



# Enabling Federated Computing in the Secondary Use of Social and Health Data

## Background paper 7.2.1

---

FinHITS — Strengthening Finnish Health Data ICT  
for Secondary Use

Project number 101126512

13 February 2025



**Co-funded by  
the European Union**

**FINDATA**

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

# 1 Table of Contents

1	Executive Summary .....	3
2	Introduction .....	3
2.1	Background.....	3
2.2	Objectives .....	4
2.3	Methods .....	4
3	Current Legislation .....	5
4	Definitions of Federated Computing.....	6
5	Key Use Cases.....	7
5.1	International Network for Data-Driven Management .....	7
5.1.1	Description of the Use Case .....	7
5.1.2	Significance of the Use Case .....	7
5.1.3	Implementation Possibilities in the Current Environment.....	7
5.1.4	Further Development Needs.....	8
5.2	Federated Learning in a European Collaborative Project .....	8
5.2.1	Description of the Use Case .....	8
5.2.2	Significance of the Use Case .....	8
5.2.3	Implementation Possibilities in the Current Environment.....	8
5.2.4	Further Development Needs.....	9
5.3	Hybrid Research on Federated Learning (Horizontal and Vertical) .....	9
5.3.1	Description of the Use Case .....	9
5.3.2	Significance of the Use Case .....	9
5.3.3	Implementation Possibilities in the Current Operational Environment .....	9
5.3.4	Further Development Needs.....	10
6	Summary of Key Challenges and Recommendations for Further Development.....	10



Finnish Social and Health Data Permit Authority

13 February 2026

## Document information

Writers	Organisation
Heikki Lanu, Head of Information Security	Findata – Finnish Social and Health Data Permit Authority
Maari Parkkinen, Development Manager	Findata – Finnish Social and Health Data Permit Authority
Hanna Tervonen, Chief Data Manager	Findata – Finnish Social and Health Data Permit Authority
Hannu Vilpponen, Head of ICT	Findata – Finnish Social and Health Data Permit Authority
Mika Hilvo, Research Team Leader	VTT Technical Research Centre of Finland Ltd.
Miikka Kallberg, Development Manager	CSC – IT Center for Science
Jaakko Lähteenmäki, Principal Scientist	VTT Technical Research Centre of Finland Ltd.
Juha Pajula, Senior Scientist	VTT Technical Research Centre of Finland Ltd.
Emmi Turunen, Product Owner	HUS
Arho Virkki, Chief Analytics Officer	Varha Wellbeing Services County of Southwest Finland

Experts	Organisation
Miikka Kallberg, Development Manager	CSC – IT Center for Science
Heikki Lanu, Head of Information Security	Findata – Finnish Social and Health Data Permit Authority
Maari Parkkinen, Development Manager	Findata – Finnish Social and Health Data Permit Authority
Hanna Tervonen, Chief Data Manager	Findata – Finnish Social and Health Data Permit Authority
Hannu Vilpponen, Head of ICT	Findata – Finnish Social and Health Data Permit Authority
Annu Kaila	HUS
Emmi Turunen, Product Owner	HUS
Harri Rantala, Data Scientist	Pirkanmaan hyvinvointialue
Tapio Pahikkala, Professor	University of Turku
Annika Pirnes, Service Manager	Varha Wellbeing Services County of Southwest Finland



Finnish Social and Health Data Permit Authority

13 February 2026

Per-Erik Gustafsson, System Specialist	Varha Wellbeing Services County of Southwest Finland
Arho Virkki, Chief Analytics Officer	Varha Wellbeing Services County of Southwest Finland
Antti Yli-Karhu, System Specialist	Varha Wellbeing Services County of Southwest Finland
Mika Hilvo, Research Team Leader	VTT Technical Research Centre of Finland Ltd.
Jaakko Lähteenmäki, Principal Scientist	VTT Technical Research Centre of Finland Ltd.
Juha Pajula, Senior Scientist	VTT Technical Research Centre of Finland Ltd.

## History of Changes

Date	Version	Editor	Status
26.3.2025	0.1	Maari Parkkinen	Draft
4.4.2025	0.1	Hanna Tervonen	Draft
3.6.2025	0.2	Maari Parkkinen	Draft
27.6.2025	0.2	Hanna Tervonen	Draft
10.7.2025	0.2	Maari Parkkinen	Draft
27.10.2025	0.2	Hanna Tervonen	Draft
4.11.2025	0.2	Maari Parkkinen	Draft
13.2.2026	0.3	Antti Piirainen	Finalised



## 1 Executive Summary

This background paper examines the possibilities of federated computing in the secondary use of social and health data, particularly in the context of secure processing environments. Federated computing enables the processing of personal data without transferring datasets, which supports data protection and the rights of data subjects. The report presents definitions of federated computing, key use cases, and the legislative framework in Finland and the EU. Additionally, it offers recommendations from an expert group for further development, emphasizing the importance of data harmonization, technical solutions, and regulatory clarity to effectively utilize federated computing in national and international research. A broad group of experts with practical experience in the challenges and opportunities of federated computing in various operational environments contributed to the report.

## 2 Introduction

### 2.1 Background

Federated computing enables the processing of personal data without the need to compile and transfer data into a single processing environment.

Federated computing is particularly useful when:

1. Extremely large datasets are involved, making data transfer challenging due to size.
2. Data includes special categories of personal information, requiring highly restricted processing environments.
3. Data from different countries is needed, but varying legislation and practices limit the transfer of citizens' data across borders.

Federated computing offers advantages. From a data protection perspective, it is beneficial that datasets can be processed without compiling them on a single platform. More limited processing of personal data also supports the rights of data subjects.

A prerequisite for implementing federated computing is the use of unified data models.<sup>1</sup> The data used in analyses must be harmonized and stored in a similar data model, allowing locally stored datasets to be analyzed using a shared algorithm. A commonly used data model for harmonizing patient records is OMOP (Observational Medical Outcomes Partnership). The Finnish FinOMOP project has produced OMOP-based standardized practices that are widely applied across the country.

This background paper was written as part of the FinHITS – Strengthening Finnish Health Data ICT for Secondary Use project (hereafter referred to as the FinHITS project), specifically under Work Package 7: Secure Processing Environment (SPE). The FinHITS project has received funding from the EU4Health program (Direct grants to Member States: for setting up services by Health Data Access Bodies – Secondary use of health data). One of the goals of the work package is to promote the implementation possibilities of federated computing in secure processing environments.

---

<sup>1</sup> Laitinen T, Virkki A, Porkka K: FinOMOP: terveystietojen kansainvälinen harmonisointi. Duodecim 2022;138:1761–3.



Finnish Social and Health Data Permit Authority

13 February 2026

## 2.2 Objectives

The objective of this background paper is to describe how federated computing can be implemented nationally and internationally in the secondary use of social and health data. The focus is particularly on secure processing environments. The aim is also to promote the effective utilization of federated computing and support the joint development of audited secure processing environments as defined in the Secondary Use Act.

## 2.3 Methods

This background paper was written in collaboration with national experts. It presents unified definitions for different implementations of federated computing and describes challenges in the national environment through key use cases. Based on these use cases, key development needs and potential solutions are listed, along with recommendations for further development.

The work on the background paper progressed in several phases. On June 6, 2024, Findata organized a preparatory meeting with the Wellbeing Services County of Southwest Finland, Helsinki University Hospital (HUS), and the Institute for Molecular Medicine Finland (FIMM). These organizations participated in a pilot study testing the possibilities of the federated model in secondary use<sup>2</sup>.

Key perspectives raised in the meeting included:

- The need for standardized practices to effectively implement federated computing.
- Streamlining and accelerating processes.
- Optimizing secure processing environments to better support federated computing, especially focusing on the exchange of anonymous parameters between environments as a primary development need.

The use cases presented in this background paper, as well as the recommendations derived from them, were compiled through collaboration among key national experts in two separate workshops, which were attended by 19 individuals. The contributors to the background paper represented a total of seven organizations: CSC – IT Center for Science Ltd., Findata – Finnish Social and Health Data Permit Authority, HUS Group, Wellbeing Services County of Pirkanmaa, University of Turku, Wellbeing Services County of Southwest Finland, and VTT Technical Research Centre of Finland Ltd.

The topics and dates of the workshops were as follows:

- Workshop 1: Current situation and key use cases (Thu 20 March, 10:00–14:00)
- Workshop 2: Further development needs and proposed solutions (Thu 14 April, 09:00–11:00)

The objectives of the workshops were to:

- Identify key use cases for federated computing
- Describe implementation and related challenges in the current operational environment
- Identify further development needs

The contents of the background paper were compiled based on the workshop work, and the paper was sent to participants a total of three times for comments and additions between and after the workshops. Comments were also requested from a high-level expert group appointed by the Ministry of Social Affairs and Health, whose task is to establish principal guidelines for

---

<sup>2</sup> Sitra: Lisänäyttöä uusista lääkeshoidoista pilot project final report, 2024. Available et: <https://www.sitra.fi/wp/wp-content/uploads/2024/04/lisanayttoa-uusista-laakehoidoista.pdf>



Finnish Social and Health Data Permit Authority

13 February 2026

anonymization, data protection, and information security in the operations of the Data Permit Authority.

### 3 Current Legislation

In Finland, the secondary use of social and health data is regulated by the Act on the Secondary Use of Social and Health Data (552/2019), hereafter referred to as the Secondary Use Act. The Act defines the purposes for which data can be used, the permit process, and the conditions for data processing.

At the time of writing, a reform of the Secondary Use Act is underway. Proposed amendments concern data permits, data requests, data processing, international research collaboration, secondary use fees, anonymization of datasets and results, and the relationship between the Secondary Use Act and clinical research.

In addition to the Secondary Use Act, other legislation also enables the secondary use of social and health data in Finland. This background paper is based on the version of the Act in force in autumn 2025 and may be updated as needed.

Under the Act, individual-level datasets may be used for scientific research, statistics, planning and investigation tasks of authorities, education, and data-driven management. In certain cases, data-driven management and oversight of social and healthcare services can be conducted without a data permit (Sections 41 and 42). For data-driven management purposes, data does not need to be processed in a secure processing environment as defined by the Act.

If the purpose is to combine data from multiple public social and healthcare data controllers (Section 6), private service providers, or data stored in the national Kanta services, Findata grants the data permit (Section 44). Otherwise, the permit is granted by the authority whose data is being requested.

Individual-level data provided under a data permit must always be processed in a secure processing environment (Section 51). If the use of individual-level data does not require a permit, the user is not obligated to use a secure processing environment.

Secure processing environments are audited by information security assessment bodies and must meet the requirements of the Secondary Use Act and Findata's regulation (1/2022). The regulation emphasizes the importance of monitoring information security, maintaining up-to-date security, and ensuring the integrity and accuracy of received datasets. Findata is responsible for ensuring the anonymity of results published based on data provided under a permit (Section 52).

The requirements for secure processing environments do not prevent the implementation of federated computing, provided that the data utilization plan covers the specific use case, no personal data is transferred outside the secure environment and the assessment body has approved the implementation. The service provider of the secure environment is responsible for ensuring that no personal data leaves the environment. A network connection from the environment or a user-specific environment may be opened if<sup>3</sup>:

- The need for the connection is justified and does not conflict with the data permit.
- The destination party is identified and deemed trustworthy.
- The connection is established, managed, and monitored by the service provider.
- The connection is protected by a firewall and only necessary traffic is allowed.
- No personal data is transferred.

<sup>3</sup> Sitra: Lisänäyttöä uusista lääkkehoidoista pilot project final report, 2024. Available et: <https://www.sitra.fi/wp/wp-content/uploads/2024/04/lisanaytto-uusista-laakkehoidoista.pdf>



Finnish Social and Health Data Permit Authority

13 February 2026

- Any significant changes to the environment are reported to the assessment body.

In addition to the Secondary Use Act, the Cybersecurity Act also applies, setting stricter requirements for monitoring information security. The EU General Data Protection Regulation (GDPR) must also be considered, especially regarding the rights of data subjects.

The European Health Data Space (EHDS) Regulation came into force on March 26, 2025. Its goal is to enable cross-border secondary use of health data through unified governance, processes, and infrastructure. The regulation covers both primary and secondary use. For secondary use, it will apply to most data categories starting March 26, 2029.

Under EHDS, secure processing environments must be used (Article 73). The technical requirements for these environments will be defined by an implementing act by March 26, 2027. The TEHDAS2 project, coordinated by Sitra, is currently developing recommendations for these requirements to support the drafting of the implementing act.

Chapter IV of the EHDS Regulation describes the processes through which individual-level data are collected and transferred into secure processing environments. However, it is expected that Member States will take varying approaches to cross-border data transfers where the goal is to gather data into a single secure processing environment. As the availability of data improves, the need for federated computing is expected to remain or even grow. Nevertheless, federated computing is not explicitly addressed in the regulation itself.

At the same time, several international projects are underway in Europe aiming to address challenges related to using sensitive datasets for research through federated computing. Most of these projects have received EU funding. They typically focus on solving one or a few specific use cases, but they do not necessarily interact with each other or take into account the requirements of the upcoming EHDS Regulation.

## 4 Definitions of Federated Computing

Federated computing refers to processing data locally without compiling all datasets into a single processing environment<sup>4</sup>. It includes both one-time computations and iterative data processing.

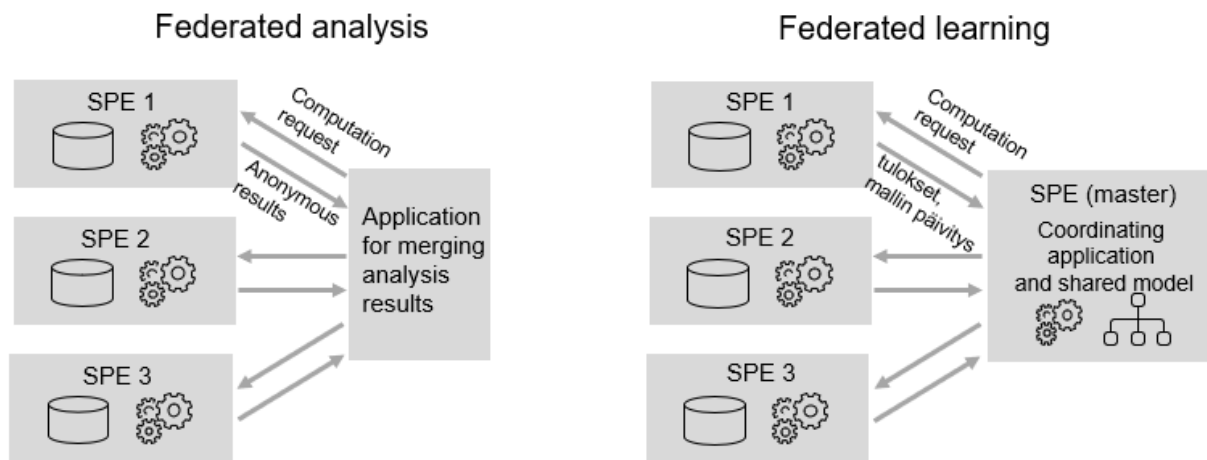
In this background paper, federated computing is defined as follows:

- **Federated processing or computing:** A general term for federated data processing.
- **Federated analysis:** One-time federated computation.
- **Federated learning:** Iterative federated computation.

---

<sup>4</sup> Federated computing is sometimes referred to as distributed computing. However, distributed computing generally refers to methods that optimise the use of computing resources, such as parallel computing or high-performance computing (HPC). For this reason, the term “distributed computing” is not used in this document.





**Figure 1.** A visual illustration (Figure 1) distinguishes between federated analysis and federated learning.

## 5 Key Use Cases

The workshops identified the following key use cases, whose implementation possibilities are examined in detail in this summary. These use cases represent examples of the most common scenarios, and their presentation aims to explore implementation possibilities and challenges in more detail. The list is not exhaustive, and other significant use cases also exist.

### 5.1 International Network for Data-Driven Management

#### 5.1.1 Description of the Use Case

The goal is to create an international network for data-driven management that enables continuous peer review and monitoring. Through analytics and data-driven management, collaboration among various actors in an international environment is facilitated. Monitoring can be conducted automatically as part of an ongoing process. Participants may include social and healthcare service providers using structurally consistent data.

#### 5.1.2 Significance of the Use Case

An international network for data-driven management would enable small-scale studies with limited sample sizes by combining data from multiple actors. For example, many complex procedures are nationally centralized in one hospital in several countries. When care needs to be developed, the partners are often foreign. This use case can also be utilized in pandemic situations and organizational governance.

#### 5.1.3 Implementation Possibilities in the Current Environment

Implementation is possible under the current legal framework, as data-driven management can be conducted outside secure processing environments. Additionally, no separate data permit is required when a social and healthcare provider uses data generated in its own operations or stored in its own registers, provided the use is necessary for producing, monitoring, evaluating, planning,



Finnish Social and Health Data Permit Authority

13 February 2026

developing, managing, or supervising services under its responsibility (Section 41 of the Secondary Use Act).

However, individual-level data cannot be combined in this context, making data structure and model harmonization a key challenge. The Secondary Use Act does not address the anonymity of results generated through data-driven management.

At HUS, a federated network implementation is being developed that technically enables the described use case. The solution is based on open-source software (a federated node responsible for running analyses and returning results) and GitHub Enterprise, which is used for managing analysis code and result logistics. In this use case, data-driven management is not fully decentralized but relies partly on centralized services. During architectural planning, it was concluded that such a system design is critical for scalability.<sup>5</sup>

#### 5.1.4 Further Development Needs

Implementing this use case requires data harmonization, effective data-driven management practices, and consideration of legislation, data protection, and information security.

## 5.2 Federated Learning in a European Collaborative Project

### 5.2.1 Description of the Use Case

The goal is to conduct federated learning in a research project that utilizes datasets from multiple EU countries without transferring the data to a shared processing environment. The research team applies for data permits separately from each country. Categorical and numerical variables are harmonized into OMOP format. Signal and image data are standardized into EDF and DICOM formats.

The research may e.g., aim to develop a European lung cancer prediction model or create an algorithm to support cardiology diagnostics, which efficiently scores different treatment methods and predicts treatment response. The computational method could later be used to support physicians in diagnostics. Since the computation is based on deep neural networks, manual validation is not possible.

### 5.2.2 Significance of the Use Case

This use case reflects a growing trend in international collaborative research. Improved prediction of treatment responses benefits patients and reduces healthcare costs. With large training datasets, diagnostic algorithms can be trained to be both sensitive and specific. Such research is best conducted internationally, as individual countries often lack sufficient data. If data from different countries cannot be transferred across borders to a shared processing environment, federated computing enables the research to proceed.

### 5.2.3 Implementation Possibilities in the Current Environment

Currently, this use case is not feasible because it requires a connection from a secure processing environment to a data domain and automatic, iterative transfer of parameters/gradients. An alternative is for the research team to adapt the algorithm separately in each country – e.g., perform federated analysis as a one-time fit and combine the models afterward.

---

<sup>5</sup> More information: <https://phems.eu>



Finnish Social and Health Data Permit Authority

13 February 2026

The project does not intend to transfer personal data, so it is technically possible under the current Secondary Use Act. However, the environments must be built to support federated learning. Even if the technical solution succeeds, challenges may arise in ensuring the anonymity of results. Algorithms must be auditable to verify that the data transferred between environments is anonymous.

The project uses harmonized OMOP-formatted data. However, local distributions are likely to differ significantly between countries due to factors such as smoking habits or data recording practices. Signal and image data are standardized into EDF and DICOM formats, but the DICOM standard has multiple versions, which creates manual work.

Currently, no secure processing environments – except HUS Acamedic – support the technical solutions needed for implementation. Even in HUS Acamedic, auditing is limited to swarm learning. Information may leak through the algorithm, for example, if the dataset includes outliers (e.g., a person over 105 years old). To implement this use case, algorithm auditing is needed to ensure privacy and confirm that the algorithm is not otherwise harmful.

#### 5.2.4 Further Development Needs

The technical implementation of secure processing environments must be developed to support federated learning. Implementation would require expanding the trust network across borders and special arrangements for algorithm auditing and result anonymization. Additionally, a pre-approval procedure should be developed to avoid the need for reviewing each individual result separately.

### 5.3 Hybrid Research on Federated Learning (Horizontal and Vertical)

#### 5.3.1 Description of the Use Case

The goal is to conduct a national study involving multiple national data controllers. The variables in the datasets vary depending on the data sources.

#### 5.3.2 Significance of the Use Case

In the example study, the dataset consists of individuals who use a large number of services. The dataset includes several entities, such as wellbeing services counties, Kela (Social Insurance Institution of Finland), and a private rehabilitation center.

#### 5.3.3 Implementation Possibilities in the Current Operational Environment

In theory, such a study is feasible. It could be implemented either as several separate projects or as a single project authorized by Findata. Regarding authorization, the controllership of the datasets or potential joint controllership must be considered.

Implementation requires data harmonization, for example, OMOP modeling, and listing the variables of datasets from different data controllers (vertical federated learning). Additionally, vertical data must be harmonized, and the method for linking research subjects must be defined (e.g., pseudonymization using consistent codes).

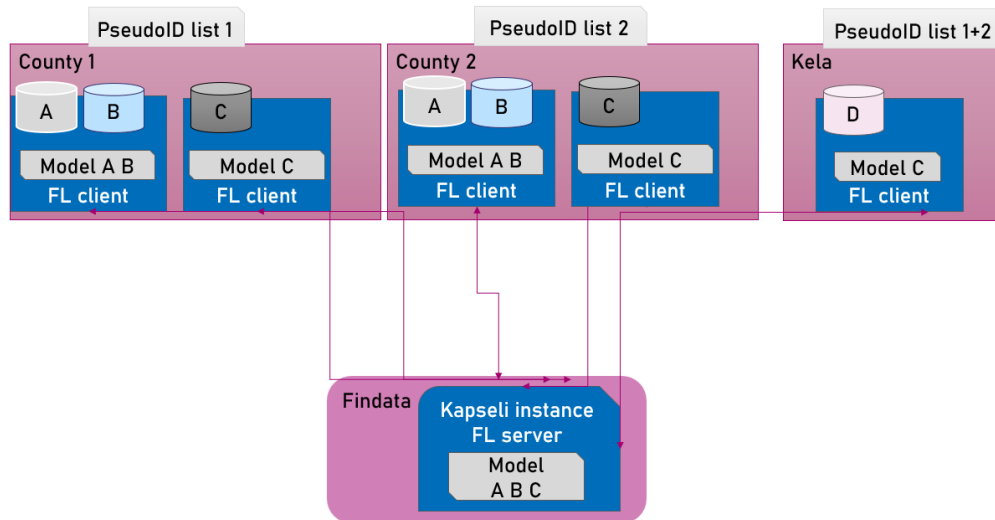
Implementation could be carried out through Findata's standardized federated learning network (Figure 2). In the hybrid model, each data owner would have a certified server where the extracted research data can be imported and defined as part of Findata's standardized federated learning network. After this, Findata's Kapseli would be defined as the centralized computing unit through which the federated learning model is executed within the defined network.



Finnish Social and Health Data Permit Authority

13 February 2026

Successful research would require standardized network and computing resources as well as unified on-demand deployment methods. Additionally, all wellbeing services counties and other data controllers must follow consistent OMOP specifications to ensure reliable analysis.



**Figure 2.** Illustration of the implementation of the hybrid use case.

### 5.3.4 Further Development Needs

To implement this use case, data harmonization and the development of technical standard solutions (network and deployment) are needed. Dividing the research across different secure processing environments could be a solution (federated computing). For example, research conducted in HUS Academic could be extended to CSC’s environment, allowing the use of CSC’s computing capacity within Academic’s secure processing environment.

Regulation do not prevent distributing computing across multiple secure processing environments, provided the data is properly authorized, no personal data leaves the secure environment, and the assessment body has approved the implementation. Implementation could involve creating virtual secure processing environment clusters between secure environments. Any potential need for additional auditing due to new technical solutions must be considered.

## 6 Summary of Key Challenges and Recommendations for Further Development

Through the examination of use cases, the following challenges (1–5) were identified.

1. There are no unified definitions for federated computing, which leads to misunderstandings. One-time federated analysis and iterative federated learning are often confused. Additionally, federated computing is sometimes referred to as distributed computing, which also includes high-performance computing – a completely different activity.

### Expert Group Recommendations for Development:



The following definitions from the background paper are recommended to clarify discussions on federated computing:

- **Federated processing or computing:** General term for federated data processing.
- **Federated analysis:** One-time federated computation.
- **Federated learning:** Iterative federated computation.

2. The relationship between legislation and the implementation of federated computing has caused confusion in both national collaboration and international permit-based projects. This background paper aims to clarify the constraints arising from the Secondary Use Act. Federated computing based on a data permit under the Act can be implemented so that personal data is not transferred outside the secure processing environment. For data-driven management, individual-level data should not be combined with other datasets under the Act. In such cases, the transferred information must be anonymous algorithms and/or results.

The requirements of the Secondary Use Act do not directly regulate federated computing. Fundamentally, federated computing enables the secondary use of social and health data without the need to combine or transfer individual-level datasets. It is therefore a preferable alternative to compiling the entire dataset in one location.

Implementation methods for federated computing should also consider the rights of data subjects under the GDPR. It is desirable that federated computing is also possible when applying the EHDS regulation to the secondary use of health data.

**Expert Group Recommendations for Development:**

- Federated computing is a recommended method and could be the primary method in some cases. Requirements for ensuring the anonymization of intermediate results exported from secure processing environments should be clarified (see also Recommendation 4). Federated computing should remain possible regardless of the intended use.
- The process for ensuring the rights of data subjects under the GDPR must be considered in federated computing use cases. Citizens must be able to obtain information about the use of their data, even when their data is used in a project employing federated computing. This must be possible even if the implementing party does not have a complete view of the individuals whose data is used in the research.
- The EHDS regulation, which came into force on March 26, 2025, and the implementing act to be issued by 2027, will regulate the requirements for European secure processing environments. These requirements must consider the enablement of federated computing.

3. Incomplete data harmonization prevents the implementation of federated computing. In cross-border data utilization, in addition to harmonization, the socioeconomic context in which the data was produced must be considered. This challenge applies not only to federated computing but to all cross-border projects.

**Expert Group Recommendations for Development:**

- Data harmonization efforts should continue and be supported both nationally and as part of the EHDS regulation implementation.

4. To enable federated computing, especially federated learning, the anonymity of information transmitted by algorithms must be ensured when data leaves the secure processing environment or when the permit does not cover combining data between secure environments. In federated learning, the transmitted information may not be anonymous, even if the original dataset is not transferred.



**Expert Group Recommendations for Development:**

- The approval criteria for algorithms and methods used in federated learning should be clarified. Collaboration between secure processing environment providers and auditors should be established.
  - Data must not be exported from secure processing environments without control. The service provider is responsible for ensuring this. A shared data transfer solution between computing environments should be adopted to improve efficiency and security.
5. HUS Acamedic has implemented the first technical solution supporting federated learning, enabling the use of Swarm Learning. Apart from HUS Acamedic, technical and organizational solutions have not yet been built into nationally audited secure processing environments. There are no practices or standardized procedures for managing a network of federated environments.

**Expert Group Recommendations for Development:**

- Cooperation between secure processing environment providers should continue and be developed to explore and compare technical solutions and implement auditor-approved solutions.
- As part of this development, technical options for integration solutions should be explored to enable connections between different environments and between the environment and user-controlled applications, in compliance with legal requirements.
- Parties must be able to ensure the security of technical solutions connecting processing environments in collaboration.

