A new approach: Accessing individual-level transaction data

Kristin Egge-Hoveid, Statistics Norway, keg@ssb.no

Abstract.

This paper describes Statistics Norway's effort in accessing individual-level transaction data from banks and grocery chains to improve household consumption statistics. Early attempts tested both the Statistics Act and GDPR and triggered controversy and criticism from the Data Protection Authority. This led to a comprehensive reassessment of legal, methodological and institutional frameworks, and a new approach based on regulation and privacy by design. Through data minimization, sampling and delayed collection, we suggest how privately held individual-level data can be used in official statistics while safeguarding trust and legitimacy.

Key words: Privately held data, data minimization, privacy measures, GDPR implementation, legal framework.

1. Introduction

Privately held third party data has become increasingly important for national statistical offices. These data sources provide granular, timely, and cost-efficient information that can enhance the quality and relevance of official statistics. While Statistics Norway (SSB) has used aggregated data from private actors for decades—for example in the consumer price index and business statistics—the use of such data linked to individual identifiers to produce official statistics represents a qualitative shift.

SSB has sought access to privately owned individual-level transaction data since 2017. The aim has been to improve the quality of official statistics on household consumption (the Household Budget Survey - HBS) by using receipt data from grocery chains and linking these to household information through register data and bank transactions.

However, SSB's early attempts to obtain transaction data tested both the new Statistics Act and the application of the General Data Protection Regulation (GDPR). In 2022, SSB imposed a legal obligation to provide data on the country's largest grocery chains and NETS Norway as payment transaction provider, through individual administrative decisions under the Statistics Act. Following controversy and complaints from data owners, the

Norwegian Data Protection Authority prohibited the data collection, concluding that individual decisions did not provide a sufficient legal basis for processing data with such intrusion on privacy[1]. This experience led to a comprehensive reassessment of how legal, methodological, and institutional frameworks can be adapted to enable the use of such data while safeguarding the legitimacy, reputation and public trust of the NSI. This reassessment may mark a shift in how we regulate and process data, when the source is privately held third-party data at individual level - where privacy and proportionality are not treated as a one-off assessment but are instead explicitly regulated and embedded in the processing model – also known as privacy by design.

2. Transaction data and household consumption statistics

Challenges related to household consumption statistics clearly illustrate the need to explore new data sources. Traditional household budget surveys (HBS) rely on self-reporting and recall questions, and are costly, infrequent, and increasingly affected by declining response rates and measurement error.

Transaction data represents a data source that can complement the traditional Household Budget Survey. At SSB, interest in transaction data emerged in connection with plans to modernize the HBS for the 2022 data collection. The goal was to move toward a system where surveys, administrative data, and privately held transaction data complement each other, with receipt data replacing the need to report grocery purchases. Groceries account for around 70 percent of the self-reported items in the traditional diaries used in the HBS.

We saw the potential of reducing the response burden by simplifying the survey and at the same time enhance the quality of the statistics. Over time, transaction data could potentially also serve as a standalone source of statistics with better timeliness than surveys.

[1] This process has also been described in Linnerud and Egge-Hoveid (2022) and Sæbø and Dimaskos (2023).

3. The 2022 data collection

After several years of preparation, including testing, data analysis, dialogue with data owners, cost-benefit analysis and thorough assessment of privacy, Statistics Norway imposed a reporting obligation through individual administrative decision pursuant to Section 10 of the Statistics Act on the largest grocery chains in Norway and the data processor for debit card transactions (NETS Branch Norway). Within six months, millions of transactions were streamed to SSB in real time. Internally, there was strong confidence in the project. SSB had a clear public mandate, broad legal authority through the revised statistical act of 2019, and the data held significant technical and analytical potential. In addition to improving the quality of the statistics, they had the potential to reduce response burden for both individuals/households and businesses and contribute to more cost-effective statistical production.

Nonetheless, controversy emerged as media coverage and public debate revealed a clear mismatch between our legal and strategic reasoning and the public expectations regarding what data a statistical authority should have access to. Reporting entities and the media questioned the purpose of the data collection, intrusion on privacy, and the real time streaming of data. Following complaints, the Norwegian Data Protection Authority prohibited further data collection, concluding that the planned collection of transaction data was too intrusive from a privacy perspective to be approved through individual decisions under section 10 of the Statistics Act. They also argued that the proportionality requirement was insufficient. The data collection was stopped, and all collected data were deleted.

The experience from 2022–2023 highlighted an important lesson: legality and legal mandates are necessary, but not necessarily sufficient. The controversy that arose did not concern data quality, definitions, or methodology - issues statistical authorities are more likely to be criticised for, but rather questions of privacy, proportionality, data security, and institutional trust. Such criticism had the potential to undermine trust in Statistics Norway as an institution, even if the legality of the data collection may have been in place.

4. Reassessment of the legal and institutional framework

Section 10 of the Statistics Act[2] provides a broad legal authorization. It establishes a basis for imposing a duty to provide information, including third-party data, when this is necessary for the development, production, and dissemination of official statistics.

However, when technological developments increase data volumes, enable more detailed information, and may intensify intrusions into the private sphere, a fundamental legal policy question arises: how should public authorities exercise such broad mandates? It can be argued that the use of new data sources, such as individual-level transaction data, increasingly challenges the distinction between what is formally lawful and what is legally and democratically legitimate.

Following the prohibition issued by the Data Protection Authority (DPA) in 2023, Statistics Norway (SSB) decided not to appeal the decision. Instead, SSB established a direct dialogue with the DPA and carried out a comprehensive reassessment of how individual-level transaction data could be collected in ways that significantly reduce privacy risks. This reassessment identified concrete measures, both legal and methodological, that enhance privacy while still preserving statistical utility.

Based on the dialogue with DPA and the reassessment of the legal and institutional framework, SSB concluded that future collection of individual-level transaction data should be regulated through *a dedicated regulation* under the Statistics Act. By choosing a dedicated regulation rather than individual decisions, SSB not only defines the obligations of others, but also sets limits on our own discretion. This marks the first time in history that SSB uses a dedicated regulation as the legal basis, representing a significant institutional shift. While an individual administrative decision binds only the data provider and can easily be changed through a new decision, choosing a regulation also imposes stricter requirements on data processing for the statistical institution. These institutional boundaries are operationalised through legally binding requirements for the design and processing of the data — what in privacy terminology is often referred to as *privacy by design* — now elevated from an internal principle to the level of regulation.

[2] *Act relating to official statistics and Statistics Norway (the Statistics Act)*

Dedicated regulation provides:

- greater predictability and equal treatment for reporting entities
- clearer boundaries for SSB
- a more robust and long-term legal basis
- democratic anchoring through public consultation

Regulation enhances democratic legitimacy by anchoring intrusive data collection in a transparent and democratically processed framework. This directly addresses the Data Protection Authority's emphasis on predictability and proportionality. This may be a necessary step in all future use of large and complex data sources when it is collected on an individual level. A general and prior proportionality assessment is often not enough. The risk is not only whether data are processed, but in how the processing is organised, sequenced, and limited both legally and in practice.

5. Data minimisation as methodological design (privacy by design)

To utilize individual-level transaction data for producing statistics on consumption across household types, income groups, and other socioeconomic categories, SSB plans to link receipts from the four largest grocery chains in Norway with payment transaction data from NETS using the shared linkage keys found in both data sources: **Time** of purchase, **location** of purchase (store) and total **amount** of the purchase.

In the 2022 data processing design, the design involved linking data for the entire population. Although several privacy measures were planned, there was no dedicated regulation in place. The proposed dedicated regulation[3] under the Statistical Act now introduces substantial data minimisation measures that are legally binding and that significantly reduce privacy risks.

First, data collection will be **delayed rather than real-time**. Transaction data will be collected with a defined time lag, reducing the possibility of continuous linkage, but still maintaining sufficient timeliness for statistical purposes.

Second, only a **small sample**—rather than full population data—will be used. A random sample of up to 10 percent of receipts and 10 percent of payment transactions linked to households (nano-samples) will be drawn, and the remaining data deleted. Approximately

[3] SSB (2026): *Høring – forslag til forskrift om opplysningsplikt for transaksjonsdata* (Norwegian only)

1 percent of the two 10 percent nano-samples are expected to be matched, corresponding to about 2.5 million receipts. On average, each Norwegian household would contribute to one receipt. Compared to traditional surveys based on self-reporting, this represents a substantial increase in data volume.

Third, **data will be collected sequentially**. SSB will never hold full datasets from both sources simultaneously. Payment data will be collected first, sampled, and reduced before receipt data are collected.

Fourth, transaction and receipt **data will be stored and processed in separate secure environments**. Processing will be highly automated, with no individual or system having simultaneous access to both environments. All activities will be monitored and logged.

Together, these measures represent a model where privacy is embedded directly in the architecture of statistical production—privacy by design.

6. Conclusion and implications for official statistics

The Norwegian experience suggests that new approaches are required when integrating privately held big data at the individual level in the production of statistics. The issue is not only questions on legal authority but how NSIs exercise broad mandates in a context where technological capabilities increase data volume, detail level and privacy intrusiveness.

Transaction data—especially receipts—are distinct from other types of data: while not the most sensitive data used in official statistics, they are data citizens do not typically expect the government to access.

In 2022, SSB applied a traditional model long used for administrative data: collecting full population data under broad legal mandates, without predefined limits on further processing, combined with real-time streaming. This, we experienced, challenged both legal frameworks, interpretations of GDPR, and institutional trust.

The new model - based on binding regulation, delayed collection, data minimization, and separated processing with privacy-preserving linkage – suggests how privately held individual data can be used in official statistics. If SSB moves toward implementation of the proposed regulation in 2026, these experiences can offer relevant lessons for other statistical offices when approaching the use of third party privately held data. Access to

such data is not merely a technical issue, but one of governance, trust, and societal acceptance.

7. Appendix References:

Linnerud, K. og Egge-Hoveid, K. (2022). *Big data for HBS – gains and lessons learned*. Statistics Norway. Paper presented at the Nordic Statistical Meeting 2022 Accessed from: https://static1.squarespace.com/static/606f36b890215d7048ddaac0/t/62ed312d3848337951e7832f/1659711790061/BIG+DATA+FOR+HBS+%E2%80%93+GAINS+AND+LESSONS+LEARNED.pdf

Act relating to official statistics and Statistics Norway (the Statistics Act): Accessed from: https://www.ssb.no/en/omssb/ssbs-virksomhet/styringsdokumenter/statistikkloven/_/attachment/inline/15f00d0d-322a-4b96-bfcb-a0159f76e2c2:165eaa37f1aae978f2a570066c4ad86830ae2094/Statistikklov_ENGELSK_red29des2020.pdf

Statistics Norway (2026): *Høring – forslag til forskrift om opplysningsplikt for transaksjonsdata* (Norwegian only). Accessed from: https://www.ssb.no/omssb/ssbs-virksomhet/hoering/_/attachment/inline/fc4f0103-8c18-4d52-9b7e-286b637a13bf:1c04e8efc208481a663e30bf87814b1cd078356f/2026-066-06_notat_2.pdf

Sæbø, H. V. og Dimaskos, X. (2023). *Use of transaction data in official statistics – experiences and challenges from Norway.* Paper presented at the ISI World Statistics Congress, Ottawa 2023. Tilgjengelig fra: https://www.isi-next.org/media/abstracts/ottawa-2023_e284ccef1ba0821348330e5e1b7de8ba.pdf