

Introducing a NACE classification index to enhance transparency, user-friendliness, and foster uniformity in its application across the European Statistical System

Matthias Flohr¹, Inês Quintanilha², Denis Dechandon²

¹Eurostat, Luxembourg

²Publications Office of the European Union, Luxembourg

Abstract

This paper introduces the NACE index which will be seamlessly integrated into an online platform, enabling users to effortlessly navigate, search and extract data from statistical classifications.

Eurostat and the Publications Office of the European Union (OP) are collaborating on a groundbreaking initiative to streamline user interaction with statistical classifications within the European Statistical System (ESS). This project aims to significantly enhance transparency, user understanding and the harmonised application of the Statistical Classification of Economic Activities (NACE), which is one of the most widely used classifications within the European Statistical System, serving as the foundation for over 40 European statistics products. Harmonising the interpretation of NACE across the ESS will increase data comparability, accuracy and clarity.

The key information asset rendering this enhanced harmonisation possible is the NACE index, which breaks down the content of each NACE position into granular entries, drawing upon the NACE explanatory notes (and agreements within the ESS on their interpretation). Each index entry corresponds to a specific economic activity. This simplifies future revisions of the NACE classification, ensuring a streamlined and adaptable system for years to come.

Technically, this project revolves around the development of a dynamic online platform that empowers users to efficiently navigate, search and export statistical classification data published by the OP. This user-friendly interface will enable users to input specific search terms or combinations of terms, making it easier than ever to find the correct categories in the relevant classifications. Additionally, the platform will offer the ability to explore various economic statistical classifications within the ESS, facilitating the export of classification data. Furthermore, the application will elucidate the structural relationships and correspondences among different statistical classifications, fostering a comprehensive understanding of the

international statistical classification system as a whole. By providing granular details, the NACE index elevates the performance of the search function within the online platform.

The European Statistical System is embarking on an exciting journey to redefine how users interact with statistical classifications. The introduction of the NACE index, supported by this innovative online application, promises to revolutionise the accessibility and utility of statistical classification data. Through improved transparency, harmonisation and usability, this project is a testament to the ESS commitment to facilitating data-driven decision-making in the European Union.

Keywords: Statistical Classification of Economic Activities (NACE), classification index, dynamic online platform, Linked Open Data (LOD), VocBench, ShowVoc

1. Introduction

NACE is the ‘statistical classification of economic activities in the European Community’ and is the subject of legislation at the European Union level, which imposes the use of the classification uniformly within all the Member States¹. It is a basic element of the international integrated system of economic classifications, which is based on classifications of the United Nations Statistical Commission (UNSC), Eurostat as well as national classifications; all of them strongly related each to the others, allowing the comparability of economic statistics produced worldwide by different institutions. *‘It is a fundamental need for any statistical system to have standard concepts, definitions and classifications. International statistical classifications are developed and adopted by international agencies to ensure that there is a standardised and consistent approach to classifying statistical data. The aim is to provide a basis for:*

- *statistics that are reasonably comparable between countries;*
- *developing national classifications for the same variable/characteristics’*

(Hancock 2013).

The NACE classification is one of the most used classifications in the European Statistical System and over 40 statistical products – *inter alia* European business statistics, national and satellite accounts – report data according to that classification. The most recent version of NACE, NACE Revision 2 update 1 (NACE Rev. 2.1)², classifies principal and secondary economic activities of statistical units (e.g. enterprise groups, enterprises, kind-of-activity units)

¹ <http://data.europa.eu/eli/reg/2006/1893/2019-07-26>

² http://data.europa.eu/eli/reg_del/2023/137/oj

according to 651 NACE Rev. 2.1 classes. **The NACE index, which is a concept not used for previous NACE versions, breaks down the content of each NACE class into granular entries, drawing upon the NACE explanatory notes and newly added entries based on continuous discussions of the ESS Standards Working Group.** This extensive additional information layer (currently over 8.500 draft entries) will harmonise the interpretation of NACE across the ESS by providing very detailed information on where to classify economy activities which will in turn increase data comparability, accuracy and clarity.

Another innovative aspect of the NACE index project is that it provides the content of NACE classes in a keyword-type format which is more machine-processable. The NACE index will be integrated into a newly developed online platform empowering users to efficiently navigate, search and export statistical classification data published by the Publication Office of the European Union (OP). The user-friendly interface will enable users to input specific search terms or combinations of terms, making it easier than ever to find the correct categories in the relevant classifications or finding related links within the classification or to other classifications. In addition to visualising the NACE classification data, the online platform will offer the ability to explore various economic statistical classifications within the ESS, such as the statistical classification of products by activity (CPA)³, the system of production statistics for mining and manufacturing (PRODOCM)⁴ and Combined Nomenclature (CN)⁵. As a long-term perspective for all of these classifications a provision of indexes could be envisaged. Since the online platform is foreseen to elucidate the structural relationships and correspondence among the different statistical classifications, a more comprehensive user understanding of the international statistical system as a whole is facilitated.

2. Theoretical framework

2.1 NACE as an economic statistical classification

The following definition of a statistical classification is set out by the 'Best Practice guidelines for developing international statistical classifications (Hancock 2013):

"A statistical classification is a set of categories which may be assigned to one or more variables registered in statistical surveys or administrative files, and used in the production and dissemination of statistics. The categories are defined in terms of one or more characteristics

³ <https://ec.europa.eu/eurostat/web/cpa/overview>

⁴ <https://ec.europa.eu/eurostat/web/prodcom>

⁵ <https://ec.europa.eu/eurostat/web/international-trade-in-goods/methodology#Goods>

of a particular population of units of observation. A statistical classification may have a flat, linear structure or may be hierarchically structured, such that all categories at lower levels are sub-categories of a category at the next level up. The categories at each level of the classification structure must be mutually exclusive and jointly exhaustive of all objects in the population of interest.”

This definition also applies to NACE, which consists of a hierarchical structure (as established in the NACE Regulation), the introductory guidelines and the explanatory notes and classifies economic activities of statistical units. The structure of NACE is described in the NACE Regulation as follows:

- a) a first level consisting of headings identified by an alphabetical code (sections),
- b) a second level consisting of headings identified by a two-digit numerical code (divisions),
- c) a third level consisting of headings identified by a three-digit numerical code (groups),
- d) a fourth level consisting of headings identified by a four-digit numerical code (classes).

The criteria used to define and delineate the classification categories of economic activities at different levels depend on several factors, such as potential use of the classification and availability of data. The criteria for detailed levels of the classifications consider **similarities in the actual production process** as the decisive classification criterion.

A statistical unit may perform one or more economic activities described in one or more categories of NACE. Since NACE classifies the principal (and potentially secondary) economic activity of a statistical unit, the principal activity which contributes most to the value added of that unit has to be identified (for further details, please refer to the NACE publication (Eurostat 2008)).

2.2 Metadata vs reference data

Metadata provides information about data, such as its structure, content, and origin. While metadata is essential for understanding and interpreting reference data, it is not a standardised and organised arrangement of words and phrases presented as lists of terms with a hierarchical structure. In the context of NACE, metadata could include details about the classification system's structure, version history, and maintenance processes.

The NACE codes and categories are considered reference data, since they refer to the standardised values, codes, or categories that provide context and meaning to other data within a system or application. In the case of NACE, it is a classification system that categorises

economic activities into different sectors, serving as a standardised reference for organising and analysing economic data across the European Union, allowing for accurate data collection, policy making, international trade and cooperation, tracking economic trends, and research and development.

2.3 Classification of reference data

In the reference data domain, especially when using standards like RDF (Resource Description Framework), SKOS (Simple Knowledge Organization System), and OWL (Web Ontology Language), classification refers to the process of organising and structuring data into categories or classes based on shared characteristics or attributes. These standards facilitate the creation and maintenance of classification systems, ontologies, taxonomies, and thesauri, which are essential tools for organising and managing reference data.

RDF serves as a foundational standard for representing data as graphs of interconnected nodes and arcs, enabling the organisation of concepts and their relationships within a particular domain. OWL extends RDF by providing a richer set of constructs for describing complex relationships and dependencies between concepts, enabling the creation of more robust ontologies. SKOS, on the other hand, is specifically designed for organising and structuring controlled vocabularies, such as thesauri, taxonomies, and ontologies, making it easier to locate, analyse, and interpret information across different platforms and applications. By utilising these standards, users can create, maintain, and share classification systems, ontologies, and taxonomies that are consistent, accurate, and interoperable, ensuring that the data remains accessible and valuable across different domains and applications.

Classification allows for the systematic organisation of data, which facilitates the ease of locating, analysing, and interpreting information. This becomes especially advantageous when dealing with expansive datasets or intricate information systems, such as NACE.

Secondly, the adoption of standards that support the organisation and classification of reference data, namely SKOS and XSKOS, ensures that the data is maintained with consistency and standardisation across various platforms and applications. The uniformity achieved through this standardisation means that data remains precise and current, allowing integration with other datasets and systems. This, in turn, promotes uniform data exchange and collaboration, regardless of the system or platform in use.

Interoperability is another facet significantly improved with standardised classifications. Enabling an uninterrupted flow of data exchange between different systems and applications,

increased data sharing, collaboration, and integration across various domains and organisations.

Lastly, the benefits of classification extend to improved analysis. When reference data is well-organised and classified, it contributes to more accurate and efficient data analysis. Ultimately, this reduces the time and effort required for data management.

2.4 The Publications Office of the European Union

The Publications Office of the European Union (OP) supports national administrations and other stakeholders dealing with reference data by promoting the use of standards and by disseminating resources like the NACE classification system, along with other statistical tools.

One of the activities of the OP is to provide consultancy to Eurostat when maintaining their classification systems and controlled vocabularies. The resources are readily available to national administrations and stakeholders, ensuring the access to current and standardised reference data, which in turn promotes data exchange and cooperative efforts throughout the European Union.

Another significant aspect of the OP's work is the support offered for the practical application and ongoing management of the NACE classification system among other resources. This encompasses consistent updates to the system and aid with data gathering, analysis, and the compilation of reports.

In parallel, the OP assists stakeholders in using the available information management systems, which include the open-source tools VocBench and ShowVoc, by providing training and periodically hosting workshops. This assistance is aimed at equipping national administrations and stakeholders with the necessary skills to effectively manage and analyse reference data.

Lastly, the OP addresses the need for multilingual datasets by translating published assets into the EU's 24 official languages when necessary. These translations are carried out by translators of the Directorate General for Translation of the European Commission, ensuring high-quality and accurate multilingual resources.

And it is in conjunction with the services provided by the Publications Office that this project of a customisable solution for managing and searching NACE has been developed for and with Eurostat.

2.5 VocBench and ShowVoc

VocBench and ShowVoc are two related information management systems that support the collaborative management of flexible controlled vocabularies, such as thesauri, taxonomies and ontologies, and classification systems that organise data within a domain. These tools are designed to facilitate the creation, maintenance and sharing of these assets, making it easier for users to manage and use reference data across platforms and applications.

VocBench is an open-source tool, a web-based application that allows users to collaboratively create, edit and maintain objects and concepts and categorise them into different classes based on common characteristics or attributes and structured sets of terms used to organise and classify information in different domains. It supports different types of assets and provides a user-friendly interface for managing the data. VocBench also includes features for versioning, workflow management and user access control, ensuring that controlled vocabularies remain consistent and accurate over time.

ShowVoc is a web-based visualisation tool that allows users to explore, analyse and maintain reference data assets created with VocBench or others (imported via RDF files). It provides a graphical representation of the hierarchical structure of the asset, making it easier for users to understand the relationships between terms and to navigate the dataset. ShowVoc also includes features for filtering and searching a complete classification system or controlled vocabulary, allowing users to locate specific terms or concepts within the asset.

Together, VocBench and ShowVoc provide a solution for managing and exploiting reference data assets, whether classification systems or controlled vocabularies, while applying standards. They enable users, with varying levels of technical expertise, to collaboratively create, maintain and share reference data, ensuring that the data remains consistent, accurate and up-to-date across platforms and applications.

2.6 ShowVoc possibilities

Just like for VocBench, the development of ShowVoc is currently coordinated by the Publications Office of the European Union, in collaboration with the University of Rome Tor Vergata in Italy and is funded by the Digital Europe Programme (DIGITAL). As an open-source application, ShowVoc provides comprehensive documentation available on the project's website (ShowVoc, 2021). The Publications Office is tasked with collating user requirements and feedback, liaising with the developers, and coordinating the development process, which includes project management, routine progress meetings, and system testing.

In conjunction with VocBench, ShowVoc was built as a data-centric portal for the EU, with a primary focus on terminology content. As the project has evolved, it has emerged as a user-friendly, open-source tool that is readily available for download and serves as a platform for publishing data resources.

While ShowVoc is more accessible to use than VocBench, it is not solely intended for novice users. It facilitates a variety of complex interactions through its functionalities, including menus of datasets, search capabilities, translation features, and alignment options. A particularly noteworthy function is its adaptable visualisation tools, offering users a range of customisable options from straightforward, legible displays to advanced views with tailored forms and adapted resource perspectives profiting from custom forms, as well as a focus on language resources by choosing the preferable language.

The platform enables users to conduct exploration and analysis through an index-based search, graphical representations, or a specialised mapping page. ShowVoc offers diverse levels of search functionality, from simple free-text searches to advanced searches that permit the incorporation of constraints and customised queries, allowing users to specify the dataset and/or the properties and labels to be presented as a result.

As a community-driven platform, ShowVoc invites contributors to submit datasets or metadata affecting specific resources. Therefore, ShowVoc enhances the utilisation of semantic technologies, as a set of tools and techniques that enable the interpretation, organisation, and effective use of data by representing its meaning and relationships, within specific domains of expertise.

Additionally, ShowVoc simplifies the publishing process by providing options to enable embedding in the stakeholders existing websites. This flexibility of highly configurable settings also permits the creation of a customisable landing page, enabling the existence of logos, headers and footers, and metadata pages, such as editorial content, descriptions, and associated documentation for download.

In conclusion, ShowVoc is an intuitive interface for users to consult their reference data assets, facilitating easy updates and searches across multilingual datasets, and democratising the publication of reference data.

3. Benefits of the simplified version of ShowVoc and the NACE index

The customisation project of ShowVoc for the Eurostat-managed asset, NACE, was initiated with a series of requirements designed to benefit from new functionalities: a user-friendly interface tailored terminologically for statisticians and non-technical users, straightforward text

definitions, easy to use search and download capabilities, targeted inclusion or exclusion of descriptions, a simplified two-tab menu for data and downloads, a customisable landing page, and the removal of ShowVoc's more complex features, such as collections (groupings of terms or concepts) or properties (set of labels and descriptions).

The Eurostat-specific ShowVoc is thus introduced with personalised footer and header designs, an advanced view, configurable labels that align with statistical terminology, new custom groups and sections, a selection of visible properties, formatted text, enriched text with links, and a search ribbon located at the bottom of the interface.

The project is currently in a phase of enhancement, incorporating additional specifications such as a default view augmented by custom project views, a single entry per dataset, collapsible and/or scrollable content to enhance readability, refined sections, and labels to prevent overlap (thereby clarifying the distinctions between activities, correspondences, and alignments), system and platform synchronisation, and an autocomplete search function.

The transformation of NACE open data and its conversion into linked data presents multiple advantages for diverse stakeholders, including researchers, policymakers, and reference data editors. It ensures that users across different fields and specialisations can benefit from semantic technologies.

In practical terms, the conversion of NACE classification into linked data, facilitating the ongoing exchange and integration of data between various platforms and applications. This guarantees the data's consistency, precision, and currency across numerous systems, allowing users to effortlessly compare information and discern trends or patterns. The improved interoperability also encourages collaboration among stakeholders, fostering continuous innovation and the refinement of classification systems.

Additionally, several interesting use cases can arise from the NACE index, which breaks down the content of NACE classes in very detailed entries and a keyword-type, more machine-processable format. The NACE index can be used as an important information source for automatic coding approaches of statistical units according to NACE classes or the artificial intelligence-based creation of correspondence tables between different versions of NACE.

References

- Carroll, J. J., Klyne, G., McBride, B. (2014). *RDF 1.1 Concepts and Abstract Syntax*. <https://www.w3.org/TR/rdf11-concepts/>
- Eurostat. (2008). *Statistical classification of economic activities in the European Community (NACE Rev. 2)*. <https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.pdf/dd5443f5-b886-40e4-920d-9df03590ff91?t=1414781457000>
- International Organization for Standardization. (2017). *ISO 15836-1:2017 - Information and Documentation - The Dublin Core Metadata Element Set*. <https://www.iso.org/standard/39077.html>
- Hancock, A. (2013). *Best Practice Guidelines for Developing International Statistical Classifications*. https://mdgs.un.org/unsd/classifications/bestpractices/Best_practice_Nov_2013.pdf
- Miles, A., Bechhofer, S. (2009). *SKOS Simple Knowledge Organization System Reference*. <https://www.w3.org/TR/skos-reference/>
- OWL Working Group. (2012). *OWL 2 Web Ontology Language Document Overview*. <https://www.w3.org/TR/owl2-overview/>
- ShowVoc. (2021). *ShowVoc: Documentation*. <https://showvoc.uniroma2.it/doc/>
- VocBench. (2021). *VocBench: Documentation*. <https://vocbench.uniroma2.it/doc/>