Creating an EU-wide Micro Data Infrastructure (MDI):
a handbook for Micro-Data Linking

Eric Bartelsman[a,b], Filippo di Mauro[c,d], Mirja Hälbig[d]

[a] *Tinbergen Institute, Amsterdam*
[b] *Vrije Universiteit Amsterdam*
[c] *National University of Singapore (NUS)*
[d] *Halle Institute for Economic Research (IWH)*

Definitions and Acronyms

Acronyms	Definitions
CIS	Community Innovation Survey
ICTeC	ICT use/ E-Commerce Survey
FATS	Foreign Affiliation Statistics
FRIBS	Framework Regulation Integrating Business Statistics
GVC	Global Value Chains
ISS	International Sourcing Statistics
LEED	Linked Employer-Employee Data
MDI	Microdata Infrastructure
MDL	Microdata Linking
NCB	National Central Bank
NSI	National Statistical Institute
SBS	Structural Business Statistics

Content

1 Introduction and Objectives

Availability of reliable micro data is universally considered by now an essential tool for research and policy work. Unfortunately, at the EU level, access to micro data is still unsatisfactory, which is at odds for instance with the recent directive on open data (EU 2019/1024)[1]. Micro data are commercially available by private providers, such as Bureau van Dyck (ORBIS) and are widely used, but coverage and cross country comparability is highly doubtful, and subscription costs are unbelievably high (in the order of several hundred thousand Euros per subscription per annum). Statistical Institutes (NSI) have progressed tremendously in the last few years in generating high quality datasets despite tight budget constraints, but data access is mostly restricted to national researchers and attempts to generate cross-country links have been done on an experimental basis and have lasted not further than the duration of specific EU financed projects (e.g. ESS MDL).

Within MICROPROD we have experimented with new micro data (e.g. intangible assets) to improve productivity measurement and shed light on its drivers in a highly technology driven economy. In doing so we have also set-up a cross country Micro Data-Infrastructure (MDI), use of which may go well beyond the productivity measurement focus. The MDI links firms' attributes across different datasets using homogenous variable definitions; when fully operational, researchers will be able to run independently their routines across the country statistics available. Since it has been set up using top notch knowledge and long experience by six (pilot) NSIs, the Micro Data Infrastructure (MDI) we have created has the potential to become a top standard, which can be broaden across the entire EU. We propose to make permanent that infrastructure and we call relevant parties to ensure adequate financing to sustain it over the long run.

This handbook describes the MDI, including guidance to micro data linking for additional NSI participants. It presents commonly faced challenges and possible solutions, as well as costs.

After introducing the concept of micro data linking (MDL) in section 2.1, we will dig deeper in the individual data sources across which the data linking needs to be implemented in order to draw a comprehensive picture of the productivity effects of technology or globalization (section 2.2). Section 3 summarizes the information content of the databases with respect to intangibles (3.1) and international orientation (3.2) and discusses ways to operationalize these concepts on the basis of the available data. In section 4 and 5, we describe common issues of linking and harmonizing the available data sources across countries, which is a prerequisite for developing stylized facts on productivity and its drivers. An outline emphasizing the need for the proposed MDI completes the report.

---

[1] https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information

2 Method and Approach

2.1 Microdata Linking (MDL)

Microdata linking (MDL) consists in connecting a variety of micro datasets using individual firm identifiers. For instance, sales (included in the Business registers) are connected with export activities (customs data) as well as firm employment (employment statistics); this provides critical inputs as we assess for instance the extent in which productivity gains are related to export activities and how they affect employment patterns.

At the EU level - under the leadership of the EUROSTAT - there is a large amount of datasets which are by now regularly collected using harmonised criteria across EU countries. They are published however in sector-aggregated format, which implies that lots of the original information at the firm level is lost. Using the same surveys and datasets, our pilot NSIs gather back together the original information and link it across surveys and datasets at the individual firm level. The result has been an incredibly rich set of information which allows us to understand for instance how a number of factors affect productivity – not only on average – but across firms (using different benchmarks, such as size and/or productivity outcomes). This is critical information, as we aim for instance to identify best practices (e.g. management, labor, financing) across firms, in order to promote policies aimed at propping up the aggregate productivity of sectors and countries.

Previous projects mostly financed by the EU have shown promising results when linking for instance trade statistics to firm performance indicators, detecting considerable heterogeneity in firms' exposure to and gains from globalization. Invariably, however, these projects have died out because of lack of financing. The proposed MDI is conceived as a permanent infrastructure. Via the MDI researchers will be able to write code modules that make use of harmonized linked micro data from a network of NSI partners. The ultimate goal is to set up, in each country, a collection of linkable datasets, for a broad set of selected indicators. Metadata describing the datasets, their variables, and characteristics of the linked sample allow common code to be run on secure data at each site.

Prospects of fully implementing the proposed MDI are good. There is an increasing number of EU NSIs involved in microdata linking. While the heterogeneity across NSIs with respect to the technical and institutional infrastructure is not trivial, the strong pressure from researchers and policy makers to iron out the differences is also strong and MICROPROD partners are firmly committed to it.

2.2 Data Sources

For most of the surveys, the collected information and data coverage are standardized by EU regulations.

Statistical Business Register (SBR)

The statistical business register (SBR) plays a central role in the production of business statistics and is the starting point for establishing statistical survey frames. The SBR contains information on identification characteristics such as ID numbers, names and addresses, demographic characteristics, economic activity, legal form and institutional sector code as well as information on control and ownership relations for enterprises, their local and legal units and enterprise groups. In MDL, the SBR serves as a 'backbone' or connection between various surveys and datasets.

Structural Business Statistics (SBS)

The Structural Business Statistics (SBS) covers the business economy including industry, construction, distributive trade and services in the European Union. Data are provided by all EU Member States, Norway and Switzerland, some candidate and potential candidate countries. The main variables within SBS are generally collected and presented as monetary values, or as counts (for example, numbers of enterprises or persons employed). Based on SBS regulatory standards, firm-level information on gross fixed capital investment by broad asset type is collected via an annual investment survey (e.g. in Hungary, Germany and the Netherlands).

Community Innovation Survey (CIS)

The Community Innovation Survey (CIS) is part of the EU science and technology statistics and provides mostly qualitative information on firm innovative activity. Surveys are carried out every two years by EU member states and a number of ESS member countries on a voluntary basis. The harmonized survey contains information on the types of innovation and various aspects of the development of an innovation, such as the type of funding and innovation expenditures.

ICT usage/ E-Commerce Survey (ICTeC)

The Community survey on ICT usage and e-commerce in enterprises is an annual survey conducted since 2002, which collects information on the use of information and communication technology, the internet, e-government, e-business and e-commerce in enterprises. Like the CIS, the EC survey contains mostly qualitative data.

International Trade Statistics

Firm-level statistics concerning exports and imports are the International Trade in Goods Statistics (ITGS) and International Trade in Services Statistics (ITSS).

International trade in goods statistics (ITGS) measure the value and quantity of goods traded between EU Member States (intra-EU trade) and goods traded by EU Member States with non-EU countries (extra-EU trade) broken down by types of goods (Combined Nomenclature) and by partner countries. The providers of statistical information differ between intra and extra EU-trade. In the first case, it corresponds to all taxable persons reporting transactions exceeding a certain threshold fixed by member states; in the second one, it corresponds to

administrative data from the customs declarations lodged by natural or legal persons in the customs administration.

International Trade in Services Statistics (ITSS) typically cover trade in services, i.e. transactions paid for the services that have taken place between the residents and non-residents.

Foreign Affiliate Statistics (FATS)

The Foreign Affiliate Statistics is distinguished into inward FATS, i.e. the activity of foreign affiliates resident in the compiling country, and the outward FATS, that is, the activity of foreign affiliates abroad but controlled by the compiling country. The FATS allow to qualitatively and quantitatively assess the degree of economic activity of a domestic enterprise abroad and identify foreign-controlled firms.

Other Sources

Another promising and interesting source is linked employer-employee data (LEED) that cover the working populations' characteristics like employment relations, income and education and socio-demographic characteristics.

The International Sourcing survey (ISS) gathers data on international organisations and sourcing of business functions in 16 European countries, covering the period 2014-2016 or 2015-2017, depending on the country. The survey results cover nearly 60 000 businesses each with more than 50 persons employed. However, since the survey is still in pilot stage, the survey design varies across countries.

Balance Sheet Data provides information on firms' financial assets and liabilities. While for nearly all pilot countries, firm financial data is available at the NSI, for some countries it is only available at the respective National Central Bank (NCB) (e.g. Germany).

3 Applications

In what follows, we provide examples on how the MDI can be concretely utilized for research purposes. The two examples included are tightly related to the ultimate objective of MICROPROD, i.e. understanding the main drivers of recent productivity trends and most notably its sluggish patterns.

3.1 Application 1: Intangibles

Intangible capital is often modeled to consist of research and development (R&D), software, patents, and branding and organization capital (Corrado, Hulten, and Sichel 2005). A subset of the above concepts is covered by standard SBS items. The SBS includes information on investment in (i) concessions, patents, licenses, trade-marks and similar rights, (ii) purchased software and (iii) total intra-mural R&D expenditure. These variables can be used to calculate

an intangible capital stock (Kaus, Slavtchev, and Zimmermann 2020; Bisztray, Vonnak, and Muraközy 2020).

The CIS covers both innovation outputs and the innovative process and inputs (type of funding, R&D expenditure) and distinguishes four innovation types: process, product, organizational, marketing, thus covering both innovative property as well as capabilities and organizational capital. Additionally, the CIS asks about the novelty of the innovation, i.e. whether it is new for the market, new to the country, developed by the firm or was adopted, and thus provides information about the innovative value.

The ICT use survey measures various dimensions of firm technology use. Besides software and databases being considered as an integral part of intangibles, the adoption of certain technologies also provides information about firms' organizational capital and ICT capabilities both in the firms' internal operations and regarding the firms' supplier and buyer relationships. The qualitative information in the survey can be used to construct an ICT intensity index which allows for variation in the underlying source variables, thereby overcoming the issue with changing survey questions and the saturation of certain variables over time (Bartelsman, Hagsten, and Polder 2018).

Additionally, firm financial data provides information on intangibles. Balance sheet data contain an accounting measure which aims to capture the entire stock of intangible capital. However, intangibles can only appear on the balance sheet of a company if their value is clearly identifiable, with the shortcomings that (i) acquired assets are much more likely to enter the firm's balance sheet, (ii) the item covers only certain aspects of the economic concept of intangibles and (iii) does not necessarily reflect the economic value of the incorporated intangible assets due to accounting principles and depreciation rules (Bisztray, Vonnak, and Muraközy 2020). Additionally, the profit and loss statement includes Sales, General and Administrative (SG&A) expenditures which is – besides Costs of Goods Sold (COGS) – the second major component of costs and includes all intangible-building activities (e.g., R&D, Advertising and IT staff expense) (Covarrubias, Gutiérrez, and Philippon 2019).

Linked employer-employee data can be used to construct a firm-level indicator of human capital that is complementary to intangible assets. For example, using data on education level and field, the number of employed data scientists could be used as a proxy for investment in software and databases.

3.2   Application 2: International Orientation

The international orientation of a firm can be, for example, captured by firm's export choices and integration in global value chains (GVCs). Research regarding the operationalization of firms' participation in GVCs is in progress in MICROPROD. This documentation will be adjusted as new results become available.

4 Legal and Administrative Issues

4.1 Data Access Conditions

The main objective of the proposed MDI is to enable external researchers access to firm-level microdata. Table 1 below shows that there is heterogeneity in access conditions across NSIs, which will need to be ironed out eventually.

Table 1: Conditions of data use license.

NSI	Conditions of data use
Statistics Finland	The license applicant may be an official body, an institution or a person in charge of a study. Data access is granted for a fixed period and specified persons.[2]
Statistics Sweden	Data access is granted to applicants from EU or EEA country, or in a third country approved for transfer by the European Commission for research or statistical purposes.[3]
Statistics Norway	Researchers, PhD students and master's degree students at an approved research institution can use the microdata.[4]
Statistics Netherlands	Access to CBS microdata is granted only to authorised institutions solely for statistical purposes.[5]
Statistics Denmark	Only Danish research environments are granted authorisation. Foreign researchers can, however, get access to micro data through an affiliation to a Danish authorised environment.[6]
Statistics France	Access to microdata is granted for research purposes.[7]
Statistics Hungary	Access to microdata is granted for research purposes.[8]
Statistics Germany	Scientific facilities assigned with independent scientific research (universities and scientific institutions) are eligible for use.[9]

---

[2] https://www.stat.fi/meta/tietosuoja/kayttolupa_en.html
[3] https://www.scb.se/en/services/guidance-for-researchers-and-universities/mona--a-system-for-delivering-microdata/terms-of-use/
[4] https://microdata.no/en/access/
[5] https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research
[6] https://www.dst.dk/en/TilSalg/Forskningsservice
[7] https://www.casd.eu/en/
[8] https://www.ksh.hu/data_access_researcher_accreditation
[9] https://www.forschungsdatenzentrum.de/en/terms-use

It is our intention to dig deeper into such differences and eventually propose a common approach, which will make access seamless regardless of the nationality of the researcher. This may well imply changes in legislation or EUROSTAT regulations.

4.2   Technological Infrastructure

A minimum requirement for NSIs to participate in the proposed MDI is to allow for remote execution of data, even better if this includes availability of remote access. All the considered pilot NSIs have either (i) installed a remote access service, allowing the researcher to access the respective NSI's server from her workstation via the web and use the data in a protected environment or (ii) allow for remote execution of sent code.

Table 2: Technical access conditions.

NSI	Remote Access or Execution	Notes
Statistics Finland	Remote Access	via FIONA
Statistics Sweden	Remote Access	via MONA
Statistics Norway	Remote Access	via RAIRD
Statistics Denmark	Remote Access	via DRS
Statistics Netherlands	Remote Access	via VPN
Statistics France	Remote Access	via CASD
Statistics Hungary	Remote Access/ Execution	see here
Statistics Germany (Destatis)	Remote Execution	via FDZ

For common code to be able to be run at each site, a common statistical software will be required. Table 3 shows, however, that there is considerable heterogeneity across NSIs with respect to the utilized software.

Table 3: Available Software across Pilot NSIs.

NSI	Available Software
Statistics Finland	R, Python Anaconda + components, Stata, SAS,SPSS Statistics, OctaveStat/Transfer Graphviz, Kdiff, Libre Office (word processing and tabulation), Lyx, Latex, MikTeX, Notepad++, Pnet, PajekRstan, Rtools, Rstudio, TramineR (connected to R-program), Strawberry, QGIS
Statistics Sweden	EquiPop, Geoda, GIT, IDLE - Python, MPlus, QGis, R, R-Studio, SAS, SAS Enterprise Guide, SAS IML Studio, SPSS, Spyder, STATA, SuperCross
Statistics Norway	SciPy, NumPy, Pandas, Statsmodels
Statistics Denmark	SAS, SPSS, STATA, GAUSS and R
Statistics Netherlands	SPSS, Stata, R, StatTransfer, MSAccess, MSWord, MSExcel, MSPowerpoint, Winzip, AdobeAcrobatReader, Windows Explorer, Blaise, WinEdt, Winbugs
Statistics France	7-zip, Adobe Reader DC, GeoDa, Ghostscript ,GSView, Gvim, HeidiSQL, Libre Office, MariaDB, Matlab, Microsoft Office, MikTeX, MySQL, Notepad, Philcarto, Python, Qgis, R, Rstudio, SAS, Scilab, SPSS, Stata, Tau-Argus, TeXnicCenter, Visual Studio Code, XLStat
Statistics Hungary	STATA, R
Statistics Germany	SPSS, SAS, Stata or partly R

Going forward, we will investigate with NSIs on possible homogenization of the software utilized.

4.3 Confidentiality Practices

Another issue concerns possible heterogeneity in confidentiality practices by country. The output of a code module could consist of output datasets, aggregated to industry or other firm-level characteristics to avoid breaking confidentiality, or tables of analytical results (i.e. regression coefficients and diagnostics). The output needs to be checked for disclosure by the NSI partners. The specifics on the kind of output that can be extracted from the data infrastructure are to be defined in collaboration with the NSIs.

The trade-off here is to coordinate on the lowest common denominator, i.e. the most restrictive disclosure policy, or to accept the heterogeneity and make it transparent to the researcher via comprehensive metadata and documentation.

5 Harmonization of Data

One of the main challenges is balancing the trade-off between dataset harmonization and loss of information. While it is preferable that the individual datasets are designed to be as similar as possible to the underlying surveys at the NSI, considerable harmonization will have to take place to make each underlying dataset comparable over time and across countries, e.g. with respect to variable names, the units and response categories for each variable, definitions of industry, size and age classes of firms, and the coding of missing values.

This section provides a summary of issues encountered in the cross-country homogenization of data. The specifics of the dataset homogenization are defined in cooperation with the NSIs.

5.1 Unit consistency

The statistical units for the linked microdata are derived from the national SBR. Even though EU regulations on Business Registers (EC 177/2008) and statistical units (EEC 696/93) define enterprise and enterprise group as the standard statistical unit, the adoption of these regulations are in different phases across the ESS.

5.2 Variable consistency

Since administrative and statistical information is collected for various purposes, the exact definition of a certain variable may vary by country. In order to be able to compare key statistics across countries, a certain degree of harmonization is required. However, choosing the common denominator may result in lower quality data. For example, employment reported in full-time equivalents is preferable to headcount measures. The former, however, is not available for all countries. The proposed solution therefore is to collect various manifestations of a variable as they are available which also provides the opportunity for robustness-checks.

5.3 Aggregation of data

Industry Coverage

One issue arises due to differences in industry coverage by various sources and countries. Since for most data sources data coverage is defined by EU regulation, differences in industry coverage should only be a minor issue.

Size-Class Coverage

Another concern are differences in the size-class coverage. While the exclusion of smaller firms is no issue in the Nordic countries, data coverage is limited to firms with more than 20 employees in Germany.

Census and non-census sources

Data sources can cover the entire firm population, i.e. they form a census, or they can cover only a part of the population, in which case they are a sample. Census data is representative of the population by definition, whereas sample-based data usually require weighting to obtain representative aggregates.

Some of the required data are samples (e.g. the CIS and ICTeC). However, all sample sources can be linked to the national business registry, thus enabling the calculation of weights to adequately reflect the population distribution. Concerns of systematic sample selection biases arise only when the presence of firms in the merged data is determined by some underlying unobserved variables.

6 Outline

The proposed MDI is a cost-efficient way to make available to researchers and policy makers EU-wide micro information. We have illustrated two examples related to the MICROPROD terms of references, but the applications are limitless.

In order to set up such research infrastructure spanning across EU NSIs, however, a number of issues need to be addressed, including heterogeneity in the technical and legal infrastructure, as well as on some harmonization issues. This notwithstanding, the strong pressure for better and more accessible data – also as reflected in recent EU directives - bodes well for a quick solution of the above issues.

References

Altomonte, Carlo, Giorgio Barba Navaretti, Filippo di Mauro and Gianmarco Ottaviano. 2011. "Assessing competitiveness: how firm-level data can help", Bruegel Policy Contribution, Issue 2011/16, November.

Bartelsman, Eric, Eva Hagsten, and Michael Polder. 2018. "Micro Moments Database for Cross-Country Analysis of ICT, Innovation, and Economic Outcomes." *Journal of Economics & Management Strategy* 27 (3): 626–48.

Bisztray, Marta, Dzsamila Vonnak, and Balazs Muraközy. 2020. "Analysis of the Importance of Intangible Capital and Knowledge for Productivity Measurement." *Mimeo.*

Corrado, Carol*, Charles Hulten, and* Daniel Sichel*. 2005. "*Measuring Capital and Technology: An Expanded Framework*," in C. Corrado*, J. Haltiwanger*, and* D. Sichel *(eds),* Measuring Capital in the New Economy, Studies in Income and Wealth*, Vol. 65, University of Chicago Press, Chicago:* 11-46.

Covarrubias, Matias, Germán Gutiérrez, and Thomas Philippon. 2019. "From Good to Bad Concentration? U.S. Industries over the Past 30 Years." Working Paper 25983. National Bureau of Economic Research.

European Commission, and Statistical Office of the European Union. 2019. *Micro Data Linking: 2019 Edition.* https://op.europa.eu/publication/manifestation_identifier/PUB_KSGQ19013ENN.

Kaus, Wolfhard, Viktor Slavtchev, and Markus Zimmermann. 2020. "Intangible Capital and Productivity. Firm-Level Evidence from German Manufacturing." *IWH Discussion Papers*, no. 1/2020.

Annex: Variable List

Business Register

Variable	Data Source	Label/Description	Format	Values
ALREADY AVAILABLE IN NORDIC DATASET				
ENT_ID	BR	Unique enterprise identification	Character	X
ENTgrp_ID	BR	Enterprise Group ID	Character	X
AD_ID	BR	Administrative ID	Character	X
Start_Ent	BR	Start date for the enterprise ID	Character	Date DDMMYYYY
End_ent	BR	End date for the enterprise ID	Character	Date DDMMYYYY
START_ENTgr	BR	Start date for the enterprise Group ID	Character	Date DDMMYYYY
END_ENTgr	BR	End date for the enterprise Group ID	Character	Date DDMMYYYY
LEGAL	BR	Legal form of the enterprise ID	Number	LL= Limited liability company - include limited liability partnerships and public corporations, SP= Sole Proprietor, PA= Partnership - exclude limited liability partnerships, GO= Government - local and central government - exclude public corporations, NB= Non profit body or mutual association, NP= Natural person(s) - include only if not involved in any economic activity – where economically active, code as sole proprietor, ND= Not defined
NACE_M	BR	Main activity of the enterprise (NACE 4-digit)	Character	2016, 3320, etc.
PUB	BR	Ownership of the enterprise (private/public)	Number	1=private, 2=public, 9=not available

Variable	Data Source	Label/Description	Format	Values
START_NACE_M	BR	Start date for the main activity	Character	Date DDMMYYYY
DEMO_REL	BR	Information on demographic relations (mergers and acquisitions etc.)	Character	0="No demographic relation in ref. year", 1="Receiving employment from other enterprise in ref. year", 2="Transfers employment to other enterprise and cease to exist in ref. year", 3=" Transfers employment to other enterprise and continues in ref. year"

Structural Business Statistics (SBS)

Variable	Data Source	Label/Description	Format	Values
ALREADY AVAILABLE IN NORDIC DATASET				
ENT_ID	SBS	Enterprise ID (identifikation number)	Character	X
SBS_12110	SBS	Turnover	Number	Amount (max. 2 decimals) (National currency)
SBS_12150	SBS	Value added at factor cost	Number	Amount (max. 2 decimals) (National currency)
SBS_12170	SBS	Gross operating surplus	Number	Amount (max. 2 decimals) (National currency)
SBS_13110	SBS	Total purchases of goods and services	Number	Amount (max. 2 decimals) (National currency)
SBS_13310	SBS	Personnel costs	Number	Amount (max. 2 decimals) (National currency)
SBS_13320	SBS	Wages and salaries	Number	Amount (max. 2 decimals) (National currency)
SBS_16110	SBS	Number of persons employed	Number	Amount (max. 2 decimals)
SBS_16130	SBS	Number of employees	Number	Amount (max. 2 decimals)

SBS_16140	SBS	Number of employees in full-time equivalents	Number	Amount (max. 2 decimals)
SBS_Type	SBS	Code to show if data (unit) is observed or imputed in SBS	Number	1= observed, 2=imputed
PLANNED ADDITION TO NORDIC DATASET				
SBS_15420	SBS	Gross investment in concessions, patents, licences, trade marks and similar rights	Number	
SBS_15441	SBS	Investment in purchased software	Number	
SBS_15442	SBS	Investment in software produced by the unit (optional)	Number	
SBS_22110	SBS	Total intra-mural R & D expenditure	Number	
SBS_13110	SBS	Total purchases of goods and services	Number	
SBS_20110	SBS	Purchases of energy products (in value)	Number	
SBS_15150	SBS	Gross investment in machinery and equipment	Number	
SBS_15110	SBS	Gross investment in tangible goods	Number	

Community Innovation Survey (CIS)

Variable	Data Source	Label/Description	Format	Values
ALREADY AVAILABLE IN NORDIC DATASET				
ENT_ID	CIS	Unique enterprise identification	Character	X
ENTGP	CIS	Enterprise part of a group (From BR)	Number	0=No, 1=yes
HO	CIS	Country of head office(From IFATS)	Character	FR, DE, IT etc.
MAREUR	CIS	Other EU/EFTA/CC market	Number	0=No, 1=yes
MAROTH	CIS	All other countries	Number	0=No, 1=yes
INPDGD	CIS	Introduced onto the market a new or significantly improved good	Number	0=No, 1=yes

INPDSV	CIS	Introduced onto the market a new or significantly improved service	Number	0=No, 1=yes
NEWMKT	CIS	Did the enterprise introduce a product new to the market	Number	0=No, 1=Yes, missing=no answer
TURNMAR	CIS	% of turnover in new or improved products introduced during 2006-2008 that were new to the market	Number	Percent (max. 2 decimals) (0-100)
INPSPD	CIS	Introduced onto the market a new or significantly improved method of production	Number	0=No, 1=yes
INPSLG	CIS	Introduced onto the market a new or significantly improved logistic, delivery or distribution system	Number	0=No, 1=yes
INPSSU	CIS	Introduced onto the market a new or significantly improved supporting activities	Number	0=No, 1=yes
INPSDV1	CIS	Who mainly developed these processes – your enterprise by itself	Number	1=tickmarked, 0=not tickmarked
INPSDV2	CIS	Who mainly developed these processes – your enterprise together with other	Number	1=tickmarked, 0=not tickmarked
INPSDV3	CIS	Who mainly developed these processes – your enterprise by adapting or modifying processes	Number	1=tickmarked, 0=not tickmarked
INPSDV4	CIS	Who mainly developed these processes – other enterprises or institutions	Number	1=tickmarked, 0=not tickmarked
RRDIN	CIS	Engagement in intramural R&D	Number	0=No, 1=yes
RDENG	CIS	Type of engagement in R&D	Number	1=Continuously, 2=Occasionally, missing=no answer

RRDINX	CIS	Expenditure in intramural R&D (in national currency)	Number	Amount (max. 2 decimals) (National currency)
RRDEXX	CIS	Extramural R&D (in national currency)	Number	Amount (max. 2 decimals) (National currency)
RMACX	CIS	Expenditure in acquisition of machinery (in national currency)	Number	Amount (max. 2 decimals) (National currency)
RTOT	CIS	Total of these four innovation expenditure categories (in national currency)	Number	Amount (max. 2 decimals) (National currency)
FUNLOC	CIS	Public funding from local or regional authorities	Number	0=No, 1=yes
FUNGMT	CIS	Public funding from central government	Number	0=No, 1=yes
FUNEU	CIS	Public funding from the EU	Number	0=No, 1=yes
FUNRTD	CIS	Funding from EU's 6th or 7th Framework Programme for RTD	Number	0=No, 1=yes
CO	CIS	Cooperation arrangements on innovation activities	Number	0=No, 1=yes
ORGBUP	CIS	New business practices for organising work or procedures	Number	0=No, 1=yes
ORGWKP	CIS	New methods of workplace organisation	Number	0=No, 1=yes
ORGEXR	CIS	New methods of organising external relations	Number	0=No, 1=yes
MKTDGP	CIS	Significant changes to the aesthetic design or packaging	Number	0=No, 1=yes
MKTPDP	CIS	New media or techniques for product promotion	Number	0=No, 1=yes
MKTPDL	CIS	New methods for product placement or sales channels	Number	0=No, 1=yes
MKTPRI	CIS	New methods of pricing goods or services	Number	0=No, 1=yes

Variable	Data Source	Label/Description	Format	Values
CIS_TYPE	CIS	Code to show if data is observed or imputed in SBS	Number	1= observed, 2=imputed
PLANNED ADDITION TO NORDIC DATASET				
ROEKX	CIS	Acquisition of existing knowledge from other enterprises or organisations		
ROTRX	CIS	All other innovation activities including design, training, marketing, and other relevant activities		
PROPAT	CIS	During the three years xxx to xxx, did your enterprise apply for a patent		
PROTM	CIS	During the three years xxx to xxx, did your enterprise Register a trademark		
PROLEX	CIS	During the three years xxx to xxx, did your enterprise license out or sell a patent, industrial design right, copyright or trademark to another enterprise, university or research institute		
PROLIN	CIS	During the three years xxx to xxx, did your enterprise license in or buy a patent, industrial design right, copyright or trademark owned by another enterprise, university or research institute		
EMPUD	CIS	percent of employees with tertiary degree		

ICT use and e-Commerce Survey (ICTeC)

Variable	Data Source	Label/Description	Format	Values

ALREADY AVAILABLE IN NORDIC DATASET				
ENT_ID	ICTeC	Unique firm id	Character	X
BROAD	ICTeC	Firm has broadband	Character	1=yes, 0=no, 9=not applicable, missing=no answer
AEBUY	ICTeC	Firm orders through computer networks (websites or EDI)	Character	1=yes, 0=no, 9=not applicable, missing=no answer
AEBVALPCT	ICTeC	% of orders through internet	Number	1 = Less than 1% 2 = 1% or more and less than 5% 3 = 5% or more and less than 10% 4 = 10% or more and less than 25% 5 = 25% or more and less than 50% 6 = 50% or more and less than 75% 7 = 75% or more 9 = not applicable missing=no answer
AESELL	ICTeC	Firm sells through computer networks (websites or EDI)	Character	1=yes, 0=no, 9=not applicable, missing=no answer
AESVALPCT	ICTeC	% of sales through computer networks (websites or EDI)	Number	Percent (0-100), -1=not applicable
IACC	ICTeC	Firm has internet	Character	1=yes, 0=no, 9=not applicable, missing=no answer
INTRA	ICTeC	Firm has intranet	Character	1=yes, 0=no, 9=not applicable, missing=no answer
WEB	ICTeC	Firm has website	Character	1=yes, 0=no, 9=not applicable, missing=no answer
MOB	ICTeC	Firm has mobile access to internet	Character	1=yes, 0=no, 9=not applicable, missing=no answer
ITERP	ICTeC	Enterprise Resource Planning	Character	1=yes, 0=no, 9=not applicable, missing=no answer
ADE	ICTeC	Automated Data Exchange	Character	1=yes, 0=no, 9=not applicable, missing=no answer
ADESU	ICTeC	to suppliers	Character	1=yes, 0=no, 9=not applicable, missing=no answer

INVREC	ICTeC	receiving e-invoices	Character	1=yes, 0=no, 9=not applicable, missing=no answer
ADECU	ICTeC	receiving orders	Character	1=yes, 0=no, 9=not applicable, missing=no answer
INVSND	ICTeC	sending e-invoices	Character	1=yes, 0=no, 9=not applicable, missing=no answer
ADEINFO	ICTeC	sending product information	Character	1=yes, 0=no, 9=not applicable, missing=no answer
ADETDOC	ICTeC	sending transport documents	Character	1=yes, 0=no, 9=not applicable, missing=no answer
ADEPAY	ICTeC	Use of ADE for sending payment instructions to financial institutions	Character	1=yes, 0=no, 9=not applicable, missing=no answer
ADEGOV	ICTeC	Use of ADE for sending or receiving data to/from public authorities	Character	1=yes, 0=no, 9=not applicable, missing=no answer
SISU	ICTeC	Sharing SCM data with suppliers	Character	1=yes, 0=no, 9=not applicable, missing=no answer
SICU	ICTeC	Sharing SCM data with customers	Character	1=yes, 0=no, 9=not applicable, missing=no answer
CRMSTR	ICTeC	share of information with other business functions	Character	1=yes, 0=no, 9=not applicable, missing=no answer
CRMAN	ICTeC	analyse information for marketing purposes	Character	1=yes, 0=no, 9=not applicable, missing=no answer
SISAINV	ICTeC	sales: management of inventory levels	Character	1=yes, 0=no, 9=not applicable, missing=no answer
SISAACC	ICTeC	sales: accounting	Character	1=yes, 0=no, 9=not applicable, missing=no answer
SISAPROD	ICTeC	sales: production or services management	Character	1=yes, 0=no, 9=not applicable, missing=no answer
SISADIST	ICTeC	sales: distribution management	Character	1=yes, 0=no, 9=not applicable, missing=no answer
SIPUINV	ICTeC	purchases: management of inventory levels	Character	1=yes, 0=no, 9=not applicable, missing=no answer

SIPUACC	ICTeC	purchases: accounting	Character	1=yes, 0=no, 9=not applicable, missing=no answer
ICT_Type	ICTeC	Observed or not observed data (unit)	Number	1=observed, 2=imputed
PLANNED ADDITION TO NORDIC DATASET				
NACE2	ICTeC	Nace code of the enterprise, 4 digits	Character	
EMP	ICTeC	Number of employed persons	Number	
PURCH	ICTeC	Total Purchases	Number	
TURN	ICTeC	Total Turnover	Number	
CUSE	ICTeC	Enterprise uses computers	Character	
EMPCUSEPCT	ICTeC	% of workers using computers	Number	
EMPCUSE	ICTeC	Persons employed using computers	Number	
ITSP2 (2012)	ICTeC	firm employs IT specialists	Character	
ITSP2 (2014-2017)	ICTeC	firm employs IT specialists (employees for whom ICT is the main job, e.g. develop, operate or maintain ICT systems or applications)	Character	
ITSPT2 (2012, 2014-2017)	ICTeC	Have provided training to develop ICT skills of personnel: for ICT specialists	Character	
ITUST (2012, 2014-2017)		Have provided training to develop ICT skills of personnel: for other persons employed	Character	
IACC	ICTeC	Enterprise has access to internet	Character	
EMPIUSE	ICTeC	Persons employed using computers with access to World Wide Web	Number	
EMPIUSEPCT	ICTeC	% of workers using computers with access to World Wide Web	Number	
IT_MOWN (2015-2017)	ICTeC	ICT functions are mainly performed by own employees	Number	

IT_MEXT (2015-2017)	ICTeC	ICT functions are mainly performed by external suppliers	Number	
ISPD (2011-2013)	ICTeC	Maximum contracted download speed of the fastest internet connection	Character	
ISPDF (2014-2017)	ICTeC	Maximum contracted download speed of the fastest fixed internet connection	Character	
PMD (2012-2013)	ICTeC	Provide to the persons employed portable computers that allow a mobile connection to the Internet for business use	Character	
EMPIACC3G (2011)	ICTeC	Persons employed provided with a portable device with at least 3G technology for accessing the Internet	Number	
EMPMD (2012-2013)	ICTeC	Persons employed, which were provided a portable device that allows a mobile connection to the internet for business use	Number	
EMPMD1 (2014-2017)	ICTeC	Persons employed, which were provided a portable device that that allows internet connection via mobile telephone networks, for business purposes	Number	
SM_ANY (2013)	ICTeC	Use any social media	Character	
SM1_ANY (2014, 2016-2017)	ICTeC	Use any social media	Character	
SM_PANY (2015)	ICTeC	Use any social media	Character	
BD (2016,2018)	ICTeC	firm analyses big data from any data source	Character	
BDOWN (2016,2018)	ICTeC	Big data analysis for the enterprise is done by the enterprise's own employees	Character	

BDEXT (2016,2018)	ICTeC	Big data analysis for the enterprise is done by an external service provider	Character	
CC (2014-2017)	ICTeC	Buy cloud computing services used over the internet	Character	
CC_HI (2014-2017)	ICTec	Buy high CC services (accounting software applications, CRM software, computing power)	Character	
CC_ME (2014-2017)	ICTec	Buy only medium CC services (e-mail, office software, storage of files, hosting of the enterprise's database)	Character	
CC_LO (2014-2017)	ICTec	Buy only low CC services (e-mail, office software, storage)	Character	
ENVRA (2011)	ICTeC	Provide to the persons employed remote access to the enterprise's e-mail system, documents or applications	Character	
RA (2016,2018)	ICTeC	Provide to the persons employed remote access to the enterprise's e-mail system, documents or applications	Character	
DI_INDEX (2014-2017)	ICTeC	Derived value: Digital Intensity index	Number	
P3D_OWN (2018)	ICTeC	using your enterprise's 3D printers	Character	
P3D_OTH (2018)	ICTeC	using printing services provided by other enterprises	Character	
RBT (2018)	ICTeC	Use industrial or service robots	Character	
RBTI (2018)	ICTeC	Use industrial robots	Character	
RBTS (2018)	ICTeC	Use service robots	Character	

Outward Foreign Affiliates Statistics (OFATS)

Variable	Data Source	Label/Description	Format	Values
ALREADY AVAILABLE IN NORDIC DATASET				
ENT_ID	OFATS	Unique enterprise identification	Character	X
ENTgrp_ID	OFATS	Unique enterprise group ID	Character	X
ENT	OFATS	Number of foreign affiliates	Number	Amount
EMP	OFATS	Number of persons employed in foreign affiliates	Number	Amount (max. 2 decimals)
TUR	OFATS	Turnover of foreign affiliates	Number	Amount (max. 2 decimals) (National currency)
CL_AREA_EE[1]	OFATS	Host country of affiliates	Character	FR, DE, IT etc.
Country_grp_OFATS[2]	ITGS	Aggregation of destination countries to groups	Number	1 = 'EU-15'
				2 = 'EU-13'
				3 = 'Other Europe'
				4 = 'Russia'
				5 = 'China'
				6 = 'Japan'
				7 = 'India'
				8 = 'Brazil'
				9 = 'USA'
				10 = 'Canada & Mexico'
				11 = 'Oceania and rest of Asia'
				12 = 'Central Am. & Rest of South Am'
				13 = 'Africa'
				14 = 'Unknown'
				15 = 'Other'
OFATS_type	OFATS	Code to show if data (unit) is observed or imputed in OFATS	Number	1=observed, 2=imputed

[1] Be aware that this variable is only in the input dataset delivered by the NSIs and not in the MDL-database
[2] Be aware that this variable is created when building the MDL-database, and should not be on the input datasets delivered by the NSIs

Inward Foreign Affiliates Statistics (IFATS)

Variable	Data Source	Label/Description	Format	Values

ALREADY AVAILABLE IN NORDIC DATASET				
ENT_ID	IFATS	Unique enterprise identification	Character	X
UCI_CO	IFATS	Country of ownership	Character	FR, DE, IT etc.
IFATS_type	IFATS	Code to show if data (unit) is observed or imputed in IFATS	Number	1=observed, 2=imputed

International Trade in Goods Statistics (ITGS)

Variable	Data Source	Label/Description	Format	Values
ALREADY AVAILABLE IN NORDIC DATASET				
ENT_ID	ITGS	Unique enterprise identification	Character	X
STAT_VALUE_ITGS	ITGS	Trade amount	Number	Amount (max. 2 decimals) (National currency)
Flow_ITGS	ITGS	Code to distinguish between import and export	Number	1=import, 2=export
CL_AREA_GEO_ITGS[1]	ITGS	Partner country (country of origin/destination) (2 letter code)	Character	FR, DE, IT etc.
Country_grp_ITGS[2]	ITGS	Aggregation of destination countries to groups	Number	1 = 'EU-15' 2 = 'EU-13' 3 = 'Other Europe' 4 = 'Russia' 5 = 'China' 6 = 'Japan' 7 = 'India' 8 = 'Brazil' 9 = 'USA' 10 = 'Canada & Mexico' 11 = 'Oceania and rest of Asia' 12 = 'Central Am. & Rest of South Am' 13 = 'Africa' 14 = 'Unknown'

				15 = 'Other'
CN08[1]	ITGS	Product nomenclature CN08 8-digit	Character	1011010, 84733080 etc.
BEC[2]	ITGS	Products aggregated to BEC-codes	Character	
ITGS_type	ITGS	Code to show if data (unit) is observed or imputed in ITGS	Number	1=observed, 2=imputed

[1] Be aware that this variable is only in the input dataset delivered by the NSIs and not in the MDL-database
[2] Be aware that this variable is created when building the MDL-database, and should not be on the input datasets delivered by the NSIs