39th Voorburg Group Meeting

Copenhagen 22-26 September 2025

Classifications and Standards
topic (2) – Upcoming classification revisions, ISIC, NACE, CPC, ATECO, COICOP

An early integrated approach using automated tools and expert validation to automatically recode from ATECO2002 to ATECO2025.

A focus on section G “Wholesale and retail trade”.

Francesca Alonzi, Annarita Mancini

Italian National Institute of Statistics - Istat Italy

Table of contents

An early integrated approach using automated tools and expert validation to automatically recode from ATECO2002 to ATECO2025.

A focus on section G “Wholesale and retail trade”.

Francesca Alonzi, Annarita Mancini

Abstract

NACE Rev. 2.1, adopted by Commission Delegated Regulation (EU) 2023/137 and into force on 1 January 2025, introduces key updates to better capture modern economic activities. While retaining the overall structure of NACE Rev. 2, the new version expands the classification by adding 36 new classes, increasing granularity in many sectors.

This research work presents a feature of the methodology adopted by the Italian National Institute of Statistics (Istat) to preliminary recode legal units from ATECO2022 to the updated ATECO2025 classification, aligned with the European NACE Rev. 2.1. To facilitate the full implementation of the new classification, correspondence relationships between the new ATECO classification and its previous version have been developed. These are presented in the form of tables, known as correspondence, conversion or mapping tables. Unfortunately the most recurrent type of relationship is complex (one or more ATECO2022 codes split into several ATECO2025 codes). In these cases, the recoding is possible only if specific information is available, i.e. information describing the economic activity carried out by the legal units themselves.

The Italian strategy, aimed at reducing the burden on enterprises and at making timely massive recoding possible, is based on an automatic reclassification procedure that has been applied at an early stage. The whole procedure is based on the development of a mapping and operational correspondence table solving in an automatic way the complex splits of the theoretical correspondence making all relationships to simple cases. The approach integrates theoretical correspondences, automated similarity-based tools, and manual expert validation to build an operational correspondence table.

This research work is intended to describe the process of development of the above-mentioned tool reporting the new NACE/ATECO section G “Wholesale and retail trade” as a case study and to analyse the main results of its application to the Italian SBR. More specifically, the aim is to verify the validity of the reclassification mapping in relation to the trade sector, which was affected by several changes from original ATECO2022 to ATECO2025.

More specifically, the research work is based on three scenarios and comparisons between the three scenarios are provided:

- scenario 1 (March 15, 2025): legal units were recoded by applying the operational mapping only;
- scenario 2 (April 1, 2025): legal units were recoded exploiting also other information, mainly statistical sources;
- scenario 3 (August 1, 2025): recoding operations also benefited from notifications by companies and profiling activities.

Keywords: correspondence tables, economic activities, NACE/ATECO, statistical classifications and standards wholesale and retail trade activities.

1. Introduction

Section G "*Wholesale and retail trade*" of the European NACE Rev. 2 has been invested with major changes in the switch to the new classification, NACE Rev. 2.1. More specifically, economic activities within this section have been reorganised according to the following decisions.

1) Activities fostered by the development of digital tools such as the exchange of digital products, streaming and downloading of contents on digital platforms, e.g. e-books, do not fall under the concept of trade described by the new section G; these activities should now be classified under section J "Publishing activities, radio broadcasting, and content production and distribution".

2) The sales channel (e.g. in stores, via stalls and markets, via Internet, through vending machines) is no longer the guiding criterion for differentiating economic activities within division 47 devoted to retail trade. Instead, the new approach is based on the type of products sold.

3) There are no new sections or divisions reserved for intermediation services activities, but new groups or classes have been introduced within the sections to which these activities pertain in order to "fit" them into their supply chain. Thus, intermediation services activities for wholesale and retail sale are integrated within section G.

All the above is reflected within the European correspondence table between original NACE Rev. 2 and derived NACE Rev. 2.1 as well as within the Italian correspondence table between original ATECO2022 and derived ATECO2025. Such tools represent the starting point for establishing the link between the old and the new classification for all statistical units in the SBR. The correspondence tables help users to understand the relationship between the old and new categories and to discern the scope of changes. When the relationship is complex (one or more ATECO2022 codes split into several ATECO2025 codes) the recoding is possible only if specific information is available, i.e. information describing the economic activity carried out by the legal units themselves.

The Italian strategy, aimed at reducing the burden on enterprises and at making timely massive recoding possible, is based on an automatic reclassification procedure that has been applied at an early stage. The whole procedure is based on the development of a mapping and operational correspondence table solving in an automatic way the complex splits of the theoretical correspondence making all relationships to simple cases. The approach integrates theoretical correspondences, automated similarity-based tools, and manual expert validation to build an operational correspondence table.

Taking into account that several methods can be applied for assigning a new activity code, this research work is intended to provide some preliminary results of the reclassification process by considering three scenarios. Scenario n. 1 was applied on March 15, 2025: at that date the operational mapping was applied to legal units in the SBR for reference period 2023; no other sources were integrated. Scenario n. 2 was applied on April 1, 2025 when the same legal units were recoded exploiting also other information, mainly statistical sources, e.g. results of the ad hoc SEA Survey. The latest scenario (scenario n. 3), dating back to August 1, 2025, also includes returns from companies and manual checks by profilers on the largest units.

The content of this work is divided into six sections. Apart from the first section that is devoted to the introduction, the second one is intended to provide some background

information concerning the adoption and implementation of the European NACE Rev. 2.1 classification as well as to provide a focus on the Italian version of the NACE, namely ATECO, including the theoretical correspondence table between the new version and the previous one. In the third section a summary on the main classification changes occurred in the trade sector (section G "Wholesale and retail trade") is provided. The core element of this research work, that is the development of an operational correspondence table, is presented in the fourth section while in the fifth section a complete overview of its potential usage is described. The major results are described in the fifth section. Finally, the last section contains the main conclusive remarks and deal with some room for improvement.

2. European and Italian background: the new NACE Rev. 2.1 and the ATECO classification

2.1. The adoption and implementation of the new NACE Rev. 2.1

The NACE classification, which provides a harmonised framework for classifying economic activities within the European Union, has recently been updated. This revision, moving from NACE Rev. 2 to NACE Rev. 2.1, aims to reflect the structural and technological changes that have occurred in the economy over the past decades, including the digital transformation, the emergence of new business models, and increased relevance of environmental and social activities.

The update enhances the clarity and relevance of categories, improves consistency with other international classifications, and better supports the production of high-quality and comparable statistics across countries and over time.

The implementation of NACE Rev. 2.1 will take place gradually across statistical domains, following a coordinated transition plan. The key phases are listed below.

- *Preparatory Phase* (2024–2025): during this period, national statistical institutes (NSIs) and Eurostat finalised the methodological framework, updated coding tools, and conducted pilot studies to test the new classification.
- *Dual coding and transition period* (2026–2027): statistical production will temporarily include both NACE Rev. 2 and NACE Rev. 2.1 codes. This ensures continuity and allows for methodological validation and linking between versions. For the reference year 2025, Member States must report data under both NACE Rev. 2.1 and the original NACE Rev. 2 framework. This includes annual indicators, international activity statistics and short-term business statistics tables (e.g. employment, wages, producer prices). Double reporting for short-term statistics continues into 2026 and 2027, ensuring a smooth shift before full implementation
- *Full Implementation* (from 2028 onwards): NACE Rev. 2.1 will become the standard classification for all new statistical production. Legacy data may be retrospectively aligned where feasible.

Each statistical domain (e.g. business registers, structural business statistics, short-term statistics, national accounts) will follow its own specific adaptation schedule within this general framework, depending on data availability, system dependencies, and policy priorities.

European national versions of the NACE shall follow the same implementation plan.

In Italy the implementation of the national classification of economic activities, namely ATECO2025 directly derived from NACE Rev. 2.1, is currently underway (see Table 1). Officially adopted on 1 January 2025, it came into force operationally on 1 April 2025 in cooperation with the major administrative bodies using the same classification for non-statistical purposes.

Table 1 - Implementation of NACE Rev. 2.1 in Italy

Milestone	**Timeline**
EU legal entry into force	1 January 2025
ATECO2025 effective in Italy	1 April 2025
Availability of the first dual coding of the SBR	By December 2025
Structural survey	2026 (reference year 2025)
Full adjustment of short-term surveys (Italy, estimate)	2028 (data transmission in NACE Rev. 2.1 and the 2025 base is scheduled for 2028)

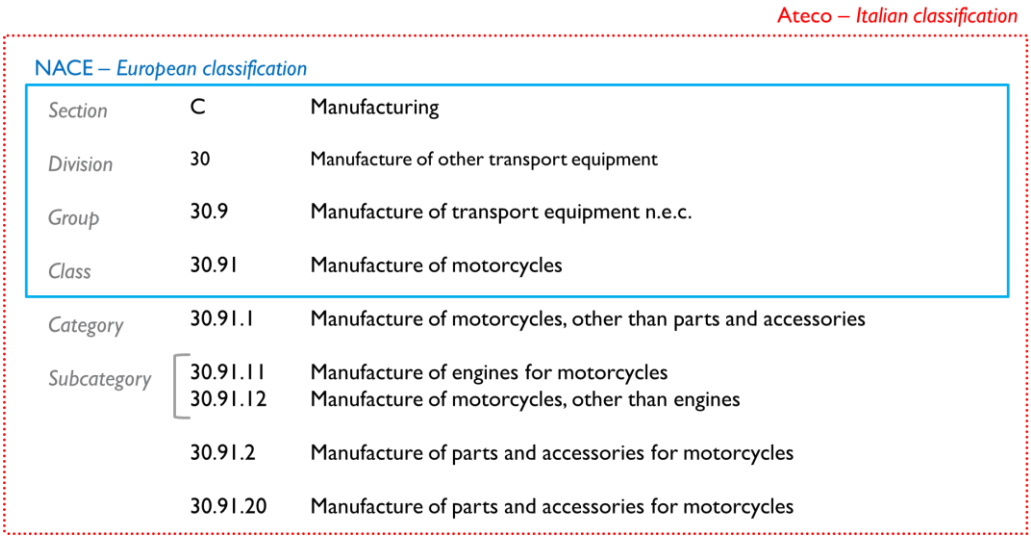
Close coordination between statistical authorities and data users is essential to ensure a smooth transition and maintain the consistency, comparability, and quality of European statistics throughout the update process.

2.2. The structure and contents of the ATECO classification

ATECO is a hierarchical classification system aligned with the European NACE at the first four levels but extended to more levels (the fifth and sixth) to better describe the national reality.

From a structural point of view, it is made of codes and headings (categories in the broad sense). Each ATECO code represents an economic activity that is shortly described by its heading; thus, the ATECO codes, enriched by their headings, aim to define in a standardised way all economic activities carried out by an enterprise, allowing enterprises with similar characteristics to be grouped together for statistical but also administrative, fiscal and other purposes.

Figure 1 - The structure of the ATECO classification: an example



Nevertheless, a short title may be suitable to describe common and easier economic activities (e.g. growing of rice) but usually it does not fit those clarity requirements that a statistical classification should respect. For this reason, the structure (codes and headings) of the ATECO classification are enriched by definitional descriptions (explanatory notes[1]) that provide supporting information about the classification codes and headings. They are statements which clearly define the category or that may assist users in determining the boundaries of the category. Explanatory notes may explain the content by giving examples of inclusions and exclusions, or provide more general rules or guidelines. Explanatory note are optional but are usually included in the classification when further definition of categories is required.

2.3. The theoretical correspondence table between ATECO2022 and ATECO2025

A correspondence provides a link between different versions of a classification or between different classifications. A correspondence details how a category in one classification relates, or links, to the new/other classification.

To facilitate the full implementation of the new classification within statistical, administrative or other types of processes, correspondence relationships between the new ATECO2025 classification and its previous version (ATECO2022) have also been made available. The correspondence relationships are presented in the form of tables, known as correspondence, conversion or mapping tables, in which each ATECO2025 code is linked with at least one ATECO2022 code.

This is therefore a theoretical mapping; for instance, when an old category splits into several categories in the new classification, all possible links are provided. Thus, such a correspondence table aims to support the recoding operations of enterprises (and other statistical units) when transitioning from the previous classification to the new one, or vice versa.

A total of 6,678 correspondence relationships were identified across the various hierarchical levels of the ATECO classification (see Table 2). Each correspondence relationship defines a link between a pair of ATECO codes: one ATECO2025 code and one ATECO2022 code.

Table 2 - Theoretical correspondence links between ATECO2022 and ATECO2025

ATECO hierarchical level	**No. of ATECO2022 codes**	**No. of ATECO2025 codes**	**No. of correspondence links**
1 Section	21	22	80
2 Division	88	87	251
3 Group	272	287	566
4 Class	615	651	1,171
5 Category	920	920	1,883
6 Subcategory	1,241	1,290	2,727
Total	**3,157**	**3,257**	**6,678**

[1] Explanatory notes are provides also at the lower level of the NACE classification.

The correspondence links are of different types; a correspondence can consist of the following relationships:

- 1:1 (one-to-one): one ATECO2022 corresponds to only one ATECO2025, and vice versa
- N:1 (many-to-one): many ATECO2022 codes merging into a single ATECO2025 code
- 1:M (one-to-many): one ATECO2022 code splitting into several ATECO2025 codes
- N:M (many-to-many): many ATECO2022 codes linked to many ATECO2025 within complex structural changes

One-to-one and many-to-one cases can be seen as "simple" matches: in such cases, an automatic recoding of ATECO2022 codes can be applied straightforward. However, the most frequent types of relationships (representing more than 70 percent of all correspondences) are of many-to-many type, which, along with one-to-many relationships, is considered "complex" for reclassification purposes. In these cases, the recoding is possible only if specific information is available, i.e. information describing the economic activity carried out by the statistical unit (information at the micro level). Especially in the early phase of recoding, this step is impractical for direct operational use when ambiguous correspondences exist, no specific information is available or when there is the need to have a tool to recode a large number of units in a short time.

3. Overview of changes in section G "Wholesale and retail Trade" and impact on theoretical correspondence tables

The old section G of NACE Rev. 2 / ATECO2022 "Wholesale and retail trade; repair of motor vehicles and motorcycles" has undergone a structural improvement in NACE Rev. 2.1 / ATECO2025 to reflect changes in trade patterns and to better delineate traditional channels from emerging ones. More specifically:

- the classification no longer separates trade activities by delivery mode; the distinction between online and physical sales is removed, simplifying the structure and avoiding artificial splits in business activity categorisation;
- as a consequence of the above point, some existing codes have been modified or replaced—such as ATECO code 47.91.10 that was reserved to online retail trade;
- division 45 of NACE Rev. 2 / ATECO2022 "Wholesale and retail trade and repair of motor vehicles and motorcycles" was suppressed;
- wholesale and retail trade of motor vehicles and motorcycles, previously included in division 45, are now classified in divisions 46 "Wholesale trade" and 47 "Retail trade" while repair activities were moved to division 95;
- new section G only includes trade of physical goods; the trade of digital goods, such as downloads and streaming services, is now classified under section J "Publishing, broadcasting, content".

Many of the above-mentioned structural changes unfortunately lead to complex relationships especially at national level: one or more ATECO2022 code split into several ATECO2025 codes. In these cases, the recoding is possible only if specific information is available, i.e. information describing the economic activity carried out by the statistical unit.

For instance, this is the case of the removal of activities codes for retail itinerant sales (i.e. retail sale via stalls and markets). All activities previously classified as retail sale via stalls and markets (e.g. classes 47.81 for food, beverages and tobacco products, 47.82 for textiles, clothing and footwear, 47.89 for other goods) have been integrated to retail sale activities in stores as well as via mail order houses or via Internet or through other sale channels.

Example 1 - NACE Rev. 2 / ATECO2022 class 47.81 “Retail sale via stalls and markets of food, beverages and tobacco products” is related to 8 possible correspondent new codes:

47.11 Non-specialised retail sale of predominately food, beverages or tobacco
47.21 Retail sale of fruit and vegetables
47.22 Retail sale of meat and meat products
47.23 Retail sale of fish, crustaceans and molluscs
47.24 Retail sale of bread, cake and confectionery
47.25 Retail sale of beverages
47.26 Retail sale of tobacco products
47.27 Retail sale of other food

Example 2 - NACE Rev. 2 / ATECO2022 class 47.91 “Retail sale via mail order houses or via Internet” is divided into 32 different correspondent NACE Rev. 2.1 classes:

47.11 Non-specialised retail sale of predominately food, beverages or tobacco
47.12 Other non-specialised retail sale
47.21 Retail sale of fruit and vegetables
47.22 Retail sale of meat and meat products
47.23 Retail sale of fish, crustaceans and molluscs
47.24 Retail sale of bread, cake and confectionery
47.25 Retail sale of beverages
47.26 Retail sale of tobacco products
47.27 Retail sale of other food
47.40 Retail sale of information and communication equipment
47.51 Retail sale of textiles
47.52 Retail sale of hardware, building materials, paints and glass
47.53 Retail sale of carpets, rugs, wall and floor coverings
47.54 Retail sale of electrical household appliances
47.55 Retail sale of furniture, lighting equipment, tableware and other household goods
47.61 Retail sale of books
47.62 Retail sale of newspapers, and other periodical publications and stationery
47.63 Retail sale of sporting equipment
47.64 Retail sale of games and toys
47.69 Retail sale of cultural and recreational goods n.e.c.
47.71 Retail sale of clothing
47.72 Retail sale of footwear and leather goods
47.73 Retail sale of pharmaceutical products
47.74 Retail sale of medical and orthopaedic goods
47.75 Retail sale of cosmetic and toilet articles
47.76 Retail sale of flowers, plants, fertilisers, pets and pet food
47.77 Retail sale of watches and jewellery
47.78 Retail sale of other new goods
47.79 Retail sale of second-hand goods

47.91 Intermediation service activities for non-specialised retail sale
47.92 Intermediation service activities for specialised retail sale
60.39 Other content distribution activities

The examples presented above (example n. 1 and n. 2) show that the recoding of units classified in NACE Rev. 2 / ATECO2022 classes 47.81 o 47.91 is possible only if specific information is available, i.e. information describing the economic activity carried out by the statistical unit (information at the micro level). Especially in the early phase of recoding, this step is impractical for direct operational use when ambiguous correspondences exist, no specific information is available or when there is the need to have a tool to recode a large number of units in a short time.

In addition, it should be noted that exist also some correspondence links involving section G that interest other sections of the NACE/ATECO classification (i.e. other than section G) creating shifts between different economic sectors. A measure of this phenomenon is provided in Table 3 where economic activities at the division level have been reclassified according to whether they:

- remain classified in the same trade section (CONFIRM);
- leave the trade section (EXIT);
- enter the trade section (ENTRY).

Table 3 - Theoretical correspondence links between ATECO2022 and ATECO2025 concerning trade activities at division level[1]

ATECO 2022		ATECO 2025		Match
39	Remediation activities and other waste management services	46	Wholesale trade	ENTRY
45	Wholesale and retail trade and repair of motor vehicles and motorcycles	46	Wholesale trade	CONFIRM
45	Wholesale and retail trade and repair of motor vehicles and motorcycles	47	Retail trade	CONFIRM
45	Wholesale and retail trade and repair of motor vehicles and motorcycles	95	Repair and maintenance of computers, personal and household goods, and motor vehicles and motorcycles	EXIT
46	Wholesale trade, except of motor vehicles and motorcycles	46	Wholesale trade	CONFIRM
47	Retail trade, except of motor vehicles and motorcycles	35	Electricity, gas, steam and air conditioning supply	EXIT
47	Retail trade, except of motor vehicles and motorcycles	43	Specialised construction activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	47	Retail trade	CONFIRM
47	Retail trade, except of motor vehicles and motorcycles	52	Warehousing, storage and support activities for transportation	EXIT
47	Retail trade, except of motor vehicles and motorcycles	53	Postal and courier activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	55	Accommodation	EXIT
47	Retail trade, except of motor vehicles and motorcycles	56	Food and beverage service activities	EXIT

[1] Theoretical links between ATECO2022 and ATECO2025 are not exactly the same of theoretical links between NACE Rev. 2 and NACE Rev. 2.1 because some correspondences have been added only for national purposes.

47	Retail trade, except of motor vehicles and motorcycles	58	Publishing activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	60	Programming, broadcasting, news agency and other content distribution activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	61	Telecommunication	EXIT
47	Retail trade, except of motor vehicles and motorcycles	68	Real estate activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	73	Activities of advertising, market research and public relations	EXIT
47	Retail trade, except of motor vehicles and motorcycles	74	Other professional, scientific and technical activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	77	Rental and leasing activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	79	Travel agency, tour operator and other reservation service and related activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	82	Office administrative, office support and other business support activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	85	Education	EXIT
47	Retail trade, except of motor vehicles and motorcycles	86	Human health activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	87	Residential care activities	EXIT
47	Retail trade, except of motor vehicles and motorcycles	95	Repair and maintenance of computers, personal and household goods, and motor vehicles and motorcycles	EXIT
47	Retail trade, except of motor vehicles and motorcycles	96	Personal service activities	EXIT
61	Telecommunications	47	Retail trade	ENTRY
62	Computer programming, consultancy and related activities	47	Retail trade	ENTRY
63	Information service activities	47	Retail trade	ENTRY
82	Office administrative, office support and other business support activities	46	Wholesale trade	ENTRY
82	Office administrative, office support and other business support activities	47	Retail trade	ENTRY

4. Developing a national operational correspondence table: data and methods

The theoretical correspondence table can be easily applied to 1:1 links or to complex links when information at the micro level, i.e. referring to each single unit, is available. In order to solve in an automatic way the cases of complex matches transforming them in simple matches (1:1), Istat developed an operational correspondence table, which assigns a single predominant ATECO2025 code to each ATECO2022 code, ensuring consistency for automated reclassification across large statistical and administrative registries.

The logic underlying the operational table is to associate each ATECO2022 code with a single, most representative, ATECO2025 code among those proposed by the theoretical correspondence table.

The choice of the ATECO2025 code to be uniquely associated to the starting ATECO2022 is the result of an integrated methodology combining in cascade the following three components.

1) Using an automatic matching algorithm that compares headings and inclusion notes (text strings) of the two classifications, ATECO2022 and ATECO2025. This algorithm assigns a similarity measure to each of the M possible ATECO2025 linked to each ATECO2022 code, resulting from the comparison of the words into which the set of strings (headings and explanatory notes) is broken down.
2) Using the results of the Survey of Economic Activities (SEA) survey (also known as ATECO survey) to support automatic unsatisfactory choices based on similarity scores resulted from the above algorithm.
3) Applying manual checks when results are still unsatisfactory.

4.1. The automatic matching algorithm to compare headings and inclusion notes of ATECO2022 and ATECO2025

The innovative approach of this procedure lies in the use of automatic matching methods based on textual analysis, aimed at simplifying a complex task that, if performed manually, would have been excessively burdensome. The initial goal of the automatic algorithm was to associate each ATECO2022 code with all ATECO2025 codes that exhibit similar informational content, based on the analysis of headings and explanatory notes of the two classifications. Each identified pair was assigned a similarity score, guiding the selection toward the pair with the highest value while respecting the constraints of the theoretical correspondence table

3.

Several methods were tested to determine the most appropriate similarity measure between the strings of the two classifications. The processing phase was carried out using the Python programming language. At first, a pre-processing of the initial data was necessary, which involved two main steps:

- removal of irrelevant words (stop words), that is, terms not meaningful for content analysis;
- conversion of words to their lexical root (stemming), to reduce them to their base forms.

Each ATECO2025 code deemed similar to the source code was associated with a similarity indicator (or score). Different types of similarity indices were used (Jaccard, Sorensen, Overlap and Levenshtein). However, these distance-based methods presented some critical issues. It was found that the words used in the descriptions of ATECO2025 did not always match those in ATECO2022, making the previously mentioned similarity indexes ineffective. Furthermore, there have been cases in which the highest level of similarity was associated with a pair not present in the theoretical correspondence table.

To address this problem, a TF-IDF (Term Frequency-Inverse Document Frequency) method was applied, which assesses the importance of terms in a document relative to the entire corpus. This approach assigns greater weight to terms that are frequent in a single document but rare across the corpus, allowing for the extraction of more relevant features for comparison.

[33] At first the analysis undertaken by Istat focused on 5-digit codes because, for statistical purposes, these are the codes subject to reclassification. But then the analyses were extended to all hierarchical levels of the Ateco respecting the same hierarchical level of the starting code.

Nevertheless, the TF-IDF method also has some limitations: it does not consider the semantic meaning of words, making it less effective with synonyms or different phrases that express the same concept. Moreover, it assumes independence between terms and focuses solely on frequency, making it difficult to capture the context or relationships between words.

In a later phase, the activity shifted towards models based on semantic concepts to analyse text similarities. Among the most promising approaches, a deep learning algorithm — a Large Language Model (LLM) — was used. Its main function was to assign a similarity score between texts capable of better capturing semantics, thereby improving upon previous distance-based methods. In this method, no text pre-processing was carried out, and the scope of search and comparison was defined, for each ATECO2022 code, by the possible ATECO2025 codes listed in the theoretical correspondence table.

The similarity score ranges from 0 to 1: the higher the value, the greater the likelihood that the match is reliable and accurate. For each complex ATECO2022 code, a list of corresponding ATECO2025 codes was generated along with their respective similarity scores.

Although the latter method was considered the most reliable, it still presented some critical issues, as it was not able to map all complex correspondences with a high similarity score. For this reason, the selection of the predominant or most representative ATECO2025 code was finally based on a combination of the tested methods, applied using a cascading logic that prioritised the highest similarity scores from the most reliable approaches.

4.2. Using the results of the SEA survey and the contribution of classification experts

In order to support automatic unsatisfactory choices based on similarity scores because they were low, the results of the SEA survey, if any, were considered, orienting on the most frequent empirical responses.

Started in April and run until July 31, 2024, Istat has conducted a survey in order to detect the economic activity carried out by enterprises and to use the information collected to reclassify the units registered in the statistical business register (SBR) according to the new ATECO2025 classification. The SEA survey collected actual business activity descriptions and coding, guiding the choice toward the most empirically frequent outcomes.

In addition, critical codes and ambiguous matches were analysed carefully by expert staff who manually reviewed the automatic choices made by using the automatic algorithm. More specifically, explanatory notes of the two classifications were checked and compared in order to assign the predominant ATECO2025.

5. **Implementing the new classification in the Italian SBR**

5.1. A step-by-step implementation by three incremental scenarios

Implementing a new classification of economic activities in the Statistical Business Register (SBR) is a complex task that, while offering a more accurate and up-to-date depiction of economic structures, can disrupt existing time series, necessitating either back-casting or double coding to mitigate this issue. In the context of SBRs, back casting is generally impractical, making double coding a primary concern, to support all statistical domains to avoid inconsistencies within the system.

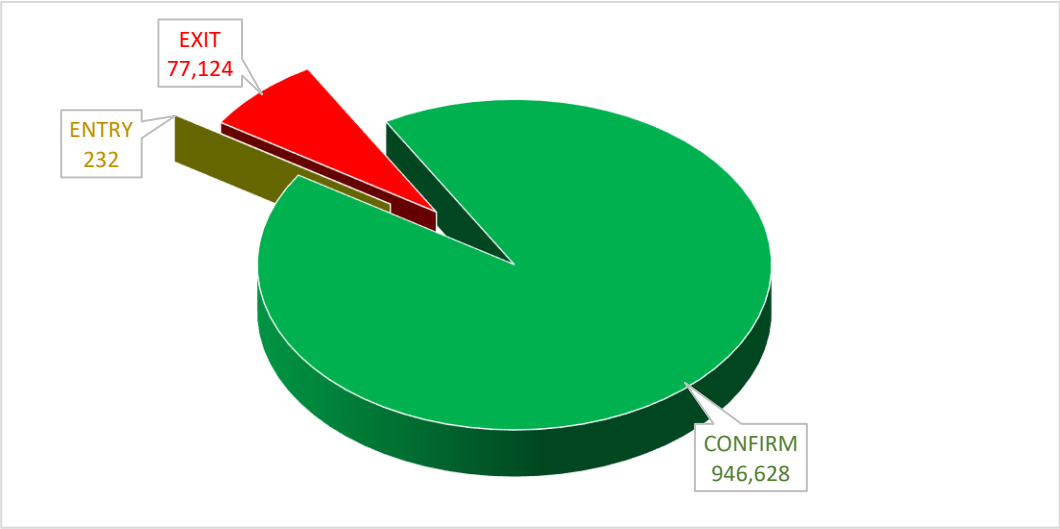
Given their central role, statistical business registers are among the first areas to implement the new classification, thereby facilitating its adoption in other business and trade statistics domains. The main step is the recoding of SBR units and management of the double coding period, several tools have been and will be developed to facilitate the coding strategy.

At the time of writing, the process of recoding is not concluded yet but the Italian operational correspondence table has already played a central role in the entire operation. In order to evaluate the usefulness of this tool, three incremental scenarios have been identified.

- Scenario n. 1 was applied on March 15, 2025: at that date the operational mapping was applied to legal units (LEUs) in the SBR for reference period 2023.
- Scenario n. 2 was applied on April 1, 2025 when the same legal units were recoded exploiting also other information, mainly statistical sources, e.g. results of the *ad hoc* SEA Survey.
- Scenario n. 3, dating back to August 1, 2025, also includes returns from companies and manual checks by profilers on the largest units.

As a starting point, all legal units active in 2023 that are classified in section G "WHOLESALE AND RETAIL TRADE; REPAIR OF MOTOR VEHICLES AND MOTORCYCLES" both according to the old NACE Rev. 2 / ATECO2022 (divisions 45, 46 and 47) and the new NACE Rev. 2.1 / ATECO2025 (divisions 46 and 47) were selected. It consists of 1,023,984 legal units. As presented in Chapter 3 where all activities were reclassified with the variable "match" all these 1,023,984 legal units were reclassified assigning the variable "match". Obviously the units classified in the old domain of ATECO2022 had the variable match CONFIRM and EXIT, while the units in the new domain ATECO2025 have the value CONFIRM and ENTRY and thus correspond to the sample of scenario n. 2 (946,860 legal units).

Figure 2 – Number of Italian legal units classified in NACE/ATECO section G by type of match (reference year 2023)



5.2. Scenario n. 1: main results

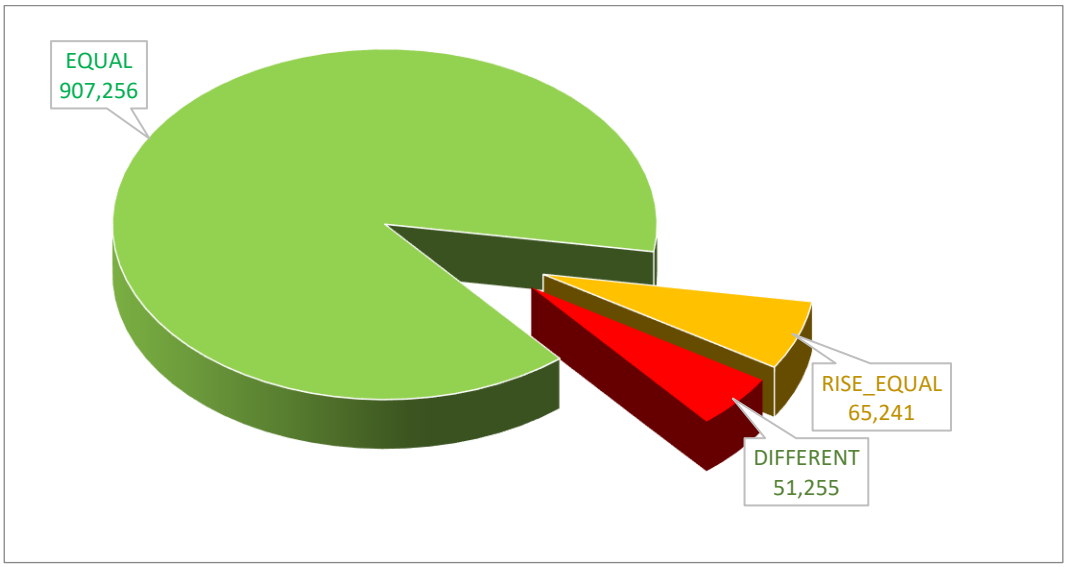
In order to evaluate the quality of the recoding derived by the application of the operational mapping, only those legal units classified in original section G were considered (the initial population consists of 1,023,752 legal units and thus correspond to CONFIRM and EXIT set). A

comparison between the new NACE Rev. 2.1 / ATECO2025 code derived by applying the operational mapping and the new code assigned in the scenario n. 2 was undertaken.

However, before showing results, it is necessary to underline that in terms of content and general principles, the operational mapping associates each ATECO2022 (original) code with the representative ATECO2025 code respecting the same hierarchical level as the starting code. The operational mapping includes all hierarchical levels of the classification.

However, there is an exception to this principle. In some cases, it was difficult to choose the most representative ATECO2025 code among all possible ATECO2025 codes belonging to the same hierarchical level as ATECO2022. In these cases, the original activities (those classified within ATECO2022) turned out to be distributed in the new codes proportionally not allowing the choice to fall reasonably in the new one of the same hierarchical level. As a consequence, in these cases an ATECO2025 code of a higher hierarchical level was chosen; the analysis of comparison between the two possible codes associated to the same unit is made considering the same hierarchical level. Thus, results show "EQUAL" cases, "RISE_EQUAL" cases and "DIFFERENT" cases (see Figure 3).

Figure 3 – Number of Italian legal units by type of comparison between scenario n. 1 and scenario n. 2



In order to further analyse the distinction provided in figure 3 between the three label of the "type of comparison" variable, table 4 provides a picture of the same set of legal units consisting of 1,023,752 legal units by the "match" variable calculated for scenario n. 1 and scenario n. 2.

Table 4 – Number of Italian legal units by type of comparison between scenario n. 1 and scenario n. 2

Type of comparison	No. of LEUs	% of LEUs
DIFFERENT	51,148	5.0
DIFFERENT	107	0.0
EQUAL	830,239	81.1
EQUAL	77,017	7.5
RISE_EQUAL	65,241	6.4
Total	**1,023,752**	**100**

As shown in the table above (Table 4) the two recoding procedures (scenario n. 1 and scenario n. 2) classify in the same ATECO code 972,497 legal units (95% of the total legal units constituting the sample of scenario n. 1) – these are those flagged as "EQUAL" and "RISE_EQUAL". The majority of them is characterised by a match type "CONFIRM".

Only a small minority of legal units (about 5% of the total corresponding to 51,255 legal units) was instead reclassified with a different ATECO2025 code[4]. Table 5 provides the matching at division level for this small sample.

Table 5 – Legal units labelled as "DIFFERENT" when comparing scenario n. 1 and scenario 2 by ATECO2025 division

ATECO2025 division in scenario n. 2	ATECO2025 division in scenario n. 1	No. of LEUs
35	47	91
46	46	2,488
46	47	172
47	46	139
47	47	48,349
61	47	16
Total		**51,255**

Taking division 46 as an example, when considering the two scenarios (scenario n. 1 and scenario n. 2), it can be notice a different reclassification in a lower hierarchical level of the ATECO2025 in 2,488 cases.

Finally, table 6 provides a complete overview of legal units recoded into new section G according to scenario n. 1 (reference year 2023).

Table 6 – Legal units recoded into new section G according to scenario n. 1 (reference year 2023)

NACE Rev. 2.1 / ATECO2025 division	No. of LEUs	% of LEUs classified in the old section G	% of LEUs classified in the new section G
46 Wholesale trade	376,423	36.8	39.8
47 Retail trade	570,312	55.7	60.2
95 Repair and maintenance of computers, personal and household goods, and motor vehicles and motorcycles	77,017	7.5	
N. of LEUs included in the initial population (scenario n. 1)	1,023,752	100	
N. of LEUs reclassified or confirmed in new section G	946,735		100

[4] In this analysis, it takes in consideration ATECO classification for the reason that the theoretical links between ATECO2022 and ATECO2025 are not exactly the same of theoretical links between NACE Rev. 2 and NACE Rev. 2.1.

5.3. Scenario n. 2: main results

As already explained, scenario n. 2 was derived by taking into account only units reclassified as "CONFIRM" or "ENTRY" from the entire population. Scenario n. 2 consists of 946,860 legal units. As you can see, the new wholesale and retail trade population (reference year 2023) lost 77,124 legal units that have moved to other economic sectors according to the new classification, while it pick up just 232 units from other sectors.

Table 7 - Legal units recoded into new section G according to scenario n. 2 (reference year 2023)

NACE Rev. 2.1 / ATECO2025 division	No. of LEUs	% of LEUs classified in the old section G	% of LEUs classified in the new section G
35 Electricity, gas, steam and air conditioning supply	91	0	
46 Wholesale trade	376,683	36.8	39.8
47 Retail trade	570,177	55.7	60.2
61 Telecommunication	16	0	
95 Repair and maintenance of computers, personal and household goods, and motor vehicles and motorcycles	77,017	7.5	
N. of LEUs included in the entire population (scenario n. 2)	1,023,984	100	
N. of legal units reclassified or confirmed in new section G	946,860		100

Comparing data presented in Table 6 and Table 7 we can say that the composition of the new domain of the sector G is almost the same. The legal units that were not classified in the new domain according to the scenario n. 1 are all of small size dimension, while those classified according to scenario n. 2 are also medium-sized. Table 8 provides the distribution by size class of all legal units belongings to the entire population (1,023,984 legal units) .

Table 8 – N. of legal units by size class and match type in the entire population (reference year 2023)

Size class	Match	No. of LEUs	% of LEUs
0-9	CONFIRM	904,505	88.3
0-9	ENTRY	166	0.0
0-9	EXIT	73,952	7.2
10-99	CONFIRM	40,380	3.9
10-99	ENTRY	64	0,0
10-99	EXIT	3,155	0.3
100-249	CONFIRM	1,127	0.1
100-249	ENTRY	2	0.0
100-249	EXIT	14	0.1
>250	CONFIRM	616	0.1
>250	ENTRY	0	0
>250	EXIT	3	0
Total		**1,023,984**	**100**

5.4. Scenario n. 3: main results

Starting from April 1, 2025, by accessing the Italian Business Portal[5] (BP) online and providing a notification to the SBR, legal units were able to ask for a change of the preliminary ATECO2025 code assigned through scenario n. 2. The respondents' notifications may result in changes or confirmations but they need to be checked and validated by expert BR staff. At the moment, this task is not completed yet; thus, all notifications asking for a different ATECO2025 code should be considered only as potential changes because they still need to be verified. Companies usually provide reliable information about their economic activities but may also incur in errors.

At the same time, largest units are being analysed by expert profilers. Manually studying the economic activities undertaken by these legal units, profilers can modify and update the preliminary ATECO2025 code assigned through the procedure known as scenario n. 2.

As already shown in Figure 2 there are 1,023,752 active legal units (reference year 2023) belonging to section G according to the original classification NACE Rev. 2 / ATECO2022 (section G being "Wholesale and retail trade; repair of motor vehicles and motorcycles"); this number is reduced to 946,735 applying scenario n. 1 while it decreases to 946,860 applying scenario n. 2 according to the new NACE Rev. 2.1 / ATECO2025 classification (section G being "Wholesale and retail Trade").

The latter two samples have been compared with the new information deriving from scenario n. 3. More specifically, with regard to the NACE Rev. 2.1 / ATECO2025 codes assigned starting from April 1, 2025, different cases may be encountered; they can be grouped as follows for our purposes:

a - **potential changes** coming from the Business Portal: on August 1, 2025 there were 3,895 changes provided directly by the companies concerning all the section of the NACE/ATECO classification; 1,086 were referred to the entire population of this research work;

b - **potential confirmations** coming from the Business Portal: on August 1, 2025 there were 66,753 confirmations provided directly by the companies concerning all the section of the NACE/ATECO classification; 14,589 were referred to the entire population of this research work;

c - online corrections managed by profilers: at 1st August 2025 there are 1,478 online interventions.

The following figures provide a deeper insight on potential changes coming from the Business Portal (case a) and online corrections provided by profilers (case c).

[5] The Business Portal (BP), known as "Statistica&Imprese", is currently the system of statistical services dedicated both to businesses and to Istat statisticians for the collection and return of information of the main surveys of official statistics. This single point of access allows bi-directional communication between Istat and the business world: using this tool all enterprises involved in business surveys can quickly and easily fulfil their information obligations receiving back a number of significant advantages, like customized statistical information feedback. Most of the updates of the SBR due to statistical sources – Structural Business Statistics (SBS) and Short-Term Statistics (STS) – are mainly obtained via BP, simplifying not only the procedures by which businesses provide statistical information, but also those collected by SBR experts.

Figure 4 - Potential changes coming from the Business Portal (BP)

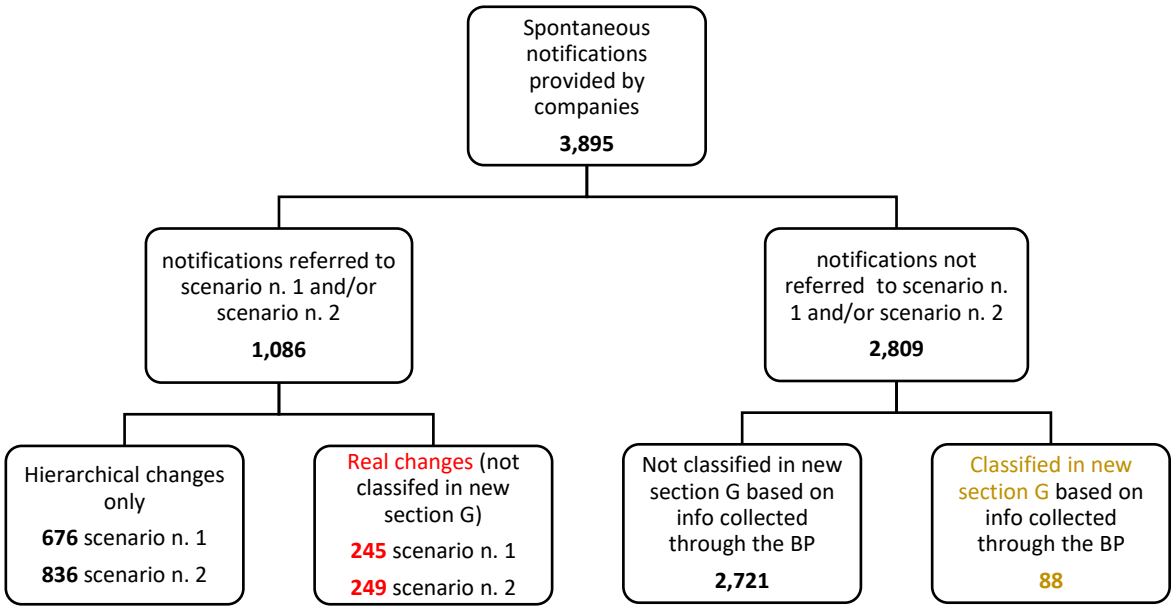
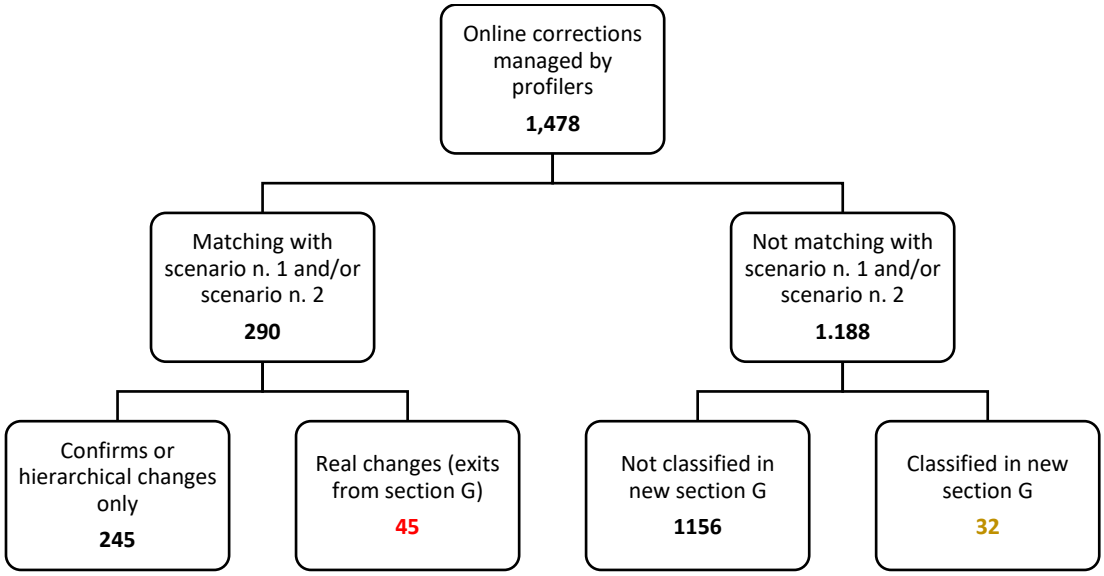


Figure 5 - Online corrections managed by profilers



For the potential changes originating from legal units' notifications or experts' corrections, it is necessary to distinguish between

- REAL CHANGE: an imputation of a NACE Rev2.1/ATECO2025 code not belonging to the new section G,
- HIERARCHICAL CHANGE: a code declared on line belonging to the new section G and the division level is the same, this means that the variation is due in some hierarchical level. For the scenario n. 1 the change for the rise code at the same level was detected and so the number of the potential changes are lower compared to the scenario n.2.

6. **Conclusions and future developments**

Implementing a new classification of economic activities poses several challenges. More specifically, the implementation of the new classification of economic activities in the Business Register has been a complex endeavour, especially due to the various expertise required from statisticians, classification experts, methodologists and IT professionals. It requires the planning and application of a multi-strategy approach to implementation based on different sources, both administrative and statistical. At an early stage, when statistical and administrative registers lack sufficient information to recode units at the micro level, specific tools may be used.

This research work presented the development of an operational correspondence table between ATECO2022 and ATECO2025 to solve in an automatic way and at macro level the cases of one-to-many splits by identifying only one ATECO2025 (the "predominant" " or "most representative") among M ATECO2025 associated to the starting ATECO2022.

Such a tool was used at national level for several purposes, also for non-statistical purposes. This research work has provided some initial results concerning wholesale and retail trade activities. More specifically, three possible scenarios have been analysed to measure the potential effects deriving from the application of the Italian operational correspondence table to recoding activities of legal units within the national SBR:

- scenario n. 1: application tout court of the Italian operational correspondence table without using other sources;
- scenario n. 2: using a combined but still early approach integrating more sources such as results from the application of the Italian operational correspondence table with results of the *ad hoc* SEA Survey;
- scenario n. 3: using a combined but still not complete recoding approach; this scenario also includes returns from companies and manual checks by BR staff and expert profilers on the largest units.

Results show that the Italian operational correspondence table has enabled rapid and standardised recoding at an early stage and that its usage is particularly useful if integrated with other sources. There are not significant differences when comparing the three scenarios as especially shown in the last paragraph. Such a result confirms that the use of an automatic recoding process based on an operational correspondence table at an early stage to have a first and preliminary structure of SBR according to the new classification was a good and productive strategy.

However, the Italian operational correspondence table suffers of some limitations as it may result in information loss, particularly for underrepresented activities. More specifically, some ATECO2025 codes were never selected as predominant, potentially leading to underestimation in certain sectors (e.g. intermediation service activities). Moreover, in some cases it sacrifices granularity by choosing codes at the highest hierarchical levels. Finally, it may be subject to statistical error due to the adoption of probabilistic approaches and practical choices.

All the above considered, the current implementation of this tool marks a first and important step in the recoding activities of enterprises. Over time, direct notifications and updates from enterprises themselves will refine the classification as shown by preliminary results offered by scenario n. 3. Thereafter, the integration of other informative sources, such as updated fiscal data or information extracted from balance sheets, will lead to an improvement in the

qualitative standards of the NACE/ATECO codes assigned to legal units and thus enterprises within the Italian SBR.

References

ALONZI F., CONSALVI M., VIVIANO C. 2024. *A Comprehensive Strategy for Implementing NACE Rev. 2.1 in the Italian Statistical Business Register*. Meeting of the Group of Experts on Business Registers organised by UNECE (15-18 October 2024). https://unece.org/statistics/events/meeting-group-experts-business-registers-0

ALONZI F., CONSALVI M., GENTILI B., VIVIANO C. 2024. *Planning a new ad hoc survey to recode units in the Italian SBR.* Eurostat NACE implementation webinar (29-30 April 2024).

ALONZI F., MANCINI A., SPERANZA A., VIVIANO C. 2025. *La tabella operativa di riclassificazione da ATECO 2007 aggiornamento 2022 a ATECO 2025*. https://www.istat.it/classificazione/ateco-2025/

ALONZI F., MANCINI A., VIVIANO C. 2025. *The preliminary recoding procedure from ATECO2022 to ATECO2025.* LXI Riunione scientifica Società Italiana di Economia Demografia e Statistica "Povertà e disuguaglianze: popolazioni e territori nell'era delle transizioni" (28-30 May 2025).

ALONZI F., VIVIANO C. 2025. *Le relazioni di corrispondenza tra le classificazioni delle attività economiche ATECO2025 e ATECO2007 aggiornamento 2022*. https://www.istat.it/classificazione/documenti-ateco/

CHRISTEN, P. 2012. Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection Springer

COHEN, W. W., & RICHMAN, J. 2002. Learning to Match and Cluster Large High-Dimensional Data Sets for Data Integration KDD 2002

COHEN, W. W., RAVIKUMAR, P., & FIENBERG, S. E. 2003. *A Comparison of String Distance Metrics for Name-Matching Task* Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web (IIWeb-03)

Commission Delegated Regulation (EU) 2023/137 of 10 October 2022 amending Regulation (EC) No 1893/2006 of the European Parliament and of the Council establishing the statistical classification of economic activities NACE Revision 2 (Text with EEA relevance).

CONSALVI M., GENTILI B., SPERANZA A., (2019), The new role of the SBR within the Italian Business Portal, Meeting of the Group of Experts on Business Registers.

CONSALVI M., FAZIO N., (2012), The Business Portal, 23rd Meeting of the Wiesbaden Group on Business Registers - International Roundtable on Business Survey Frames, Washington.

DEVLIN, J., ET AL. 2019 *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* NAACL https://arxiv.org/abs/1810.04805

LEVENSHTEIN, V. I. 1966. *Binary codes capable of correcting deletions, insertions, and reversals.* Soviet Physics Doklady

RAJARAMAN, A., & ULLMAN, J. D. 2011. *Mining of Massive Datasets* Cambridge University Press

Regulation (EC) No 1893/2006 of the European Parliament and of the Council of 20 December 2006 establishing the statistical classification of economic activities NACE Revision 2 and amending Council Regulation (EEC) No 3037/90 as well as certain EC Regulations on specific statistical domains Text with EEA relevance

WOLF, T., ET AL. 2020. *Transformers: State-of-the-Art Natural Language Processing* EMNLP

Francesca ALONZI, Istat - Italian National Institute of Statistics, francesca.alonzi@istat.it
Annarita MANCINI, Istat - Italian National Institute of Statistics, annarita.mancini@istat.it