

Connecting Primary and Secondary Use of Health Data in Europe: Toward the Conditions for Meaningful Data Fluidity

Summary of the xShare WP4 workshop findings held in Brussels in March 2026 during the IHE Connecthaton

Abstract

The integration of primary and secondary uses of health data is a central ambition of the European Health Data Space (EHDS). While Europe has made significant progress in developing interoperability standards and data infrastructures, the effective reuse of health data remains limited in practice. Drawing on a multi-stakeholder workshop bringing together clinicians, public health authorities, and technical experts, this proceeding paper examines the persistent barriers to data reuse and the conditions required for achieving "data fluidity." Particular attention is given to the role of common data models which enable large-scale analytical reuse but also expose upstream challenges related to data quality, semantic consistency, and contextual interpretation. The analysis shows that interoperability is fundamentally a socio-technical challenge, requiring alignment between clinical workflows, governance structures, and data infrastructures.

1. Introduction

Across Europe, healthcare systems are undergoing rapid digital transformation, resulting in an unprecedented expansion in the volume and diversity of health data. Electronic health records, disease registries, laboratory systems, and digital applications collectively hold the promise of improving patient care while enabling research, innovation, and more responsive public health systems. The EHDS seeks to realize this promise by creating a framework through which health data can be securely and efficiently reused across borders and domains.

Yet, despite this progress, a striking disconnect persists between the **availability of data** and its **effective use**. Data is abundant, but its reuse remains constrained, often requiring substantial additional effort to render it usable for purposes beyond immediate clinical care. This gap has led to increasing recognition that interoperability—long framed primarily as a technical challenge—must be reconsidered in broader terms.

The discussions analysed in this paper, drawn from a workshop on connecting primary and secondary use of data, provide a rich set of insights into this issue. They reveal that while infrastructures such as federated data networks and common data models, including OMOP, have made important strides in enabling secondary use, they also bring into sharper focus the limitations of current data generation practices. In doing so, they point toward a more comprehensive understanding of what is required to achieve genuine data fluidity.

2. From Data Capture to Data Reuse: Persistent Misalignments

One of the most consistent themes emerging from the workshop is the tension between how data is **produced in clinical settings** and how it is **expected to be reused**. Clinical data is primarily generated to support patient care, and its structure reflects this purpose. Much of the most clinically meaningful information is embedded in narrative text, shaped by the logic of diagnosis, uncertainty, and professional judgment. As one participant succinctly observed, “the granularity is in the text.”

This creates a structural mismatch with the requirements of secondary use, which depend on standardized, computable data. Efforts to bridge this gap have taken multiple forms. On the one hand, there have been attempts to increase the use of structured data entry at the point of care. On the other, technological solutions such as natural language processing have been developed to extract structured information from unstructured text. Both approaches, however, encounter important limitations. Structured data entry can impose additional burdens on clinicians, while automated extraction introduces issues of accuracy and interpretability.

These tensions become particularly visible in the context of common data models such as OMOP. By transforming heterogeneous local datasets into a standardized structure, OMOP enables large-scale observational research and federated analytics. However, the process of mapping clinical data into such a model is neither trivial nor neutral. It requires decisions about how to interpret and represent data, often in the face of ambiguity or incomplete information. As a result, the quality and usability of the resulting dataset remain deeply dependent on the conditions under which the original data was captured.

The frequently cited expression “garbage in, garbage out” captures this dependency, but it also risks oversimplification. The issue is not merely one of “bad data,” but of **data that is not aligned with its intended reuse**. Addressing this misalignment requires more

than technical fixes; it calls for a reconsideration of how data is generated, by whom, and for what purposes.

3. The Central Role of Clinical Practice

Any effort to improve data reuse must therefore engage directly with clinical practice. The workshop discussions make clear that clinicians operate under significant constraints, including time pressure, cognitive load, and competing priorities. Within this context, additional requirements for structured data entry or reporting are often perceived as burdensome, particularly when their benefits are not immediately visible.

Participants repeatedly emphasized the importance of **avoiding additional workload for clinicians**, noting that data initiatives that fail to account for clinical realities are unlikely to succeed. At the same time, examples from countries such as Denmark suggest that these challenges are not insurmountable. When data collection is closely aligned with clinical workflows, and when clinicians receive meaningful feedback—such as benchmarking or decision support—engagement can be significantly improved.

This has important implications for infrastructures designed to support secondary use. Whether data is ultimately analysed through registries, federated networks, or models such as OMOP, its quality and completeness are shaped at the point of care. **Systems that treat data reuse as a downstream activity, disconnected from clinical practice, risk perpetuating the very problems they seek to solve**.

4. Semantic Complexity and the Limits of Standardization

Beyond issues of data capture, the workshop highlights the enduring challenges of semantic interoperability. Standardized terminologies such as SNOMED CT and LOINC are widely regarded as essential tools for enabling data sharing and reuse. However, their implementation in practice reveals significant complexity.

Clinical concepts are often open to multiple interpretations, and coding systems may offer several possible representations for the same phenomenon. Even when standards are adopted, differences in coding practices between institutions—and even between individual practitioners—can lead to inconsistencies. As one speaker noted, “there will always be discussion” about the appropriate code.

These issues are further amplified when data is mapped into common data models. In the case of OMOP, for example, local codes must be translated into standardized vocabularies, a process that involves both technical and interpretive work. Decisions made during this process can have significant implications for downstream analyses, affecting comparability across datasets and the validity of conclusions.

This underscores a critical point: **standardization does not eliminate ambiguity; it redistributes it**. Effective semantic interoperability therefore depends not only on the

adoption of standards, but also on the governance mechanisms, expertise, and collaborative processes required to apply them consistently.

5. The Problem of Context

A related challenge concerns the role of context in data interpretation. Clinical data is inherently situated, reflecting the circumstances under which it was produced, including the clinical question being addressed, the patient population, and local practices. When data is abstracted from this context—whether in the form of indicators, aggregated datasets, or standardized models—important nuances may be lost.

This issue was repeatedly raised during the workshop, particularly in relation to the interpretation of indicators and cross-site analyses. Even when data is harmonized structurally, as in OMOP-based networks, differences in context can lead to divergent meanings. The question “do we all mean the same?” becomes central, especially in federated settings where analyses are conducted across multiple institutions.

Addressing this challenge requires more than technical solutions. While metadata frameworks and provenance tracking can help, they must be complemented by domain expertise and collaborative interpretation. Without this, there is a risk that data-driven insights may be misleading or misapplied.

6. Fragmentation and the Limits of Integration

The persistence of fragmentation across European health systems provides the backdrop for these challenges. Differences in organizational structures, regulatory frameworks, and technological infrastructures result in highly heterogeneous data landscapes. This heterogeneity is evident not only between countries, but also within them.

In this context, tools such as OMOP are often used to enable a degree of analytical integration, allowing data from different sources to be queried in a consistent manner. However, as the workshop discussions make clear, such approaches do not eliminate fragmentation; they provide ways of working with it. The process of transforming and harmonizing data introduces its own complexities, requiring coordination, validation, and ongoing maintenance.

The experience of public health authorities during the COVID-19 pandemic illustrates the stakes involved. The need for timely, high-quality data exposed the limitations of existing systems, which were often not designed for rapid, large-scale data sharing. This has reinforced the importance of building infrastructures that are not only interoperable, but also resilient and adaptable.

7. Governance and the Question of Alignment

Underlying many of the issues discussed is a broader challenge of governance. The workshop highlights the extent to which data reuse depends on coordination between multiple actors, including clinicians, hospitals, public health agencies, researchers, and technical organizations. Each of these actors operates within its own set of incentives, constraints, and priorities.

Infrastructures such as federated networks and OMOP-based platforms require a high degree of alignment to function effectively. This includes agreement on data models, coding practices, analytical methods, and governance processes. Achieving such alignment is not simply a matter of technical design; it requires sustained collaboration and shared ownership.

The EHDS provides an important framework for facilitating this alignment at the European level, particularly through its legal provisions for data access and sharing. However, the workshop discussions suggest that legal frameworks must be complemented by operational and organizational mechanisms that enable coordination in practice.

8. Rethinking the Boundary Between Primary and Secondary Use

A final theme concerns the distinction between primary and secondary use of data. While this distinction is embedded in policy and regulation, it appears increasingly difficult to sustain in practice. Public health surveillance, for example, depends on timely access to clinical data, while research relies on longitudinal patient records.

At the same time, many infrastructures designed for secondary use, including OMOP-based systems, are built as layers on top of primary systems, relying on data extraction and transformation processes. This separation can lead to duplication, delays, and missed opportunities for feedback.

The workshop discussions point toward the need for a more integrated approach, in which data is captured once and reused across multiple purposes. Achieving this will require rethinking system architectures, workflows, and governance models, with a view to reducing the distance between data generation and data use.

9. Conclusion

The ambition of creating a European health data ecosystem that supports both care and innovation is both timely and necessary. The workshop analyzed in this paper provides a clear message: while significant progress has been made in developing technical infrastructures, the key challenges lie elsewhere.

Data quality, semantic consistency, contextual interpretation, and governance alignment all play a decisive role in determining whether data can be meaningfully reused. Infrastructures such as OMOP demonstrate what is possible when data can be

harmonized and analyzed at scale, but they also highlight the limitations of current approaches to data generation and management.

Achieving “data fluidity” will therefore require a shift in perspective. Rather than focusing solely on interoperability as a technical objective, it must be understood as a broader socio-technical process, involving the alignment of systems, practices, and actors.

Only through such an approach can the promise of the EHDS be fully realized, transforming abundant data into meaningful and actionable knowledge.

“We have a lot of data… but meaningful use is still missing.”