Bridging Health and Socio-Economic Data: Nordic Experiences in Cross-Domain Data Access

Ville Koskinen, Statistics Finland, ville.koskinen@stat.fi

Nikolaj Burmeister, Statistics Denmark, nbu@dst.dk

Paper

Background and purpose

Across the Nordic countries and Europe more broadly, a persistent divide between health data holders and providers of socio-economic data—most notably National Statistical Institutes (NSIs)—continues to pose significant challenges for data-driven research and analysis.

Researchers, public authorities, and private-sector actors increasingly require access to combined datasets to address complex societal and health-related questions. However, differences in institutional processes, regulatory interpretations, governance structures, and long-established traditions across national bodies often make such data integration difficult, time-consuming, and opaque.

This seminar explores these challenges through national experiences from Finland and Denmark, with a particular focus on initiatives aimed at bridging the gap between NSIs and health data holders.

Towards Interoperable Register-Based Research Infrastructure

The Nordic countries, have an exceptional foundation for register-based research. Population, education, labour market, income, social benefit, business and health records provide a rich basis for analysing society over time and across life domains. Yet these resources are not always governed as elements of a shared research infrastructure. Health data, in particular, are treated as a separate domain, with their own legislation, access procedures, technical environments and institutional logic.

This separation is understandable: health data are sensitive and require strong legal, ethical and technical safeguards. But it creates problems when analysis questions cut across health, work, education, income, social security and public

services. Ageing, disability, labour market participation, inequality, fiscal sustainability and public service reform all require health information linked to expenditure, employment, income and wider social conditions. Data-intensive research and AI development cannot rely on repeated, project-by-project extraction and harmonization[1].

The issue is whether the current sectoral model allows health data to be used - under appropriate safeguards - together with other societal data. The development of the European Health Data Space (EHDS) makes this issue urgent. It will require Member States to build stronger national capabilities for health data access, interoperability and secure use. But implementation choices matter. EHDS can become a bridge between health data and the wider register-based research infrastructure, or it can create a separate health data silo if developed too narrowly within one sector.

Health data should not be seen as an island. It is one dimension of a wider social and economic data ecosystem. The central challenge is to build research infrastructures that allow sensitive data from different domains to be combined lawfully, securely and efficiently.

The experience from Finland: Two infrastructures, two logics - one research need

The Finnish institutional setting illustrates this challenge. In official statistics and register-based data production, health data are largely organized separately from other population, labour market and socioeconomic data. The Finnish Institute for Health and Welfare produces many key health and social care statistics and maintains health-related registers, while Statistics Finland is responsible for population, labour market, education, income, business and macroeconomic statistics. Statistics Finland's Research Services, established 2010, provides unit-level microdata for scientific research and analysis, largely based on data collected for statistical purposes. In parallel, the Social and Health Data Permit Authority Findata, established 2019 under the Act on the Secondary Use of Health and Social Data, guides applicants, grants permits where the Act gives it competence, compiles, combines and pre-processes permitted data.

The two organizations reflect different but complementary logics. Statistics Finland has a long experience of producing, harmonizing, documenting and maintaining register-based datasets for repeated statistical and scientific use, while Findata is less a producer of continuously maintained research datasets and more a coordinating access and permit authority for data held by multiple social and health controllers. In practice, health data are governed mainly through project-specific access procedures, while socioeconomic data are increasingly organized into reusable research datasets.

Within their mandates, both systems may work well. The problem emerges at the boundary between them. Separate remote access environments, metadata catalogues, permit services and development funding may be justified

[1] Terveystiedon tulevaisuus tekoälyn aikakaudella, Kallioniemi, Porkka 2026 (The Future of Health Data in the Age of AI, in Finnish) https://www.sitra.fi/julkaisut/terveystiedon-tulevaisuus-tekoalyn-aikakaudella

organizationally, but together they make cross-sectoral research more complex than necessary. They can also weaken quality and comparability, as extraction, linkage and harmonization work is repeated across projects. Over time this becomes an administrative burden and limits the evidence base available for policymaking.

From project-specific extraction to reusable research data structures

Data governance often distinguishes between primary and secondary use. Primary use refers to the original administrative, operational or service purpose for which data are collected; secondary use refers to later reuse for research, statistics, evaluation and policy analysis. Secondary use, however, should not mean secondary importance in system design.

If analytical reuse, interoperability and long-term public value are considered only after operational systems have been built, the result is often fragmented data that are difficult to use responsibly and efficiently. Research and policy analysis should not be an afterthought. This does not mean weakening data protection; it means designing systems, metadata, governance and access procedures so that responsible reuse is possible from the start.

Many socioeconomic research questions require stable, longitudinal and population-wide data structures: education and work histories, income and taxation, benefits and pensions, family relations, migration, housing, region and public services. Health information can then be analysed within this broader life-course context. For example, a study on the costs of chronic illness may require health care contacts and diagnoses, but also tax data, social transfers, disability pensions and employment outcomes.

If these data must be separately extracted, combined and documented, the same work is recreated project by project. This may be tolerable for specialized one-off studies, but it is poorly suited to recurring research and policy needs, where comparability, continuity and cumulative learning are essential.

Ready-made datasets are research infrastructure

A more efficient solution is to develop ready-made research datasets for recurring and well-understood needs. This does not remove the need for permits, secure environments or clearly defined purposes. It means that commonly used data structures are created, harmonized, documented and updated in advance, so they can be reused under controlled conditions for approved research and analysis. Benefits are practical as well as scientific: less repeated extraction work, better documentation and comparability, lower marginal costs for public authorities, and more predictable access for externally funded research.

Statistics Finland already has extensive experience with this model. It produces about 70 ready-made research datasets covering individuals, households, enterprises and population characteristics. Their value is not only faster access, but also consistency: the same definitions, classifications, linkage procedures and documentation can be used across projects. This improves comparability, reduces duplication and lowers the additional cost of each new project.

The same logic could be applied more systematically to selected health and social care data. In many socioeconomic and health economic studies, researchers do not need the most detailed clinical information, complete diagnostic histories, free-text records or highly specific treatment details. They often need robust, well-documented and regularly updated indicators - such as broad diagnosis groups, service contacts, medication categories, disability-related indicators and long-term care information.

Reusable datasets can also support data protection. Detailed diagnosis codes, for example, could be grouped into broader categories when sufficient for the research purpose, reducing the need to handle unnecessarily granular data. In this sense, reusable datasets are not a relaxation of data protection, but one way of implementing it more effectively.

The missing institutional layer

The current Finnish model leaves an important institutional gap. The need for reusable datasets combining health and socioeconomic information is clear, but responsibility for producing and maintaining them is not. Findata coordinates permits and access to social and health data, but systematic production of reusable, harmonized and regularly updated cross-sectoral datasets requires capabilities and resources beyond permit administration alone.

Statistics Finland and its Research Services have expertise in building reusable, harmonized and well-documented datasets. Health and social care register holders, such as the Institute for Health and Welfare and the Social Insurance Institution of Finland, have the source data and sector expertise to define meaningful variables. Findata has a central role in permit and access governance. What is missing is a clear mandate, sustainable funding and an operational division of labour.

This matters because demand is recurring. Advances in AI and data-intensive research will increase the need for timely, well-documented and securely governed data. For the public sector, it is inefficient to treat each recurring need as a new data production process.

From sectoral data systems to shared research infrastructure

The Nordic countries have some of the world’s most comprehensive administrative data resources, but governance is often organized along sectoral lines. This is necessary for administration, but it becomes a barrier when research questions cut across health, taxation, education, labour markets and social security. The next step should not be to weaken data protection or centralize all data in one institution, but to build shared research infrastructure across institutional boundaries.

A practical way forward is to start with a limited number of recurring cross-sectoral research needs where public value is clear. For these, authorities could jointly define reusable data structures, documentation, access rules, updating procedures and cost sharing. In this division of labor, national statistical offices should not be seen merely as providers of background variables. Their core

expertise lies in turning administrative records into stable, documented, linkable and reusable research data infrastructures. Health and social care data holders would contribute source-data expertise, while health data access bodies would ensure lawful and secure access to sensitive health and social care data.

Such a model would also support European and national data space initiatives. A health data space disconnected from the wider socioeconomic data ecosystem will meet only part of the needs of research and policymaking. For many purposes, the goal should be interoperability between health data and other core societal data - not because health data are less sensitive, but because health is embedded in social and economic life.

The principle should be simple: health data must be strongly protected, but not unnecessarily isolated. In Nordic welfare states, questions about health are often also questions about work, income, education, social security and public services. Building bridges between these data domains is therefore a public investment in better research, policy evaluation and decision-making.

Danish Efforts to Integrate Data Access

The Danish presentation will outline a current initiative to strengthen collaboration between Statistics Denmark and the Danish Health Data Authority in relation to applications for data access.

Denmark possesses extensive data on both health and socio-economic factors, and there is considerable potential in making these datasets more accessible and interoperable. Access to microdata for research and analysis has been available for more than 40 years.

The combination of health data with socio-economic data has taken place under the auspices of Statistics Denmark. Consequently, an applicant seeking data access would require approval from both Statistics Denmark and the Danish Health Data Authority (DHDA). The DHDA would then transfer the relevant data to SD, after which a project could be established under SD's supervision.

This process has proved rather cumbersome for many applicants, primarily because the DHDA does not control all health data. In the Danish context, health data are fragmented across multiple authorities, each with differing policies and interpretations of current legislation. Moreover, there are discrepancies in policies and regulations between SD and DHDA.

SD and DHDA have taken initiative to implement a seamless connection between this portal and SD's existing application portal, Denmark's Data Portal. The collaboration focuses on creating a more unified user journey, including shared interpretations of regulation, common security standards, improved metadata exchange, and clearer information on pricing and access conditions.

A contract for this collaboration was agreed in April 2026, and work will commence in June 2026. This paper therefore outlines both the challenges encountered—and overcome—as well as the work that lies ahead in the coming years.

Main challenges

The collaboration between SD and DHDA was initiated more than five years ago, when a steering group was established to develop the One Contact Point for Health Data (OCP). Supported by private funding, this group has worked with private consultants and a broad range of stakeholders to define the core policies and purpose of the OCP.

From SD's perspective, the steering group has primarily focused on reorganising the health-related authorities. However, throughout the process, there has been a clear intention to strengthen collaboration with SD. In recent years, this has also led to parallel dialogues between SD and DHDA, as well as between the respective ministries.

During these discussions, several key questions have raised concerns. The main issues include the following:

A. What is health data?

It remains unclear where health data begin and end. As part of the EHDS, attempts have been made to define health data, but these definitions have been supplemented by socio-economic data required for contextual purposes.

This reflects a divide between health professionals and those from other academic disciplines. Many economists, for instance, regard health data as background variables within socio-economic studies—effectively reversing the perspective of health professionals.

In essence, health data and socio-economic data are intrinsically linked, and a bridge between the relevant authorities is required.

B. Who combines data?

> Given the need to integrate health data with socio-economic data, it is essential to determine responsibility for this task.
>
> An agreement has been reached between SD and DHDA that the collaboration will build upon the existing division of labour. Accordingly, the OCP will continue the DHDA policy of transferring health data to SD.
>
> SD will establish projects under the framework of Denmark's Data Portal and will remain responsible for output control.

C. What can data be used for?

> SD operates under a strict legal framework whereby data access may only be granted for research and analysis. It is strictly prohibited to obtain information on identifiable individuals or to use data for administrative purposes, such as clinical treatment. In essence, only secondary use is permitted, not primary use.

By contrast, DHDA and its OCP are designed to accommodate both primary and secondary uses of data. This reflects the nature of health sciences, where data use often combines patient treatment with research activities.

This issue has been addressed by allowing DHDA to collect a limited set of socio-economic data from public registers. This process does not involve SD, thereby ensuring compliance with SD's legal framework. However, it does create duplication of effort, as these registers are also collected and quality-assured by SD before being used for official statistics.

DHDA can offer its socio-economic data for primary purposes as well as for uses falling outside SD's legal framework. A corresponding legal framework is currently under development.

Reaching these agreements has required extensive negotiation. Various options were considered, including merging the two portals. However, due to differences in purpose, it was ultimately decided to keep the portals separate. This allows DHDA to develop a portal suited to the broader needs of health professionals, while enabling SD to remain within its legal remit, focusing on secondary data use.

When health professionals require data from SD, they may begin their application via the OCP portal; however, during the process, they will be redirected to SD's portal (Denmark's Data Portal).

The work ahead

A contract for closer collaboration has been signed, and private funding (€1.5 million) has been secured to develop a seamless application process between Denmark's Data Portal and the OCP portal currently under development. The project will run from 2026 to 2027, potentially extending into 2028.

The collaboration is structured around several workstreams, with particular focus on the following:

A. User journey across the two portals

The starting point will be to define the user journey. Some users may enter via Denmark's Data Portal but require approval and data from the OCP, while others may begin via the OCP and need to be redirected to Denmark's Data Portal. The objective is to make these user journeys as seamless as possible.

User journeys will be defined collaboratively. The first workshop on this topic will take place in June 2026. At the NSM conference, participants will be updated on progress in this area.

B. Standardised rules and regulations

To enable seamless user journeys, it is essential to clarify the elements involved and the services that can be offered.The collaboration is expected to focus on three key areas:

- Which institutions can be authorised, and what are the requirements for authorisation?

  This is critical, as access to Denmark's Data Portal is limited to users affiliated with authorised institutions. DHDA and SD currently apply different authorisation frameworks, which must either be aligned or replaced with an alternative solution.

- What should be included in the application process?

  Application procedures must be broadly aligned. Application forms should therefore be developed jointly. While there is already significant overlap, a key challenge lies in differing interpretations regarding the scope of applications (e.g. specific hypotheses versus hypothesis-generating projects or broader research programmes).

- A common interpretation of data minimisation

  At present, SD provides access to broader datasets than DHDA (e.g. full population data). Establishing a shared interpretation of data minimisation is therefore highly relevant. Encouragingly, convergence has already begun in this area.

C. Exchange of metadata

All applicants require an overview of available metadata. Denmark's Data Portal has already developed a "Data Safari", which includes some health data. This must be expanded to incorporate all datasets from the OCP, with clear indication that access requires OCP approval.

Similarly, the OCP must receive metadata from Denmark's Data Portal to ensure that users initiating applications via the OCP portal have a comprehensive overview of available data.

Summary of the Danish experience

The budget allocated to the project is likely insufficient to meet all the needs of users and organisations. Nevertheless, it is adequate to establish the fundamental components and to gain valuable experience.

Importantly, the project represents a necessary step towards closer collaboration and more consistent interpretation of rules and regulations. This will undoubtedly benefit users and ensure that Danish society derives greater value from its data resources.

There is widespread awareness of the forthcoming EU data spaces. To meet these requirements, further development of the collaboration may be necessary, as the current model does not readily scale to an international level. This remains a challenge to be addressed.

Conclusions and lessons learned

The Finnish and Danish experiences point to the same basic conclusion: health data and socio-economic data cannot be governed as separate worlds if they are to support high-quality research, policy evaluation and public decision-making.

First, cross-domain data access should be seen as institutional as well as technical issue. Even where organisations remain separate, user journeys, application procedures, reusable datasets, metadata exchange and regulatory interpretations can be better aligned.

Second, reusable data structures should be treated as research infrastructure. Preparing well-documented datasets in advance can reduce repeated extraction work, improve comparability, support data minimisation and make access more predictable for researchers.

Third, collaboration must respect different legal mandates while reducing unnecessary friction for users. A workable model should build on the complementary roles of statistical institutes, health data authorities and data holders.

Fourth, progress should start from concrete, recurring use cases where combined health and socio-economic data are clearly needed and public value is evident. This creates practical experience, trust and immediate benefits.

Finally, European data spaces make these lessons more urgent. If health data spaces develop separately from wider socio-economic infrastructures, they will only partly meet the needs of research and policy analysis. The Nordic countries are well placed to show how sensitive data can be strongly protected while still enabling health to be analysed in its broader social and economic context.

Keywords: Microdata, collaboration, Nordic experiences.