NSM NORDIC STATISTICAL MEETING 2026

From Batch to Streaming: Preparing Automated Testing for Faster Official Statistics

Martin Handeland Skjæveland, Statistics Norway, martin.skjaeveland@ssb.no

Abstract

Statistical offices face an increasing demand to produce faster and more frequent statistics, while maintaining high standards of quality. Traditional batch deliveries from data providers are inherently vulnerable to delays and errors. Moving from batch processing to continuously streamed data from APIs offers opportunities for short-term official register statistics. However, it also poses its own set of risks and challenges, and to realise the potential of the streaming approach, automated quality checks are required to identify errors, missing information, and inconsistencies. Statistics Norway (SSB) is currently modernising the data collection process from the *a-ordning*, a collaborative system shared by three agencies that provides information about employment, remuneration in cash and kind, and taxes. [1] Data from the a-ordning are hereafter referred to as the *a-melding*.

In the process of developing a new system for collecting the a-melding, we consider automated testing a key element in ensuring efficiency, timelines, and reliability. Though collected in one integrated system, the a-melding is shared with SSB through multiple API-services. For example, information related to employers' national insurance contribution bases appear in multiple API-services at the levels of employment relationship, establishment, and enterprise. Such relationships enable reconciliation checks to assess completeness and consistency of information on jobs, individuals, enterprises, and incomes across API-services.

To support continuous monitoring, reconciliation checks must be fully automated. We present practical experiences from implementing such tests using automated testing frameworks, initially on historical data. These experiences contribute directly to improving the quality and timeliness of the a-melding data product, which is crucial for many statistics, including employment statistics, earnings statistics, and national accounts.

Keywords: automated testing, API, data reconciliation

[1] The Norwegian Tax Administration (NTA) administers the a-ordning on behalf of SSB and the Labour and Welfare Administration. Data for all industries and individuals are reported monthly in a one a-melding, and this is the main source for producing statistics on earnings and the labour market.

1. Introduction

For over ten years, statistics on the Norwegian labour market have flourished on data from the a-ordning, hereafter referred to as the a-melding. This rich data source is a fundament for many statistics, as it provides essential information on a wide range of variables for all jobs each month. The Tax Administration is currently working on new API-sharing services, allowing us to acquire data in near real time, as opposed to receiving them in monthly batches. This provides great new opportunities for faster dissemination of monthly statistics, but realisation of these opportunities rests on our ability to continuously survey the integrity of the data collection process.

Since the a-melding is one comprehensive data source with several levels of entities, a large number of records, and many variables, assessing validity and integrity is challenging. Streaming data, meaning that we collect all data per unit and event, instead of receiving just the latest event data per unit in batches, significantly increases the amounts of records, thus enhancing the challenge.

Increasing data frequency makes manual controls at a comparable level impractical due to the required resource intensity. To prepare for this approach to collect data, we see automated testing as essential. However, it should be emphasised that the aim of this paper is not to present tests that cover all aspects of data quality, such as identifying severe reporting errors. Instead, we focus on specifying tests detecting faults arising from data collection and data processing failures, even though automated testing should be applied to other areas as well.

Although we are still in the initial phase of data collection, analysing historical data allows us to establish test expectations. These expectations should also apply when collecting the same data from APIs, providing a useful basis for understanding how tests are likely to function in practice.

2. The a-melding

The a-melding is a service used by employers to report information about income, tax, and employment. The service entered into force on 1 January 2015 with the aim to simplify and coordinate reporting to the Norwegian Tax Administration (NTA), Statistics Norway (SSB), and the Labour and Welfare Administration in one a-

melding. NTA administers the scheme on behalf of the other public agencies. Employers provide the a-melding monthly, which is then utilised by authorities to make tax assessments, collect national insurance contributions, process cases concerning benefits, and more. As such, it works as a register of the labour market.

The a-melding is a direct source for statistics on employment, earnings, and national accounts to name a few. Moreover, a-melding data are made available by SSB for researchers as well as other institutions through the microdata service microdata.no.

Throughout the paper, for space limits, we focus especially on parts of the a-melding relevant for statistics on the labour market and earnings even though it comprises other areas as well.

Every year, 260,000 employers submit information concerning 4.6 million employees and pensioners. The deadline for submitting the a-melding is the 5th of each month, or the following working day if the date falls on a weekend or public holiday. Employers can also submit changes, which increases the amount of data reported significantly.

2.1 The structure of the a-melding

The a-melding contains a wide range of variables on different entities and levels. The levels from top to bottom are:

1) Employer/enterprise
2) Local unit
3) Employed/income recipient

An enterprise is an organisational unit producing goods or services with a certain degree of autonomy in decision-making (Statistics Norway, 2025). At the enterprise level, we will find sums of different variables like employer's national insurance contributions (pay roll tax) and withholding tax. These are the amounts that the employer needs to pay to the tax authorities.

Each employer/enterprise can carry out one or more activities and therefore consists of one or more local units (establishments). A local unit is a geographically delimited unit where economic activity is carried out, typically within a specific industry. At this level, information on the basis for employers' national insurance contributions is

recorded, which forms the foundation for the aggregated amounts reported at the enterprise level.

On the lowest level, we find the income recipient, most often an employee. Examples of information reported on this level are remuneration, withholding tax, occupation, working hours, start and end date of employment, and leaves.

Each enterprise and local unit is identified by a unique organisation number, while persons are identified by national identity numbers.

2.2 Data sharing practices from the Tax Administration to SSB

Each reporting period (month) is gradually populated with data as reports are submitted to the Tax Administration. Batches are currently provided to SSB on the 10th each month, five days after the deadline of submission for the enterprises in the previous month.

We are now taking steps toward more frequent data collection through streaming from several different APIs provided by the Tax Administration where the most important services are referred to as:

- Enterprise (Opplysningspliktig)
- Local unit (Underenhet)
- Income (Inntekt)
- Employment (Arbeidsforhold)

As indicated by their names, the first two services offer information on the enterprise and local unit, while the others contain information concerning the individual as an employee or recipient of an income.

3. Testing scope

Data validation is nothing new to national statistical institutions. We understand data validation in a broad sense: the process of ensuring that data meet specific quality standards before used in statistics production or analyses. In this paper we primarily address the challenges of validating data consistency across systems and datasets. We emphasise data reconciliation, understood as “the process of comparing and

verifying information across systems to ensure data integrity, data accuracy, and data consistency" (IBM, 2026).

Furthermore, we focus on validating "source data" and "input data" as referred to in Statistics Norway's (2026) description of steady states of data. However, although beyond the scope of this paper, it is advantageous to expand these validation procedures to "processed data" as well as "output data" to validate the whole chain from collecting data to dissemination of statistics.

4. Testing strategy

A robust strategy for automated testing requires a set of tests that cover multiple aspects of the data's properties. At the same time, the test suite should be transparent, easy to understand, and straightforward to maintain. To meet these requirements, we must be selective in deciding which tests to include.

Test results should be produced with minimal human intervention and as little manual verification as possible. While in-depth exploration of data is often necessary, this should be done on the assumption that the data are complete and internally consistent.

The design of tests should be aligned with the specific types of errors the checks are intended to detect. Reporting errors differ from errors arising in data collection. Errors related to data collection are the primary concern because they would compromise the integrity of the data before any downstream processing takes place. Although significant reporting errors must be detected and subsequently addressed—either through follow-up with reporting units or through corrective measures during data processing—ensuring the integrity of the data collection process must come first. This is the focus of the paper.

One challenge in designing a test suite is that it is not always clear what requirements the data should be expected to meet. Automated tests rely on explicit expectations, yet even when clear relationships exist in the data, minor exceptions are often unavoidable. As a result, reconciliation checks would have to tolerate a degree of discrepancy. A recurring question is therefore where to set appropriate thresholds. Furthermore, we do not want to raise alerts too frequently, as this would increase the need for manual follow-up. Although some manual inspections are

necessary, limited resources must be targeted at the controls that have greatest impact on data quality.

Historical experience from more than ten years of using the a-melding in statistical production will be highly valuable in striking a balance between too few and too many checks. In addition, robustness can be improved through triangulation, where several complementary tests are used. While a single failed test may not justify an alert, multiple failures occurring simultaneously might. Outputs should be logged continuously, and results could be made available to users in various ways, for example through email notifications and more detailed presentation in some dashboard.

We have identified various categories of tests that will be applied to the a-melding. Each test comes with its own advantages and are meant to complement each other. Here we focus on reconciliation checks.

5. Automated reconciliation checks

Given that we must acquire data from multiple APIs to obtain the full a-melding for each enterprise, data collection from these APIs must happen in parallel. Moreover, information elements that are linked and related adhering to separate APIs allow us to assess the consistency of data collected from the different services. Therefore, such checks serve as a tool to assess whether collection procedures from separate APIs work properly. For example, it can help us indicate if elements in one API are missing by linking and comparing with data elements from other APIs.

An advantage of this approach is that, as we collect the a-melding from all APIs in parallel, automated checks can run as frequently as data are streaming in without too complex and high-maintenance designs. Thus, we can detect potential failures quickly. We have identified two approaches to testing internal consistency.

5.1 A sub-entity should always join with a higher-level entity

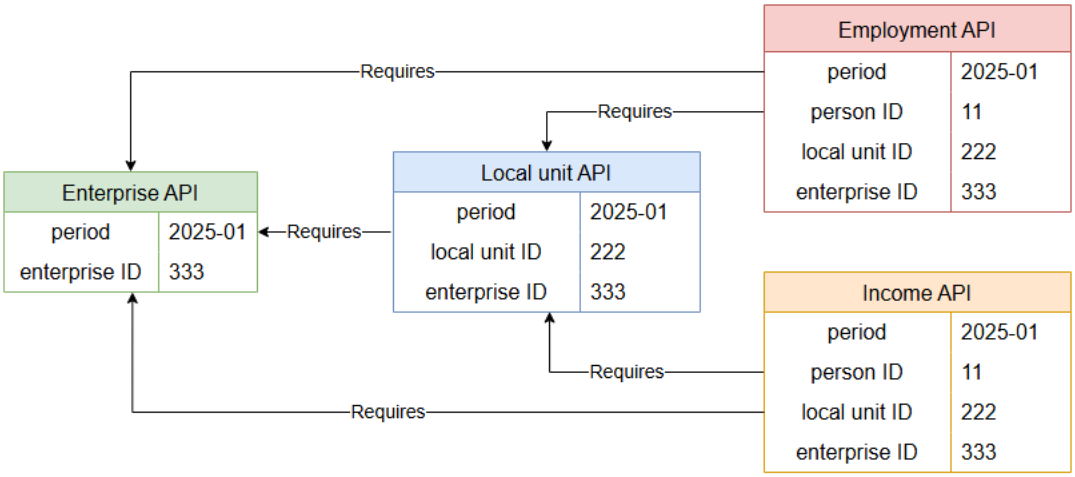
As noted earlier, reporting in the a-melding are made on multiple entities in a hierarchical structure. This means that for every reported income relationship or employment, all related higher-level entities must also exist. Specifically, the income recipient or employed is always associated with a local unit, and that local unit links to an employer.

From this hierarchy, the following requirements follow:

- If there is a local unit, income, or employment relationship reported, there must also be a corresponding record for the employer in the Enterprise API.
- For each reported income and/or an employment relationship, the local unit to which the income or employment relationship belongs must also exist in the Local unit API.

In short, lower-level entities can never exist in isolation; they must always be supported by the relevant higher-level entity in the hierarchy. If higher-level entities are missing, this would indicate that we are missing information on these entities, for instance because of errors in the data collection process for an API.

Figure 1: Illustration of relations between the a-melding APIs

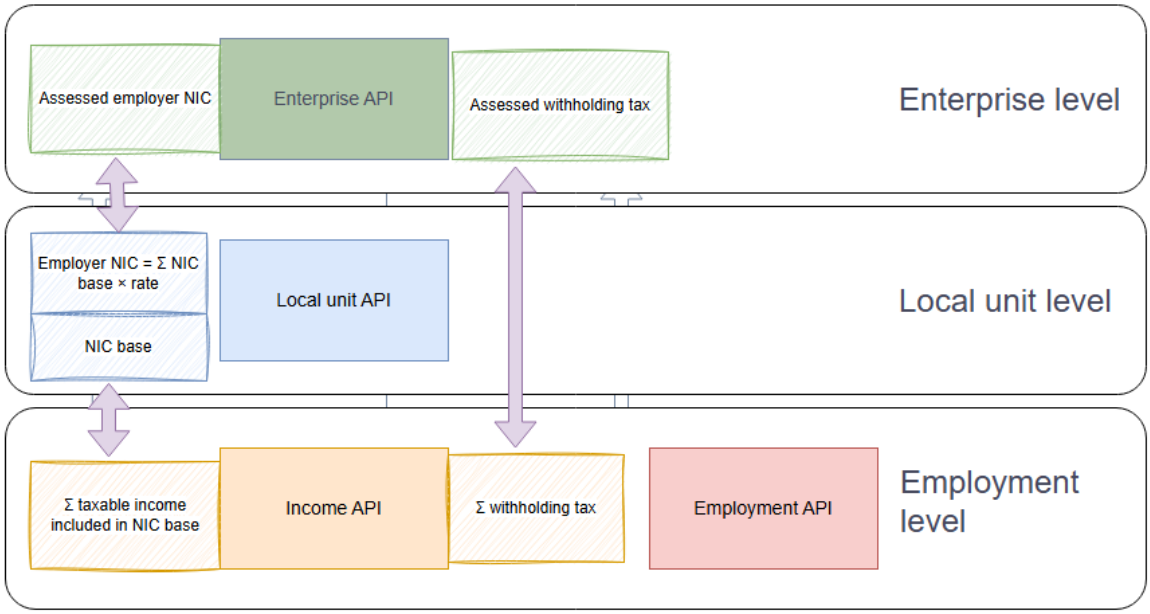


5.2 Aggregates should be internally consistent

Another way to assess internal consistency is to compare aggregates that, by construction, should be identical—or at least close—to each other. Such checks are particularly powerful when they rely on variables derived from the same underlying information.

Two especially useful variables in this respect are employers' national insurance contributions (NIC) and withholding tax. This is because both variables are calculated from reported remunerations, meaning that reported totals should be reproducible from lower-level data.

Figure 2: Illustration of relationships between aggregates across APIs



Historical data show that in more than 99 percent of cases (enterprise X period), these sums add up. Nonetheless, we do see deviations especially when it comes to NIC. In cases where special rules apply discrepancies will often occur. In other cases, the employer has simply calculated wrongly. Such errors are often accompanied by business rules that have been broken and the reporting must be corrected. We see a more consistent relationship between lower-level and higher-level amounts for employer NIC and withholding tax, perhaps related to more uniform requirements and stricter business rules.

From this, we know there are exceptions to the norm, not necessarily caused by incomplete data. There could also be errors in what is reported or just legitimate deviations in the data because of various exceptions. In the context of testing the data collection process, it is important to separate inconsistencies resulting from reporting and errors coming from mishandling of data.

One way to deal with this is triangulation—evaluate each unit by multiple tests of different aspects of data. For example, one employer reporting paying a higher NIC than what to expect from calculation of contribution basis and contribution rates for its local units, is not enough to fail the test as this could simply be a reporting error. But if this deviation is combined with a mismatch of sum withholding tax and lower-level summed withholding tax, it raises a stronger suspicion. Moreover, it would be particularly suspicious if multiple types of mismatches occur for several employers at the same time.

6. Discussion

We have presented the powerful potential of automated reconciliation checks as a tool for securing consistent and accurate streamed data. Other tests should work complementary to reconciliation checks. For example, although we are missing all data reported within a time interval, our data could be completely consistent, thereby passing all reconciliation checks. Useful complementary tests could for example be stability checks; testing whether there are changes from one period to the next not explained by trends, seasonal components, etc. Missing data could also be detected by checking for gaps in between timestamps; from historical experiences we expect a-meldings to be reported any and all days of the year, so no reporting on any day would raise suspicion. Such types of tests could also be automated.

A particular advantage of reconciliation checks is that these checks are available continuously, while stability checks for example, are difficult to do before the reporting deadline as data are completer and more comparable to data from previous periods. For a limited period, in a migration phase, we will have access to both the batch-based system as opposed to streaming data. Then, fine-grained comparisons are available that would help us find very discrete errors. However, we must still wait until we have received the monthly batch. In the meantime, we can do a lot of internal reconciliation tests while migrating from batch to streaming.

Reconciliation checks can on the other hand detect whether we are missing or receiving faulty data from one API while being provided with complete and consistent data from the others.

Changes in legislation may affect the accuracy of certain consistency checks. For example, authorities introduced a temporary additional NIC in 2023 subsequently abolished in 2025. During this period, the NIC base was defined differently, requiring minor adjustments to the aggregation logic used in the internal reconciliation checks. In situations such as this, leaning on a range of tests ensures stronger robustness. However, it also shows that reconciliation tests require maintenance, underlining the importance of keeping it simple to reduce costs.

An interesting further development is to do complete end-to-end testing by expanding data reconciliation efforts to the statistics products or at least more comprehensive surveys of the coherence between input data and published

statistics. They are not supposed to be completely equal in aggregate because of population delimiting for example, but a degree of coherence is expected. Moreover, reconciliations could be conducted across data products within the organisation. There is for example a strong relationship between income reported in the a-melding and income in the tax return register, a source for statistics on tax for personal tax payers.

References

IBM. “Data Reconciliation.” *IBM Think*. Accessed April 28, 2026. https://www.ibm.com/think/topics/data-reconciliation.

Statistics Norway (SSB). *“Definitions of Unit Types.”* Last modified September 4, 2025. https://www.ssb.no/en/metadata/definisjoner-av-statistiske-enheter.

Statistics Norway (SSB). *Steady States of Data in SSB, Version 3.0*. Notater 2026/05. Oslo: Statistics Norway, 2026. Accessed April 28, 2026. https://www.ssb.no/en/teknologi-og-innovasjon/metoder-og-dokumentasjon/steady-states-of-data-in-ssb-version-3.0/_/attachment/inline/2331a955-e22c-4f01-83b9-71c5cce3bdcf:2c0815657f403e56e8209b4a4551fcd7aeb08cbb/NOT2026-05_ENG.pdf.