



Microdata.no – A new way to analyse cross-national registry data in the Nordics?

*Kenneth Schnelle, Statistics Norway, sck@ssb.no, Rune Gløersen, Statistics Norway, rgl@ssb.no,
Johan Sjøberg, Statistics Norway, sjj@ssb.no*

Abstract.

The research and analysis service microdata.no provides users from pre-approved institutions with immediate access to Statistics Norway's (SSB) register data through a privacy-by-design analysis platform. Automatic statistical disclosure control (SDC) enforces statistical confidentiality by returning only anonymised outputs. Consequently, end users in microdata.no do not process personal data, which removes the need for additional privacy protection measures.

We present an extension that enables secure, multi-owner data sharing and cross-register linkage. External data controllers (currently the Norwegian Institute of Public Health (NIPH) and the Norwegian Patient Registry) can upload data into dedicated datastores within microdata.no where they control their own data.

This architecture could open opportunities for collaboration between Nordic national statistical institutes, which could become data controllers offering data for analyses in microdata.no, enabling comparable studies on executed variables across the Nordic countries. The same SDC rules will apply at query time across all participating datastores, giving end users a unified, low-barrier environment for reproducible, cross-register analyses. Equally, the same analysis could be reused and executed on harmonised samples/population frames across the Nordic countries.

The data controller enters into a data processing agreement with the owners of microdata.no, who act as the data processor. This agreement ensures compliance with GDPR requirements to protect personal data from unauthorized access, as well as the confidentiality obligations outlined in the Norwegian Statistics Act. These contractual agreements, being legally compliant with GDPR and national statistics acts, could also form the legal basis for cross-Nordic cooperation using microdata.no as a common infrastructure.

A data-admin interface provides a variety of services to help data owners manage data provisioning and monitor the usage of their data.



This contribution shows how extending microdata.no from a single data provider solution to a governed, multi-tenant platform can (i) widen access safely, (ii) reduce time-to-data for researchers, and (iii) support cross-border, register-based research in the Nordic countries.

Keywords: Multi-register, cross-border, data-sharing, research

1. Introduction

Access to cross national register data in the Nordic countries is currently constrained by fragmented legal frameworks, decentralised data ownership, and time-consuming approval processes. Several attempts to realise a one-stop-shop concept for cross-Nordic research have failed, leaving researchers to submit parallel applications to each data holder, navigate differing access rules, and conduct analyses within isolated environments, which limits comparability and slows down research.

The microdata.no service builds upon concepts that are unique in offering strong research functionality while ensuring privacy for data subjects. Consequently, end-users are not responsible for processing restricted data, leaving immediate access to analyse all data controlled by microdata.no without prior application. We believe that a coordinated use of microdata.no by Nordic statistical institutes could bring a useful solution to significantly speed up access to restricted Nordic data for comparative research and analytics.

2. About microdata.no

Available since 2018, microdata.no is a metadata-driven platform, that give pre-approved institutions easy and immediate access to microdata for research and analyses. Microdata.no is fully compliant with the requirements outlined for Secure Processing Environments (SPEs) in the European Data Governance Act. This includes compliance with the more specific requirements outlined by the European Health Data Space (EHDS) programme through the Secure Provision and Use of Health data (SPUHiN) project. In addition to the common security aspects for prohibiting any type of unauthorised access to the data kept within the SPE, microdata.no has unique capabilities to ensure

- that any output presented to the end users are either metadata information or anonymised statistics/analysis results



- that it is impossible for the end-user to export the real data which is processed within system back-end
- that the end user cannot in any way access or inspect the back-end real data
- that the end user is unable to manipulate the analytical tool in order to reveal confidential information, and that the user can be stopped if their behaviour indicates attempts of such misuse.

These capabilities enable end-users to be granted immediate access to the service as soon as the employer has signed an institutional user contract. Furthermore, that the end-users can analyse any data available by the service without any further application or approval for the purpose of the analysis. Close to 80 pre-approved institutions as defined by Eurostat and the Norwegian Research Council as research institutions, as well as public administration and private consultant agencies (during commissioned work) subscribe at a low cost (ref 1), giving access for more than 700 employees and 200 students to the platform. Currently we have recorded approximately 500 publications and more than 80 master theses where microdata.no is cited (ref 2). The end-users can freely link, tabulate and perform analysis on more than 700 different variables from register data managed by Statistics Norway.

Microdata.no is owned and maintained in cooperation by the Norwegian Agency for Shared Services in Education and Research (Sikt) and Statistics Norway (SSB) and has been developed with fundings from the Research Council of Norway (RCN) and their National Financing Initiative for Research Infrastructure (ref 3).

Further development in the currently on-going microdata.no 3 project is partly financed by the Norwegian Digitalisation Agency's co-financing scheme.

3. Legal framework

The legal framework that regulates the usage of microdata.no can be described in four dimensions:

1. The legal framework that regulates privacy protection for the registered persons and business stated by the GDPR, Statistical Act and other specific legal regulations.



2. The user agreements which outline the conditions for use, where the employer must sign an employer agreement stating the responsibility of the institution, under which every employee must sign an end-user agreement.
3. The contracts and agreements that set up the obligations and conditions for the shared ownership between the owners Sikt and SSB.
4. The service and data processing agreements that regulate the conditions for providing data for linkage and analyses in microdata.no. The data processing agreements are tailored to the legal framework mentioned in item 1 above.

These contractual agreements, being legally compliant with GDPR and national statistics acts, form the legal basis for cross sectorial cooperation to use microdata.no as a common infrastructure for research and analyses of protected data in Norway.

4. The cooperation with Norwegian health data holders

The service microdata.no has been further developed to enable other public data owners to provide and manage their own data within the service, so that end-users can link and analyse their data with data from SSB. NIPH has in close cooperation with the owners of microdata.no been testing their possible use of the service for this purpose. NIPH is responsible for several national health registers, including the Cancer Registry of Norway (CRN) and the Norwegian Patient Registry (NPR). All these health registers are regulated by the Norwegian Health Registry Act. In addition to developing the technical capability of the service to provide for different data owners to manage their own data, we have developed a comprehensive set of templates and other framework concerning the necessary data processing agreements needed (i.e. the fourth dimension of the legal framework). This framework also includes testing and risk assessment procedures and documentation.

5. Multi data owner services

A set of tools (Microdata-tools) has been developed to help data owners to control and manage their own data (ref 4). Data to be uploaded into microdata.no must be curated according to specifications set out by the service (ref 5). This includes metadata that must comply with the requirements of the Microdata Information Model (ref 6). A validation tool is available to check formats and consistency between data and



metadata before uploading data to the microdata service that currently runs on the secure on-site infrastructure managed by SSB.

After the production of data and metadata files, the data administrator accesses a secure data transfer service, that enables encryption and transfer of the files directly to their own designated server controlled by the microdata service.

After decryption of the data inside the SPE of the microdata service, each entity-identification is replaced by a unique microdata pseudonymous identifier, while the original identification is completely removed. Then the data are uploaded into the designated data store within the microdata service. From here, each data owner manages their own data through a specific data administration user interface. This allows for the data administrators that has been granted such role and access by the data owner, to test and verify that the data are correct, to publish the data upon their own decision for end-use and also, if needed, to withdraw access to any of their published datasets whenever necessary.

6. Detection of end-user behaviour

User activity within the service is continuously monitored to detect patterns that may indicate attempts to compromise confidentiality. This includes automated logging and analysis of user behaviour. In addition, all users are required to formally acknowledge that any attempt to circumvent or breach the confidentiality mechanisms of the service is strictly prohibited and may constitute a violation of the terms of use.

We believe that the capabilities of the microdata.no service, given the compliance with relevant European legal framework in combination with quite harmonized Nordic statistical acts constitutes a realistic framework for a similar cooperation between the Nordic statistical offices, as we have established between NIPH and SSB.

7. Use Case: Nordic Data Integration and Cross-Country Analysis

The microdata.no service provides opportunities to utilize Nordic data for research and analyses within a common analytical platform with built-in privacy control. Once the legal framework is established, mostly any register-based data reused for official statistics could be made available for researchers without any further application



procedures. This enables comparative analyses across countries, for example by examining differences and similarities in social, economic, or health-related conditions.

To achieve this, the partners must standardise metadata and curate data according to the microdata.no protocol. This protocol is compliant with all FAIR principles, making data accessible, findable, interoperable and reusable. In addition, timespan of timeseries, time of observation and other measures to ensure comparability must be aligned. Such standardisation sounds challenging, but the basic requirements are already taken care of within the European Statistical System to provide comparable statistics across Europe.

Furthermore, the service would allow researchers to combine data across countries to study larger populations, which is particularly relevant in cases where the number of observations within a single country is too limited to produce statistically significant results. This situation often arises for rare conditions or diagnoses.

In addition, the system is designed to support transparency and reproducibility. All analytical steps can be documented and replicated within the platform, facilitating verification of results and enabling other researchers to reproduce the analyses under the same conditions.

8. Additional work to provide for international use of microdata.no

Microdata.no has been developed to be used for international cooperation. The user interface is in English, and the syntax of the analytical programming language and scripts is in English. However, some remaining actions must take place to provide for international use.

Microdata.no has built in multilingual capabilities, but currently all metadata about the Norwegian data remains to be translated to English.

Further development is needed to open the service for international use. We have been waiting for the EIDAS2 developments to utilize national services to authenticate foreigners. This is not yet in place. However, we can take measures to accommodate for authentication within the European eduGAIN network or investigate other possible measures for secure user authentication and authorisation.



Being in a third round of major development activities, part of these obstacles can be prioritised, given an interest in using the service. One important goal in the cooperation with NIPH is that we will develop functionality where data exploration using microdata.no or even using microdata.no to define the population requested upon application could be used to speed up the time-to-data for research purposes.

In a longer-term perspective, a common use of microdata.no could lead to a joint interest in other developments where microdata.no could add value for secondary use of statistical microdata.



9. Appendix: List of references

- (ref 1) <https://www.microdata.no/en/prices/>
- (ref 2) <https://www.microdata.no/bibliografi/>
- (ref 3) <https://www.forskningsradet.no/en/financing/what/infrastructure/>
- (ref 4) <https://github.com/statisticsnorway/microdata-tools>
- (ref 5) <https://www.microdata.no/wp-content/uploads/2022/06/Key-Architecture-Choices-in-Microdata.pdf>
- (ref 6) [https://www.microdata.no/wp-content/uploads/2022/02/FAIR Data Versioning in Microdata.pdf](https://www.microdata.no/wp-content/uploads/2022/02/FAIR_Data_Versioning_in_Microdata.pdf)