR) must continue to apply, including those relating to the legal basis, special categories of data, security of processing and, where necessary, the data protection impact assessment.

6. Harmonisation of procedures and documentation

Finally, it is proposed to draw up a standardised template for anonymisation documentation, which can be used by hospitals, universities, research institutes and clinical trial sponsors.

The template should include at least the following:

- a description of the dataset and the purposes of the processing;
- the anonymisation techniques applied and the reasons for their selection;
- assumptions regarding the parties who might attempt re-identification and the means they might reasonably use;
- the results of isolation, linkability and inference tests;
- an assessment of the impact on scientific quality;
- complementary security measures and conditions of access;
- the procedures and timeframe for reviewing the assessment.

Harmonised documentation would facilitate the comparability of assessments, collaboration between research centres and the transparency of decisions taken.

Final proposal

In light of the considerations set out above, it is proposed that the EDPB assess the advisability of supplementing its guidance on anonymisation with an implementation document dedicated to health research, accompanied by practical examples, risk assessment criteria and documentation templates.

This addition should pursue three main objectives: to ensure a more consistent application of anonymisation criteria in the healthcare sector; to strengthen the ability to identify re-identification risks arising from data integration and new technologies; and to preserve, as far as possible, the scientific value of the information processed.

The proposal does not aim to lower the level of protection required by European legislation, nor to automatically regard data that has undergone processing techniques as anonymous. Rather, it aims to promote a more practical and verifiable methodological approach, capable of distinguishing genuine anonymisation from other risk mitigation measures.

In conclusion, the development of sector-specific guidelines and shared operational tools could help to strengthen legal certainty for operators, foster scientific cooperation and support the development of European clinical research, whilst ensuring that technological innovation proceeds in a manner consistent with the protection of individuals' fundamental rights.

Parma,

Dr. Cesare Costi, JD
ccosti@ao.pr.it
cesare.costi@unipr.it

Italian has been adopted as the original language of this document