Conseil national
de l'information statistique

Montrouge, 6 June 2025 – N° 88/H030

CNIS meeting on official statistics matching

Paris, 28 May 2025

Summary

Opened by Jean-Luc Tavernier, Director General of INSEE, and three years after the January 2022 conference on the same subject, this meeting responded to a recommendation from the consultation group which, under the aegis of the CNIS, had specified the tools envisaged by INSEE to develop matching, improve quality, harmonisation and security compared to what had been practised until then. Approximately 170 people participated.

Firstly, Christel Colin (Director of Demographic and Social Statistics) and Corinne Prost (Director of Methodology and Statistical Coordination) from INSEE explained the role played by the introduction of the non-significant statistical code (CSNS) in the development of matching within the official statistical service (SSP). This has led to consideration of its applications, particularly to avoid increasing the number of variables within the same database, which complicates analysis. Data matching can shed light on certain socio-economic phenomena, supplement fields of analysis, and improve production processes, provided the matched administrative or private data is relevant and of good quality. As it is not possible to answer every question, matching must be targeted towards the desired insight while respecting the principle of minimisation (only collecting the data necessary to achieve the desired outcome).

To fuel discussion on the issues surrounding matching, three examples were presented and discussed by a user from outside the official statistics service:

- The value of matching and its complexity were illustrated by Laurent Lequien, Deputy Director of Transport Statistics at the Statistics Department of the Ministries of Territorial Planning and Ecological Transition. Matching the vehicle registration system and the results of vehicle technical inspections with different sources of vehicle characterisation data, then, via the CSNS, with the demographic file on housing and individuals, or even data on subsidies for the purchase of low-pollution vehicles, provides detailed information on the profile of road vehicle users. This type of long-term work also provides a better understanding of the redistributive issues involved in the ecological transition. Clément Malgouyres, a researcher at Crest and the Institute for Public Policy, testified to the qualitative leap they have enabled in understanding the vehicle fleet and the public policies that affect its dynamics. He emphasised the benefits for researchers of having access to enriched files and, more broadly, to unmatched administrative databases such as those of the Services and Payment Agency (ASP).

- To illustrate the contribution of methodological matching, Yves Jauneau, head of the Labour Market Synthesis and Trends Division at INSEE, explained how it shed light on the recent divergence between the Labour force Survey and employment estimates. They made it possible not only to quantify the underreporting rates for certain categories of employed persons, but also to study the impact of the collection method on these rates. Magali Dauvin, an economist at the French Economic Observatory (OFCE), praised and emphasised the value of such work in improving employment forecasts.

- Another example presented was the InserSup system. Pierrette Schuhl, head of the ministerial statistics department for higher education and research, pointed out that it was in response to an interministerial request from 2021 in application of the 2020 research programming law. The result of matching student files with those of the labour market, it provides employment rates for most degrees and thus helps young people with their career choices. It will be gradually expanded (self-employment rates by the end of 2025, wider coverage of degrees, etc.). However, it needs to be supplemented by surveys, for example on jobs abroad. Nagui Bechichi, creator of the Suptracker website, which promotes this data, emphasised the importance of this tool, which links higher education and professional integration, both for guidance and for evaluating supply or reform policies.

Following these examples, Olivier Lefebvre, project manager for INSEE's RESIL project, explained what is contained in this Statistical Register of Individuals and Housing, emphasising the absence of the NIR (Social security personal number) and any 'thematic' data (income, marital status, occupation, etc.), and the matching service it provides to the official statistics. The audience was able to appreciate how far the project had come since its first presentation in 2022, in accordance with the provisions of the Consultation group and the legal framework set out in the decree of 7 January 2024. The decree does not allow researchers direct access to the enrichment service; however, they will be able to benefit from it through partnerships with SSP entities, if these lead to the production of official statistics, as is the case for co-constructing surveys; they will also benefit from the extension of the range of statistical files available.

To continue the day's discussions, Bertrand du Marais, President of the CNIS, led the final session dedicated to the contributions of a reference framework for matching. This framework, presented by Corinne Prost, was requested by the SSP to ensure that data matching complies with the principles of necessity (similar to the opportunity notices for statistical surveys), proportionality and minimisation of the data used, as well as the principle of transparency for all data matching carried out. Corinne Prost emphasised the importance of coordinating practices within the SSP and making these practices visible. Anthony Guérout, representing the Association of Mayors of France at the CNIS, discussed the needs of local authorities, which are both producers and major consumers of reliable, accurate and up-to-date data for developing, implementing and monitoring the impact of public policies at the local level. In addition to IT security issues, he emphasised the essential protection of individual freedoms. In this regard, he noted that during the population census, citizens were surprised rather than concerned about the collection of individual data concerning them, believing that the town hall already had this information. On behalf of the ministerial statistical services, Christelle Minodier emphasised the great usefulness of having a common framework for better harmonisation of practices and for organising dissemination to researchers at the Secure Data Access Centre (CASD). She mentioned the valuable use of RESIL in providing a reference population for calculating rates (prevalence, beneficiaries, non-take-up, etc.). François Clanché, Director of INED (National Institute of Demographic Studies), expressed the hope that the reference framework for official statistics enabling the development of linkages would open doors for research, but that sufficient visibility would still need to be given to new productions. He conveyed the researchers' request to clarify, without undue delay or limitation, what their possibilities for access to linkages would be. Dominique Meurs, professor of economics at Paris Nanterre University, reinforced this request by highlighting the contribution made to research on administrative data in the Netherlands by matching facilities, with flexible data access mechanisms and short turnaround times that are particularly suited to young PhD students.

Discussions with participants throughout the day revealed keen interest in the potential offered by the RESIL service and, more broadly, by the pairings made by public statistics, whether for knowledge purposes, to reduce the burden of surveys or for methodological work. While they reduce the survey burden by limiting the number of questions – a use case recommended

by the European Statistics Code of Practice – they raise ethical questions that justify the need to ensure compliance with the principles of transparency, proportionality and minimisation. The contribution of the reference framework for public statistics matching is therefore recognised. With regard to transparency, the CNIS will contribute by posting the list of matches on its website, which will be detailed on the SSP websites, following the example of INSEE, which will publish the matches made with RESIL. With regard to minimisation, it is a question of exercising discernment in the choice of variables selected when enriching a file by matching it with other sources, because 'not all data is valuable' and the cost of analysis is high. However, this principle raises concerns among researchers that they may be limited in their work, as they cannot necessarily identify in advance the variables that are most relevant to their analyses. Researchers are also keen to ensure that the wealth of statistical files resulting from matching does not lead to a reduction in the availability of files at the Progedo-Quetelet centre (production files for research – FPR) in favour of the CASD alone, access to which can be costly. To preserve the statistical confidentiality of the FPR, one solution is to make samples extracted from the comprehensive files available, a route that has already been taken. The conference recalled on this occasion that guaranteeing the IT security of data represents a cost, but one that is necessary to maintain public confidence in official statistics.

The debate also focused on how the principle of proportionality, which producers are required to respect under the GDPR, can be assessed during a CNIS review, both for a given project and in light of a set of pairings.

The day also highlighted the limitations of matching, which is why it is more often used to supplement a survey questionnaire than to replace it. The opportunity it offers to improve the comparability of European statistics, insofar as it allows for a better understanding of the populations covered by a reference source when matched with a more comprehensive database, remains to be explored.

At the end of the meeting, Bertrand du Marais, President of the CNIS, proposed that the CNIS review the implementation of the reference framework for official statistics matching in two years' time and consider a proposal for a reference framework for researchers to carry out matching themselves.