

## Annex B. National firm-level data sources

### Finland

The outcomes from Finland were generated thanks to the co-operation with Statistics Finland.

The data is available for the years 2008 to 2018. The final dataset covers the business world in Finland.

The final dataset used for the analysis is compiled of the following sources that contain corresponding information:

- The *Statistical Business Register* includes the following variables used in the scale-up analysis: firm-level identifier, start date, end date, the field of activity, legal form.
- *Structural Business Statistics* includes information on turnover, value-added, EBITDA<sup>2</sup>, purchases, total employment costs, number of full-time equivalent (FTE) positions.
- *International Trade in Goods Statistics* reports information on exports and imports for each trading firm by product and country. Data for International Trade in Goods Statistics are collected by Finnish Customs
- *FOLK* is a linked employer-employee register with information on wage, education, occupation, employee age, gender, nationality, last promotion, etc.

The Statistical Business Register is at the core of microdata linking. Data are aggregated on the level of the firm with corresponding shares from the employee- or product-level datasets and linked based on firm-level identifiers available in all data sources.

Employment is measured as headcount at the end of the reference year.

### Italy

The analysis for Italy was made possible thanks to the collaboration with researchers from the Bank of Italy.

The sample used includes all corporations observed in the Company Accounts Data System (CADS)<sup>3</sup> database for at least one year between 2001 and 2018. For the period under analysis, the universe of incorporated firms (around 650 000 entities) is covered, which account for about 70% of the total revenues of the private non-financial sector. The data are carefully analysed as this database is used extensively by banks for credit decisions.

CADS contains detailed balance sheet information but the number of employees is reported only for a small number of companies. CADS is therefore complemented with Italian Social Security Institute (*Istituto Nazionale per la Previdenza Sociale*, INPS) data, which report the number of employees for the world of Italian firms, calculated as an average headcount per year.<sup>4</sup>

## Portugal

Portuguese data is available for research in-house and is compiled and redistributed by the National Statistical Institute Portugal (*Instituto Nacional de Estatística*, INE).

The data covers the universe of public-owned and private firms operating between 2010 and 2018. The final database is composed of the following inputs:

- *Quadros de Pessoal* (Personnel Records) contain information on all firms in Portugal excluding self-employed workers. The database provides the annual accounts of all enterprises incorporated under Portuguese law that are legally required to file their annual accounts with the Ministry of Employment of Portugal. These annual accounts typically contain the main figures on turnover, number of employees, date of establishment of an entity, as well as industry and location of a firm.
- The same database, *Quadros de Pessoal*, contains all employees working in firms with at least one employee. The dataset covers all employees in all establishments of a firm and is reported in October of each year. The employee-level variables include wages (normal and overtime pay on an hourly basis), occupation classification, education level and individual characteristics such as age, gender and nationality of a worker.
- Financial data are extracted from the *Integrated Business Accounting System* (IBAS), which comprises information reported by firms, from tax authorities and business register data. The data thus covers the population of companies. Information for firms that do not report in time is replaced by information from the business registry.
- Information on trade, based on customs data, covers firms engaged in exports or imports activities within or outside of the European Union and is defined at a firm-product-destination level. The following information is reported: number of exported and imported products defined with the 8-digit Combined Nomenclature (CN) classification, code of destination or source country, the volume of transactions and nature of flows (exports or imports).

Each firm is identified by a unique identification number, which allows tracing the firm in the cross-section and panel data.

Employment is measured as headcount in the firm in October of each year.

## Slovak Republic

The output for the Slovak Republic is an outcome of co-operation with the Institute of Financial Policy (IFP), a branch of the Ministry of Finance of the Slovak Republic. The IFP provided part of the dataset for the OECD's inhouse research. For confidential data, the IFP co-operated with the OECD to produce the final output.

The IFP collects historical data on firms from several sources such as the Statistical Office, the Social Insurance Agency and the registry of financial statements and compiles the individual sources into one database. The final dataset covers a full business population of approximately 200 000 firms. The firms with different accounting standards such as the self-employed are excluded from the dataset.

The database starts in 2004 but the most reliable data are available starting from 2013. Financial statement reporting was voluntary for firms up until 2014, hence the 2004-12 dataset does not contain a full sample of firms and/or variables containing only a subset of firms. Therefore, the scale-up analysis focuses on the period 2013-18. Due to this limitation, it is not possible to compare non-scalers with scalers before, during and after scaling for the Slovak Republic. Data on firm employment from the Social Insurance Agency are available only from 2014 onwards.

The following sources of data are used:

1. The Statistical Office business register dataset, which provides information on the sector, location, ownership, employment size category, incorporation date and termination date.
2. Firms' financial statements, which report information from their yearly balance sheet, including value-added, profits, assets, loans and current/non-current liabilities.
3. Records from the Social Insurance Agency on the firm's total number of employees, the wage structure within the firm and the number of foreign-born employees.

Datasets are merged using a unique firm identification number.

The measure of employment reflects the headcount of full-time personnel in the firm.

## Spain

The data for Spain is provided by researchers at the Bank of Spain. It covers the firms over the period 2003 to 2018.

The micro-level firm dataset used for this report is built using the data from financial statements that all firms in Spain are required by law to submit annually to the Commercial Registry (*Registro Mercantil*). Firms are obliged by law to provide accurate information on their financial situation, making the financial statements a reliable source of accurate information. The Commercial Registry regularly transfers to the Bank of Spain digitalised raw data on the financial statements submitted by firms. The Statistical Department of the Bank of Spain then processes and cleans this raw data according to exhaustive statistical and accounting criteria, resulting in the *Central de Balances Integrada* (CBI) dataset. This database is only available for inhouse economists and external researchers working in collaboration with the bank's staff on selected investigation projects.

Despite the continuous efforts to improve the coverage of the dataset, it does not cover private sector firms as it excludes companies that submit information late (after the regular submission deadline) or via paper format. One of the efforts to expand the coverage includes the acquisition of the SABI database.<sup>5</sup> The SABI database also uses the financial statements submitted by firms to the Commercial Registry but its main potential advantage lies in covering large- and medium-sized firms that submit their statements either late or on paper.

Employment data for Spain measure full-time equivalent (FTE) employment over the given year.



**From:**  
**Understanding Firm Growth**  
Helping SMEs Scale Up

**Access the complete publication at:**  
<https://doi.org/10.1787/fc60b04c-en>

**Please cite this chapter as:**

OECD (2021), "National firm-level data sources", in *Understanding Firm Growth: Helping SMEs Scale Up*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/a22a6994-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.