OHDSI Italy – National Node

Observational Health Data Sciences and Informatics – European Chapter

Response to the Public Consultation on

EDPB Guidelines 1/2026 on Processing of Personal Data for Scientific Research Purposes

Submitted by: Prof Lucia Sacchi, on behalf of the OHDSI Italy National Node

Lead Institution: University of Pavia, Italy

June 2026

1. About OHDSI Italy

OHDSI Italy is the national node of the Observational Health Data Sciences and Informatics (OHDSI) network in Italy, operating within the OHDSI Europe Chapter. Established in 2022 and coordinated by the University of Pavia, the Node brings together 39 active organisations including academic institutions, teaching hospitals, IRCCS research centres, and SMEs, representing over 23 data partners with real-world health data assets harmonised to the OMOP Common Data Model (CDM) [https://www.ohdsi-europe.org/national-nodes/italy].

The Node actively participates in multi-site federated studies under the OHDSI network, including the FALCON-Bladder study (37 sites, 16 countries), the EHDEN Mega-Study on Drug Shortages (published in The Lancet Public Health, 2025), and the Italian OMOP_SEPSIS machine learning prediction study. It contributes to European governance frameworks through participation in TEHDAS2 project and has developed a shared DPIA template for OMOP CDM adoption distributed across Italian data partners.

OHDSI Italy's aggregated infrastructure encompasses longitudinal real-world health data on several million patients, standardised under OMOP CDM v5.4, and operated in compliance with GDPR, Italian data protection legislation, and EHDS-aligned governance frameworks. The Node submits these comments in its capacity as a federated real-world data research network with direct operational experience of the issues addressed in the Guidelines.

2. General Observations

OHDSI Italy welcomes the EDPB's initiative to issue dedicated guidance on the processing of personal data for scientific research purposes. These Guidelines address a genuinely complex area of GDPR application, and the effort to provide concrete examples and operational clarity is appreciated.

As a federated real-world data network operating across multiple Italian institutions and participating in European multi-site studies, OHDSI Italy has direct operational experience of the interpretive gaps and practical ambiguities that these Guidelines are intended to resolve. Our comments focus on the sections most directly relevant to distributed health data research infrastructures, specifically: the scope of 'scientific research' (§2), research data infrastructures (§2.2), presumption of purpose compatibility (§3.1), storage limitation (§3.2), lawfulness of secondary use of health data (§4, in particular §4.4.3), attribution of responsibility in multi-institutional settings (§7.3), and appropriate safeguards for federated processing (§8).

3. Specific Comments

3.1 Section 2 – Scope of 'Scientific Research'

The six key-indicative factors proposed in §2.1 provide a useful operational framework for qualifying activities as scientific research under the GDPR. OHDSI Italy supports this approach, which aligns with how the OHDSI network has historically defined its research activities: methodologically grounded, ethically overseen, reproducible, and oriented toward generating generalisable knowledge that benefits public health.

However, we observe that two aspects of the framework require clarification in the context of distributed observational research.

First, factor (i) (methodical and systematic approach) references 'formulation and testing of a hypothesis' as the paradigmatic form of scientific inquiry. In large-scale federated observational research using OMOP CDM, which often begins with exploratory data characterisation, phenotype development, or population-level descriptive analyses, the research value is real but may not be framed around a pre-specified hypothesis. The Guidelines acknowledge that 'exploratory or preliminary' research may have 'a stated objective as opposed to a hypothesis', which is welcomed. We recommend that this clarification be made more prominent, and that the Guidelines explicitly state that characterisation studies, cohort generation studies, and data quality assessments conducted within a methodologically standardised framework qualify as scientific research even absent a confirmatory hypothesis.

Second, factor (iv) (autonomy and independence) requires that 'researchers processing the personal data have academic (e.g. a PhD title) or scientific qualifications'. In multi-institutional federated networks, the data processing pipeline typically involves data engineers, informaticians, and ETL (extract-transform-load) specialists who operate under the scientific direction of qualified researchers but might not themselves hold academic titles. Consequently, these technical professionals must be formally authorized to directly access and examine the original source data, as direct interaction with the raw datasets is an indispensable prerequisite for engineering highly effective, accurate, and high-quality ETL pipelines. We recommend that the Guidelines clarify that the qualification requirement applies to the scientific leadership of the research project, not to every individual involved in technical data processing operations, consistently with the controller/processor distinction developed in Section 7.

3.2 Section 2.2 – Research Data Infrastructures

The Guidelines' recognition in §2.2 that the processing of personal data within research data infrastructures may be motivated by scientific research purposes, and that the key-indicative factors apply to the 'relevant processing operations that are necessary for operating the research infrastructure' rather than to individual downstream studies, is a critically important clarification.

OHDSI Italy operates precisely such an infrastructure: a federated network of OMOP CDM-standardised data warehouses, each maintained at the level of individual data partner institutions, accessible for approved research projects without patient-level data transfer. The Guidelines' Example 4 (public agency making pseudonymised data available for approved research projects) closely mirrors this model, and we are encouraged that the EDPB endorses it.

We nonetheless identify the following gaps that require further guidance.

First, the Guidelines do not address the situation, common in OHDSI-type federated networks, where the 'infrastructure operator' role is distributed rather than centralised. In the OMOP/OHDSI model, there is no single entity that 'holds' all the data; instead, each data partner retains data locally and participates in federated queries. The entity that coordinates the network (e.g. the Node lead institution) does not access patient-level data, but only aggregated results. The Guidelines should provide guidance on whether and how the infrastructure-level scientific research qualification applies to this distributed topology, and how responsibilities are allocated among network coordinator, data partners, and individual study coordinators.

Second, the Guidelines are silent on the role of common data model standardisation, including ETL processes, concept mapping (e.g. local drug codes to RxNorm, diagnoses to SNOMED-CT), and data quality assessment, as ancillary processing operations qualifying under §2.3. These operations are resource-intensive, require expert judgment, and are essential preconditions for federated research. We recommend that the Guidelines explicitly recognise that data harmonisation to a common data model constitutes an ancillary processing operation motivated by scientific research purposes, provided it is conducted within an infrastructure governed by the factors in §2.1.

3.3 Legal Basis for Secondary Use of Routinely Collected Clinical Data: The Consent Paradox in Federated Research Networks

This section addresses what OHDSI Italy considers the most critical gap in the current Guidelines, and the single most significant barrier to lawful, large-scale observational health data research in Italy and several other Member States. We urge the EDPB to address it with specific, unambiguous guidance in the final version of the Guidelines.

3.3.1 The Problem: Consent as the Default Legal Basis for Research in Italy

The GDPR establishes a sophisticated framework for the secondary use of health data for scientific research. Article 5(1)(b) presumes purpose compatibility; Articles 6(1)(e) and 6(1)(f) provide legal bases grounded in public interest and legitimate interest; Article 9(2)(j) provides a derogation for special categories of data for scientific research; and Article 89(1) establishes the safeguard architecture that makes these provisions operable. Read together, the GDPR clearly envisages that large-scale observational research on routinely collected health data can and should be conducted without requiring individual patient consent for each research project, provided appropriate safeguards are in place. In practice, however, OHDSI Italy member institutions, all operating within the Italian legal framework, face a structural situation in which individual patient consent has historically been treated as the most legally certain pathway for the research use of routinely collected clinical data. This is not because Italian researchers misunderstand the GDPR; rather, it reflects the way in which the GDPR's research provisions have been implemented and operationalised in the Italian legal framework and in institutional practice, resulting in a system that has often made the secondary research use of health data more difficult to pursue in a clear and consistent manner.

The specific Italian legal situation is as follows. Article 110 of the Italian Data Protection Code permits the processing of health data for scientific research without consent where the research is carried out on the basis of Union or Member State law, including certain biomedical or health research programmes, or where, due to specific reasons, informing data subjects or obtaining consent is impossible, involves disproportionate effort, or risks making the research impossible or seriously impairing its objectives. Following an amendment in 2024, such cases no longer require prior consultation with the Italian Data Protection Authority (Garante della Privacy). Controllers must instead comply with the safeguards identified by the Data Protection

Athority, including a favourable reasoned opinion of the competent ethics committee, appropriate measures to protect data subjects' rights and freedoms, documentation of the reasons why consent or individual information could not be obtained, and a data protection impact assessment to be made public, or published in extract form where justified.
In addition, the Italian framework now provides a distinct pathway for research hospitals institutions (IRCCS). The Garante has clarified that public and private IRCCS may rely on Article 110-bis(4) of the Italian Data Protection Code for the use, for scientific research purposes, of personal data originally collected in the context of healthcare. According to this provision, research processing by IRCCS does not constitute further processing incompatible with the original clinical purpose. Where IRCCS rely on this pathway, they must carry out a data protection impact assessment and publish it on their website, in full or in extract form where justified.
Earlier Italian Data Protection Authority authorisations and related deontological rules provided part of the previous operational framework for scientific research, but the current Italian position must now be read primarily considering the amended Article 110 and the safeguards identified by the Garante in 2024.
The result is that Italian hospital research units, when wishing to use routinely collected clinical data for observational studies, are frequently advised by their Data Protection Officers and legal counsel to obtain individual patient consent. This may also occur in retrospective studies on data already collected for care, in studies using pseudonymised OMOP CDM data, and in participation in multi-site federated studies where no patient-level data ever leaves the institution.

The operational impact is strong:

- Retrospective studies covering historical patient populations may be significantly constrained, as re-contacting patients can be impracticable or impossible particularly where patients can no longer be reached or have died. Prospective studies may require systematic consent collection at the point of care, creating operational burdens for clinical staff and selection biases that may affect robustness and representativeness of research results.
- Participation in international OHDSI federated studies, where the scientific protocol is defined centrally and individual sites contribute aggregate results only, requires each Italian data partner to independently navigate a consent or derogation pathway that other European partners do not face.
- The resulting fragmentation contributes to the underrepresentation of Italian institutions in European real-world evidence studies, to the detriment of Italian patients whose health data could inform clinical guidelines and regulatory decisions.

3.3.2 Why the Current Guidelines Do Not Resolve This Problem

The Guidelines confirm (§3.1.1) that the presumption of purpose compatibility under Article 5(1)(b) GDPR applies to further processing for scientific research, and that no separate compatibility test under Article 6(4) is required. This is welcome. However, the Guidelines also state (§21–22) that the controller must still verify that the legal basis applicable to the initial processing is suitable for the further processing and acknowledge that this 'applies in particular if the initial legal basis for the primary processing is consent or a legal obligation'.

This formulation, while legally accurate, risks being read by Italian institutions and their DPOs as confirming the status quo: because clinical data in Italy is collected on the basis of Article 6(1)(c) GDPR (legal obligation under healthcare law) combined with Article 9(2)(h) GDPR (healthcare provision), and because the Guidelines indicate that legal obligation as a basis may not straightforwardly extend to research use, the practical reading will continue to be: obtain consent.

Moreover, the Guidelines do not address the upstream problem: the fragmented national legal framework under Article 9(2)(j) GDPR. While the EDPB cannot directly modify national substantive law, it might issue a clear interpretive signal encouraging Member States to assess whether their national implementing legislation under Article 9(2)(j) GDPR genuinely enables secondary research use of health data, consistently with the GDPR's research-enabling objectives.

3.3.3 Recommended Guidance

We would suggest the EDPB to include the following clarifications in the final Guidelines:

- **Clarification 1:** Explicit statement that Article 6(1)(e) GDPR (public interest) constitutes a legally valid and appropriate basis for the secondary research use of routinely collected clinical data by institutions, including teaching hospitals and IRCCS research centres which, by virtue of their nature as public bodies, pursue general research objectives in the public interest, as set out in legislation (e.g. D.Lgs. 517/1999 and corresponding regional laws), without requiring a separate national legislative act beyond that statutory framework. The EDPB is also invited to clarify the relationship where relevant under Union or Member State law, between Article 9(2)(g) and Article 9(2)(j) GDPR in public-interest health research settings.
- **Clarification 2:** Explicit statement that the transition from Article 9(2)(h) GDPR (healthcare provision) to Article 9(2)(j) GDPR (scientific research) for special categories of data constitutes a permissible further processing under the presumption of purpose compatibility in Article 5(1)(b), provided appropriate safeguards under Article 89(1) are adopted, and that this transition does not require a new national legislative act or individual consent where the institution operates under an existing statutory research mandate.
- **Clarification 3:** A recommendation to Member States that national implementing legislation under Article 9(2)(j) GDPR be designed to genuinely enable secondary research use of health data, and that Member States are invited to assess whether frameworks that effectively require individual consent for all research use, in the absence of prior authorisation from the supervisory authority, achieve an appropriate balance between data protection and the GDPR's research-enabling objectives under Article 89.
- **Clarification 4:** Guidance that, for OMOP CDM-based federated research where no patient-level data is transferred outside the institution and only genuinely aggregate results (counts and summary statistics meeting minimum size thresholds, such that recipients cannot reasonably identify individual data subjects, consistently with CJEU T-557/20) are shared, the processing is lawful under Article 6(1)(e) and Article 9(2)(j) GDPR on the basis of the institution's statutory research mandate and the safeguards required by Article 89(1), without requiring individual patient consent.

These clarifications would not create new rights or obligations under the GDPR; they would make explicit what the GDPR's research provisions already permit. Their inclusion in the final Guidelines would provide Italian (and other) DPOs with the authoritative EDPB interpretation needed to move away from consent as the de facto default for observational research on routinely collected health data. The above clarifications should in no way be read as limiting the obligation of data controllers to adequately inform data subjects about the processing of their personal data in accordance with Articles 12–14 GDPR. Making information on the processing of personal data available, including through publication and layered notice mechanisms, remains an essential requirement for the data controller as a means of ensuring

transparency and enabling the effective exercise of data subjects' rights (see further Section 3.8 below).

3.4 Section 3.2: Storage Limitation

The Guidelines' recognition (§28–29) that storage beyond the fulfilment of the original research purpose is permissible where the data is intended for future research — including in infrastructure contexts where the specific future projects are not yet defined, provided the research area is specified — is operationally important for longitudinal data warehouses.

OHDSI Italy notes that the requirement to 'substantiate how it intends to use the personal data in future scientific research projects' (§29) may in practice be difficult to apply to infrastructure-level data holdings, where the range of plausible future studies is broad and legitimately so. We recommend that the Guidelines clarify that for research data infrastructures operating under an established governance framework (ethics oversight, access controls, DPIA, data use agreements), the substantiation requirement is satisfied at the infrastructure governance level rather than requiring project-by-project anticipation of future use. This is consistent with the approach taken in Example 4 and with the logic of broad consent under §4.1.2.1.

3.5 Section 4.4.3 – Derogations Under National Law (Article 9(2)(j) GDPR)

Section 4.4.3 acknowledges that Union and Member State laws may provide derogations from the prohibition on processing special categories of data for scientific research. The Italian example illustrates how national implementation under Article 9(2)(j) GDPR may remain operationally fragmented even where research pathways do exist.

Following the 2024 amendment of Article 110 of the Italian Data Protection Code, the processing of health data for scientific research without consent no longer requires prior consultation with the Garante in cases where informing data subjects or obtaining consent is impossible, involves disproportionate effort, or risks making the research impossible or seriously impairing its objectives. Instead, controllers must comply with the safeguards identified by the Garante, including a favourable reasoned opinion of the competent ethics committee, appropriate measures to protect data subjects' rights and freedoms, documentation of the reasons why consent or individual information could not be obtained, and a data protection impact assessment to be made public, or published in extract form where justified.

In parallel, the Garante has clarified that public and private IRCCS may rely on Article 110-bis(4) of the Italian Data Protection Code for the use, for scientific research purposes, of personal data originally collected in the context of healthcare. Where IRCCS rely on this pathway, they must carry out a data protection impact assessment and publish it on their website, in full or in extract form where justified.

Accordingly, the Italian issue is not best characterised as the absence of any derogation, but rather as the coexistence of differentiated pathways: a clearer statutory route for IRCCS under Article 110-bis(4), and a more case-specific pathway for other healthcare institutions under Article 110, consent where feasible, or other applicable Union or Member State legal provisions. This may create uncertainty for federated OMOP CDM-based research, since the legal pathway can vary depending on the category of participating institution, even where patient-level data remain local and only aggregate results are shared.

In multi-site studies spanning multiple Member States, each data partner institution must therefore assess the applicability of its relevant national legal pathway independently, and the

requirements attached to those pathways, including ethics review, transparency measures, DPIA obligations and specific pseudonymisation safeguards, may vary significantly.

We recommend that the Guidelines: (a) acknowledge that, in federated multi-site research, the legal assessment under Article 9(2)(j) GDPR is necessarily conducted independently by each data partner under its applicable national law; (b) encourage Member States to adopt clear, operational and proportionate national rules that facilitate federated research across institutions through standardised safeguard requirements; and (c) clarify that the future application of Article 53(1)(e) of the EHDS Regulation (Regulation (EU) 2025/327) may contribute to reducing this fragmentation.

3.6 Section 7.3 – Joint Controllers in Multi-Institutional Research

The guidance on joint controllers in §7.3 is highly relevant to federated research consortia such as OHDSI Italy. The Guidelines confirm (§141) that entities participating jointly in drafting a research protocol may be considered joint controllers even if actual data processing is conducted only by some of them.

OHDSI Italy has direct experience with the complexity of this determination. In network studies, the study coordinator drafts the protocol and analysis code, data partners execute the analysis locally on their own data, and aggregate results are returned. The Guidelines' Example 25 (non-interventional observational clinical study with a research consortium) usefully addresses a scenario close to this model.

We recommend two clarifications:

First, the Guidelines should address the situation where participation in a multi-site OHDSI study consists of executing a standardised, pre-specified analysis package (e.g. an ATLAS-generated study package) without any active participation in the determination of research purposes or protocol design. In this case, the data partner institution is executing pre-defined analytical instructions rather than determining purposes or means, and we submit that this resembles a processor relationship more than joint controllership. The Guidelines should provide clearer criteria for distinguishing between joint controllership (active participation in protocol determination) and processor-type relationships (execution of externally defined analytical instructions) in the specific context of standardised federated network studies.

Second, the requirement under Article 26(1) GDPR to document joint controller responsibilities in a formal agreement is operationally burdensome at network scale, where a single coordinating institution may run dozens of studies involving varying subsets of a network of 20+ data partners. We recommend that the Guidelines acknowledge that a framework agreement at network level — covering standard terms for joint controller responsibilities across multiple studies — is a valid and proportionate compliance mechanism, provided it is supplemented by study-specific addenda covering study-specific purposes and data flows.

3.7 Section 8: Appropriate Safeguards

3.7.1 Pseudonymisation in OMOP CDM Environments

In this regard, the EDPB's own Guidelines on Pseudonymisation (January 2025) provide directly relevant guidance on pseudonymisation techniques in health research contexts that should be read in conjunction with the present Guidelines. We draw attention to a specific operational challenge arising in OMOP CDM-based infrastructures: the OMOP CDM specification assigns a system-generated integer identifier (person_id) to each patient, replacing direct identifiers. This constitutes a form of pseudonymisation. However, in multi-site federated studies, the person_id values are institution-specific and not shared across data

partners. The aggregate results shared with the study coordinator contain no patient-level identifiers.

This architecture provides strong re-identification protection at the federated results level, but the local OMOP CDM itself remains pseudonymised (not anonymised) data under the GDPR. The Guidelines should acknowledge this architecture explicitly and confirm that operating a local OMOP CDM with institution-generated person_ids, sharing only aggregated results with study coordinators, constitutes an effective pseudonymisation safeguard under Article 89(1) GDPR, provided appropriate access controls and data use agreements are in place.

We also note the specific challenge arising from temporal precision in OMOP CDM. Date fields in OMOP CDM (e.g. condition_start_date, drug_exposure_start_date) include day-level precision, which may constitute a re-identification risk for rare conditions or rare events. Some national implementations apply date suppression or date shifting techniques. The Guidelines should address the legitimacy of partial date suppression as a proportionate pseudonymisation measure under Article 89(1) GDPR and confirm that its application does not require separate legal justification beyond the general pseudonymisation obligation. Furthermore, the EDPB is invited to clarify under which conditions personal data may be pseudonymised, as a preparatory step for future scientific research within a standing research infrastructure, even where no specific study is yet underway. The Guidelines could also usefully clarify, where relevant under Union or Member State law, how such preparatory processing should be assessed in relation to the applicable legal basis, the relevant derogation under Article 9 GDPR, and the safeguards required by Article 89(1). This clarification is of direct operational importance for institutions operating OMOP CDM data warehouses as standing research infrastructures.

3.7.2 Federated Databases and Secure Processing Environments

Section 8.5 lists as an appropriate safeguard ‘setting up a secure processing environment or an access point through which researchers can gain access to research data without any need for local storage’, and ‘rules and measures when providing personal data to other researchers, including … use of federated databases’. OHDSI Italy strongly endorses this recognition.

We recommend that the Guidelines provide further operational guidance on the characteristics that a federated database architecture must have to qualify as an appropriate safeguard, specifically: (a) no transfer of patient-level data to the coordinating institution; (b) local execution of analytical queries at each data partner site; (c) sharing of only aggregate results (cell counts, summary statistics, model coefficients) that meet minimum cell size thresholds; (d) governance controls on which studies may access which data partners; and (e) audit logging of all access and query executions. These characteristics, which define the OHDSI distributed network model, should be explicitly recognised as satisfying the secure processing environment safeguard.

3.7.3 DPIA for Multi-Site Federated Research

The Guidelines (§8.2) correctly identify the DPIA as the starting point for adopting appropriate safeguards. In the context of federated research networks, a question arises as to whether each participating institution must conduct a separate DPIA for each study, or whether a network-level DPIA supplemented by study-specific risk assessments constitutes an adequate and proportionate compliance approach.

OHDSI Italy has developed a shared DPIA template for OMOP CDM adoption that has been distributed across Italian data partners. We recommend that the Guidelines explicitly acknowledge that a network-level or infrastructure-level DPIA framework, developed under the responsibility of the network coordinating institution and adopted by individual data partners through a formal governance mechanism, constitutes an appropriate compliance approach for federated research infrastructures, consistently with the joint controller framework under Article 26 GDPR.

Furthermore, to ensure proportionality and prevent administrative overburdening, when study-specific DPIAs are required for research projects leveraging an already assessed OMOP-based infrastructure, they should not replicate the full technical and organizational assessment. Instead, by explicitly referencing the foundational infrastructure-level DPIA, these study-specific assessments should focus exclusively on unique project elements—such as specific research objectives, the precise dataset, potential variations in the legal basis, and study-specific risks—highlighting only the necessary additional safeguards

3.8 Section 5 – Transparency Obligations: Practical Challenges in Secondary Use

The Guidelines' treatment of transparency obligations (§5) is detailed and addresses most scenarios. We draw attention to one practical challenge specific to OMOP CDM-based secondary use of routinely collected clinical data: the obligation to inform data subjects under Article 14 GDPR when data originally collected for care is subsequently used for research.

In a hospital with 500,000+ patient records, direct individual notification for each new research study is operationally infeasible and, under Article 14(5)(b) GDPR, may not be required where it would involve disproportionate effort. The Guidelines' provisions on indirect notification (§89) are helpful but need to be operationalised more concretely for hospital data warehouse contexts.

We recommend that the Guidelines provide specific guidance on a layered transparency approach for hospital-based OMOP CDM research: (a) a general, standing transparency notice at institutional level covering the secondary use of clinical data for scientific research; (b) a study-level notice published on a dedicated webpage for each approved research study; and (c) individual notification only where feasible without disproportionate effort. This layered approach, if explicitly endorsed by the EDPB, would provide a defensible compliance pathway for institutions that operate large-scale research data warehouses.

3.9 Section 2 and Section 8 – AI Model Training and Validation as Scientific Research in the Public Interest

The EDPB Guidelines should explicitly clarify in Section 2 that the use of personal data for the training and validation of AI and machine learning models may qualify as scientific research under the GDPR, including where such activities retain an exploratory character. Such clarification is particularly important considering the broad interpretation of scientific research reflected in Recital 159 GDPR and the examples provided in §2.1, which encompass technological development and innovation carried out for societal benefit.

A clear recognition of these activities as scientific research would allow controllers to rely on the GDPR provisions specifically designed to facilitate research in the public interest, including Articles 9(2)(j), 5(1)(b), and 89.

Without such clarification, there is a risk that AI model development carried out for public-interest research purposes may be incorrectly characterised as technical or commercial development activities, rather than as scientific research. This would undermine the GDPR's objective of fostering scientific progress while ensuring a high level of protection for

fundamental rights and freedoms. Legal certainty is therefore essential to enable the responsible development of innovative data-driven methodologies that contribute to public health and scientific advancement.

Furthermore, the Guidelines should recognise in **Section 8** that modern privacy-enhancing technologies (PETs) can constitute appropriate safeguards within the meaning of Article 89(1) GDPR. In particular, federated learning architectures, as mentioned among the examples of safeguards in **§8.5**, can significantly reduce the need for transfers of patient-level information, as personal data remain under the control of the originating institution and only model updates or parameters, rather than raw personal data, are exchanged. Of course, this introduces greater complexity, requiring additional technical safeguards and precise organizational-legal frameworks (e.g., defining data processors and owners).

4. Conclusions

OHDSI Italy appreciates the EDPB's effort to provide comprehensive guidance on scientific research processing. The Guidelines represent a substantial step forward in clarifying the application of GDPR to real-world health data research. We submit our comments in a constructive spirit, with the objective of ensuring that the final Guidelines are operationally workable for federated research networks of the type that are increasingly central to European health data science.

Our key recommendations may be summarised as follows. The first four, marked as priority, address the legal basis issue that we consider the most urgent gap in the current Guidelines:

PRIORITY RECOMMENDATIONS: Legal Basis for Secondary Research Use of Health Data:

- **1.** Confirm that Article 6(1)(e) GDPR (public interest) is a valid and sufficient legal basis for secondary research use of routinely collected clinical data by institutions which, by virtue of their very nature as public bodies, pursue general research objectives in the public interest, as set out in legislation, and clarify where relevant under Union or Member State law, the relationship between Article 9(2)(g) and Article 9(2)(j) GDPR in public-interest health research settings.
- **2.** Confirm that the transition from Article 9(2)(h) to Article 9(2)(j) GDPR for further processing of health data is covered by the presumption of purpose compatibility under Article 5(1)(b) and does not require individual consent where Article 89(1) safeguards are in place (§3.3.3, Clarification 2).
- **3.** Recommend to Member States that national frameworks effectively requiring individual consent for all research use of routinely collected health data be assessed for their compatibility with the GDPR's research-enabling objectives under Article 89 (§3.3.3, Clarification 3).
- **4.** Confirm that pseudonymised OMOP CDM-based federated research with genuinely aggregate result sharing meeting minimum size thresholds may be carried out under Article 6(1)(e) and Article 9(2)(j) GDPR, without individual patient consent, where the applicable Union or Member State legal framework is satisfied and appropriate safeguards under Article 89(1) are implemented (§3.3.3, Clarification 4).

Further recommendations:

- **5.** Clarify that exploratory characterisation studies and data quality assessments within standardised methodological frameworks qualify as scientific research under factor (i) (§2.1).

- **6.** Clarify that the researcher qualification requirement under factor (iv) applies to scientific leadership, not to all technical data processing staff (§2.1).
- **7.** Explicitly recognise ETL, vocabulary mapping, and data quality assessment as ancillary processing operations motivated by scientific research (§2.3).
- **8.** Clarify that infrastructure-level governance satisfies the storage limitation substantiation requirement for multi-purpose research data warehouses (§3.2).
- **9.** Acknowledge the challenge of fragmented national implementation under Article 9(2)(j) GDPR in multi-site federated research, including situations in which different categories of institutions within the same Member State may rely on different legal pathways. The Italian example, including Articles 110 and 110-bis(4) of the Data Protection Code, illustrates how this fragmentation may affect federated OMOP CDM-based studies, and why clearer and more operational EU-level guidance remains necessary pending the fuller application of the EHDS framework, including Article 53(1)(e) of Regulation (EU) 2025/327 (§3.5).
- **10.** Distinguish between joint controllership and processor-type relationships in standardised federated network studies and endorse framework agreements for network-level joint controller compliance (§7.3).
- **11.** Explicitly recognise the distributed network architecture (local execution, aggregate results only, no patient-level transfer) as satisfying the secure processing environment safeguard (§8.5).
- **12.** Endorse a layered, institution-level transparency approach for hospital OMOP CDM research warehouses (§5).
- **13.** Clarify under which conditions personal data may be pseudonymised as a preparatory step for future scientific research within a standing research infrastructure, even where no specific study is yet underway, and how such preparatory processing should be assessed in relation to the applicable legal basis, the relevant derogation under Article 9 GDPR and the safeguards required by Article 89(1) (§3.7.1).
- **14.** Clarify that the systematic training and validation of AI models for public interest purposes constitutes scientific research under Section 2, and explicitly recognise federated learning and other PETs as appropriate safeguards under Section 8 (§3.9)

OHDSI Italy remains available to provide further technical input or to participate in any stakeholder consultation that the EDPB may organise in connection with the finalisation of these Guidelines.

On behalf of OHDSI Italy National Node

Lead Institution: University of Pavia (Prof. Lucia Sacchi)

Co-Lead Institution: Biomeris

Contact: lucia.sacchi@unipv.it

Website: https://www.ohdsi-europe.org/national-nodes/italy

Date: June 2026