Mapping competitiveness with European data

DAVIDE CASTELLANI AND ANDREAS KOCH

BRUEGEL BLUEPRINT 23

Mapping competitiveness with European data

DAVIDE CASTELLANI AND ANDREAS KOCH

BRUEGEL BLUEPRINT SERIES

BRUEGEL BLUEPRINT SERIES

Volume XXIII
Mapping competitiveness with European data

Davide Castellani and Andreas Koch

Editorial coordinator: Stephen Gardner
Production: Michael T. Harrington
Cover: Bruegel/Michel Krmek

BRUEGEL
33, rue de la Charité
1210 Brussels,
Belgium
www.bruegel.org

ISBN: 978-90-78910-36-7

MAPCOMPETE is a project designed to provide an assessment of data opportunities and requirements for the comparative analysis of competitiveness in European countries.

This project is funded by the European Union.

LEGAL NOTICE: This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 320197. The views expressed in this publication are the authors' alone, and do not necessarily reflect the views of the European Commission.

CENTRO STUDI LUCA D'AGLIANO

SciencesPo
DEPARTMENT OF ECONOMICS

The project leader is László Halpern for CERS-HAS. The leaders of the six teams are: Carlo Altomonte (Bocconi University) for Bruegel; Giorgio Barba Navaretti (University of Milan) for Ld'A; Gábor Békés for CERS-HAS; Andreas Koch for IAW; Lionel Fontagné for Paris School of Economics and Philippe Martin for Science Po.

Supporting institutions are the National Bank of Belgium, Banque de France, Banco de España, Deutsche Bundesbank, Banca d'Italia, Magyar Nemzeti Bank and the Italian National Institute of Statistics (ISTAT).

Contents

Acknowledgements

This report is truly the result of a collective effort by many senior and junior researchers at the six institutions that are part of the MAPCOMPETE project (Bruegel, Centro Studi Luca d'Agliano, CERS-HAS, IAW, Paris School of Economics and Sciences Po). The authors would like to acknowledge that parts of this report were initially drafted by other researchers. In particular:

- Section 2.1 builds on the 'Competitiveness Indicators Report' (deliverable 3.1 of the MAPCOMPETE Project) by Carlo Altomonte, Marco Antonielli, Michael Blanga-Gubbay and Silvia Carrieri (Bruegel);

- Sections 2.2 and 2.3 build on two technical reports describing the state of the art at the sectoral, regional and aggregate level and at the microeconomic level (deliverables 2.1 and 2.2 of the MAPCOMPETE project) by Davide Castellani, Silvia Cerisola, Giulia Felice, Emanuele Forlani and Veronica Lupi (LdA). The collection of information on availability and computability of indicators used in these sections also benefited from the contribution of Chiara Angeloni (LdA), Marco Antonielli (Bruegel), Zsuzsa Holler (CERS-HAS) and Gianluca Santoni (PSE). Many officials and researchers within national statistical institutes, national central banks and other institutions were also a key part of this task. A specific acknowledgement of these contributions is provided in Box 2.3;

- Chapter 3 is largely based on a report on the 'General considerations of the matchability of datasets within and across countries and regions' (deliverable 4.2 of the MAPCOMPETE project) by Andreas Koch and Katja Neugebauer (London School of Economics, formerly IAW). Carlo Altomonte and Elena Zaurino (Bruegel) contributed to chapter 3.3;

- Chapter 4 was written by Gábor Békés with Zsuzsa Holler (CERS-HAS);

- Giorgio Barba Navaretti (LdA) and Carlo Altomonte (Bruegel) provided valuable insights into section 5.

Finally, the authors would like to thank Lauro Panella (European Commission) and Jan Hagemeier (Central Bank of Poland) for acting as discussants on a previous version of this report and for providing very fruitful insights.

Davide Castellani and Andreas Koch

Perugia and Tübingen, January 2015

About the authors

Davide Castellani is professor of applied economics at the University of Perugia and a research fellow at CIRCLE (Sweden) and LdA (Italy). His research focuses on the determinants of the firms' internationalisation choices and their impact on international technology transfer and economic performance. His work has been published in international journals including the *Journal of Economic Behaviour and Organisation*, *Journal of International Business Studies*, *Journal of International Economics*, *Oxford Economic Papers*, *Research Policy*, *The Review of World Economics* and *The World Economy*. He has been involved in a number of research projects funded by the European Commission under various EU Framework Programmes.

Andreas Koch is a research fellow at the Institute for Applied Economic Research (IAW). His work focuses on the continuous structural changes at the interfaces of regions, firms, working environments and technological development. He has been engaged in exploiting and analysing new micro-level data (firms and individuals) in Germany and on the European level in applied projects and from a methodological point of view. He is an assistant lecturer at the Department of Geosciences at the University of Tübingen and of *New Economic Papers – Economic Geography* (NEP-GEO).

Foreword

Reality for policymakers, and the decisions they make, are to a great extent products of the statistics available to them. It is not just the coverage and harmonisation of data that are important. It is also the type of data that matters. As Europe attempts to put itself back on the path to growth, the need for clear data on competitiveness, for an accurate statistical underpinning beyond some broad macroeconomic broad indicators and for new insights from new ways of looking at economic data has arguably never been greater.

This volume – a product of the MAPCOMPETE EU-funded project, in which Bruegel participates – provides an important service for researchers and policymakers by examining the availability and usefulness in Europe of indicators of competitiveness. At the country, sector and regional levels, the authors find that Europe is served rather well. In addition, micro-data, which could be used to tell us about competitiveness at the firm level, is generated by EU member states. A previous project initiated by Bruegel, EFIGE, examined the characteristics of firms that succeed globally and showed why firm-level information is needed (www.efige.org/). But it can be hard in practice for researchers to access the micro-data and to use it to create bottom-up indicators of competitiveness.

This is an area in which policymakers should intervene. The matchability and accessibility of data should be improved. The authors of this Blueprint set out a number of practical ways in which this can be done to some extent in the short term, but a longer-term approach is also required to build an effective European statistics framework that will support broad growth and competitiveness objectives. This volume shows how it can be done.

Guntram B. Wolff, Director of Bruegel

Brussels, February 2015

Executive summary

There is widespread agreement that improving competitiveness throughout Europe is at the heart of the structural resolution of past and future crises. However, agreement is likely to stop there. Although many studies and reports from different international and national institutions measure the competitiveness of firms, regions or nations, there is no common single definition of competitiveness, and no consensus on how to properly and consistently measure competitiveness across countries and/or over time. Moreover, even though a number of aggregate indicators (eg real effective exchange rate, unit labour costs, export share and prices) are available and broadly used, many suffer from measurement errors, not necessarily delivering the same ranking across countries or across time. Finally, there is no single and/or harmonised dataset allowing the different facets of competitiveness to be captured in an internationally comparative perspective.

This is the case for existing indicators of competitiveness, prevalently defined at the macro (national or industry) level. However, most of the current policy debate about competitiveness also neglects a large body of economic literature suggesting that the performance of countries is greatly affected by the performance of firms. Research has increasingly shown that the statistics typically used for policy design are frequently insufficient and misleading. In particular, standard statistics, essentially based on average figures, are unable to represent adequately the ability of a country or a sector to compete in the global market. Also, it has been shown that international operations of firms are not adequately represented by standard trade statistics because international investment and the fragmentation of production are increasingly important features underlying competitiveness[1].

Understanding firms' competitiveness is thus central to the policy discussion: the relevance of firms' heterogeneity in terms of their size, productivity, innovation

1. A first good reference at the international policy level that goes in this direction is the Competitiveness Research Network , set up by the System of European Central Banks.

activities and internationalisation strategies means that policy needs to be designed around diverse firm characteristics and strategic responses rather than around an invariant *representative firm*. To this end, the usual measures of competitiveness based on aggregate data need to be harmonised and made comparable for different countries and years, and also to be complemented with additional indicators built-up from micro-data (which we label as 'bottom-up' indicators). This need for more micro-based indicators of competitiveness is however frustrated by the lack of clarity on what could be the best sources of information and on the access conditions. Although the universe of existing data from the different official and non-official data providers is very rich, and technical progress has extended significantly the potential uses of this data, there are still major restrictions in terms of the extent to which a researcher is actually able to access the data and to compute the indicators of competitiveness he/she needs.

MAPCOMPETE, a support action for the European Commission carried out by a consortium of European research institutes (see www.mapcompete.eu), has been designed to address the challenges discussed above, with special reference to providing an assessment of data opportunities and requirements for the comparative analysis of competitiveness in European countries at the macro and the micro level. This report picks up some of the main issues of the MAPCOMPETE project and provides an inventory and an assessment of the data related to the measurement of competitiveness in Europe. By doing so, this Report, and the associated meta-database available at www.mapcompete.eu – which provides detailed information on data accessibility and computability of more than 150 indicators – can be a key handbook for a researcher interested in measuring competiveness, or for policymakers interested in the feasibility and in the quality of alternative competitiveness measures. This Report also identifies the opportunities emerging from recent progress made in scientific research and facilitated by different data providers who increasingly make their data available to research. Finally, this inventory allows us to identify the main issues that need to be addressed by policymakers in order to improve data accessibility for the economic analysis of competitiveness in Europe.

State of affairs

An inventory of the indicators that can be built to measure competitiveness in Europe requires several steps of analysis. The first step is an evaluation of the existence of the necessary raw data and the computability of the competitiveness indicators, which frequently involves combining different data sources or families. Second, an assessment is done of the accessibility of data in individual countries. Third, an

appraisal is carried out of the extent to which data for different countries can be matched and/or 'bottom-up' indicators of competitiveness can be compared for different countries.

Our overall conclusions are:

1. Competitiveness indicators are available at the country, sector and regional level (eg unit labour costs, price indices, REER, trade balance data, aggregate productivity) and are generally computable for relatively long time series in most EU28 countries. These macro indicators are also generally easily accessible via Eurostat, national statistical institutes, national central banks or other data providers, and can usually be compared across EU countries.

2. Availability of micro-data, and therefore *computability of bottom-up indicators, is also rather good for many countries*. This implies that, within countries, it is possible, in principle, to match different databases.

3. There is, however, *a major problem in accessing both specific databases and even more matched data in many EU countries*. The report highlights many legal, non-legal (such as unclear procedures, restrictions on the nationality of data users) and technical barriers severely limiting the access to data and consequently the ability of researchers to construct bottom-up indicators that are not generally constructed by statistical agencies.

4. Furthermore, if we consider building up cross-country statistics from micro-level data, which should be the final aim of any meaningful assessment of European competitiveness, the quality of European statistics is at the moment rather poor, due to limited harmonisation, matchability and accessibility of data. The possibility to build pan-European micro-level databases to assess the state and the dynamics of competitiveness in the whole region is limited, notwithstanding the considerable efforts of the European Statistical System (ESS) to coordinate national statistical institutes (NSI) to harmonise the methodology, the scope and the legal framework for data collection and processing.

Policy: what should be done?

This report shows that the information on measures of competitiveness currently available to researchers is insufficient. Aggregate data, which is easily accessible and widely available, does not allow researchers to provide the answers that policymakers

need. Micro-data for individual countries is mostly inaccessible to external researchers, and the situation is even worse when one tries to compare figures based on micro-data which is comparable across countries: only a few cross-country harmonised firm-level surveys are available, mostly for only one or a few years. There are almost no examples of matched data across countries, and internationally comparable figures can be gathered only from a few micro-distributed data exercises. This is very different from, for example, the United States, where micro-level data that is matchable and comparable for different states has existed since at least the mid-2000s. This implies that we lack the proper information to assess the status of competitiveness at the European level, compared to the situation in the United States.

The first-best solution to overcome these bottlenecks would be to change the national and EU-level rules of data content, availability, matching and access. Some important steps have been (or are in the process of being) taken in this direction: the efforts undertaken by the ESS towards greater harmonisation of data and the construction of pan-European datasets; the reduction of the burden on enterprises in collecting and providing internal data; the provision of a common ESS infrastructure framework for the production and compilation of business statistics with an appropriate legal background and new administrative mechanisms allowing for the sharing of information, services and costs among ESS partners; the definition of consistent data requirements and of a common data quality framework, which will enable the linking and matching of statistics obtained through the regular collection of global business statistics. However, the timeline for completing all these measures is far too long. Therefore, such long-term actions aiming at changing regulations need to be complemented by more short-term measures. As these are viable, but still only second-best solutions, we will call them 'workarounds'.

The first workaround is to exploit the availability of improved methods and techniques, such as matching after separate processing (eg the Distributed Micro-Data (DMD) approach) or the imputation of missing or unavailable data. Projects exploiting the DMD approach, such as European Central Bank's CompNet or Eurostat's ESSLait, are providing important insights into new aspects of competitiveness by producing micro-aggregated statistics going beyond the first statistical moment of the distribution of firms' competitiveness indicators. However, if not properly supported by policy, these initiatives may prove to be one-shot exercises, while instead they need to be refined, constantly updated and carried out in a timely way in order to provide more up-to-date figures for policy decisions.

The second workaround can be to improve techniques of matching and accessing micro-level data, either by improving architectures for matching data (eg by involving 'matching institutions', among which a natural candidate could be a Directorate General of the European Commission) or for data access by researchers (eg by improving techniques of data anonymisation). In most countries, access to micro-data would be practically and legally feasible for external researchers, but it is easier for the data providers to restrict access. We claim that restricting access is cost- and responsibility-efficient for the data providers, but very inefficient for researchers and policymakers in general. But if these are the real issues behind the restrictions to data access, there are available solutions. Data access does not need to be free for all researchers. Researchers can contribute from their research funds to cover the cost of setting-up the infrastructure for data access and anonymisation. Nevertheless, EU support could play a crucial role, especially for smaller member states, which might not be able to afford to bear the fixed costs of setting up new infrastructures and developing the necessary capabilities, such as language skills and economics knowledge, which are crucial in order to foster cooperation and build a truly European infrastructure for accessing micro-level data.

The third workaround is to support multi-scope cross-country surveys, which allow researchers to gather information on a wide range of firms' activities and performance indicators, in order to enable them to assess their contribution to overall competitiveness. The Community Innovation Surveys and the International Sourcing Surveys, coordinated by Eurostat, are interesting examples in this direction, although they both focus on specific aspects of competitiveness. The EFIGE survey, administered by a consortium of research institutions and supported by the EU FP7 programme, is another case in point, which instead takes into consideration a greater number of aspects of competitiveness. However, to make this solution effective, there is a need for greater harmonisation and coordination, in order to concentrate resources on fewer surveys. These should cover many aspects of competitiveness and they should be based on a greater number of firms followed constantly over time, so that the dynamics of firm competitiveness can also be accurately assessed.

In summary, developing national capabilities to better service micro-level data is the most cost-effective and sustainable way to generate new indicators of competitiveness. Once these permanent structures are in place, the access by individual researchers to micro-level data or projects based on the distributed micro-data approach should be more feasible. At the same time, given that setting up these capabilities for all EU28 countries will take time and, in some cases, legislation, we also recommend unification and extension of corporate surveys piloted under various

projects funded by the European Commission's Seventh Framework and Horizon 2020 programmes. Carefully crafted annual surveys will allow new measures of competitiveness to be constructed and, at the same time, provide a greater understanding of its dynamics even in the short term.

1 Introduction

There is widespread agreement that improving competitiveness throughout Europe is at the heart of the structural resolution of past and future crises. Firms increasingly base their choices on parameters related to competitiveness, and the European Commission continuously monitors external imbalances using quantitative measures of aggregate competitiveness. For these reasons, a number of international institutions, such as the European Commission, the European Central Bank, the World Bank and the World Economic Forum, are committed to producing regular comparative reports on competitiveness at the national level (see European Central Bank, 2012; European Commission, 2013; World Bank, 2013; World Economic Forum, 2013), at the regional level (see Annoni and Dijkstra, 2013) or based on aggregated firm-level data (see CompNet Task Force, 2014).

Despite the availability of numerous publications and reports on the issue, there are some serious challenges in terms of the conceptualisation and measurement of competitiveness. First, although many studies and reports measure the competitiveness of firms, regions, or nations, there is no commonly shared single definition of competitiveness. Second, there is no consensus on how to properly and consistently measure competitiveness for different countries and/or over time. Even though a number of aggregate indicators (eg real effective exchange rates, unit labour costs, export share and prices) are available and broadly used, they can suffer from measurement errors, and do not necessarily deliver the same ranking for different countries or across time. Third, there is no single and/or harmonised dataset that enables the different facets of competitiveness to be captured in an internationally comparative perspective.

Although there has been an explosion of available information in terms of digitalised datasets, the ability to effectively exploit these data repositories has been hampered by two main factors. First, there is a clear tendency towards the use of a restricted set of economic indicators, mostly designed when the richness and detail of available data was much less than today. In particular, most of the policy debate about competitiveness neglects a large body of economic literature suggesting that the

performance of countries is greatly affected by the performance of their firms.

Second, research has increasingly shown that the statistics commonly used for policy design are frequently insufficient and misleading. A proper assessment of the statistics in use often requires the construction of alternative indicators. This is particularly relevant for competitiveness indicators. In particular, standard statistics, essentially based on averages, are unable to adequately represent the ability of a country or of a sector to compete in the global market. Also, it has been shown that the international operations of firms are not appropriately represented by standard trade statistics because international investment and fragmentation of production are increasingly important features underlying competitiveness.

Understanding firms' competitiveness is thus central to the policy discussion: the relevance of firms' heterogeneity in terms of their size, productivity, innovation and internationalisation strategies means that policy needs to be designed around diverse characteristics and strategic responses, rather than an invariant *representative firm*. Usual measures of competitiveness based on aggregate data need to be complemented with additional indicators built-up from micro-data (which we label as 'bottom-up' indicators). However, this need for more micro-based indicators of competitiveness is frustrated by the lack of clarity on the best sources of information. Although the universe of existing data from the different official and non-official data providers is very rich and technical progress has significantly extended the uses that can be made of this data, there are still major restrictions in terms of the extent to which a researcher is actually able to access the data and to compute the indicators of competitiveness he/she needs.

MAPCOMPETE, a support action for the European Commission performed by a consortium of European research institutes (see www.mapcompete.eu), has been designed to address these challenges, with special reference to providing an assessment of data opportunities and requirements for the comparative analysis of competitiveness in European countries. Based on an inventory of the existence, availability and accessibility of data that measures the different dimensions and aspects of competitiveness, both at the macro and micro levels, the project also seeks to identify the main potentials and drawbacks of the current data landscape. It seeks to outline future options and pathways for generating and providing data on competitiveness.

This Blueprint picks up some of the main issues from the MAPCOMPETE project. Its principal aim is to provide an inventory of the data related to the measurement of

competitiveness in Europe, mainly on the basis of micro-level data. By doing so, the Blueprint, and the associated meta-database[2], which provides detailed information on data accessibility and computability for more than 150 indicators, can be a starting point for a researcher interested in measuring competiveness, or for policymakers interested in the feasibility and in the quality of alternative measures. It also aims to identify the opportunities emerging from recent progress made by scientific research and facilitated by different data providers who increasingly make their data available to research. Finally, this inventory allows us to identify the main issues that need to be addressed by policymakers in order to improve data accessibility for the economic analysis of competitiveness in Europe.

The report is organised as follows. We first introduce, in chapter 2 (section 2.1), some general considerations on the measurement of competitiveness by developing and presenting a system of indicators organising the field into different areas. Chapter 2 also contains an extensive inventory of the available data both at the macro level (2.2) and at the micro level (2.3) mainly produced by public data providers, such as EU national statistical institutes and national central banks. This inventory highlights whether information on competitiveness is available in EU countries, whether, especially for micro-level data, it can be combined to compute the relevant indicators of competitiveness and to what extent an external researcher (ie not affiliated with the data provider) can access the data. Chapter 3, then focuses on the availability of micro-data comparable across countries. It briefly reviews issues related to the matching of micro-level data within and between countries, illustrates the Eurostat experience in facing an increasing demand for micro-data comparable across countries (3.2), and contains both an inventory and some illustrative examples of datasets that contain information on previously unconnected areas or that gather information from different countries (section 3.3). Chapter 4 presents some final considerations on how to improve access to micro-data related to the measurement of competitiveness in the future. Chapter 5 offers some policy recommendations[3].

2. The MAPCOMPETE meta-database, which allows searching for meta information on availability and accessibility of data needed to build indicators of competitiveness for the 28 EU countries, is available at http://www.mapcompete.eu/.
3. The Annex (section 6 of this Blueprint) provides the more technical details on the indicators of competitiveness and detailed tables.

2 Mapping competitiveness indicators in the EU countries

2.1 Indicators of competitiveness

With the purpose of improving the toolbox of competitiveness indicators, the *Mapping European Competitiveness* (MAPCOMPETE) project provides an assessment of data opportunities and requirements for the comparative analysis of competitiveness in European countries. Existing competitiveness indicators have been surveyed in order to provide a critical assessment and a selection of indicators to be used in the data-mapping exercise. This section introduces the methodology, the assessment and the results of this survey and serves as a manual to interpret the findings of sections 2.2 and 2.3.

2.1.1 Methodology

Competitiveness indicators cover almost all aspects of market performance. Price and quality, the ability to innovate, the structure of the labour market, the level of international integration of markets, and qualitative conditions of countries' business environments are frequently evoked in discussions of competitiveness. In fact, there is no shared definition of competitiveness or consensus on how to measure it. We decided, in line with Altomonte *et al* (2011), to consider competitiveness as related to the ability of firms in a given country – not the country itself – to mobilise and efficiently employ (also outside the country's borders) the productive resources required to offer those goods and services for which other goods and services can be obtained (domestically or internationally) at favourable rates of substitution (or terms of trade).

This definition was inspired by a large body of economic literature suggesting that the performance of countries is greatly affected by the performance of firms. Understanding firm competitiveness is thus central to the policy discussion: the relevance of the heterogeneity of firms in terms of their size, productivity, internationalisation strategies and so forth means that policy needs to be designed around diverse characteristics and strategic responses rather than around an invariant *representative firm*.

In light of the definition, we conducted a systematic investigation of existing competitiveness indicators in the economic literature, policy papers and other sources. Focusing on the performance of firms affected the search in two ways.

First, we focused in particular on indicators that aggregate information from firm-level data, which we label as *bottom-up* indicators. These indicators can be useful complements to the macro-indicators, constructed with aggregated data. Indeed, one of the major contributions of MAPCOMPETE is to highlight where the existing 'standard' competitiveness indicator toolbox can be enriched with harmonised and complementary bottom-up indicators.

Second, recognising that firms compete not only on price, we gave special attention to non-price competitiveness indicators. This induced a view of competitiveness that has *sustainable growth* as the underlying concept.

Despite taking this direction, the lack of a common understanding of competitiveness in the policy debate motivated us to further specify our analysis. In our conceptualisation, *indicators* of competitiveness are distinguished from *drivers* of competitiveness. In theory, the difference is striking: indicators tell us if firms, countries, sectors or regions perform well compared to each other; drivers tell us what determines this performance. However, in practice this difference is less obvious: indicators and drivers are sometimes used in the same context to denote different aspects of competitiveness; in other cases, indicators are not used as outcomes but rather as determinants.

In this chapter, we deal primarily with indicators rather than with drivers of competitiveness. In a commentary published in the *Financial Times*[4], Risto Penttila, chief executive of the Finnish Chamber of Commerce, made a very compelling argument, which supports our choice:

> *"Either the World Economic Forum is wrong or Europe is in deep trouble. The latest competitiveness rankings from the Swiss think-tank list Finland as the most competitive country in the EU. At first, the country's business leaders thought someone was pulling their leg. But the news was real. If Finland is the best the EU can offer, we should all be very concerned. (...) The report's authors define competitiveness as 'the set of institutions, policies, and factors that determine the level of productivity of a country'. But Finland's experience shows that having well-*

4. 'If Finland is the best Europe can do, we should be worried', *Financial Times*, 24 June 2014.

functioning institutions is not a cure-all. The country ticks all the boxes: well-protected property rights, good schools, reliable infrastructure, predictable macroeconomic policies. It is one of the biggest spenders on research and development in the world. Yet the productivity of Finnish industries has plummeted since 2009."

2.1.2 Classification logic and selection of indicators

Organising competitiveness indicators around several *concepts* helped us to assess them against their primary objective, comparing similar indicators, and finding complementarities. We use the following six concepts:

1. Productivity
2. Market share
3. Prices and costs
4. Innovation and technology
5. Firm dynamics
6. Global value chains

These six concepts describe complementary aspects of competitiveness. We do not aim to prioritise, but rather to organise them. In order to take into account the firm-level dimension for each concept, we introduce and report also indicators that can only be built up from firm-level or, more generally, micro-level data. For practical reasons, we label these indicators as bottom-up. All other indicators are based on aggregated data and can be defined at country, sector or regional level.

The indicators concepts were ultimately helpful for choosing a subset of the indicators that should be used in the subsequent part of the project, such as the data mapping, which will be illustrated in sections 2.2 and 2.3. Within each concept we propose a selection based on:

- Indicators' adequacy for the competitiveness concept;
- Reliability and efficiency of the statistical techniques; and
- Complementarity of the indicators.

Reducing the number of usable indicators entails a loss of information. However, many indicators within each category are highly correlated with each other, or can be easily

summarised by other indicators. Moreover, selecting the indicators also helped to highlight the areas in which available indicators still offer an unsatisfactory picture of the competitiveness concept.

More than 140 indicators were collected within our survey. The table below reports the number of surveyed indicators in each category.

Table 2.1: Concepts and indicators

Concept	Number of indicators
Productivity	18
Trade competitiveness	21
Prices and costs	15
Innovation and technology	43
Firm dynamics	8
Global value chains	32
Others	7

More detailed information on the concepts and indicators of competitiveness, including a technical assessment of the different aspects highlighted in this chapter, such as adequacy, reliability of the statistical techniques, complementarity, macro vs. micro dimensions, within each category, is provided in section 6.1.

2.2 Mapping the macro-level indicators

For each indicator that was computable with aggregate data, we have identified the relevant level of disaggregation (national, sectoral and regional). Out of a total of 43 indicators, 41 can be computed at the national level, 32 at the sectoral level and 16 at the regional level. Therefore, we map data availability for 89 indicators. For expositional purposes, we arrange the 89 indicators into groups of relatively homogenous types of measures.

Data is presented in two-way tables in which each row represents one indicator, and each column is one country. In each cell of these tables we report a number from 0 to 2, which summarises the extent to which data for a given indicator is available in each country. In Box 2.1, we discuss the criteria that we use to assign these scores. These tables can be read (and commented) both along the rows and down the columns. In

other words, one can highlight the availability of data for each indicator across countries, or of each country across all indicators. We believe the former is more informative for the aim of this report, which is to provide an overview of the availability of comparable competitiveness indicators in different countries.

It is worth mentioning that some indicators can be computed from more than one source and the different sources could imply different coverage in terms of countries, time spans and/or sectors and regions. The results presented in this chapter are based on the authors' *a priori* choices of the most appropriate source for each indicator. In particular, we assigned a higher priority to data sources which were more exhaustive in terms of the information they provide about countries (ie we assigned cross-country comparability a higher priority). If two (or more) sources provide the same country coverage, we preferred the one with the longer time series.

Detailed tables and comments are provided in the Annex (Section 6.4). Here, we summarise the main conclusions from this task.

For the 89 indicators of competitiveness at country, sectoral and regional levels for the EU28 countries, our analysis shows that the degree of computability for the macro-indicators is quite good. However, there are some exceptions. It is possible to group the exceptions in three main categories: i) by country, ie, there are countries for which data availability is particularly scarce for the majority of indicators; ii) by indicator, ie there are indicators on which information is particularly scarce for the majority of countries and levels of aggregation; iii) by level, ie there are levels of aggregation on which information is particularly scarce for several indicators for the majority of countries.

In terms of exceptions by country, most EU28 countries show a good level of computability for a relevant number of indicators at different aggregation levels. Information is scarcer for Croatia and Greece than for other countries.

In the second category of exceptions, information on indicators of firm dynamics (such as entry and exit rates) is quite heterogeneous across countries and levels of aggregation, but in general only half of the countries show the highest level of computability. The indicators belonging to intangible assets and financial activity are computable for only a few countries and/or quite short time intervals. The information on R&D expenditure and output is in general quite good with the exception of license and patent revenues from abroad as percent of GDP and EU Summary Innovation Index (SII), which are computed and comparable across countries for all EU countries since

BOX 2.1: THE DEGREE OF COMPUTABILITY OF THE COMPETITIVENESS INDICATORS AT NATIONAL, SECTORAL AND REGIONAL LEVELS

The indicators of competitiveness at national, sectoral and regional levels are usually computed from easily accessible data. Therefore, the key dimension from which one can evaluate the extent to which an indicator can be computed (degree of computability) is the time span for which the indicator is available. We defined three levels of the degree of computability based on the length of the time series. In the general case, ie when data is available on an annual basis, we assign:

- The value 2 if data is available for a given country since 2000 (or earlier).
- The value 1 if data is available only for a about a decade, but not for the more recent years. We operationalise this threshold as between 2000 and 2008.
- The value 0 if data is available only for a very limited time span (eg from 2008 onwards), or not available at all.

In some special cases, for instance when availability is subject to discontinuity over time, we assigned the degree of computability according to the following scale:

- The value 2 is assigned when the indicator X is available every two years, for a time span of at least 10 years.
- The value 1 is assigned when the indicator X is available every two years, but for a time span of less than 10 years.
- The value of zero is assigned when the indicator X is available for a time span of five years or less and not continuously.

In the case of Community Innovation Survey data (Indicators from I_027 to I_036):

- The value 2 is assigned when four waves are available.
- The value 1 is assigned if three waves are available.
- The value 0 is assigned if fewer than three waves are available.

We refer to annual data unless mentioned otherwise.

only 2004 (2006 for Spain and Greece) for the former, and since 2008 for the latter. The good availability of these innovation indicators can mostly be attributed to the data from the Community Innovation Survey (CIS), which is available for a longer time span in many European countries.

As for the third category of exceptions, by level of aggregation, it should be mentioned that in general information for the indicators at the sectoral and regional level is both

scarcer and less homogeneous than at the aggregate level across countries and indicators. In particular, the indicators' computability is high at the aggregate level, but quite limited at both sectoral and regional levels for those indicators belonging to labour productivity and Total Factor Productivity, innovation activity, SMEs and R&D expenditure and output. The indicators' computability is high at the aggregate level but quite limited at the sectoral level for those indicators belonging to trade competitiveness (Group 4), while the indicators' computability is high at the aggregate level, but quite limited at the at the regional level for those indicators belonging to innovation activity, all firms (Group 8).

2.3 Mapping the bottom-up indicators

2.3.1 Methodological issues

As we have illustrated, indicators of competitiveness can be calculated at national, sectoral and regional level by aggregating firm-level data, ie. by applying a *bottom-up* approach. Firm-level data allows researchers and policymakers to define a multitude of indicators that can be used to describe phenomena such as differences in regional productivity, the entry and exit rate in a specific market or international competitiveness (eg the intensive and extensive margin of trade).

This section provides an overview of the availability and accessibility of data needed to compute a series of *bottom-up* indicators of competitiveness for the EU28 countries.

We discuss both the degree of *computability* of different indices and the degree of *accessibility* of firm-level data which is necessary to compute the related indicators. While the computability concerns the quality and time coverage of indicators, accessibility concerns limitations on access to firm-level data[5]. This information is extracted from the meta-DB (section 6.3) and will be fully searchable, jointly with the other meta-data, via a webtool at www.mapcompete.eu.

It is worth mentioning that this section focuses on indicators that are well-established in the literature on competitiveness, as reviewed in section 2.1, and that can be computed from micro-level databases collected mainly by national central banks (NCB) and national statistical institutes (NSI). Surveys, projects or commercial databases can also offer internationally comparable indicators/data on competitive-

5. As mentioned above, at this stage we mainly rely on official firm-level data collected by central banks and national statistical offices.

ness. Some of these sources provide information on a variety of firm characteristics associated with competitiveness (eg World Bank's BEEPS database or the EFIGE survey data), other focus on specific aspects such as internationalisation and productivity (CompNet), managerial practices (LSE), innovation and finance (FINNOV), economic and financial performance (Amadeus, CompNet, MicroDyn), firm and employment dynamics (MicroDyn and OECD DynEmp) corporate linkages (eWho Owns Whom), cross-border investment projects (FDI Markets). These sources will be discussed in chapter 3.

Some of the bottom-up indicators are considered along several dimensions such as type of firm (all firms, exporter, importers, foreign-owned firms, domestic multinationals, etc), level at which data can be aggregated up (country, sector and region), and underlying distribution (average, median, variance, etc)[6]. For each index, three levels of aggregation are considered: country, sector and region. The mapping of micro-level databases includes information on firms' industrial sector (usually NACE Rev. 1.1 or Rev. 2) and geographical location (usually NUTS2 region). The mapping allows users to know if a given competitiveness index is computable by aggregating up data at sector, and/or regional level for each country. Moreover, the bottom-up approach allows scholars to determine competitiveness measures which are not confined to averages. The existence of databases with population (or surveys) of firms makes it possible to define additional moments of the competitiveness' indices. In this perspective, when mapping micro-level databases, we make sure that the median, standard deviation and various percentiles of the distribution can also be computed.

The bottom-up competitiveness indicators that we have considered can be grouped in the following conceptual areas:

- Productivity, which for expositional purposes is presented in two distinct sets of tables: Labour productivity (including Unit Labour Cost) and Total Factor Productivity (TFP);
- Firm dynamics;
- International activities;
- R&D and other activities.

6. For example, we consider the possibility to define the average, the median, and the standard deviation of TFP for exporting firms.

Labour productivity

This area includes information which is used to calculate the labour productivity index as value added per worker. The index is defined for different type of firms such as domestic firms, exporters, importers, multinationals, affiliates of foreign multinational firms, foreign and domestic-owned exporters. Moreover, in this category we also consider the firms' unit labour cost. Regional and sectoral dimensions are taken into account, as well as the possibility to define different points of index distribution. Summary results are reported in Table 6.14.

Total Factor Productivity (TFP)

Similarly to labour productivity, for each country, we collected information on the availability of the data that is necessary to calculate firm-level TFP. In addition, the decomposition of TFP proposed by Olley and Pakes (1996) and the decomposition of TFP growth proposed by Foster *et al* (2001) are also considered. Regional and sectoral dimensions are included, as well as the possibility to define different statistical moments of index distribution. A full list of indices can be found in the annex. Summary results are reported in Table 6.15.

Firm dynamics

Another source of competitiveness is the rate of turnover of firms (ie the entry and exit rate), and the average growth rate of firms. Therefore, the data mapping includes information on firms entering and exiting the market, survival rates after different time periods, average firm size (relative to age), dispersion of firms by size and growth rate. Summary results are reported in Table 6.16.

International activities

In this area, we mapped the availability of information on trade activity at the firm-level. This group includes data on the number of export destinations, number of exporting firms (total and by destination), number of products exported (total and by destination). In addition, different measures of the intensive and extensive margins of trade are included, as well as firm-level estimates of quality (unit value of exports). Information on the number of foreign-owned firms as a share of all firms, and the share of domestic multinational firms (MNFs) to total firms (by country, sector and region) are also collected. Summary results are reported in Table 6.17.

R&D and other activities

This area includes some additional information on firm-level competitiveness, such as firm-level expenditure of R&D and the level of tangible assets. Summary results are reported in Table 6.18. For expositional purposes, we group R&D and tangible assets with some of the indicators of international activities (firm-level estimates of quality, share of foreign owned firms to total firms, share of domestic multinational firms) into a category that we label 'R&D and other activities'.

Table 2.2: Computability criteria

Thresholds	Degree of computability	Colour code
Good time span and good matchability. Observations at least since the year 2000.	2	Green
Observations only after the year 2000; matching different datasets is basically possible, but associated with some problems.	1	Yellow
No matchability and/or only few years of data (from 2006).	0	Red
With the available information it is not possible to assess the time span and/or the matchability.	9	Grey

Table 2.3: Accessibility criteria

Thresholds	Accessibility code	Colour code
Public or available on site.	2	Green
With restriction, but possible under certain conditions.	1	Yellow
No way (eg 'Confidential' or dependent on nationality status).	0	Red
Conditions not reported by data provider.	9	Grey

BOX 2.2: CRITERIA FOR COMPUTABILITY AND ACCESSIBILITY OF INDICATORS

Computability: The degree of computability of an indicator depends on the span of time coverage and the quality of data. In particular, we consider whether the different sources of firm-level data necessary to calculate the related indicator can actually be matched. For example, data can be easily matched if firms have a unique identifier in different databases. As reported in Table 2.2, we define four levels of computability, and for each level we assign a numerical and a colour code.

Green indicates the highest degree of computability for an indicator, yellow suggests medium level of computability, and red low (or no) computability. Grey is used for an indicator for which it is not possible to assign a degree of computability because of the lack of information.

Accessibility: The degree of accessibility of an indicator is defined by the conditions that regulate the access to firm-level data. Similarly to computability, we report four levels of accessibility, and for each level it is assigned a numerical code and a colour. It is important to underline that the degree of accessibility describes the restriction in the access to firm-level data, which are necessary to calculate a given indicator. For example, in the case of labour productivity index, the degree of accessibility indicates the conditions of access to firm-level data on employment and value added.

Green indicates the highest degree of accessibility for an indicator, yellow suggests limited accessibility, and red restricted accessibility. Grey is used for an indicator for which it is not possible to assign a degree of accessibility because of the lack of information.

BOX 2.3: THE INFORMATION GATHERING PROCESS

Gathering information needed to compile the MAPCOMPETE MetaDB for bottom-up indicators proved to be challenging. The first problem was to find a suitable contact within each country. Our first-best option was to contact someone within the national statistical institute (NSI) in each EU28 country and gather information from them. A few months into the project, this proved highly complicated, so we decided that we would gather information through contacts within national banks, exploiting a collaboration with the ECB Competitiveness Network - CompNet

(http://www.ecb.europa.eu/home/html/researcher_compnet.en.html). CompNet is a project to build bottom-up indicators exploiting data accessible by national banks. As a matter of fact, some of the indicators that MAPCOMPETE considers relevant bottom-up indicators have been actually computed within CompNet.

With the help of Filippo di Mauro and Paloma Lopez-Garcia at the ECB we were able to find contact persons in each of the 28 EU member states. In some cases, those contact persons were able to help us fill the MAPCOMPETE MetaDB and in other cases they referred us to people within the NSI. In cases in which we could not find a personal contact, we compiled the information based on publicly available information. After a first round of data collection, we drafted a first version of this report and sent it to contact persons within NSIs and NCBs for validation. In cases in which we had no direct contact, we sent the draft to a generic contact email within the NSI. This prompted replies from NSIs and NCBs, which allowed us to further integrate the information collected. At the end of this process, we were able to report on 25 out of the 28 EU countries: Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and the United Kingdom.

In Austria, Denmark and Spain the information could not be verified by the NSI.

From Cyprus, Greece and Luxembourg we were not able to gather enough information from publicly available sources and the contact persons that we had identified were not able to help us, so these countries are not included.

We would like to thank all the people that, within each country, helped us gather the information needed to compile the MAPCOMPETE MetaDB.

Country	Contact persons	Institution
Belgium	Catherine Fuss	National Central Bank
Bulgaria	Svetoslava Filipovich	National Statistical Institute
Croatia	Blaženka Vukeli	National Statistical Institute
Czech Republic	Kamil Galuscak	National Central Bank
	Pavel Hájek	National Statistical Institute
	Zuzana Cabicarová	National Statistical Institute

Estonia	Aavo Heinlo	National Statistical Institute
	Jaanika Merikyll	National Central Bank
Finland	Satu Nurmi	National Statistical Institute
France	Philippe Brion	National Statistical Institute
Germany	Sven Blank	National Central Bank
Hungary	Peter Harasztosi	National Central Bank
Ireland	Keith McSweeney	National Statistical Institute
	Iulia Siedschlag	ESRI
Italy	Stefania Rossetti	National Statistical Institute
Latvia	Lilita Laganovska	National Statistical Institute
Latvia	Inga Malasenko	National Statistical Institute
	Sandra Vitola	National Statistical Institute
Lithuania	Vera Bezviuk	National Statistical Institute
Malta	Anne Maria Caruana	National Statistical Institute
Netherlands	Harry Goossens	National Statistical Institute
Poland	Jan Hagemejer	National Central Bank
	Karolina Szlesinger	National Statistical Institute
Portugal	Maria Conceição Veiga	National Statistical Institute
	Ana Cristina Soares	National Central Bank
Romania	Virginia Balea	National Statistical Institute
Slovakia	Tibor Lalinsky	National Central Bank
Slovenia	Urška Cede	National Central Bank
	Barbara Dremelj Ribi	National Statistical Institute
Spain	Juan Carlos Farinas	Universidad Complutense Madrid
Sweden	Eva Hagsten	National Statistical Institute
United Kingdom	Daniele Bega	HM Revenues and Customs

2.3.2 Availability and accessibility of micro-data in EU countries

Austria[7]

The data needed to compute bottom-up indicators derives from two main data-sources. The first source is a firm sample with detailed balance sheet data collected by the statistics department of the Österreichische Nationalbank (OeNB). In recent years, the sample has been approximately 8,000 firms per year, representing 35 percent of total employment. The sample is clearly biased towards larger enterprises. The database starts in the early 2000s. The rather low number of firms is because only larger corporations have to publish their balance sheets. The OeNB collects additional balance sheet data from firms receiving larger loans from banks. This is the reason why the OeNB firm sample, small as it is, is larger than the one collected by Bureau Van Dijk which covers fewer than 3,000 firms per year for Austria ('Sabina database').

The second data source is OeNB-Statistics Austria micro-data on exports, imports and FDI.

Labour productivity – Labour productivity is computable only for the non-representative sample of firms for which balance Sheet data is available at OeNB and only from the early 2000s. Under these conditions, micro-aggregated labour productivity (average, median, other moments) is computable for all firms (I_001_04) and for exporters (exploiting the matching with OeNB-Statistics Austria micro-data on exports, imports and FDI). Micro-aggregated ULC (average, median, other moments) for all firms (I_013_02) is also computable under with the above-mentioned constraints. This information, however, is not accessible.

TFP – Under the conditions already explained for labour productivity, Micro-aggregated TFP (average, median, other moments) is computable for all firms (I_003_03) for exporters (I_003_05) and for importers (I_003_06). Olley and Pakes TFP decomposition (I_004_01) and Foster decomposition of TFP growth (I_005_01) are also computable. This information, however, is not accessible.

Firm dynamics – As far as we have been able to reconstruct, only dispersion by firms size (I_055_01) and share of fast-growing firms (which we refer to as gazelles[8])

7. Please note that the information provided was collected mostly from publicly available sources and has not been verified by the National Statistical Office.
8. These are generally defined as firms displaying growth rates significantly above the average firm (see, for instance, Henrekson and Johanson, 2010).

(I_056_01) are computable (the conditions mentioned above still apply). This information, however, is not accessible.

Internationalisation – Through the OeNB-Statistics Austria micro-data on exports, imports and FDI, possibly matched with balance sheet data collected by the OeNB, the following indicators are computable: average, median and other moments of value of export per exporting firm (I_009_02), number of exporting firms (extensive margin) (I_048_01), average, median, other moments of export sales as a share of total turnover (intensive margin) (I_047_01), number of importing firms (extensive margin) (I_048_01) and average, median, other moments of imported intermediates as a share of total cost of materials (intensive margin) (I_050_01). This information, however, is not accessible.

R&D and other activities – Information on R&D expenditure is available in the balance sheet data collected by the OeNB. Thus, R&D expenditure – mean (I_023_04), R&D expenditure (% of turnover) – mean (I_023_05) and asset tangibility (I_059_03) are computable under the above-mentioned restrictions. This information, however, is not accessible.

Accessibility

The sources quoted above are not publicly available.

Belgium

The data needed to compute bottom-up indicators derive from two main sources. The first source is BelFirst database collected by Bureau Van Dijk. BelFirst is publicly available (conditional on a fee payment). It includes information on firms' balance sheets. Sector classification is identified by a NACE code at 5-digit level. Most of the series of data start from 1995.

The second data source is the National Bank of Belgium (NBB). NBB collects data on both balance sheets, trade at firm-level (Transaction Trade dataset), and FDI (Survey on Foreign Direct Investments). Sector classification is identified by a NACE code at 4-digit level (both rev1.1 and rev.2). Production data are disaggregated at CN8 product level.

Labour productivity – Labour productivity is computable in all its versions. Existing data allows the calculation of micro-aggregated labour productivity (average, median, other moments) for all firms (I_001_04), domestic firms (I_001_05), exporters

(I_001_06) and so on. Access to firm-level data is confidential. Only the labour productivity (I_001_04) and the unit labour cost (I_013_02) are fully accessible, because the necessary data is available in BelFirst. In the case of indicator by export status (eg, I_001_06), or ownership (eg, I_001_09), the indices are computable but not accessible.

TFP – Similar to labour productivity, TFP can be easily calculated for a long time span. Existing data allows the calculation of micro-aggregated TFP (average, median, other moments) for all firms (I_003_03), domestic firms (I_003_04), exporters (I_003_05), and so on. Again, access to firm-level data at NBB is confidential. The TFP indices are all computable but only the indices for all firms (I_003_03) and the TFP decomposition index (I_004_01, and I_005_01) are accessible, because the necessary data are available in BelFirst.

Firm dynamics – The entry rate is poorly computable (I_051_03) because of the lack of entry and exit information both in BelFirst and the NBB database: if a firm enters these databases, it does not necessarily mean that the firm is a brand new one (the same for exit). The only reliable source is CompNet database, where this indicator is already computed. Similarly, the exit rate (I_052_03), and the survival rate (I_053_01) are not clearly computable. Instead, indicators on firms' growth are easily computable. The dispersion of firms by size (I_055_01) is computable and accessible through BelFirst, while average firm size by age (I_054_01) and the share of gazelles (I_056_01) are computable but not accessible (entry and exit data are in the NBB database).

Internationalisation – NBB has a rich dataset that collects information on trade activity at firm-level. All the indices listed in Table 6.17 such as the average (median, variance, and other moments) of number of export destination per exporting firm (I_043_01) are computable. The intensive (eg, I_047_01) and extensive (eg, I_045_01) margin of trade are also computable. However, NBB data is confidential and the indices are not accessible.

R&D and other activities –R&D data is not available at NBB and in BelFirst. Moreover, the R&D expenditures are poorly reported in annual accounts, and only for the largest firms. However, R&D data is available from 1998 to 2011 at Belspo (Federal Public Planning Service Science Policy)[9]. Instead, it is possible to calculate (with NBB data) the share of foreign owned firms (I_041_03), and the share of domestic multinationals

9. This information has been retrieved from the website (http://www.belspo.be/belspo/index_en.stm). In principle, micro-level data at belspo should be identifiable by VAT number and thus matchable with NBB data. R&D data is potentially accessible at belspo (conditional on a project submission). However we were not in position to verify such information on matchability and accessibility.

(I_042_03). In addition, it is possible to calculate both the level of tangible assets (I_059_03) and the average level of unit values (I_62_01) for exported goods. However data is confidential with the exception of I_059_03 (computable on BelFirst).

Accessibility

NBB data is confidential and restricted, and use is allowed only to NBB members (or affiliates). NBB data on firms' balance sheets is the same data provided by Belfirst, and this source is available on payment of a fee.

Bulgaria

Firm-level data can be recovered from three main data sources. The first is the Information System Business Statistics (ISBS) integrated database, containing the annual reports (a set of accounting and statistical questionnaires) of all economically active enterprises in Bulgaria[10]. The second source is the Statistical Business Register. Last, firm-level trade (custom data on trade with third countries) data is reported in SAD (Single Administrative Document). In addition, trade data is collected by the National Revenue Agency (NRA) of Bulgaria (intra-EU trade), and by the Customs Agency (extra-EU trade). Bulgaria started from 2010 to collect data on foreign/domestic ownership of firms and multinationality status (ie if a firm has affiliates abroad) through a particular 'Report on enterprise group'[11].

The National Statistical Institute of Bulgaria (BNSI) maintains the ISBS and the Statistical Business Register. The BNSI is also responsible for both intra and extra trade in goods statistics. Finally, the presence of a unified identification code (EIK) for each enterprise in Bulgaria allows firm-level data from different sources to be linked. The only exception is the foreign trade database which contains the EIK from 2008, so even if data is available, the trade related indices are not computable at firm level.

The time span starts from 2001 and data was collected using NACE Rev.1 from 2001 to 2003, NACE Rev.1.1 until 2008, and NACE Rev.2 since then. Geographical location is identified with a NUTS3 code.

10. The system provides online collection of annual reports of all the economically-active enterprises, containing a set of accounting and statistical questionnaires, for both large and small (with net receipts from sales of up to 100 thousand BGN) enterprises. The questionnaires for large firms are more detailed.
11. The data on ownership of the firms is available in Stat. BR since before 2000; the data concerning ownership of enterprise group is available from 2010. In the summary tables, we define the computability of indices by ownership considering information on ownership for enterprise group.

Labour productivity – Indicators in this group are not all perfectly computable. It is possible to measure only the labour productivity (I_001_04), and unit labour cost (I_013_02) for all firms from 2001 using survey data. In the case of importers (I_001_07) and exporters (I_001_06) the degree of computability is lower (data from 2008). The index can be calculated both at sectoral and regional levels.

TFP – TFP index and its decompositions are computable from 2001 (I_003_03, I_04_01, and I_005_01) with survey data. The degree of computability of TFP by trade status is low (eg, I_003_05 from 2008).

Firms' dynamics – There is information on firms' dynamics in Bulgaria from 2005. Furthermore, it is possible to compute indices on firms' dispersion (I_055_01) and share of gazelles (I_056_01) from 2001. For the existing data, accessibility is limited.

Internationalisation – Concerning internationalisation, Bulgarian databases provide information on external trade from 2008. All the internationalisation indicators are computable from 2008.

R&D and other activities – R&D data has been collected since 2001, as well as data on tangible assets. Information on the unit values of exports has been collected since 2008 (I_62_01), while information on firms' ownership starts only in 2010.

Accessibility

All the sources mentioned above are restricted, and access is strictly regulated by the Protection of Secrecy (chapter 6, of Statistical Act).

The micro-data from different statistical fields is accessible, if it is possible and does not conflict with existing regulations, and after a decision of the Commission appointed under Art.10 of the 'Rules for providing of anonymised data on scientific and research purposes'. These rules govern the provision by BNSI of micro-data and the procedure for obtaining them. The rules are based on, and in accordance with, requirements of national and relevant EU legislation. See https://unstats.un.org/unsd/dnss/docViewer.aspx?docID=2772. See also indicator 15.4 in http://www.nsi.bg/sites/default/files/files/pages/LegalBasis_e/BG_report_FINAL.pdf.

Croatia

Firm-level data for Croatia is derived from the Croatian Bureau of Statistics (CBS). The main sources are Structural Business Statistics and Community Innovation Surveys (CIS) compiled for Eurostat, complemented by data on international trade collected by the same office. Firm-level balance sheet information is not available in Croatia, with the exception of for turnover and R&D expenditures, which are collected for the CIS.

Sectoral disaggregation is NACE, 4-digit, while regional disaggregation depends on specific variables/datasets.

Labour productivity – No indicator is computable, because firm-level information on value added and number of employees is not available.

TFP – As above, no indicator is computable, also because firm-level information on value added and number of employees is not available.

Firms' Dynamics – Entry rate (birth rate) (I_051_03) and exit rate (death rate) (I_052_03) are computable since 2008. However, in 2014 the CBS has started to follow more accurately firms' survival. Real births are available from 2010 onwards, and only survival for 1-3 years is observable. The other indicators are not computable because of the lack of information on firms' ages and number of employees[12].

Internationalisation – Average, median and other moments of value of exports per exporting firm, total (I_009_02) and average, median, other moments of export sales as a share of total turnover (intensive margin) (I_047_01) are not computable because the information on value of production sold abroad is not available, while average, median, other moments of imported intermediates as a share of total cost of materials (intensive margin) (I_050_01) is not computable because of the lack of firm-level information on material costs. Percent of exporting firms in total number of firms (extensive margin) (I_046_01) and percent of importing firms in total number of firms (extensive margin) (I_049_01) are computable only since 2008 because the information on total number of firms is only available since that year. All the other indicators are computable since 1991.

12. For firms that started up in 2010 and later, there is information on firms' age, and for all active companies there is information on the number of employees (for certain years). Breakdown by size is feasible.

R&D and other activities – Asset tangibility (I_059_03) is not computable because information on tangible fixed assets and total assets is not available, while firm-level estimates of quality (I_070_01) is not computable because firm-level data on value of production sold abroad is not available. R&D expenditure – mean (I_023_04) and R&D expenditure (% of turnover) – mean (I_023_05) are computable for 2006, 2008 and 2010 through CIS. Share of foreign-owned firms in total firms (by country, sector, region) (I_041_03) is computable from 2008 and share of domestic MNFs in total firms (by country, sector, region) (I_042_03) is only computable for 2013, since information on multinational status of the firm has just started to be collected.

Accessibility

Access to most data is restricted. Data collected for CIS (turnover and R&D expenditure) can be accessed under certain conditions (for scientific purposes according to the Ordinance on the methods of statistical data protection and Ordinance on Conditions and Terms of Using Confidential Data for Scientific Purposes).

Czech Republic

The main databases for the Czech Republic are the Business Register (named RES) and the External Trade Database. Both datasets are collected by the Statistical Office (CZSO), but are also available at the National Central Bank (NCB)[13].

For the period up to 2007, the Business Register includes companies with 20 or more employees. From 2008, the Business Register considers only firms with 50 or more employees (smaller sample). The External Trade Database available at the NCB is a smaller version of the full dataset at CZSO (data on 1,000 biggest exporters and 1,000 biggest importers). According to the reported information, the Business Register starts from 2002, while the External Trade Database is available from 1999.

The NCB also collects firm-level data on FDI inflows (about 5,700 firms). Information on foreign ownership is also available in the Business Register (50 or more percent of equity). In addition, statistics on outward foreign affiliates (about 500-600 Czech firms with significant foreign affiliates) are collected and available at the NCB, and data has

13. The Business Register includes all companies (legal persons), self-employed persons (natural persons) and authorities, that is 2.8 million entities. The CSZO administers data concerning international trade with goods. Data on international trade with services is collected by the Czech National Bank.

been harmonised since 2007. Indicators can be defined at NACE rev.2 classification (2 digits) from 2005 (or 2007). Regional disaggregation is not reported.

The External Trade Database at NCB can in principle be matched with the Business Register because the national firm identifier ICO is available in both databases. However, the Czech National Bank is not authorised to provide micro-data originating from CZSO. Finally, note that in the External Trade Database, the main identifier is DIC (tax ID), while ICO (national firm ID) is a secondary identifier, and thus some combinations are not feasible.

In conclusion, the main issue for the Czech Republic is not the availability of underlying variables, but the unclear accessibility of Custom Data. Finally, it is worth mentioning that some of the indicators can be retrieved from the CompNet database.

Labour productivity – Labour productivity indicators are computable from 2002 (or 2005 for exporting firms) with a harmonised classification (NACE rev2). Data on multinational status needed for indicators I_001_08 and I_001_09 (domestic and foreign multinationals) is available only from 2007 and only for a restricted sample of firms.

TFP – The same considerations of labour productivity indicators apply to TFP indicators.

Firms' dynamics – Firm dynamics indicators are computable through the Business Register. However, information on firm deaths is not reported: thus indicators I_052_03 and I_053_01 are not computable.

Internationalisation – All the indicators on internationalisation are computable, through the Business Register and Custom Data. The Business Register allows us to compute directly I_009_02, while for the other indicators it is necessary to merge the two sources.

R&D and other activities – R&D indicators (I_023_04, and I_023_05) are computable using CIS for 2000, 2001, 2004, 2006, 2008, and 2010. The share of foreign-owned firms is computable from 2005 (I_041_03), while the share of multinationals is computable from 2007 (I_042_03). Tangible asset level is computable.

Accessibility

Business register data can be accessed both at the NCB and CZSO. For access, an external researcher has to provide a research project and pay a fee. Data can be accessed both on-site and with CDs (depending on the agreement). According to NCB,

custom data is available only for NCB employees, and the NCB does not report the conditions to use FDI, and outward FATS data. Access conditions for the External Trade Database at CZSO are regulated by special contract of confidentiality, and the access is only granted for research purposes (on payment of a fee).

More details are available at
http://www.czso.cz/eng/redakce.nsf/i/statistical_data_for_scientific_research_purposes.

Denmark[14]

Firm-level data in Denmark is from Statistics Denmark (the central authority on Danish statistics). In order to describe indicators' computability, we collected information on different data sources such as the Industrial Accounts Statistics, the External Trade in Goods, or the FIDA database. The first of these includes balance-sheet information, the second contains the trade statistics (Intrastat and Extrastat), while the FIDA database is an employer-employee database that encompasses labour cost and some balance-sheet items. In addition, we consider the Business Demographics and the Foreign Owned Enterprise databases.

All the databases report information on firms' industry that is compatible with NACE classification. Regional location is collected in the Industrial Accounts Statistics and in the FIDA database. However, the computable indicators, as the *internationalisation* indices, can be defined at regional level merging the different databases. In principle, it seems that all the mapped databases can be merged given that several ID codes are reported for each firm, but we have not had confirmation from Statistics Denmark (see footnote 14).

Labour productivity – The labour productivity indices are computable from 1995, 1997 (by import/export status) and 2004 (by ownership). Indicators I_001_08 and I_001_09 are not computable (since the information on multinational status is missing).

TFP – Similar to labour productivity, TFP indices are computable from 1995, 1997 (by import/export status) and 2004 (by ownership). Indicators I_003_07 and I_003_08 are not computable (missing the information on multinational status).

14. Please note that the information provided was collected from publicly available sources. Despite several attempts to contact Statistics Denmark, we could not verify and integrate this information. In particular, we are not in position to verify the details and the extent to which different sources can actually be matched.

Firms' dynamics – It is possible to calculate the entry and exit rate (I_051_03, and I_052_03), and survival rate and average firm size (I_053_01, and I_054_01) index using different data sources (FIDA or Business Demography). Indices on firms' dispersion and share of gazelles (I_055_01, and I_056_01) are computable. For these two indices, Statistics Denmark has a specific database (*Gazelles in Denmark)*.

Internationalisation – All the trade indicators can be computed from 1997.

R&D and other activities – R&D expenditure (I_023_04), R&D intensity (I_023_05), and firms' ownership (I_041_03) indicators are computable. Conversely, the share of domestic multinationals (I_042_03) is not computable. Finally, tangible assets and the firm-level index of quality are computable too.

Accessibility

Data is accessible for persons affiliated to Danish institutions which are recognised by Statistics Denmark, conditional to the approval of a project. In principle, foreign researchers can access data if they have an affiliation with a Danish institution. Affiliation can only take place if the authorised institution is willing to take the responsibility for the foreign researcher, making sure that all rules governing access to micro-data are observed. Data can be accessed on site or remotely. See more information at http://www.dst.dk/en/TilSalg/Forskningsservice.aspx

Estonia

Firm-level data can be recovered from three main data sources: (i) Business Register merged with custom data, (ii) Central Bank data, and (iii) R&D database. While the first two databases are available at the Central Bank, the latter is collected by (and available at) Statistics Estonia (SE). In addition, Statistics Estonia collects information on economically active enterprises in a database named the Statistical Profile: it is updated from official Business Register and statistical surveys. Data in the Statistical Profile and in the other surveys, such as R&D survey (as CIS), can be linked for micro-analysis. The Statistical Profile database is available also for the Central Bank.

The main data source is the Business Register merged with custom data, which is available at both institutions.

Firms are classified according to NACE rev.2 classification at 3 or 4 digit level (only at 2 digits for R&D surveys). Part of the time series starts from 1995, while others start

from 2003. Regional aggregation is not reported (since Estonia itself is a NUTS 2 region); R&D is estimated also at the NUTS3 level[15]. It is important to underline that even if all the disaggregated groups are possible within the available variables, the confidentiality rule requires that firm-level information cannot be discovered if fewer than three firms belong to the group (and one firm dominates the group). Given that Estonia is a small country this is not unlikely.

Labour productivity – All the labour productivity indices are computable since 1995, although the indicators by export/import status and foreign/domestic ownership are computable only since 2003[16].

TFP – Similarly to labour productivity, all the TFP indices are measurable within the limits mentioned above. The decomposition indexes are computable since 1995.

Firms' dynamics – All the indices about firms' dynamics are computable.

Internationalisation – The competiveness indexes on trade activity are computable from 2003.

R&D and other activities – R&D data are available (I_023_04) at the National Statistical Office from 1998. Similarly, R&D intensity (I_023_05) is computable, merging R&D surveys with the Business Register. Information about foreign ownership is available from 2003. Finally, data on firms' tangible assets and export unit value is available.

Accessibility

Data is at SE, and the availability of micro-data for scientific purposes is regulated by legal acts and can be used in the safe centre (see http://www.stat.ee/legal-acts). In addition, all the sources mentioned above are highly confidential, so accessibility rules are quite restrictive.

Finland

Finnish data is available from different sources. Most data is collected by the National Statistical Office, while the database on foreign trade statistics is collected by the

15. In the case of some big corporations, R&D value is connected to their headquarters, not to the unit performing the R&D.
16. The indicators by export/import status are in principle computable through the CompNet database since 1995, however we were not in position to verify details and access conditions.

Finnish Custom Office. The different sources can be matched, so that computability of indices is guaranteed. Firms are classified according to NACE rev. 2. Regional disaggregation is possible. The unit-level data is confidential. Total number of firms is publicly available.

Labour productivity – All the labour productivity indices are computable. Access to the data is limited.

TFP – All the TFP indices are computable. Access to the data is limited.

Firms' dynamics – It is possible to calculate the entry and exit rate index (I_051_03, and I_052_03), as well as survival rate and average firm size (I_053_01, and I_054_01). However, because of mergers and acquisitions, the quality of data might not be good and the degree of computability is reduced. However, indices of firms' dispersion and share of gazelles (I_055_01, and I_056_01) are computable. Access to the data is limited.

Internationalisation – Trade indicators can be computed. However, the coverage of indicators is different according to the data source and the thresholds of registered transactions, meaning the degree of computability is reduced (I_009_02, I_043_01, I_043_02, I_044_01, I_047_01, and I_050_01). These issues do not arise with the overall numbers (and percentage in the total number of firms) of importers and exporters (I_045_01, I_046_01, I_047_01, I_048_01, and I_049_01). Access to the data is limited.

R&D and other activities – R&D expenditure (I_023_04, I_023_05), firm ownership (I_041_03, I_042_03) are computable, although the computation of tangible assets (I_059_03) and firm-level estimates of quality (I_070_01) could imply some possible problems. Access to the data is limited.

Accessibility

Data is accessible at the Research Laboratory or via the remote access system conditional on a user licence, access agreements and a fee payment. See more details at http://www.stat.fi/tup/mikroaineistot/index_en.html.

France

Micro-level data is available from three different databases. First, FICUS – *Système Unifié de Statistique d'Entreprises* (FICUS – SESA) up to 2007 contains balance-sheet data (from fiscal forms), with other information and identification number from

business registers. Then, the ESANE (FARE) (since 2008) reports information of the same kind (balance-sheet data and other information from social data or business registers; ownership is available through a merge with specific surveys or administrative data (LIFI). Finally, the *Déclarations Douanières* – administrative data collected by the DGDDI (directorate of the ministry of economy) reports trade statistics (at firm level). All three databases are available at the National Statistical Office (INSEE).

Users have to be careful about the meaning of the firm unit (legal status in these databases).

As of July 2014, data up to 2012 is available.

Firms are classified according to NACE classification at 2 digits (rev.1 from 1994 to 2007, and rev.2 from 2008 to 2012); geographical location can be identified with a NUTS 2 code. The historical series go from 1994 to 2007 and 2008 to 2012 (with the Nace-Rev2). Data is partial for 2008 (beginning of the new system).

Labour productivity – Almost all labour productivity indices are highly computable, as well as unit labour cost. Labour productivity indices by ownership are computable from 2008. Data to calculate the competitiveness indices is highly confidential but access is feasible.

TFP – Almost all TFP indices are computable for a relatively long time series, with the exclusion of statistics by ownership (available since 2008). The relative underlying data is confidential, so that the degree of accessibility is limited.

Firm dynamics – All the competitiveness' indices on firms' dynamics are computable. Data is available on the FICUS or ESANE databases. However, data is confidential.

Internationalisation – All the measures on trade activity are computable. Data is available through *Déclarations Douanières* by DGDDI.

R&D and other activities – Indicators of R&D expenditure (I_023_04, I_023_05), tangible assets (I_059_03) and export unit value (I_070_01) are computable. Ownership data has been collected since 2008 (I_041_03, I_042_03).

Accessibility

All the sources mentioned are highly confidential, but micro-level data will be

accessible with the new system by submitting a research proposal and conditional on approval by a committee. Details on accessibility can be found at http://www.casd.eu/.

Germany

German bottom-up indicators can be computed based on data from several datasets, the most important of which are: (i) the Financial Statements Statistics; (ii) the Micro-database Direct Investment (MiDi); (iii) Germany's International Trade in Services from the Deutsche Bundesbank; (iv) a panel on manufacturing firms based on 'Official Firm Data for Germany' (AFiD) provided by the Federal Statistical Office (Destatis); and (v) data on employment at establishment level by the Federal Employment Office. Finally, some of the indicators can be retrieved directly from the CompNet database (at ZEW).

Data is classified with a NACE code (2 or 3 digit level) both in rev.1.1 and rev.2 (from 2008). In the mapped database it is not possible to recover information on the exported quantities and the ownership of firms abroad (ie if a German firm controls firms abroad).

Despite the general and good accessibility of the micro-level data at each institution, matching data between those institutions is nearly impossible because of privacy protections. Within a specific project – KombiFiD (www.kombifid.de) – data from the three above-named institutions was matched for a limited number of firms. However, all firms had to be asked for their written consent to agree to the matching and the data was only matched for one specific year. The matched dataset had to be deleted after three years. This restriction causes a limited computability for some of the indicators, despite good availability of the original variables needed to calculate the indices.

For example, AFiD panel can be merged with other firm-level databases from Destatis. However the same AFiD is not easily matchable with IAB Establishment Panel at BA. This issue raises a trade-off between time coverage and the number of computable indicators. The AFiD panel starts to be complete from 2002, while BA data covers a longer time span (from 1975). However, the data contained in the AFiD panel allows identification of more indicators because the AFiD is richer in information than BA data. In addition, we are not able to map (at the moment) a detailed dataset on international trade activities (for manufacturing firms) at the Deutsche Bundesbank.

In light of this, the report and the summary tables in the Annex describe the indicators that can be constructed with data at Destatis, in order to maximise the number of computable indicators.

Labour productivity – The aggregate values of labour productivity and unit labour costs are computable in the mapped databases, with the exclusion of the indicators by import (I_001_08) and multinational status (I_001_08, I_001_09). Some of the indicators are available in CompNet (by sector, NACE rev.2 2 digit).

TFP – The same considerations we made for labour productivity apply to TFP indicators. In addition, Olley and Pakes, and Foster decomposition are computable from 2002.

Firm dynamics – All the indicators on firms' dynamics are computable and information is accessible at Destatis.

Internationalisation – Using the information in the AFiD database, it is possible to calculate exports per firm (I_009_02), and the extensive and intensive margin of exports. However, indicators by destination and number of exported products are not computable for two reasons. First, trade data by destination and number of products are available only at the Bundesbank, but merging is not allowed. Second, Bundesbank collects only data on trade in services. For the same reasons, indicators on import activity for manufacturing firms are not computable.

R&D and other activities – Indicators of R&D and tangible assets are computable with the mapped databases. The multinational firm status and unit value of exports are not computable given that the necessary information is not available in the mapped databases.

Accessibility

Most of these datasets are available in general under certain conditions at the respective institutions. Destatis, the Federal Employment Office (*Bundesagentur für Arbeit*, BA) and the Bundesbank all have dedicated Research Data Centres[17] which offer on-site or remote access (or direct access via Scientific Use Files) to many of their micro-level datasets, according to the German laws of privacy protection. Data is accessible to researchers, but only at the BA can foreign researchers get access to the data without cooperating with a partner from Germany.

Data from the Deutsche Bundesbank is accessible only at the Research Centre (in

17. See www.forschungsdatenzentrum.de for the Destatis' Centre, see www.fdz.iab.de for the Federal Employment Office's Centre and http://www.bundesbank.de/Navigation/DE/Bundesbank/Forschungszentrum/forschungszentrum.html for information about the Research Data Centre of the Deutsche Bundesbank.

Frankfurt am Main). The use of data from the Deutsche Bundesbank is subject to special confidentiality conditions. Because of legal requirements, individual data cannot be made generally available. However, this data is made available under strict conditions and for clearly defined academic research purposes. Bundesbank has a visiting researcher programme at the Research Centre.

In the case of BA, the FDZ offers three ways of data access for researchers. These differ according the degree of anonymity of the data and the terms of data use: (i) on-site, (ii) remote data access, and (iii) Scientific Use File (rare). In all the three cases, the researchers have to present a research project that has to be approved by FDZ. In the case of on-site access, there is the possibility to apply for financial support[18].

The research data centre of the Destatis offers four different forms of access to selected micro-data of official statistics: (i) public use files, (ii) scientific use files, (iii) safe centres, and (iv) remote execution. They differ with regard to both the anonymity of the data, and the form of data provision. The scientific use files are well-suited for large parts of the scientific data analyses. Foreign users, who are not employed by German institutions, may work with the data both at the research centre and via remote executions. More details can be found at
http://www.forschungsdatenzentrum.de/en/datenzugang.asp.

Hungary

The data used to compute bottom-up indicators for Hungary is derived from six sources. First, company income tax return data of double-entry bookkeepers is collected by the National Tax and Customs Administration of Hungary (NAV)[19]. Tax return data includes information connected to balance sheets and profit and loss statements. Second, there is product-country-year level trade data based on survey data and data collected in customs procedures. For years prior to EU accession, trade data covers all transactions and it is based on customs declarations. Since 2004, trade data consists of Extra- and IntraStat statistics. Extrastat is based on customs declarations while IntraStat is based on a survey which covers companies with an annual intra-EU trade turnover of above the yearly determined exemption threshold. Information on R&D is reported in the Innovation Database (based on the Community Innovation Survey) and the research and development (based on R&D surveys of the HCSO) database of

18. More details are at http://fdz.iab.de/en.aspx.
19. NAV transmits the data to the Hungarian Central Statistical Office (HCSO) and HCSO makes it available for research purposes.

the Hungarian Central Statistical Office. Finally, the Business Register records information on firms' year of creation/destruction[20]. All the databases are maintained and made available in a safe research room at the HCSO, subject to agreements with HCSO.

Labour productivity – Almost all labour productivity indices are computable, with the exclusion of aggregates for domestic multinationals (I_001_08), and affiliates of foreign multinationals (I_001_09) given that data on multinational status is not available. In addition, it is possible to compute also the unit labour cost[21]. These indicators are accessible also through CompNet. Data is available from 1992.

TFP – It is possible to calculate all the TFP indices, and the two decomposition terms. Because of the absence of information on multinational status it is not possible to define TFP for domestic multinationals (I_004_01) or affiliates of foreign multinationals (I_005_01). Computable indicators are accessible also through CompNet[22]. Data is available from 1992.

Firm dynamics – According to the mapping, firm dynamics indicators are all computable from 1992. Note that there are caveats in calculating age of firms, especially for the early years, since the Business Register is truncated at 1992.

Internationalisation – According to the mapping, indicators of internationalisation are all computable from 1992.

R&D and other activities – Indicators on R&D since 1999 can be retrieved from the innovation and the research and development databases (CIS observations are biannual). Data is merged with tax return data to obtain I_023_05. Information on firms' multinational status is not available. Tangible assets and unit value of exports are computable.

20. The sources of Business Register data are: own data collections of the HCSO, database of the National Tax and Customs Administration, register of the Court of Registration, Central Office for Administrative and Electronic Public Services, Hungarian State Treasury, etc.Firms are classified according to NACE 4 digit code (rev1.1 from 2003, rev.2 from 2008), and geographical location is defined by NUTS3. In the trade database, location is not reported. However it can be retrieved from the balance sheet data. In addition, in the Business Register, location information is not available for all the firms.
21. Note that there are caveats in calculating value added for certain sectors (oil and tobacco industries).
22. TFP indicator in CompNet is the Wooldridge augmented Levinsohn-Petrin (GMM estimated) and TFP distribution is shown for all firms, exporters and non-exporters and by firm size. No distinction according to ownership. Decomposition is Olley-Pakes type and Forster without entry and exit in CompNet.

Accessibility

The Hungarian matched data was created by the CSO by assigning an anonymised identifier to each company, which is consistent between years and databases. Data protection, required by the law, is a key element in the operations of the CSO. Therefore, variables that provide a direct possibility to reveal the identity of a company (eg name of the company, address of the headquarters or tax number) were deleted. Technically, the data is stored on a server in separate files according to topics. Merging the different databases using the ID numbers assigned by the CSO is performed by the researcher.

The matched database is accessible only to researchers with an agreement with CSO, such as the Hungarian Academy of Sciences or some ministries. Access is granted after registering the project at the CSO. The accessibility of the matched database is restricted to a safe research room inside the building of the CSO where researchers can work on the data on site and save their results. Note that accessibility is still limited and occasionally quite slow. The researcher who works with the data has to be in the research room in Budapest and needs be affiliated with a partner.

Ireland

Different source of data, all collected by the Central Statistics Office Ireland (CSO), are taken into account: the Census of Industrial Production (CIP), the Annual Services Inquiry (ASI), the Merchandise Trade Data (MTD) and the Business Expenditure on Research and Development Survey (BERDS).

In these databases, firms are classified according to NACE classification (4 digit, rev.1 and rev.2); geographical location is identified with a NUTS3 code. Historical series are mostly available from the middle of the 1990s.

Labour productivity – All the labour productivity indices are computable. The indices by ownership are available from 1996, while all the other indicators are computable from 1991.

TFP – All the TFP indices are computable, even if the data for capital stock presents some difficulties in the calculation[23]. There are restrictions on the use and publication of results.

23. No capital stock data is available in CIP or ASI. Capital stock could be calculated based on capital investments and disposals using the perpetual inventory method. Starting stocks could be obtained by breaking down previous year's end of year industry-level capital stocks obtained from CSO to the firm level using the firm's share of industry-level fuel use.Firms' dynamics – All the indices on firms' dynamics such as entry rate (I_051_03), exit rate (I_052_03) or survival rate (I_053_01) are computable from 1991.

Internationalisation – All the indicators of competiveness on international activities are computable[24].

R&D and other activities – R&D indicators (I_023_04 and I_023_05) can be computed from 2002 on biannual basis. Ownership and tangible assets are computable. Unit values of exports require merging of two datasets[25].

Accessibility

Access to the data is in principle possible, but subject to stringent conditions. Firm-level data can be accessed on-site only, while the use and publication of results is subject to statistical office approval.

Italy

The firm-level data considered here is provided by the National Statistics Institute (Istat). Istat collects firm-level data through different databases: the Business Register (ASIA), the Register of Domestic and Global Groups, the Business Demography, the Surveys on Firms' Accounts, the International Trade in Goods[26] (linked to firms), the Survey on Foreign Affiliates and on Foreign Controlled EU Enterprises, the Survey on R&D Expenditure and the Balance Sheets Panel. At the ADELE Laboratory (Laboratory for Elementary Data Analysis) researchers can use, under certain conditions, micro-level data collected in the surveys.

Firms are classified according to NACE nomenclature at 4 digit level (rev.1 from 2001 to 2007, and rev.2 after 2008); geographical location is identified with a NUTS2 code or NUTS3.

Labour productivity – All the indicators on labour productivity are computable from 2001. The accessibility is feasible through ADELE. However, the indices by export/import status and ownership/multinational are not accessible, because different data sources cannot be merged with balance-sheet data at the ADELE laboratory. Similarly, data for unit labour costs is not accessible from ADELE (since two

24. Information on firms in the service sectors is not available.
25. This variable could be constructed using a merged dataset of industry enterprise census with customs data.
26. This database is the result of merging a database on international transaction with the business register. It is created for Eurostat statistics.

data sources need to be merged). In addition, indicators can be retrieved from the CompNet database.

TFP – All TFP indicators are computable, but only the aggregate index (I_003_03) and the TFP decompositions (I_004_01 and I_005_1) are accessible (all the information is in the Surveys on firms' accounts). Similar to labour productivity, the indices by trade, ownership and multinational status are not accessible given that data sources at the ADELE laboratory are anonymised. However, indicators can be retrieved from the CompNet database.

Firm dynamics – Firm dynamics indicators are computable from 2001, but not accessible because statistics are calculated with the Business Demography (which is not available at ADELE laboratory). While it would be possible to compute and access the firm dynamics statistics, using the Business Register, ISTAT indicates the more reliable figures are those calculated with the Business Demography, according to Eurostat guidelines.

Internationalisation – All the indicators of internationalisation are computable, but data is not accessible to researchers (elementary trade data is not available at the ADELE laboratory).

R&D and other activities – R&D data is available from 2001 from the R&D survey, and the correspondent indicators are computable. Similarly, indicators on ownership and tangible assets are computable, but accessible for the period 2001-08 (more recent data is not available at ADELE yet). Finally, the average unit value of exports is not computable given that exported quantises are not available at firm level.

Accessibility

Firm-level data is confidential and restricted. The Business Register (except for Business Demography) and micro-data stemming from surveys is available to the users at the ADELE Laboratory (Laboratory for Elementary Data Analysis). However, it should be stressed that identification codes of single units are not available to external researchers; thus it is not possible to merge data from different surveys without a specific agreement with Istat (research protocol)[27]. Databases with the full population

27. See for example project Istat – Micro3. For further information about ADELE laboratory see http://www.istat.it/en/information/researchers/analysis-of-individual-data.

are not accessible to researchers, but descriptive statistics from these databases are available on request.

Latvia

The firm-level data considered here is provided by the Central Bureau of Statistics of Latvia (CBS). CBS collects firm-level data through different databases among which are the Annual Enterprise Survey, the Business Register and State Revenue Service data (SRS)[28].

The three databases can be merged through a unique identifier. The CBS of Latvia also collects monthly data on exports and imports (Custom data) from 2005 without information on firms' location[29]. We were not in a position to verify the matchability of detailed trade data with other databases from CBS. The Business Register reports import and export status by firms.

Information on Latvian multinational firms is missing, while foreign ownership is reported.

Firms are classified according to NACE nomenclature at 4 digit level (rev.1 from 1997 to 2005, and rev.2 after 2005): because of the implementation of NACE rev.2, the data series are comparable from 2005. Geographical location is identified with a NUTS 2 code (as already mentioned above, this information is not available for Custom data). For each year, the preliminary data version is available around ten months later, while final data is available 18 months later (eg for data for January 2014, the preliminary version is available around October 2014 and the final version in June 2015).

Since that data is harmonised and comparable from 2005, we report in the summary tables a degree of computability equal to one, even if the indicators can be computed in the previous years.

Labour productivity – All the labour productivity indicators are computable from 2005. The mapped data does not allow indicators for multinational firms to be computed because this information is not available.

28. The SRS includes annual financial statements of enterprises and employers' declaration on salary tax.
29. Foreign trade data for EU member states is collected by the Intrastat system using monthly statistical surveys. Foreign trade data for the third countries is compiled on the basis of information taken from customs declarations.

TFP – All the TFP indicators are computable from 2005. The mapped data do not allow indicators for multinational firms to be computed because this information is not available. The Business Register reports only information on statutory capital, so that it is difficult to retrieve information for tangible fixed assets.

Firm dynamics – Indicators of firm dynamics are computable only through CompNet. The mapped data allows computation of the entry rate (I_051_03), dispersion of firms (I_055_01) and the share of gazelles (I_056_01).

Internationalisation – The entire set of internationalisation indices is computable.

R&D and other activities – Variables 'R&D expenditure' and 'Turnover' are not matchable, and the indicator on R&D intensity (I_023_05) is not computable. Similarly, it is not possible to compute the indicator on multinational firms (I_042_03), because this information is not available. As mentioned above, tangible assets are not available (I_059_03). Unit value of export is computable.

Accessibility

Information on the value of exports (imports) by destination and product are not accessible because it is confidential. Other data is in principle available on request, conditional on a fee payment.

Lithuania

The firm-level data considered here is collected by Statistics Lithuania and includes several firm-level surveys, as well as balance-sheet data, tax declarations, the Business Register and customs declarations.

Data is usually classified according to NACE classification (4 digit), while international trade data can also be classified according to CN at 8 digits. As for regional disaggregation, Lithuania is itself a NUTS 2 area; only added value, number of employees, labour cost and turnover can be aggregated at NUTS 3 level.

Labour productivity – All the indicators are computable. Micro-aggregated labour productivity (average, median, other moments) – all firms (I_001_04) is available from 2000 to 2012, while the others only since 2004-05.

TFP – All the indicators are computable since 2005.

Firm dynamics – All the indicators are computable since 2005.

Internationalisation – All the indicators are computable since 2005, except average, median, other moments of imported intermediates as a share of total cost of materials (intensive margin) (I_050_01), which is not computable because the value of imported inputs is not available.

R&D and other activities – R&D Expenditure – mean (I_023_04) and asset tangibility (I_059_03) are computable from 2000 to 2012, while the other indicators are computable only since 2005.

Accessibility

Firm-level data is confidential. By the Law of Statistics, micro-level data could be used for research purposes. Confidential statistical data may be provided for scientific purposes to be used in a manner that makes it impossible to directly identify the respondents based on the data, and where the research establishments ensure data protection.

Malta

Information for Malta was retrieved taking into account several datasets, all compiled by the National Statistical Office of Malta (NSO). Although many indicators are in fact computable, data is usually available only for the last few years[30].

Malta is itself a NUTS 2 area, so regional disaggregation is not available. As for sectors, NACE rev. 2 (2 digit) disaggregation is available.

Labour productivity – All indicators are computable in principle, but only for a few years. (I_001_04), micro-aggregated labour productivity (average, median, other moments) – exporters (I_001_06) and micro-aggregated ULC (average, median, other moments) – all firms (I_013_02) are computable since 2007. Micro-aggregated labour productivity (average, median, other moments) – domestic firms (I_001_05), micro-aggregated labour productivity (average, median, other moments) – domestic multinationals (I_001_08), micro-aggregated labour productivity (average, median, other moments) – affiliates of foreign multinationals (I_001_09), micro-aggregated

30. Structural business statistics started to be compiled in 2007, while the Business Register Questionnaire containing information on foreign/domestic ownership is available only since 2010.

labour productivity (average, median, other moments) – foreign owned exporters (I_001_10), and micro-aggregated labour productivity (average, median, other moments) – domestic owned exporters (I_001_11), are computable only since 2010.

TFP – Information on tangible and total assets is not collected, thus TFP indicators are not computable.

Firm dynamics – All the indicators are computable only since 2010, with the exception of dispersion of firm by size (I_055_01) and share of gazelles (I_056_01), which are computable since 2007.

Internationalisation – All the indicators are fully computable since 1995. Exceptions are: percent of exporting firms in total number of firms (extensive margin) (I_046_01) and percent of importing firms in total number of firms (extensive margin) (I_049_01), computable since 2010; average, median, other moments of export sales as a share of total turnover (intensive margin) (I_047_01), and average, median, other moments of imported intermediates as a share of total cost of material (intensive margin) (I_050_01), which are computable since 2007.

R&D and other activities – Firm-level estimate of quality (I_070_01) is fully computable since 1995. R&D expenditure – mean (I_023_04) is computable since 2000. R&D expenditure (% of turnover) – mean (I_023_05) is computable since 2007. Share of foreign-owned firms in total firms (by country, sector, region) (I_041_03) and share of domestic MNFs in total firms (by country, sector, region) (I_042_03) are computable only since 2010. Asset tangibility (I_059_03) is not computable because tangible and total assets data is not collected.

Accessibility

All the information is accessible on request for research purposes, except data on foreign/domestic ownership.

The Netherlands

The Netherlands is rich in micro-level databases that allow researchers to compute competitiveness indicators. All the mapped databases are provided by Statistics Netherlands (CBS)[31]. The Dutch data reports information to define sectoral (NACE

rev1.1 and rev.2) and regional (NUTS 3) aggregation. The only variable (at micro-level) for which we did not find a source is the 'total assets'.

The main issue in the mapped databases is related to the matchability of data from different sources. According to the information reported, we are not able to assess if it is possible to merge data collected in different databases. Even if most of the underlying variables are collected, the computability is uncertain.

Labour productivity – According to the collected information, it is possible to compute only the labour productivity index for all firms (I_001_04) and the unit labour cost (I_013_02). For all the other indices we are not able to state computability, given that we have no information on the data merging. See Table 3.15.

TFP – We can compute only the TFP index for all firms, and the decomposition indices (I_004_01 and I_005_01). Similarly to labour productivity, we are not able to state computability, given that we have no information on the data merging.

Firm dynamics – All the indices are computable from 1993 or 2000, depending on the data source (General Business Register or Annual Structural Survey, respectively).

Internationalisation – Some of the internationalisation indices are computable, if they involve just the use of the Survey on International Trade in Goods. Conversely, we cannot define the computability of indices by import status because we have no information on matchability of data from different sources.

R&D and other activities – The R&D indices are computable from 2003, while unit value of export from 1990 (I_070_01). Conversely, we cannot report the computability for the index I_37_07, I_38_09, and I_059_03.

Accessibility

In general, many indicators of competitiveness are available to both domestic and foreign researchers. Access to micro-level data follows explicit rules, and specific charges apply. According to CBS: *"All datasets in the Centre for Policy Related Statistics' micro-data catalogue are available for authorised external researchers to do their own*

31. For an overview of existing data at Statistics Netherlands see the catalogue at http://www.cbs.nl/NR/rdonlyres/0C40DD86-7AF3-4179-B74C-1B476A6A5387/0/120119catalogusmicro-data.pdf.

research using these datasets. The catalogue does not contain all the datasets Statistics Netherlands uses to compile its statistics. CBS datasets not (yet) included in the catalogue may be made suitable for use by external researchers as custom-made datasets. The catalogue (classified by theme) includes documentation reports of the most recent version of datasets immediately available for use. This documentation contains a description of the contents and structure of the dataset. The enclosures referred to in this documentation are available only in Dutch and on request". More details can be found at: http://www.cbs.nl/NR/rdonlyres/50625EDE-3274-4D7C-B19B-5E5D0F239E2F/0/131112dienstencatalogusosra2014eng.pdf.

Poland

Information on Polish firm-level data has been provided by the Central Bank of Poland (NBP) and Central Statistical Office of Poland (NSO). The main source is the NSO for both balance sheet data and innovation data (NSO database in accordance with the Frascati Manual). The balance sheet database reports total revenues, revenues from exports (total), and all the cost variables as well as the assets and liabilities. Firm-level data is collected quarterly for firms with over 50 employees, and annually for firms with more than 10 employees. Sectoral classification has a break in 2009 (NACE rev.1.1/ rev.2 switch), but the NACE identifiers can be traced back at the firm level to 2007.

Balance sheet data covers the period 1995-2011 and includes value of imports and exports; however detailed trade data (ie quantities, products, destinations) is available as custom data from 2004 at CN8 classification[32]. Customs data is available at both the Ministry of Finance and the NSO[33]. Information on the year of firms' creation and death can be retrieved from the Business Register (REGON). Finally, firms' IDs are unique for all databases at NSO but information is anonymised, so that the data cannot be matched with other data sources at NSO by external researchers. Moreover, the customs data can in principle be merged with the balance sheet data but not at the NBP because both sources provide anonymised data with incompatible ID codes.

Information about the conditions for access to the micro-level datasets have not been reported.

32. Export status can be inferred using financial statements but it would be less reliable than customs data.
33. The customs data at the NBP is the same data as held by the NSO and ministry of finance (primary origin of the data). The accessibility of the customs data is limited to the NBP.

Labour productivity – Almost all labour productivity indices are computable, with the exclusion of aggregates for domestic multinationals (I_001_08) and affiliates of foreign multinationals (I_001_09) because of the lack of information on ownership and multinational status[34]. In addition, it is possible also to compute the unit labour cost from 2002. These indicators are accessible also through CompNet. Data is available from 1995.

TFP – Both the TFP indicators and the two decomposition terms are computable. However, due to the absence of information on ownership and multinational activities, it is not possible to define TFP for domestic multinationals (I_004_01) and affiliates of foreign multinationals (I_005_01, see footnote 34). Computable indicators are accessible also through CompNet. Data is available from 1995.

Firm dynamics – All the indicators on firm dynamics are computable using balance sheet data. The indicators on firm dynamics can be computed using balance sheet data for firms with over 10 employees. Otherwise, for the indicators I_051_03, I_052_03, I_053_01, and I_054_01 the relative information on firms' entry and exit are imputed and reported in the regional register (REGON) at NSO[35, 36]. Conversely, dispersion of firm by size (I_055_01) and share of gazelles (I_056_01) are computable from 1995 by using balance sheet data at NSO.

Internationalisation – All the internationalisation indicators are computable from 2002 or 2005 (eg I_009_02), however it was not possible to collect information on data accessibility.

R&D and other activities – R&D indicators are computable, as are unit value and asset tangibility. Ownership information is not collected (I_042_03).

Accessibility

According to the information that we were able to gather, we can only state that the

34. I_001_09 is computable if firms with foreign capital as affiliates of foreign multinationals are considered.
35. The REGON database cannot be matched by external researchers with other data sources at NSO. REGON is not available at NBP. Data for firms with more than 10 employees is available since 2002.
36. At the Central Statistical Office of Poland, data on business demography (birth rate, death rate, survivals, gazelles) is computed in accordance with the rules contained in Annex IX of Regulation no 295/2008 of the European Parliament and of the Council concerning structural business statistics. Data is prepared on the basis of the statistical business register which is updated on the basis of additional sources (not used by the REGON database) and as such is appropriate for business demography. Data on business demography of Poland (according to Annex IX) is available since 2008.

rules of statistical confidentiality are determined by the Law on Official Statistics issued on 29 June 1995. In theory, access to micro-data is possible only under specific conditions, but the practice shows that access to individual data beyond CSO and NBP is nearly impossible.

Portugal[37]

The firm-level data considered here is collected by the National Statistical Institute (INE). The mapping covers two datasets: Integrated Business Accounts System (which covers, through Simplified Business Information, all balance sheets at firm level) and International Trade in Goods (Intrastat and Extrastat data – firm-level database).

The Integrated Business Accounts System includes all firms from 2004 to 2012. Trade data ise collected in Intrastat and ExtraStat. Intrastat reports trade data for the firms with transactions above an annual exemption threshold (defined according to annual coverage rates established in the EU legislation) and data is available since 1993. Estimations of non-response and below thresholds are made, but not at firm-level (only aggregated data by commodity and partner country). Extrastat series includes all transactions available since 1993 (the compilation of data is based on customs declarations – administrative data from the Portuguese Customs and Taxes Authority).

In addition, the Annual Business Survey, which is a database with around 50,000 enterprises from 1996 to 2004, is available at INE.

Firms are classified according to NACE classification rev 1.1/rev 2 (5 digits) and at second level of NUTS.

Labour productivity – The labour productivity indicators are computable only from 2005 (to 2012). In the case of aggregated index (for all firms, I_001_04) the indicator can be obtained also from CompNet. The aggregates of labour productivity by ownership and multinational status are not computable since the information on foreign/domestic ownership is not available.

TFP – Indices on TFP are computable from 2005 to 2012. In the case of aggregated index (for all firms, I_003_03) the indicator can be obtained also from CompNet, as well as the TFP decomposition index (I_004_01, and I_005_01). The TFP index by ownership

37. According to the collected information, we are not in position to describe the data details, matchability or accessibility conditions.

and multinational status are not computable.

Firm' dynamics – Firm dynamics are all computable and comparable from 2004. Data from previous years exists, but is not comparable with new series. The information is compiled based on the Integrated Business Accounts System.

Internationalisation – Similarly to previous indices, international competitiveness indicators can also be computed from 2005.

R&D and other activities – R&D indicators are computable from 2005 at NACE 2 digit level. Information on ownership (foreign/domestic ownership of the firm) is not available. Similarly to R&D, tangible assets (I_059_03) and unit value of exports (I_070_01) are also computable from 2005.

Accessibility

We are not in position to describe in details the accessibility conditions. However, in principle data seem accessible.

Romania

For Romania, the sources of data considered here are collected by the National Statistical Office (NSO) of Romania. The main data sources are Structural Business Statistics (SBS), the Business Register and the Foreign Trade Statistics (FTS). All the three sources can be merged.

The SBS includes data on firms' balance sheet and ownership status from 2002; SBS collects also information on production sold abroad (exports) but it does not cover the whole population of exporters. Information on the value of imports and exports (at firm level) is recorded in the FTS and the data is available from 2007. Data on destinations and products exported (as well as quantities) is still not available: NSO is at the time of writing collecting the information and working on the raw data.

Labour productivity – Basic indicators of labour productivity are computable (all, exporting and non-exporting firms) from 2002. Notice that export status from 2002 is derived from the Structural Business Statistics (using the value of production sold abroad), but the information cannot be precise. Conversely, the indicator by import status (I_001_07) cannot be computed given that the information on import activity (as export) is reported only in the FTS that is not harmonised with the Structural

Business Statistics. The other indicators are only computable from 2007. Unit labour cost is computable from 2002.

TFP – The same caveats of labour productivity apply to the computability of TFP's indicators. Indicators by export status can be recovered using information in SBS, while indicators by import status cannot (import activity is only in FTS). Indicators by ownership status and international activity are available from 2007. Both Olley and Pakes (OP) decompositions and Foster decompositions can be computed from 2002.

Firm dynamics – All the indicators of firm dynamics are computable from 2002.

Internationalisation – Some of the internationalisation indices are computable from 2002. However, the indices available from 2002 (I_009_02, I_045_01, I_041_02, and I_041_02) rely on SBS and therefore are not representative of the population. From 2007, FTS starts to include trade data for most of the firms with detailed set of information, such as quantities and number of products exported, and destinations (similarly for imports). However, FTS is still in the phase of collecting and working on the raw data. FTS data is at time of writing not available and not harmonised with SBS.

R&D and other activities – Only asset tangibility, and R&D indicators are computable (from 2002). The indicators for ownership and multinational presence (I_041_03 and I_042_03) are computable from 2007. The unit value index (I_070_01) is not computable given that data on exported quantises in FTS have still to be validated.

Accessibility

Data is not accessible because a safe environment for data security is not yet in place.

Slovakia

For Slovakia, databases considered here are collected by the Statistical Institute of the Slovak Republic and the National Bank of Slovakia. The former institution compiles the Annual Report on Production Industries that targets non-financial corporation (ie firms with 20 and more employees or turnover higher than €5 million) and the individual trade data (from customs offices). The Bank of Slovakia compiles the annual reports on inward and outward foreign direct investment, and the register of organisations[38].

38. Notice that also in this case the balance sheet data, such as value added, is available only for companies with 20 and more employees or turnover higher than €5 million.

Firms are classified according to NACE classification (4 digits). The historical series are in principle collected from 2000 to 2011, even if the real availability and comparability may differ.

Labour productivity – Aggregated indexes of labour productivity and unit labour cost (I_013_02) are computable both from mapped databases (annual reports on production industries) and CompNet. Data to calculate labour productivity indices by export status[39] based on customs data is available from 2004. Labour productivity per exporter (I_001_06) can be calculated using balance data on sales abroad (collected within reports on production industries) from 2000. Data to calculate labour productivity indices by domestic/foreign ownership[40] is available from 2008.

TFP – Similarly to labour productivity, aggregated TFP indexes are computable both from mapped databases (annual reports on production industries) and CompNet. Data to calculate labour productivity indices by export/import status, and by domestic/foreign ownership is available from 2004 and 2008, respectively. However, TFP per exporter (I_003_05) can be calculated from 2000 using balance sheet information on sales abroad. OP and Foster decompositions are available from 2000.

Firm dynamics – Data for firm dynamics have been collected in principle since 2000 but the availability has to be verified sector by sector[41]. Conversely, dispersion of firms by size (I_055_01), and the share of gazelles (I_056_01) are computable from 2000.

Internationalisation – The indices of internationalisation can be computed from 2004 with individual (firm) trade data[42].

R&D and other activities – R&D data is computable from 2000 only for firms with an R&D unit. The ownership indicators (I_041_03 and I_042_03) can be computed by merging annual reports on production industries with the register of organisations (from 2008). Indexes on tangible assets and unit value of export can be computed from 2000 and 2004, respectively.

39. Match annual reports on production industries with individual trade data.
40. Match annual reports on production industries with the register of organisations.
41. Indicators on entry and exit rate for NACE 2 digit level can also be obtained from the CompNet database.
42. Simple indices of export status and export activity (without destination or product decomposition) can be also calculated using total sales abroad covered in reports on production industries.

Accessibility

The firm-level databases are not available online, and access is confidential: the rules of access have not been specified.

Slovenia

The databases considered for Slovenia are the Slovenian Business Registry (SBR), the Annual Reports of Direct Investments, the IntraStat and ExtraStat database, and the Research and Development Activity database. It should be noted that all companies in Slovenia, whether limited or unlimited liability companies (including listed companies), economic interest groupings and main offices of foreign business entities, are legally obliged to submit their annual reports to the Agency of the Republic of Slovenia for Public Legal Records and Related Service (AJPES). An additional source is the Slovenian companies' annual reports used for the CompNet project[43]. All databases are available at the Statistical Office of the Republic of Slovenia (SURS). All mentioned databases have unique ID identifier so it is possible to merge micro-level databases.

Firms are classified according to NACE classification (rev1 from 1995 to 2004, rev2. from 2005 to now), and location is identified by NUTS3 code[44].

Labour productivity – The aggregate values for labour productivity and unit labour costs are computable in the mapped databases, and are also available through Slovenian companies' annual reports. Similarly, unit labour costs are computable. Indexes I_001_08 and I_001_09 are computable only from 2008 because the information on the multinational status of a firm (ie if a firm controls enterprises abroad) was not collected before.

TFP – All the TFP indices are computable, although I_003_07 and I_003_08 only since 2008 because the information on the multinational status of firms was not reported before.

Firm dynamics – All indices for firm dynamics are computable, even if some, such as the entry and exit rate, are computable only from 2004 because the year of firms' deaths is reported from 2004.

43. The AJPES data is regularly used for national statistical purposes by other institutions, and includes the Slovenian companies' annual reports.
44. Trade data was collected according to NACE rev1 until 2007.

Internationalisation – Similarly to labour productivity and TFP, all the indices of internationalisation are computable from 1995.

R&D and other activities – R&D indices are computable from 1995. Index I_041_03 and I_038_08 are computable from 2003 and 2008, respectively. Finally both asset tangibility and quality index are computable from 1995.

Accessibility

All the micro-data is accessible at SURS and is restricted only for research purposes. See http://www.stat.si/eng/drz_stat_mikro.asp.

Spain

The databases considered for Spain are the Industrial Economics Survey, the Harmonised Demographics of Companies, the Central Business Register, the Inward FATS, the CIS and the Pitec database. All the data sources are provided by the Spanish National Statistical Office (INE). Information provided in this section has been compiled from publicly available sources, mainly the web site of INE (http://www.ine.es/). Officials at INE were contacted, but they could not help us in verifying the information. Information provided here should be used according to the conditions indicated at the following URL:
http://www.ine.es/ss/Satellite?c=Page&p=1254735849170&pagename=Ayuda%2FINELayout&cid=1254735849170&L=1#.

If computable, the indicators can be calculated from 1993-2012, with the exception of the R&D measures (from 1998), entry and exit (1999-2013), and survival at different lifetimes (2004-11). Among the databases mentioned, the Industrial Economic Survey reports information for the manufacturing sector only[45]. Industry is identified by a NACE Rev. 1.1 code (switch with Rev. 2 is in 2009).

Labour productivity – All the labour productivity indicators are computable as is unit labour cost. For indicators I_008_01 and I_009_09, it is not possible to define the computability given that it is not clear how to recover reliable information on the multinational status of a firm.

45. The Industrial Economic Survey includes from 2008 all firms with more than 50 employees and a stratified sample of firms with fewer than 50 employees. From 1993 to 2007, the database includes all firms with more than 20 employees and a stratified sample of firms with fewer than 20 employees.

TFP – Like labour productivity, all the TFP indicators and relative decompositions are computable. For indicators I_003_08 and I_003_09, it is not possible to define the computability given that it is not clear how to recover reliable information on the multinational status of a firm.

Firm dynamics – All the firm dynamics indicators are computable. However, the computability of I_=50_04 (the average firm size relative to entry, by age) cannot be defined, because there is no reliable information on year of a firm's creation.

Internationalisation – Most of the internationalisation indices are computable from 1993. However indicators that require information on exported quantity, number of products exported and destination markets (I_043_01, I_043_02, and I_040_1) cannot be computed because such data has not been mapped.

R&D and other activities – R&D indicators are computable, as well as asset tangibility. For the other indicators, the computability has not been reported given that the availability of the underlying data is still not properly mapped.

Accessibility

In the case of the Industrial Economics Survey, only other statistical institutions (Statistical Institutes of Autonomous Communities) are provided with micro-data files. As for the CIS and the Pitec databases, it is possible to access firm-level data anonymised on the INE website through a specific procedure. Researchers must submit a request by filling out the required fields in the tab 'Solicitud de descarga de BBDD'. Once the request has been evaluated and approved, the researcher will receive within 72 hours an email providing a username and password, valid for three months. Except for anonymisation of a set of variables, the files available on the website correspond with the original files.

Sweden

The databases we consider for Sweden are the Structural Business Statistics (SBS), the International Trade Survey[46], R&D Survey and the Business Register. All the databases are collected by Statistics Sweden (SCB). Firms are classified according to NACE classifications; the revisions 1 and 2 of NACE classification are both reported in

46. International trade statistics changed when Sweden joined the EU.

the transition period 2006-10. Firms' location has not been mapped. However, if location of firms is available, according to SCB this information is difficult to use because plants might for instance report the addresses of their head offices.

Firm-level data can be merged through a firm ID, although in case of sample surveys, overlaps can be smaller than original surveys. All the indices are highly computable.

Labour productivity – All the labour productivity indices are highly computable, as well as unit labour cost. Almost all the indices are computable from 1980, while indices by trade status are computable from 1995 (eg I_001_06 and I_001_07).

TFP – All the TFP indices are highly computable. Similar to labour productivity, TFP indices are computable from 1980 with SBS, while TFP indicators by trade status are computable from 1995 using the international trade surveys.

Firm dynamics – All the competitiveness indices on firm dynamics are computable.

Internationalisation – All the measures on trade activity are computable. Data is available from 1995.

R&D and other activities – Indicators on R&D expenditure (I_023_04, I_023_05) are reported in R&D surveys; tangible assets (I_059_03) and export unit value (I_070_01) are computable too. Ownership data has been collected since 1980 (I_041_03, I_042_03).

Accessibility

All firm-level data is restricted but data can be accessed by European researchers via remote access, conditional on a confidentiality check and an administrative charge.

United Kingdom

The databases considered for the United Kingdom are the Annual Respondent Database (ARD), the Annual Inquiry into Direct Investment in the UK (AFDI), the Business Enterprise Research & Development (BERD) database and trade statistics from HM Revenue and Customs (HMRC).

The first three databases are collected by the Office for National Statistics, but the first two are available through UK Data Service (UKDS). The ARD can be merged with AFDI

and BERD using the IDBR code[47]. The database resulting from the merging of ARD, BERD and AFDI classifies firms according to SIC industrial classification.

With the exception of export and import status, trade data can be retrieved from trade statistics at HMRC, which is custom data on firms' trade activities. Import and export declarations from and to countries outside the EU are available from 1996-2012, while trade with EU countries is available only from 2008 to 2012. Firms are classified according to SITC2 and HS4 classification (in addition CN8 nomenclature is reported). In principle, HMRC data can be merged with external sources (such as ARD). However, it is necessary to describe the data that a researcher would like to obtain and the HMRC Datalab Team will consider each dataset on a case by case basis[48].

Labour Productivity – All the labour productivity indicators and the unit labour cost are computable from 1995.

TFP – All the TFP indicators and the relative measure of decomposition are computable from 1995.

Firm Dynamics – All the indicators on firm dynamics are computable from 1995. The exit rate (I_052_03) is not computable given that data on firms' deaths is not available in the mapped databases.

Internationalisation – All the internationalisation indices are computable. However, it is important to underline some critical aspects. At first, the indices for the extensive margin of trade (both imports and exports) are computable from 1995 because the ARD database reports all the necessary information. According to the mapped databases, the other indicators (ie the intensive margins) are constructed with HMRC data. This implies that foreign trade data within the EU is available from 2007 while trade data outside EU is available from 1996. Then we made the choice to define the computability of these indices not perfect (in yellow).

R&D and other activities – All the R&D indicators are computable, as well as indicators on multinational status and ownership. Unit values can be calculated with HMRC data. The caveats of internationalisation indices apply also to unit value index.

47. Inter-departmental Business Register (IDBR).
48. Key identifiers have been removed from the HMRC Datalab datasets as part of the anonymisation process so matching will have to be undertaken by HMRC. http://www.hmrc.gov.uk/datalab/data.htm#6.

Accessibility

All the sources are available via the submission of a research project to the appropriate institution (UKDS, ONS, and HMRC Datalab). In addition, the HMRC Datalab requires a short training course, which includes legal issues as well as statistical disclosure control of output. At the moment the Datalab is only open to UK-based institutions and by law HMRC is only allowed to share the data if it serves one of HMRC's functions. Data is available only on-site.

2.3.3 Concluding remarks

The picture is remarkably different in each country when we analyse the computability and the availability of a set of competitiveness indexes that can be calculated through a *bottom-up* approach (ie using firm-level data). Table 2.4 provides a synthetic overview of the computability and accessibility for selected bottom-up indicators, which we use to provide a summary of our main findings.

First, the degree of computability is rather good for a wide span of indicators for many countries. In particular, Table 2.4 (left panel) shows that in Belgium, Denmark, Estonia, Finland, France, Hungary, Ireland, Slovenia, Spain, Sweden and the UK, most of the selected indicators are computable for a relatively large number of years. However, computability is relatively low across the board in Croatia, the Czech Republic, Malta, Portugal and Romania.

Second, indicators for labour productivity, TFP and international activities have the highest degree of computability, given that they require the use of basic items from balance sheet/business register data and trade statistics, respectively. It seems more problematic to merge information from the balance sheet/business register with a foreign-ownership flag, so that productivity for affiliates of foreign multinationals cannot be computed for Croatia, Denmark, Germany, Hungary, Latvia, Poland and Portugal. Indicators of firm-level estimates of quality, which require information on both value and quantity of exports by firm, are also not (or are poorly) computable for a relatively high number of countries. Finally, for indicators of firm dynamics, it turns out that computability is better for entry rates than for exit rates.

The mapping of computability of bottom-up indicators suggests that if scholars or policymakers need to define a competitiveness indicator through a *bottom-up* approach, they might face three main situations:

Table 2.4: Compatibility and accessibility by country for selected bottom-up indicators

	Computability											Accessibility										
	I_001_04	I_003_03	I_001_06	I_001_09	I_051_03	I_052_03	I_046_01	I_047_01	I_023_04	I_041_03	I_070_01	I_001_04	I_003_03	I_001_06	I_001_09	I_051_03	I_052_03	I_046_01	I_047_01	I_023_04	I_041_03	I_070_01
	Micro-aggregated labour productivity – all firms	Micro-aggregated TFP – all firms	Micro-aggregated labour productivity – exporters	Micro-aggregated labour productivity – foreign multinationals	Entry rate (birth rate)	Exit rate (death rate)	% of exporting firms in total no. firms (extensive margin)	Export sales as a % of total turnover (intensive margin)	R&D Expenditure – mean	Share of foreign owned firms in total firms	Firm-level estimates of quality	Micro-aggregated labour productivity – all firms	Micro-aggregated TFP – all firms	Micro-aggregated labour productivity – exporters	Micro-aggregated labour productivity – foreign multinationals	Entry rate (birth rate)	Exit rate (death rate)	% of exporting firms in total no. firms (extensive margin)	Export sales as % of total turnover (intensive margin)	R&D Expenditure – mean	Share of foreign owned firms in total firms	Firm level estimates of quality
Austria	1	1	1	9	9	9	9	1	1	9	9				9	9	9	9			9	9
Belgium	2	2	2	2	9	9	2	2	1	2	2	2	2									
Bulgaria	2	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
Croatia					1	1	1		1	1		9	9	9	9				9	1		9
Czech Rep.	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1
Denmark	2	2	2		2	2	2	2	2	1	2	1	1	1	9	1	1	1	1	1	1	1
Estonia	2	2	1	1	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1
Finland	2	2	2	2	1	1	2	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1
France	2	2	2	1	2	2	2	2	2	1	2	1	1	1	1			1	1	1	1	1
Germany	1	1	1		2	2	2	2	1	1		1	1	1	9	1	1	1	1	1	1	9
Hungary	2	2	2		2	2	2	2	2	2	2	1	1	1	9	1	1	1	1	2	1	1
Ireland	2	1	2	2	2	2	2	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1
Italy	1	1	1	1	1	1	2	2	1	1		1	1	1						1		9
Latvia	1	1	1		1		2	2	2	1	2	1	1	1	9	1	9			1	1	
Lithuania	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
Malta	1		1	1	1	1	1	1	1	1	2	1	9	1		1	1	1	1	1		1
Netherlands	2	2	9	9	2	2	9	9	1	9	1	1	1	9	9	1	1	9	9	1	9	1
Poland	2	2	2		1	1	1	1	2	1	1	1	1	1	9	1	1	1	1	1	1	1
Portugal	1	1	1		1	1	1	1	1		1	1	1	1	9	1	1	1	1	1	9	1
Romania	1	1	1	1	1	1	1	1	1	1												9
Slovakia	2	2	2	1	2	2	1	1	2	1	1											
Slovenia	2	2	2	1	2	1	2	2	2	1	2	1	1	1	1	1	1	1	1	1	1	1
Spain	2	2	2	2	2	2	2	2	2	2	9					9	9	9		2	9	9
Sweden	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1
UK	2	2	2	2	2		2	1	2	2	1	1	1	1	1	1		1	1	1	1	1

1. The data to calculate the indicator is available and the indicator is computable;
2. The data to calculate the indicator is not available and the indicator is not computable;
3. The data to calculate the indicator is available **but** the indicator is **not** computable.

Cases (1) and (2) are straightforward: an indicator is computable (or not) if the underlying data is available (or not). The third case is the most interesting and challenging. We observe that many indicators require matching of different databases. Our assessment is that it is not infrequent that researchers face problems at this point, since some data sources cannot be matched, ie there is not a unique identifier that allows users to combine information from two or more datasets, or there are restrictions limiting the possibility to combine different data sources.

In order to compute a competiveness index (*bottom-up*), it is not only necessary that all the required variables are available. If data stems from different sources it is important to match these sources in a unique database. This procedure is easier if the same institution collects all the databases.

Once it is ascertained that an indicator is computable, the researcher needs to assess if the data is actually accessible to someone who is not affiliated to the institution(s) providing the data. Table 2.4 (right panel) highlights that access to micro-level databases is not an easy task for researchers, because of confidentiality restrictions, rules of access based on the nationality of the researcher (or the institutions to which he/she is affiliated) or based on discretionary choices. In many cases, access is guaranteed to researchers under certain conditions and a submission of a research proposal. Bottom-up indicators are not accessible in Romania, because a safe environment for ensuring secure access to the data is not yet in place. In countries such as Ireland, access is subject to stringent conditions, is possible only on-site, and publication of results is subject to the approval of the National Statistical Institute.

In some countries, such as Austria and Slovakia, it was not possible to ascertain the access rules. In some countries, nationality rules apply. In Belgium, data is collected by the National Bank of Belgium, and is accessible only to NBB members (or affiliated). In Denmark, the procedure for accessing the data is clearly defined, and thus this would qualify Denmark as demonstrating best practice in terms of data accessibility, but access is allowed only to researchers affiliated to Danish institutions. Similarly, in Hungary data is easily accessible only to researchers who have an agreement with the CSO. In the UK, access to HMRC Datalab is open only to UK-based institutions.

The best practices in terms of data accessibility are those in which data can be accessed remotely, with no constraints on affiliations or nationality, and with a clearly formalised procedure that has no (or little) room for discretion over who the data provider can give access. In this perspective, Sweden appears to demonstrate best practice, since data can be accessed remotely, conditional on a confidentiality check and an administrative charge. Similarly, in Finland and France, there is a rather clear procedure to allow access to micro-data to external researchers, also via remote connections. In France, access requests need to be approved by a committee, and this creates some room for discretion. In Slovenia micro-data is accessible for research purposes, but only at SURS. In the Netherlands access to micro-data is also relatively easy, although it was not possible to ascertain if remote access is possible.

Germany has also some of the desirable features, such as the possibility of remote access, but in this case, there is a problem of computability, since data is often provided by different institutions and cannot be merged. In some countries, access to data varies according to the type of data. For example, in Italy, only the data from the surveys can be made available to external researchers, while micro-data with the full population of firms is not accessible. In the Czech Republic, business register data can be accessed relatively easily, while for other types of data, such as custom data and FATS data, conditions are more stringent. Malta allows access to firm-level information for research purposes, except for data on foreign ownership and capital. In Latvia, data is available upon request, except for data on trade by destination and product, which are confidential.

In conclusion, the availability of an indicator depends on different factors that influence computability and accessibility. The computability of an indicator relies on different factors such as the existence of the right data and the possibility to merge data from different sources if necessary. The accessibility of data depends on the rules of access and their clarity. The existence of large datasets is not a sufficient condition to guarantee the availability of an indicator. The best practices we observed rely on data existence, ease of merging data from different sources, and clarity in the rules of access.

3. Bottom-up competitiveness indicators comparable across EU countries: challenges and responses

In chapter 2 we showed that many bottom-up indicators of competitiveness require the matching of data from datasets within the same country, and this affects both the degree of computability and accessibility of indicators for European countries. In this chapter, we take two additional steps. We provide an overview of the most important challenges and actual advancements in matching micro-level data from different sources (section 3.1). This helps in understanding why it can be complicated to compute bottom-up indicators. We then analyse the challenges of building bottom-up indicators that are comparable across countries. We first illustrate how the European Statistical System (ESS) is facing the demand for micro-data comparable across countries (section 3.2), and then present an overview and some concrete examples of matched data and cross-country firm-level surveys and datasets in Europe (section 3.3).

3.1 Data matching: background, terminology and challenges

Linking data from different sources has become increasingly popular. For a long time, linking was restricted to aggregate data based on common or harmonised concepts, but now links are increasingly being made between data from different sources and institutional contexts (eg administrative data) with diverging underlying concepts and, more importantly, including micro-level data from, for example, firms, individuals or transactions. These general developments have been accompanied and driven by a changing political, legal and technological framework in many European countries, which has gradually improved the accessibility of previously restricted datasets and which has made it technically feasible to work with these datasets.

Linking data from different sources containing different records or including information on different subjects and issues is interesting for policy-oriented, comparative scientific research for several reasons (see, for instance, Borgman, 2010; Christen, 2012; Herzog *et al*, 2010; Winkler, 2006):

- More complex research questions can be addressed: for example, linking data on employers with data on their employees might permit conclusions to be drawn about the role of certain groups of employees, or about employment stability, for the productivity of firms (see Bender *et al*, 2008). On a more general level, the integration of data from administrative sources (register data) and survey data might significantly widen the scope and depth of potential analyses (see Bakker, 2010). Furthermore, longitudinal analysis might be made possible or facilitated.

- Accuracy, reliability, and quality of existing data can be improved by cross-checking, monitoring and validating information from different sources. Moreover, missing information in one dataset might be completed by using information from another dataset. There is also the potential to address and understand the reasons for survey non-response, and to identify and treat measurement and representation errors in register data (Bakker, 2010).

- The burden on respondents, the bureaucratic effort and the overall costs of data collection and analysis can be vastly reduced without compromising quality, and the hidden potential of administrative data can be leveraged.

However, there are also a number of challenges and limitations. These involve technical aspects such as data quality *within* existing datasets and diverging data quality *between* datasets. Data harmonisation is an important issue in this respect. The major obstacles to free matching of data are often legal restrictions or ethical issues preventing the linking of data. Privacy and non-disclosure are pivotal issues in this respect. However, against the background of the increasing availability of micro-level data, computer science and research in social science have developed a series of techniques and workarounds that are able simultaneously to leverage the potential of matched data and to guarantee the preservation of privacy.

3.1.1 What is data matching?

A series of concepts and definitions exists on to the matching of datasets from different sources. Data linkage or record linkage denotes *"simply the bringing together of information from two records that are believed to relate to the same entity"* (Herzog *et*

al, 2010, p. 1), such as the linking of information on addresses from a mailing list with information on phone numbers from a telephone directory, or information on firms' employment figures from labour statistics with information on the firms' balance sheets. The terms data matching or statistical matching are used to refer *"to a series of methods whose objective is the integration of two (or more) data sources referring to the same target population. The data sources are characterised by the fact they all share a subset of variables (common variables) and, at the same time, each source observes distinctly other subsets of variables. Moreover, there is a negligible chance that data in different sources observe the same units (disjoint sets of units)"* (Zio, 2012)[49].

Linking data from different sources is not a new idea. Theoretical contributions and early applications of data matching and record linkage techniques date back to the 1940s and they can be observed in large-scale census collections and in the health sector. Newcombe *et al* (1959), for instance, relate differentials in family fertility to hereditary diseases by linking data from health records and a register of handicapped children to birth and marriage records. Subsequent developments were affected quite substantially by the upcoming discipline of computer science, with a special focus on technical and methodological questions (eg Fellegi and Sunter, 1969). In recent years, there has been a continuous convergence between statistics and computer science in this respect.

Important factors facilitating and supporting these recent developments at the interface of data matching, statistics and social sciences are (1) the rapid and exponential advancements in information technology, particularly with respect to hardware capacity (processors, memory, storage); (2) the continuous 'discovery' and opening up of data and data repositories, particularly at official data providers like the Statistical Offices, or as 'Big Data', and their activation for scientific research, and (3) the development of techniques and methodologies enabling access to and the processing of confidential data without violating privacy and nondisclosure aspects related to the data (see, for instance, Schiller and Welpton, 2013).

3.1.2 Data quality as the basic precondition for data matching

Data quality is a crucial determinant of any effort to link data from different sources, because it defines the credentials which define the potential and the limits of matching datasets. If the quality of a dataset is poor with regard to potential identifiers, matching

49. For a more detailed discussion of these terminological issues see, for instance, Christen (2012).

can be hampered or even precluded; it is likely that also the quality of a matched dataset based on this data will be poor, although the process of matching can improve data quality in several aspects: *"If data would be of perfect quality, then data matching could be accomplished through straightforward database join operations* [deterministic matching] *and no sophisticated indexing techniques or approximate comparison functions would be needed"* (Christen 2012, p. 40). In some cases, matching of data is also used in order to improve, complement or cross-check the content of data of poor quality on a specific subject.

Data quality is a complex and multi-dimensional concept and it is described by several criteria (see Christen 2012, p. 39f; Eurostat, 2003; UNECE, 2007), the most important of which are:

- *Accuracy, integrity and reliability*: What is the origin of the data? By whom have they been collected, surveyed, compiled and/or changed? What are the framework conditions of the data collection and compilation? Are there any commercial interests involved? Is the information contained in the data believable?

- *Completeness*: This aspect concerns both records and the attributes of records (variables). How many missing values are there in the data? Why are values or attributes missing? Are there any thresholds with regard to the coverage of statistical units?

- *Consistency, coherence and comparability*: The issue is relevant both within and between datasets used for matching. Have there been changes in the coding of attributes over time? Are there duplicate records in the database? Is an original database (to be matched) grounded in different sources? Are the data or published results from the data comparable to similar data? Are the concepts comparable to other datasets?

- *Timeliness and punctuality*: At what exact point in time was the data recorded? How great is the time lag between reference point and clearance of data? How old is the data?

- *Relevance and interpretability*: Are the data and the issues covered relevant to economic analysis? Are the contents of the databases meaningful and can they be used in a reasonable way?

- *Accessibility:* Are there any restrictions on access to the data, eg for certain user groups or for specific segments of the data? Do distinct regulations on data access exist? In what respect is the data sensitive to non-disclosure?

- *Clarity and documentation:* Is precise and accessible documentation of the data available? Are metadata available in a standardised format (eg SDMX, ESMS; see SDMX, 2008; European Commission, 2009a)? Are test data or scientific use files (SUF) available?

Several factors have an impact on the quality of data. The following are of particular relevance with regard to matched data (Christen, 2012):

- *Origin of data from multiple sources*: if data originate from different organisations with different backgrounds (eg different disciplines), this will affect consistency of databases and has to be handled with caution.

- *Subjective judgement of data production*: not all potentially relevant aspects are recorded in the data to be matched, which might hamper the matching potential.

- Data matching is a process consuming *time, money and computing resources*. Particularly the latter have become much more easily available and tools have become more and more powerful. But as many datasets grow simultaneously (eg Big Data), more resources and novel techniques are always needed.

- In linking data from different sources, a *trade-off between security and accessibility* is frequently needed.

- The inherent technical features of datasets are an important factor affecting consistency of data from different sources. This refers both to the coding of data and to data representations (eg relational databases).

- *Input rules might be restrictive and/or bypassed,* which might hamper data quality. For example, in a register survey of firms, there might be a complex system of allocating the firm to an industry sector. Thus, many respondents might revert to a simple solution and fill in, for example, simply 'manufacturing' instead of 'manufacturing of chemical products'.

- Last but not least, both *data needs and the technical systems for data collection and storage change* over time. This might cause changes in the structure and

contents of datasets, with certain attributes disappearing and new ones being added to the data.

In summary, linking data from different sources has plenty of potential, but the quality of the original data influences the quality and the validity of the resulting matched data. Thus, data harmonisation, which is described in the next section, is an important feature of data matching and matchability.

3.1.3 Harmonisation of data

Harmonisation of existing data on different levels of aggregation is part of the technical process of data matching, and is also a potential avenue towards the creation of comparable cross-country data necessary for cross-country research[50]. The general objective of data harmonisation is to improve data quality and to make the datasets to be merged more comparable with respect to their central characteristics/variables (Granda and Blasczyk, 2010).

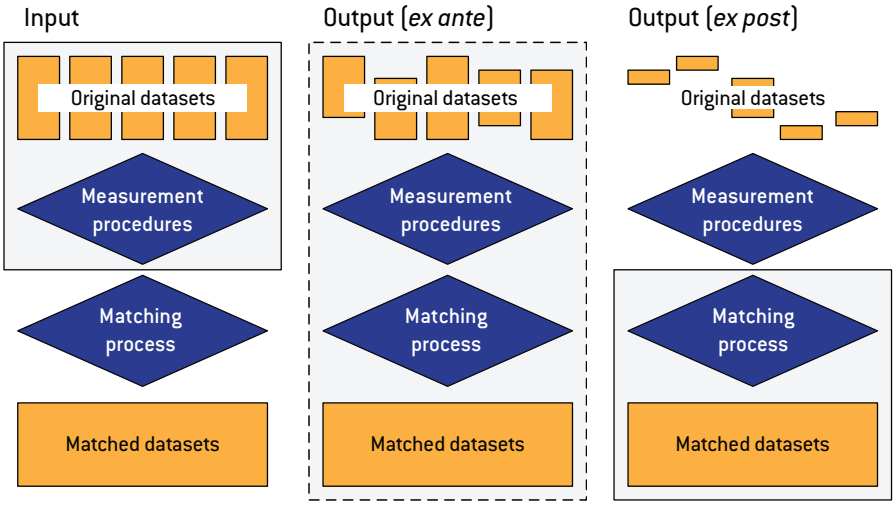
Data harmonisation itself offers several benefits. It provides a common basis for standardised data, it decreases data redundancy and costs of data exchange, and it ensures data compatibility and comparison (TID, 2012). Generally, harmonisation and standardisation of datasets can be performed at different stages of the matching process, with the two main forms of harmonisation being input harmonisation and output harmonisation (CHINTEX, 2001; Kallas and Linardis, 2008; Burkhauser and Lillard, 2005; see Figure 3.1)[51].

The basic characteristic of input harmonisation is that standardisation starts before any process of matching, ie the inputs of the matching process are harmonised right from the beginning. Input harmonisation thus *"aims to achieve standardised measurement processes and methods in all national or regional populations. Comparability is realised through standardisation of definitions, indicators, classifications and technical requirements"* (Granda and Blasczyk, 2010, p. 1). Input harmonisation is always *ex-ante* harmonisation, while *ex-ante* harmonisation is

50. For a more detailed discussion of these terminological issues see, for instance, Christen (2012).
51. Another way would be the creation of new cross-country data from scratch, eg through new cross-national surveys. There are several examples of such datasets that have been created during recent decades. Most, however, take into account the micro-level of individuals, but firms or establishments have been rather neglected. Notable exceptions are the Community Innovation Survey (CIS), the EFIGE Survey, or the Continual Vocational Training Survey (CVTS). For a critical overview, see Burkhauser and Lillard (2005).

implemented before data are surveyed or compiled, and *ex-post* harmonisation refers to already existing records (Kallas and Linardis, 2008).

Figure 3.1: Input and output harmonisation



Source: CHINTEX, 2001, p. 3, modified.

Output harmonisation, on the other hand, is characterised by a standardisation process starting only with or even after the matching of the original data. Output harmonisation *"uses different national or regional measurements possibly derived from non-standardised measurement processes. These measurements are 'mapped' into a unified measurement scheme. Thus, only the statistical outputs are specified, leaving it to the individual countries/regions to decide how to collect and process the data necessary to achieve the desired outputs"* (Granda and Blasczyk, 2010, p. 1). Some authors (eg CHINTEX, 2001; Kallas and Linardis, 2008) further divide output harmonisation into *ex-ante* and *ex-post* output harmonisation.

In practice, input harmonisation is mostly applied in cross-country surveys based on standardised measures, such as the Community Innovation Survey or the EFIGE firm-level survey. Most other applications work with output harmonisation, either of the *ex-ante* or of the *ex-post* form, whereas various interim forms exist.

In the context of an entire matching process, as discussed earlier, data harmonisation is a part of the pre-processing of single databases only in the case of output harmonisation. In the case of input harmonisation, the databases to be linked would

already be harmonised, at least in many respects. Then, harmonisation would be shifted to the earlier conceptual steps, for example of survey design. When more than two databases are to be matched, information from the matching between some of them can be used for the further matching process (Christen, 2012).

As perfect harmonisation is rarely possible, particularly in frameworks based on output harmonisation, an adjustment for measurement errors has to be taken into account.

Harmonisation of data may concern several issues and elements of the data to be harmonised (see, for instance, ESSnet-ISAD, 2007, p. 42):

- Statistical units,
- Reference periods,
- Populations (coverage),
- Variables (in case of differences in definition),
- Classifications,
- Metadata.

For many of these issues, international standards already exist, for example for classifications of industries or products. Concerning metadata, the SDMX framework defines standards for the international exchange of metadata and is applied by several international organisations such as Eurostat, the World Bank and the OECD (SDMX, 2009; Vale, 2009, p. 28). Particularly for Europe, the European Commission has set up a recommendation 'on reference metadata for the European Statistical System' (the ESMS, see European Commission, 2009a), which refers to the European Statistics Code of Practice (Eurostat, 2011) and is based on the SDMX framework.

The limits of data harmonisation are mainly defined by national institutional frameworks or by existing technical rules and standards, which are generally hard to overcome. In particular, fundamental concepts of statistical units such as firms, establishments or employees are often defined slightly differently in different countries (see Broersma *et al*, 2010, for an example of employer-employee data in the Netherlands and in Germany).

3.1.4 Privacy and non-disclosure

An important issue for the analysis of micro-level data is privacy and confidentiality of information on single statistical units, particularly individuals, households, enterprises or administrations (see UNECE, 2007a, for an overview). The legal conditions on non-

disclosure are generally a national matter and they differ widely between European countries, although some harmonisation efforts have been pursued already, for example the European Commission Regulation (EC) No 831/2002 on Community Statistics, concerning access to confidential data for scientific purposes (see European Commission, 2002) or more recently the European Statistics Code of Practice (Eurostat, 2011).

Regulation 831/2002 applies to access to a series of Pan-European micro-level datasets, for which it sets out procedures for access to confidential data (see Santos and Museux, 2005)[52]. Beyond the datasets covered by this regulation and its amendment in Regulation 1000/2007, access to micro-level data on a European level is theoretically granted, but in practice it is rather restricted, as stated in the European Statistics Code of Practice (Eurostat, 2011, p. 8): *"Access to micro-data is allowed for research purposes and is subject to specific rules or protocols."*

With regard to the scientific analysis of micro-level data, there is a trade-off between the perception of privacy and the risk of identification of sensitive information (such as on individuals' health complaints or on firms' business strategies), and the interest in and need for scientific research (Santos and Museux, 2005). Matching data from different sources might create additional challenges for privacy protection, as the quality and the quantity of information on single observations (ie individuals or firms) generally increase when linking data from different sources.

Many data-holding institutions in European countries (and worldwide) have introduced techniques allowing for the analysis of micro-level data without violating rules of nondisclosure, thus guaranteeing the confidentiality of the respective data. Some of these techniques will be discussed in the next section.

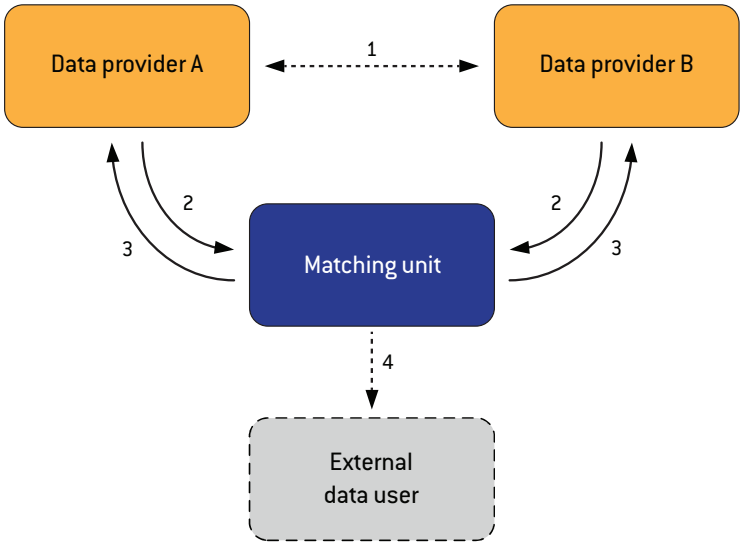
3.1.5 Potential solutions and workarounds for data and matching restrictions

One approach to overcome at least some of the challenges of matching processes and of matched datasets are so-called matching architectures. These techniques are primarily intended to prevent misuse of data to be matched. For example, databases to be matched can be sent to a trusted matching institution before being sent to

52. These datasets are the European Community Household Panel (ECHP), the Labour Force Survey (LFS), the Community Innovation Survey (CIS) and the Continuing Vocational Training Survey (CVTS). More recently, Regulation 831/2002 was amended by Commission Regulation 1000/2007 which includes further datasets, namely the Structure of Earnings Survey (SES), the European Union Statistics on Income and Living Conditions (EU-SILC) and the Adult Education Survey (AES).

researchers for analysis (see Figure 3.2). The matching unit then only matches the identifiers, whereas researchers later do not get the identifiers but only the contents of the matched data (for an example, see Brook *et al*, 2008).

Figure 3.2: A simple architecture for matching of confidential data ('three-party protocol')



Source: Based on Christen (2012), p. 193, modified.

As the involvement of the third party (the matching unit) causes some disclosure and security risks (eg collusion of the data provider with the matching unit), the process can also be performed without a matching unit, and the data providers can communicate directly with each other.

Confidentiality issues can also be addressed at the level of data access. As many micro-level datasets contain sensitive information, for example with regard to individuals' or firms' characteristics, which can be directly linked to the respective firms or individuals, issues of privacy and non-disclosure are pertinent. Most often, there are country-specific legal restrictions governing the non-disclosure of the data. Without accessing micro-level data directly, however, a reasonable analysis of the data is often not possible. Therefore, several solutions for researchers to get access to original or slightly anonymised data without the risk of de-anonymisation have been developed in recent years (DWB, 2012).

Generally, these solutions range along a continuum from 'no access at all' to 'restricted access' and 'full access'. Whereas the first and the last alternatives are irrelevant in the present context, various alternatives have been developed with regard to the provision of partial or restricted access to micro-level data.

Restrictions (and thus, the necessary non-disclosure and confidentiality of data) can be either realised by limiting the data to a restricted sample (eg a Scientific Use File), through the anonymisation of sensitive parts of the data (eg identifiers, addresses, names), or by restricting access to these sensitive attributes of records. In this context, data providers have developed a series of techniques to regulate access to micro-level data. One way is through on-site access to the original data: the researcher has to visit a physical data storage environment (safe centre) in which the legal and technical aspects of confidentiality can be taken into account (DWB, 2012; Brandt, 2012).

Another solution applied by several national statistical institutes and data archives is the concept of remote access. The researcher sends the syntax of his programme for data analysis[53] to the data provider, which runs the programme on the basis of the real data. Ultimately, the researcher has only access to the results (which are, moreover, checked for potential disclosure and privacy issues) and does not see the micro-level data itself (DWB, 2012).

Some institutions are able to provide a more advanced remote access, allowing the data user to access the (anonymised) data from anywhere without being able to access sensitive characteristics. This is the case in the Netherlands and Sweden, for instance, and is being assessed by a project in Germany, the Morpheus Project (see Höhne and Höninger, 2013). This project analyses an anonymised dataset stored on a server located at a statistical institute (it is not possible to download the data). After running the programmes, researchers receive the results of their analysis as well as a corresponding quality assessment, which allows for an evaluation of the validity of the results.

To improve access to different micro-datasets, Eurostat has launched some projects with international partners: the 'Decentralised Access to EU Micro-data Sets' project (completed 31 January 2010) and the 'Decentralised and Remote Access to confidential data in the ESS' (DARA) project (Brandt, 2012).

53. Most data holders also provide some type of dummy data which simulates most of the characteristics of the real data and which helps the researcher to prepare operative programmes.

Schiller and Welpton (2013) present a solution for the current European Union Remote Access Network (EU-RAN), established by the Data Without Borders project. This project plans to allow access to detailed confidential data from around the EU to researchers from within their own country of residence, which would eliminate travel time and costs. Their proposal builds on five general principles (Schiller and Welpton, 2013):

- Access must be distributed;
- Access should come from a single point;
- Access must be secure;
- Access must be compatible;
- Researchers must be able to work collaboratively.

To put it simply, the solution from Schiller and Welpton (2013) uses a remote access which only requires simple VPN (virtual private network) software[54]. Figure 3.3 illustrates the principle of EU-RAN. Data providers (usually from different member states) make data available, which always remains within the institutions or at least within the country of origin in order to comply with national legal requirements. On the reverse side, researchers or other users have (restricted) access to the data via secure connections from either anywhere, at the data providing institution itself, or within a specifically equipped safe centre.

The fact that researchers have access to the data does not necessarily imply that they can download the data. Therefore it is necessary to provide a 'virtual working environment' which includes analytical software and applications that allow results to be generated, prepared and presented. The purpose of the information platform with metadata is the provision of information and a general support.

One possible option for the future is the MiCoCe (micro-data computation centre) concept, whereby only small parts of the data are moved into the working memory of the MiCoCe, and are later deleted. Secure connection systems are used (see Schiller, 2013).

54. This system provides a secure encrypted connection between the user and the server with the data, as widely used for financial or military services.

3.1.6 The distributed micro-data approach

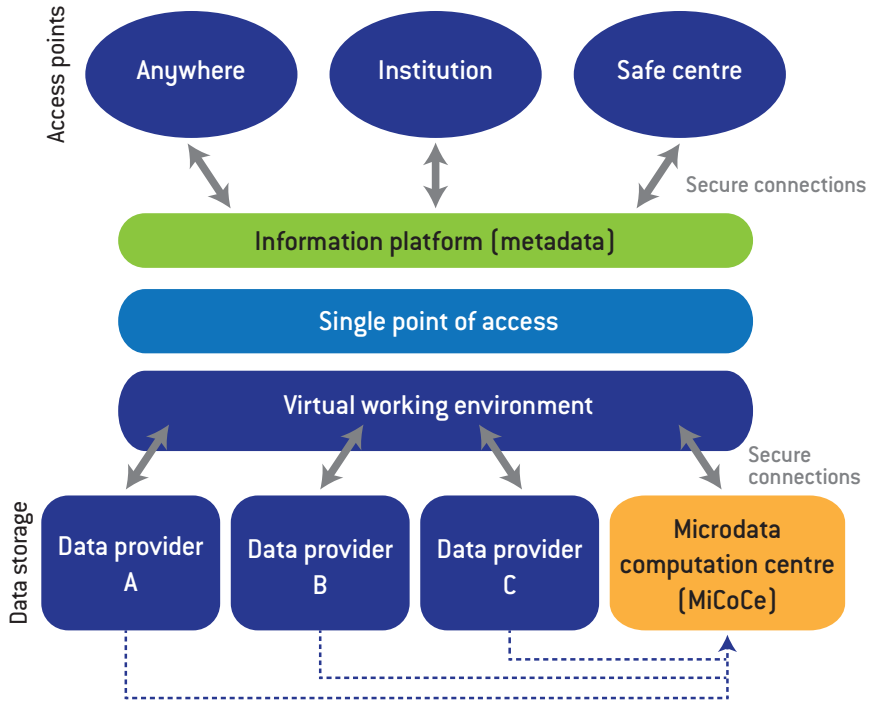
An example of accessing micro-level data in a cross-country perspective is the distributed micro-data approach, which was introduced mainly by Eric Bartelsman, John Haltiwanger and Stefano Scarpetta about a decade ago (see Bartelsman *et al*, 2004, 2005, 2009, 2009a; and Bartelsman and Hamilton, 2004). This approach is mainly based on an *ex-post* output harmonisation. The main motivation for the underlying procedure is that although micro-level data exists in many countries and is even accessible within and comparable across countries, it cannot be combined in one location.

The basic principle of the approach is to analyse the national micro-level data separately, but on the basis of a common and harmonised methodology. The comparative analysis is then based on a joint evaluation of the results (eg indicators, tables) generated on the basis of the separate micro-level datasets. Three main stakeholders are involved in this approach (see Bartelsman and Hamilton, 2004):

(1) The data providers, which might be national statistical offices (NSOs) or other institutions holding micro-level data, such as labour market agencies or any other institution holding sensitive data.
(2) A research data centre, which is the 'broker of information' between data providers and data users. This centre might, for instance, be in charge of collecting and publishing metadata, of controlling the output with regard to nondisclosure or of mediating mutual requests between the two parties. The centre may be a regular (national) research data centre at one of the participating NSOs or it might be a new institution, created explicitly for a comparative project.
(3) The data users or researchers, who conceptualise, design and conduct analyses using the data from the respective data providers. The researchers themselves, however, do not see the original data, but only the produced results, eg cross-country tables.

Bartelsman and other researchers apply this procedure to data on various subjects. Bartelsman *et al* (2008, 2009), for instance, address the problem of comparative analyses of firms' productivity in different countries. Another subject addressed on the basis of the distributed micro-data approach is firm dynamics. Bartelsman *et al* (2004, 2005, 2009a), and Koch (2008) and Vale (2006), address and create harmonised concepts that allow for a general analysis of micro-level data on firm entry and exit. Haltiwanger *et al* (2008) and Broersma *et al* (2010) address the question of job flows on the basis of a comparative harmonised analysis of micro-level data from

Figure 3.3: The EU-RAN remote data access architecture



Source: Schiller and Welpton (2013).

two or more countries. All these issues are of great relevance for the analysis of competitiveness.

Examples of recent and ongoing projects making use of the distributed micro-data approach are CompNet (see ECB, 2013) and EU KLEMS (O'Mahony *et al*, 2008; O'Mahony and Timmer, 2009). Within Work Package 10 of EU KLEMS[55], a series of economic indicators, particularly relating to productivity, have been assembled from micro-level data from different European countries.

In the light of the still remaining severe restrictions on the accessibility of micro-level data, particularly when it comes to cross-country perspectives, the distributed micro-

55. Within the FP6-funded EU KLEMS project, both aggregate and micro-level data on various economic topics have been collected and analysed using a cross-country comparative approach. For further information, see www.euklems.net.

data approach seems to be an adequate instrument for working around these restrictions. Although it is certainly not equal to analysing matched micro-level data from different countries and/or sources, it enables researchers to take into account heterogeneity both within populations (eg of firms, individuals) and between populations (eg countries, regions, sectors).

3.2 The European Statistical System (ESS) and the challenging demand for micro-data

3.2.1 The origins of the ESS

Historically, data collection and analysis were national issues. Statistical institutes were set up, collected data and later also managed data collected by other authorities, such as customs or tax authorities. This process was first altered when the need for harmonised European data arose and Eurostat became prominent. In this second stage, data collection remained in national hands, but aggregates were supplied to Eurostat. We are now in a new stage, when data collection and international access to survey and administrative data becomes ever more important.

Before the emergence of the current European Statistical System (ESS) there were considerable differences between member states, both in terms of the methods and concepts used and in the quality of the statistics produced. National statistical

BOX 3.1: A BRIEF HISTORY OF THE DEVELOPMENT OF THE SYSTEM OF EUROPEAN STATISTICS

The foundation of the European Coal and Steel Community (ECSC) in the early 1950s is considered the beginning of the need for harmonised and comparable European statistics. The 1957 Rome Treaty on the European Economic Community (EEC) subsequently marked the birth of European legislation on statistics. However, the approach to statistical methods remained primarily based on the goodwill and cooperation of the NSOs during the first decades of the EEC (for a short summary, see European Commission, 2009). In the 1990s, common rules on the transmission, production and unification of data were mainly set out in Council Regulations No. 1588/90, No. 322/97 and by the Commission Decision 97/281/EC. At the same time, European policy became directly based on statistics, with the most noteworthy example being the convergence criteria for European Monetary Union. This development is closely related to the more general expansion of the statistical legislation.

institutes (NSIs) were supervised by their governments and were free to decide objectives and methods to produce a variety of statistics. The harmonisation of statistics has been implemented (and is still far from complete) gradually in parallel with the enlargement of the European Union.

NSIs now collect, edit and store micro-data from several sources to meet national needs and EU requirements. While they have to provide detailed, quality statistics to researchers and policymakers, they are also obliged to protect the confidentiality of the information. Traditionally, NSIs publish aggregate information at the macro or sector level, and currently most of the information transmitted to Eurostat is in the form of aggregate numbers, or simple frequency or magnitude tables. As a consequence, data protection methods for aggregate, tabular data are well established in all EU member states (Hundepool *et al*, 2010). However, in recent years, the demand for micro-data for research purposes gradually increased, setting new challenges for data protection.

The provision of statistics to Eurostat by the NSIs is a cost-effective solution for Eurostat, but it puts a heavy burden on NSIs (Sverdrup, 2005). Balancing the available resources between the needs of Eurostat and national providers is often problematic because of the increasing demand for detailed, quality statistics at the EU level. All NSIs dedicate a substantial part of their resources to meet the EU requirements. This is especially true in small countries, where NSIs work mostly to serve the needs of the EU.

Hence, we are at a new stage of data collection, which has been also induced by the widespread use of micro-data and proposals from economists on how firm-level data should be used to compare competitiveness, labour markets and other economic features in different countries. Ideally, in a European research area, scientists can access data from all countries, datasets will be matched while preserving confidentiality and micro-data based measures will be created in a unified form to obtain comparable measures.

Data harmonisation methods build on principles established by other international organisations – especially the United Nations and the Organisation for Economic Cooperation and Development – but an important difference is that while the standards set by other international organisations are generally authoritative but not obligatory, the EU can impose legal obligations on member states (see Shearing, 2013), though the EU system remains decentralised, with Eurostat in a coordinating role. This decentralised structure is a plausible solution, since the system must be able to incorporate national statistical systems which developed independently and have

different organisational structures (Grünewald, 2001), but it has major disadvantages.

The European Statistical System (ESS) is a partnership between Eurostat and the NSIs and authorities in the member states (and a few other countries) responsible for the compilation of European statistics. The ESS Network ensures the availability of reliable and comparable European statistics for all member states. The basic principles and rules on how the ESS should function are established in the statistical law of the European Union[56], which came into force in 2009 and, at the time of writing, was being revised and amended[57]. This framework regulation provides the legal framework for the development, production and dissemination of European statistics, but neither specifies the types of statistics produced nor the concepts and methods used. Details of the production and dissemination of European statistics are covered in sector-specific Eurostat regulations and corresponding guidelines.

Eurostat itself collects mainly aggregate data[58]. National, regional and sector-level statistics are produced separately by the statistical authorities of European countries under Eurostat's supervision. Harmonisation of concepts and methods and the reliability and timeliness of data are guaranteed by formal and informal means[59].

Member states conduct various surveys, including standardised surveys recommended by Eurostat, to meet the need for European indicators. They also make use of

56. European Commission Regulation (EC) No 223/2009.
57. On January 27th 2015, the European Council has released information about an agreement reached with the European Parliament on new rules aimed at ensuring the quality and reliability of EU statistics. The draft regulation aims at strengthening governance of the European statistical system (ESS). The amending regulation requires that heads of NSIs have the sole responsibility for deciding on processes, statistical methods, standards and procedures, and on the content and timing of statistical releases and publications for all European statistics. Similarly, the director general of Eurostat must have the sole responsibility for deciding on processes, statistical methods, and on the content and timing of statistical releases and publications by Eurostat. The amending regulation also reinforces a legal for more extensive use of administrative data sources for the production of European statistics without increasing the burden on respondents, NSIs and other national authorities. According to the proposal, NSIs should coordinate relevant standardisation activities and receive metadata on administrative data extracted for statistical purposes. Free and timely access to administrative records should be granted to NSIs, other national authorities and Eurostat, but only within their own respective public administrative system and to the extent necessary for the development, production and dissemination of European statistics. For more information, see http://www.consilium.europa.eu/en/press/press-releases/2015/01/european-statistics--rules-improve-data-policymakers.
58. Notable exceptions are some micro-level cross-country surveys conducted by Eurostat and implemented in most EU countries as, for instance, the Community Innovation Survey (CIS) or the European Labour Force Survey (EU-LFS).
59. 'The European Statistical System', chapter in Handbook of Methodology for Modern Business Statistics (2014), http://www.cros-portal.eu/sites/default/files//General%20observations-05-T-European%20Statistical%20System%20v1.0_0.pdf.

administrative data, but practices vary widely. Differing practices in the use of micro-data go hand in hand with differences in national legislation governing the treatment of micro-data. As a result, there are several comparability issues for the raw data (see section 3.1). Furthermore, as the main objective is to serve Eurostat at aggregate level, access to micro-data at EU level is not a priority. Consequently, confidentiality and access regulations remain in national hands and vary greatly.

Since the current system is regarded as inflexible and unable to appropriately adapt to changing user needs, there is an intention to move away from the separate production of statistics towards a more integrated system[60]. For instance, the European Commission decided to improve the accessibility, harmonisation and applicability of European statistics (see, for instance, European Commission, 2001, and Lamel, 2002).

An important step towards this goal was the implementation of the European Statistical System Networks of Excellence (ESSnet) addressing the need for synergies, harmonisation and dissemination of best-practice methods within the ESS[61].

Subsequently, ESSnet projects were designated as networks *"of several ESS organisations aimed at providing results that will be beneficial to the whole ESS"* (Eurostat, 2013). One central characteristic of an ESSnet project is the connection of a wide range of expertise throughout the ESS organisations in order to develop specific actions which would benefit the whole European system. Using such a method, it is not necessary that all EU member states participate in every ESSnet project, results of which are shared with the rest of the EU countries (see Table 3.1 for a selection of recent ESSnet projects).

60. 'Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade', COM(2009) 404 final. Recent effort related: ESS VIP programme.
61. The initiative started with the implementation of the Centres and Networks of Excellence (CENEX) in 2005. The first CENEX (pilot) project on Statistical Disclosure Control (SDC) started at the end of 2005, lasted twelve months and involved statistical offices from eight European countries (Hundepool, 2007).

Table 3.1: Selection of ESSnet projects

Name	Organisation	Detail
Admin Data (Use of Administrative Data)	Office for National Statistics, UK	Explores the possibilities of the use of admin data for business statistics.
Consistency	Eurostat	Aims at the *"achievement of a streamlined framework for business-related statistics"*.
DARA (Decentralised and Remote Access to Confidential Data in the ESS)	DESTATIS, Germany	Establishes a secure channel from a safe centre within an NSO to the safe server at Eurostat.
Data Warehouse	Statistics Netherlands	The overall objective is to provide assistance in the development of more integrated databases and data production systems for business statistics in ESS member states.
EGR (EuroGroupRegister)	Statistics Netherlands	Developing an improved EGR business model (version 2.0).
ESSnet on Profiling (Profiling of Large and de la Statistique Complex Multinational Enterprise groups)	Institute National	The Profiling project aims at facilitating the profiling of large and complex multinational enterprises.
ESSLait (ESSnet on Linking of Micro-data)	Statistics Sweden	The general concept is to improve and apply the methodology for data linking and ICT impact analysis that was developed in the ESSLimit and ICT Impacts studies.
GEOSTAT 1B	Statistics Norway	The GEOSTAT action is about developing guidelines for datasets and methods to link census 2010/2011 statistics to a common harmonised grid, building on the network and work made by partners in the European Forum for GeoStatistics.
Global Value Chains	Statistics Denmark	Devises ways on how to use data within the ESS to measure economic globalisation and the internationalisation of businesses.
MEMOBUST	Statistics Netherlands	The main objectives of this project are the identification of best practices and the development of a common methodology and ESS guidelines supporting the production of business statistics.
NET-SILC2	CEPS/ INSTEAD[62]	Aim of Net-SILC2 is to develop a methodology for the analysis of the EU-SILC data.

Source: Bruegel.

62. http://www.ceps.lu/?type=module&id=53.

With regard to these criteria, it is obvious that any ESSnet project has only a supporting character and can never be a stand-alone venture.

3.2.2 The current modernisation of European business and trade statistics

One of the first ESSnet programmes was adopted in December 2008 with a term of five years from 2009-13 and was called 'Modernisation of European Enterprise and Trade Statistics' (MEETS, see European Economic Community, 2008). The aim of MEETS, which included various projects, was the adaptation of business statistics to new needs, including the adjustment of the statistical system to the production of statistics and to the reduction of the burden on enterprises in collecting and providing internal data. MEETS was intended to contribute to the following objectives (European Economic Community, 2014):

- To review priorities and develop indicators for new areas;
- To achieve a streamlined framework for business-related statistics;
- To support the implementation of a more efficient way of producing enterprise and trade statistics;
- To modernise INTRASTAT[63].

To reach these targets, the European Commission spent €42.5 million. MEETS consists of several smaller studies, including different ESSnet projects which directly or indirectly contribute to it (European Commission, 2011a, see also Table 3.1)[64].

In addition to MEETS, Eurostat has started the FRIBS project (Framework Regulation Integrating Business Statistics) which aims to satisfy the need for the integration of global business-related statistics into a single cross-cutting legal framework (European Commission, 2012). The project started in 2011 with a five-year duration. It was launched to meet the objectives of the European Statistical Programme 2013-17 (European Commission, 2011a).

Specifically, the European Commission plans to provide a common infrastructure tool for the production and compilation of business statistics and to define consistent data

63. INTRASTAT is a unique database founded on the EU Regulation No. 3330/91 which regulates the collection of information and the production of statistics on trade in goods between countries of the European Union (European Commission, 1991).
64. In addition to ESSnet projects, a number of external studies conducted by national statistical institutes or external experts have also been commissioned (European Commission, 2011).

requirements and a common data quality framework. This will make the linking and matching of statistics obtained through the regular collection of global business statistics possible, providing greater added value to the collection of information. Therefore, FRIBS tackles several issues (European Commission, 2012; Statistikrat der Bundesanstalt Statistik Österreich, 2013), such as:

- The lack of full methodological consistency in different domains of business statistics;
- The differences in surveys on business statistics and their diverging periodicities across Europe;
- Non-harmonised use of administrative sources in EU countries;
- Improvement in the exchange of micro-data between the member states of the ESS;
- The high burden on enterprises in terms of reporting intra-EU trade statistics; and
- Lack of data linking across business-statistical domains.

Along with the MEETS and FRIBS programmes, the European Commission released several additional recommendations and practice guidance. One of the first initiatives in this respect was the installation of the 'Foreign Affiliates Statistics System' (FATS, see European Economic Community, 2007). This database measures commercial presence in foreign markets through affiliates and therefore describes the overall activity of foreign affiliates residing in a given target country (Eurostat, 2009).

Inward and outward FATS data is available on an annual basis. Although the rules for uniform data collection were established only in 2007, data goes back to 1996[65]. Data collection is done by the statistical offices of the member states and data is then aggregated by Eurostat. This system is also used for many other databases (eg ITSS or ITGS, see below).

Another implementation of a common European database is the Single Market Statistics System (SIMSTAT), started in 2011 and following the previous INTRASTAT database (European Statistical Advisory Committee, 2012). This database is of particular importance because the collection of INTRASTAT data generates around 50 percent of the administrative burden from official statistics (Radermacher, 2013).

SIMSTAT uses principles of modern design for trade statistics, which opens up the possibility of gradually replacing the import survey of, for example, ITSS by a combined

65. Between 1996 and 2006, data was collected on a voluntary basis and thus is not complete in terms of country coverage or uniformity.

dataset. Moreover, it provides opportunities to improve the databases and to simplify the reporting burden for enterprises (European Statistical Advisory Committee, 2012). One of these modern approaches is the linking of data from the 'International Trade in Services Statistics' (ITSS) and the 'International Trade in Goods Statistics' (ITGS) to existing business registers such as the 'Structural Business Statistics' (SBS, see European Statistical Advisory Committee, 2012, and Granner, 2013).

In addition to these simplifications, SIMSTAT shall also provide access to detailed micro-data at the firm level on intra-EU exports (Granner, 2013). The data constituting the ITSS, the ITGS and the SBS systems is collected by member states and later aggregated by Eurostat.

In January 2013, the ESS Committee introduced the ESS.VIP Programme. This programme implements a joint strategy for a more integrated statistical system and a more efficient European database, which was approved by the ESS Committee in 2010 (Museux *et al*, 2013). Its main purpose is the development of a common ESS infrastructure framework with an appropriate legal background and new administrative mechanisms allowing for the sharing of information, services and costs among all ESS partners (European Committee, 2013). The following ESS.VIP programmes were proposed for 2014 (Museux *et al*, 2013; European Committee, 2013):

- ESS.VIP project ESBRs (European System of Interoperable Statistical Business Registers): Their purpose is to obtain better business statistics through the interoperability of consistent business registers. The programme runs until 2017 (Liotti, 2013).
- ESS.VIP component Data Warehouses: Focuses on the improvement of the data and metadata infrastructure. More specifically, solutions are developed covering the reference enterprise data warehouse architecture and to improve the connectivity for member states of their data warehouses to the ESS data warehouse (Museux *et al*, 2013).
- Seasonal Adjustment: Is a re-launch of the former Seasonal Adjustment User Group. Contributes to the harmonisation of business statistics among member states (see http://www.cros-portal.eu/content/seasonaladjustment).
- Free and Open Source Software (FOSS): This project contains several different approaches which aim at improving access to the generated databases. Some of its aspects are shared services, the Data Warehouse, a Communication Network and the European Statistical Data Exchange Network (Museux *et al*, 2013).

Dependent on the performance of these projects and the available budget, the

European Commission plans to launch several other projects (Museux *et al*, 2013, and European Committee, 2013).

To sum up, there is an intention at the EU level to meet the increasing demand for micro-level data for research purposes, but there are many open questions about practical implementation. Despite that fact that collaborative projects provide guidance and assistance to the member states, substantial differences between member states remain. Most countries provide access to confidential micro-data for scientific purposes, but both the set of available databases and the conditions of access vary in different countries.

3.3 Cross-country and matched datasets in Europe – overview and examples

3.3.1 Overview

Table 3.2 gives an overview of examples of cross-country and matched datasets in Europe and beyond. Four types of matched datasets, projects or institutions providing support and access for matched data can be distinguished:

Type 1: Multi-country harmonised micro-data collections
This type of cross-country dataset comprises collections of data from different countries which are compiled on the basis of a harmonised methodology. This is the case with, for example, systematic and regular collections of available data (such as the firm-level data provided by Bureau von Dijk) or with cross-country surveys based on a harmonised methodology and harmonised questionnaires.

Type 2: Micro-aggregated statistics
These are collections of aggregate data (eg on sectoral and/or regional levels) which have been compiled from micro-level data on the basis of a harmonised methodology, mainly distributed micro-data approaches. Examples are the CompNet database or the OECD's DynEmp data.

Type 3: Specific projects dedicated to matching micro-level data
This type of matched micro-level data is based mostly on singular projects with a specific, mostly topical aim. Usually, the resulting datasets can be replicated for the specific purpose of the project, but it cannot be used outside the project because of technical and/or legal restrictions.

Type 4: Coordination actions and collections of meta-data
Type 4 is not about matched cross-country micro-level data itself, but comprises initiatives which have the aim of organising, supporting and/or

facilitating the access and the matching of micro-level data from different countries (sometimes, such initiatives also exist within countries). Examples for such initiatives are the Data without Boundaries (DwB) or the German KombiFiD projects.

In Section 3.3.2 below, illustrative best-practice examples for each of the above four types of matched data/institutions will be described and discussed.

3.3.2 Examples of cross-country (and) matched datasets in Europe

To illustrate the types of recent data matching efforts, we briefly outline five examples. The EFIGE dataset is an example of a multi-country harmonised micro-data collection (Type 1); the dataset being synthesised within the CompNet project is an example for a micro aggregated dataset (Type 2); the project 'Combined firm-level data for Germany' (KombiFiD) serves as an illustration of what has been labelled 'specific projects dedicated to matching micro-level data' (Type 3); and the 'Data without Boundaries' DwB project is an example of a coordination action aiming at facilitating data access in general (Type 4). Finally, the Global Value Chain project is an example of a combination of a multi-country survey (Type 1) and micro-data linking (Type 3).

3.3.2.1 EFIGE

The EFIGE dataset[66] is dataset generated within the EFIGE (European Firms in a Global Economy: internal policies for external competitiveness) project, which was supported by the European Commission's 7th Framework Programme, coordinated by Bruegel and carried out between September 2008 to August 2012 by academic and international institutions and national central banks in Europe[67]. The dataset provides representative and comparable samples of manufacturing firms in seven European countries. It includes about 3,000 firms for each of Germany, France, Italy and Spain, more than 2,200 firms for the United Kingdom, and about 500 firms for each of Austria and Hungary.

The EFIGE survey, for the first time in Europe, included a broad array of questions that allow several crucial issues related to competitiveness to be addressed. The questionnaire generated both qualitative and quantitative data on firms' characteristics and activities, for a total of about 150 variables covering six broad areas:

66. The complete name is 'EU-EFIGE/Bruegel-UniCredit Dataset' (Altomonte and Aquilante, 2012).
67. See http://www.efige.org/ for details of partners.

Table 3.2: Examples of cross-country and matched datasets in Europe and beyond

Name	Provider	Type	Content and aims	Countries	Time span	Update frequency	Availability
Amadeus – European Company Data	Bureau van Dijk	1	Amadeus is a database of comparable financial and business balance sheet information on Europe's biggest 500,000 public and private companies by assets. Amadeus includes standardised annual accounts (consolidated and unconsolidated), financial ratios, sectoral activities and ownership data. The database is suitable for research on competitiveness, economic integration, applied microeconomics, business cycles, economic geography and corporate finance.	43 countries	1990-present	Weekly updates	Access can be acquired by purchase
Community Innovation Survey (CIS)	Eurostat	1	The Community Innovation Survey (CIS) based innovation statistics are part of the EU science and technology statistics. Surveys are carried out with two years' frequency by EU member states and number of ESS member countries. Compiling CIS data is voluntary to the countries, which means that in different surveys years different countries are involved.	EU countries	2000, 2004, 2006, 2008, 2010	Biannually	CIS microdata can be accessed via CD-ROMs (scientific-use files) and in the Safe Centre at Eurostat, Luxembourg.
CompNet	ECB	2	The CompNet database is an outcome of the work of the Compnet project, organised by the ECB with the participation of the national central banks of EU countries. The objective of CompNet is to develop a more consistent analytical framework for assessing competitiveness, which allows for greater correspondence between determinants and outcomes. The CompNet database contains various indicators of competitiveness resulting from the analysis of (national) micro-level data based on a harmonised methodology (DMD approach).	EU	2012-present	Annually	https://www.ecb.europa.eu/secure/comtrade/login.html

Table 3.2: Examples of cross-country and matched datasets in Europe and beyond, *continued*

Name	Provider	Type	Content and aims	Countries	Time span	Update frequency	Availability
Data without Boundaries (DwB)	EU Commission	4	Data without Boundaries aims to enhance transnational access to official micro-data for researchers. Programme participants cooperate with NSIs and European data archives to create an integrated model of transnational micro-data access. As part of the project, a comprehensive, structured meta-database providing information on official micro-data available for research purposes in Europe as well as on the procedures for requesting access to these data are being built.	Europe	2011-15	–	www.dwbproject.org
DIECOFIS – Development of a System of Indicators on COmpetitiveness and FIScal Impact on Enterprise Performance	EU Commision	3	The project's objective was to explore the linkage between fiscal policy, enterprise performance and competitiveness on the national, as well as, the EU level. For this purpose, single data sources on enterprises were systematised into an integrated database which allows for creating micro-founded indicators describing the impact of fiscal policy on enterprise performance and competitiveness. Furthermore, micro-simulation models were developed to analyse the effect of national or EU-wide fiscal policies.	EU	2001-2003	Completed	http://statmath.wu.ac.at/stat4/hackl/diecofis/-
Dun & Bradstreet Who Owns Whom	Dun & Bradstreet	1	Who Owns Whom is a set of annual directories published by GAP Books in partnership with Dun & Bradstreet (D&B). It allows the identificaion of relationships between companies, suppliers and customers worldwide and provides detailed information about more than 3.5 million companies including their corporate structure, ownership, etc. The information is divided into seven geographic regions and facilitates the establishment of appropriate networks or the taking of profitable business decisions based on competitor analysis.	International	1958-present	Annually	Access can be acquired by purchase

Table 3.2: Examples of cross-country and matched datasets in Europe and beyond, *continued*

Name	Provider	Type	Content and aims	Countries	Time span	Update frequency	Availability
DYNEMP	OECD	2	Firm employment dynamics are at the heart of the process of creative destruction, of the reallocation of resources across firms and of productivity growth. However, the data required for international comparative analysis over time is scarce and often difficult to access. To fill this gap, the OECD DYNEMP project has developed a new cross-country database of micro-aggregated firm-level data from administrative data sources, mainly national business registers.	International	2001-11	Annually	http://www.oecd.org/sti/dynemp.htm
EFIGE	European Commission	1	The EFIGE dataset is a result of the EFIGE project (European Firms in a Global Economy), funded within the EU's 7th Framework Programme. The data is based on a harmonised survey in seven European countries and it contains information on different aspects of firm performance, ownership, employment, innovation, international activities, and competitiveness.	7 European countries (Germany, UK, Austria, Hungary, France, Italy, Spain)	2010	Only one cross-section	http://www.efige.org/efige-data-released/
Enterprise Surveys	World Bank	1	Enterprise Surveys collect fully comparable firm-level survey data on about 80,000 firms in 122 countries (with a focus on World Bank client countries). Including non-global surveys, the total number of observations is about 110,000 in 135 countries. The ES has are intended to become the main source of comparable firm data across countries and through the years with the aim to build comprehensive panel data sets. Currently the panel data comprises 79 countries.	Emerging countries	2005-present	Each country is surveyed every three to four years	Data is freely available after registration on www.enterprisesurveys.org

Table 3.2: Examples of cross-country and matched datasets in Europe and beyond, *continued*

Name	Provider	Type	Content and aims	Countries	Time span	Update frequency	Availability
ESSLait	ESSNet/Eurostat	2	The general concept of this project is to improve and apply the methodology for data linking and ICT impact analysis that was developed in the ESSLimit Project and the ICT Feasibility Study, and to generate micro-aggregate datasets for future use. The earlier projects demonstrated that a wealth of information can be extracted by micro-data linking, through which already available data can be analysed in completely new ways.	EU	2013	Finished	http://dragon155.startdedicated.com/ons_drupal/
ESSLimit	ESSNet/Eurostat	4	Update and further advance the methodology for ICT impact analysis developed in the Feasibility Study (grant no. 49102.2005.017-2006.128) to include additional data and datasets for an enlarged set of countries. Harmonisation of the methodology is primordial. Study the possibilities and willingness in the participating countries to redesign their survey strategy covering the datasets planned to be used in this project in such a way that it takes into account not only the individual surveys and constraints like the response burden but also the successful exploitation of linked micro-data for impact analysis.	EU	2010-12	Finished	http://dragon155.startdedicated.com/ons_drupal/taxonomy/term/32
EU KLEMS	European Commsion	2	The EU KLEMS Database results from the corresponding EU KLEMS project, which aimed at creating a database on measures of economic growth, productivity, employment creation, capital formation and technological change at the industry level for all European Union member states from 1970 onwards. This work was intended to provide an input to policy evaluation, in particular for the assessment of the goals concerning competitiveness and economic growth potential as established by the Lisbon and Barcelona summit goals. The database aimed at facilitating the sustainable production of high quality statistics using the methodologies of national accounts and input-output analysis.	EU	1970-2011 Project ran from 2003-08 and is finished.	March 2011 is the latest release	www.euklems.net

Table 3.2: Examples of cross-country and matched datasets in Europe and beyond, *continued*

Name	Provider	Type	Content and aims	Countries	Time span	Update frequency	Availability
EU linked employer/ employee data (ESES)/SES	Eurostat, London School of Economics	1	ESES is a unique linked employer-employee micro-database that compares the wage and employment structures of EU countries. It is a large enterprise survey, comprising all employees working in enterprises with 10 or more employees and almost all sectors (except agriculture, forestry, fishing, public administration, and defense). It is conducted every four years in all 28 EU member states, in all candidate countries and EFTA countries and provides variables on earnings levels, employees' characteristics and the employer. Employees and employers are matched via a unique ID. The ESES is available for reference years 2002, 2006 and 2010 (next 2014).	EU, EU candidate countries, countries of the European Free Trade Association (EFTA).	2002, 2006, 2010	Every four years	Data is available in anonymised form via CD-ROM, or remotely via the LEED-LISSY system. However, the data available through LEED-LISSY is limited with respect to countries and years. The entire data in unanonymised form is available only at the Safe Centre at Eurostat's premises in Luxembourg.
EuroGroups Register	EU Commision	1	The EuroGroups Register (EGR) has been established as a network of registers in Member States and on the EU level, the Business Registers of NSIs (and in future the corresponding databases at NCB/ECB) and the central EGR at Eurostat. When the EGR network becomes fully operational it should serve as a unique survey frame and form the basic tool for improving many statistics related to globalisation.	EU	2008-present	The EGR frame (reference year T) is available at T + 16 months.	The EGR frames are accessible to all NSIs and national central banks responsible in the ESS of producing official statistics and are disseminated via eDamis.
FDI markets (FT)	Financial Times	1	FDi Markets provides information on companies globalising through FDI. Part of the service is an online database of crossborder greenfield investments across all sectors and countries worldwide. The investment project database provides real-time monitoring of crossborder investment projects which allows filtering investment opportunities, understanding investment flows and patterns etc. Also available is a company database, comprising profiles of all companies investing overseas. Besides, there is an Investor Signals Module which functions as early warning signal and indicates whether a company may be considering investment.	International	2003-present	Real-time monitoring	http://www.fdimarkets.com/

Table 3.2: Examples of cross-country and matched datasets in Europe and beyond, *continued*

Name	Provider	Type	Content and aims	Countries	Time span	Update frequency	Availability
Innodrive	European Commision	3	The aim of this FP 7 research project was to improve our understanding by providing new data on intangible capital and new evidence on the contributions of intangible capital to economic growth. The study intended to improve information about the capital embodied in intellectual assets (eg human capital, R&D, patents, software and organisational structures) and it aimed at unovering the growth potential associated with intangible capital accumulation in manufacturing, service industries and the rest of the economy.	EU 27 and Norway	–	Project conducted from 2008-2011, now finished	http://www.innodrive.org/
Integrated Public Use Microdata Series International	Minnesota Population Center, National Statistical Offices, and international data archives.	3, 4	IPUMS-International provides harmonised census microdata originating from publicly available census samples from 79 countries and comprising about 560 million person records. IPUMS is consistently documented across countries and over time and allows researchers to select a set of variables out of a broad range of population characteristics which is most suited for their respective analysis. The data is provided for scholarly and educational purposes only and can be accessed by previously approved researchers free of charge via a web extraction system.	International	1850-present	Annually	IPUMS-International makes these data available to qualified researchers free of charge through a web dissemination system. https://international.ipums.org/international/index.shtml
International Wage Flexibility Project (IWFP)	Federal Reserve Bank of New York or the Federal Reserve System	3	Within the scope of the International Wage Flexibility Project (IWFP), wage changes, wage rigidities and its causes and consequences were analysed. Thirteen country research teams explored distributions of individual wage changes relying on 31 micro-datasets on individual earnings. A new methodology was developed for estimating nominal and real wage rigidity and for correcting the distribution of wage changes for measurement errors. Together with the rich data information gathered, the new methodology allows investigating how distinct data features affect resulting estimates of wage rigidity.	16 countries, 15 EU countries and the US	Once	Never again	–

Table 3.2: Examples of cross-country and matched datasets in Europe and beyond, *continued*

Name	Provider	Type	Content and aims	Countries	Time span	Update frequency	Availability
KombiFiD (Combined Firm data for Germany)	Federal Statistical Office, Federal Labour Office, Deutsche Bundesbank	3	KombiFiD was a feasibility study conducted between 2008 and 2011 aiming to assess the potentials, the obstacles and the benefits of matching official firm-level micro data from different institutions in Germany. Administrative and survey data from three official providers, which is in principle not matchable due to legal restrictions, has been prepared for matching by obtaining the written consent of a sample of firms. The result was a sample of more than 16,500 firms which could be used for analyses of various research questions. Due to legal restrictions, the time frame was limited until the end of 2014.	Germany	Once	Never again	Data available to researchers until 31/12/2014
Luxembourg Employment Study Database (LES)	Cross National Data Center Luxembourg	1	LES contains harmonised data from labour force surveys for 16 countries at two time points – about 1990 and 1995 – and US 2000. After these two waves of LES data, we decided not to extend the LES Database to later years.	International (12 countries)	1990, 1995	Finished	Remain on our servers and, if you wish to access them, you may.
Luxembourg Income Study Database (LIS)	Cross National Data Center Luxembourg	1	LIS is focused on income microdata, contains harmonised datasets collected from multiple countries over a period of decades. The LIS datasets contain data on market income, public transfers and taxes, private transfers, household characteristics, labour market outcomes, and, in some datasets, expenditures. The datasets include household- and person-level microdata. LWS is focused on wealth microdata, contains a smaller number of harmonised datasets. The LWS datasets include variables on assets and debt, market and government income, household characteristics, labour market outcomes and, in some datasets, expenditures and behavioural indicators. The LWS datasets contain household-level microdata.	International	1980-present	Waves	1. LISSY : A remote-execution system that allows research using the LIS or LWS microdata.2. Web Tabulator: An online table-maker. 3. LIS Key Figures: Two sets of national indicators

Table 3.2: Examples of cross-country and matched datasets in Europe and beyond, *continued*

Name	Provider	Type	Content and aims	Countries	Time span	Update frequency	Availability
Luxembourg Wealth Study Database (LWS)	Cross National Data Center Luxembourg	1	The Luxembourg Wealth Study Database (LWS) is the first cross-national database of harmonised wealth microdata in existence. The LWS datasets include variables on assets and debt, market and government income, household characteristics, labor market outcomes and, in some datasets, expenditures and/or behavioural indicators.	International (12 countries)	1994-present	–	See preceding row.
World Management Survey	Centre for Economic Performance (LSE)	1	Over the last decade, a large consortium of universities (LSE, Stanford University, Harvard Business School, University of Cambridge) has undertaken a large survey research program to measure management practices systematically across firms, industries, and countries.	Europe, North America, Latin America, and Asia	2004-present	2004, 2006, 2010	http://worldmanagementsurvey.org/?page_id=183
MicroDyn	wiiw (Wien)	2	The MICRO-DYN centralised database is an attempt to reconcile and combine aggregated firm-level data from statistical offices in a number of European countries in one dataset. The final dataset contains data (27 indicators on, e.g. firm characteristics, employment and productivity) from the national statistical offices of 10 countries and was supplemented with data from the Amadeus database for eight additional countries. The data generated from the Amadeus database was put separately since it is in many ways not comparable and should only be used with caution jointly in the analysis with data from statistical offices.	18 European countries	1985-2009 (partially)	Finished	http://www.micro-dyn.eu/files/wp7/MicroDyn_Database_Description.pdf

Table 3.2: Examples of cross-country and matched datasets in Europe and beyond, *continued*

Name	Provider	Type	Content and aims	Countries	Time span	Update frequency	Availability
Nordic Exports of Goods and Exporting Enterprises	Nordic Council of Ministers	2	This report addresses some of the questions raised by economic globalisation. What are the strongholds of the Nordic countries, in terms of the type of goods that are exported and in terms of geographical markets? How are the Nordic countries performing on the main emerging markets, and are exports to these markets indeed where the Nordic countries' export of goods is growing? How important are SMEs compared to the larger enterprises in the export of goods of each of the Nordic countries? And how have SMEs and large enterprises coped with the crisis, both in terms of domestic employment and in terms of export market performance? The new databases that enable this analysis consist of business and international trade data which are linked at micro level (enterprise level) in each country.	Scandinavia	2008-2012	finished	Open Access: http://www.norden.org/en/publications/open-access
OECD Innovation Microdata Project	OECD	2	Research teams from around 20 countries worked in a coordinated way using similar data cleaning methods and econometric models on their national data sets and producing harmonized tabulations with results on innovation surveys. Harmonization of innovation surveys using the distributed micro data analysis method	20 countries, worldwide	Project conducted from 2006-2009, Data starts in the 90s	finished	Book is output
WIOD	EU Commission	1	The World Input-Output Database (WIOD) provides time-series of world input-output tables for forty countries and a model for the rest-of-the-world, covering 1995 to 2011. These tables have been constructed on the basis of officially published input-output tables in conjunction with national accounts and international trade statistics. It also provides data on labour and capital inputs and pollution indicators at the industry level.	27 EU countries and 13 other major countries	1995-2011	finished	http://www.wiod.org/new_site/data.htm

- Structure of firms (company ownership, domestic and foreign control, management);
- Workforce (skills, type of contracts, domestic vs. migrant workers, training);
- Investment, technological innovation, R&D (and related financing);
- Export and internationalisation processes;
- Market structure and competition;
- Financial structure and bank-firm relationships.

Most questions relate to the year 2008, with some questions requesting information for 2009 and previous years in order to build a picture of the effects of the crisis, and the dynamic evolution of firms' activities. An interesting characteristic of the EFIGE dataset is that, on top of the unique and comparable cross-country firm-level information contained in the survey, data can be matched with balance-sheet figures.

EFIGE data has been integrated with balance-sheet data drawn from the Amadeus database managed by Bureau van Dijk, retrieving nine years of usable balance-sheet information for each surveyed firm, from 2001 to 2009. This data in particular enables the calculation of firm-specific measures of productivity and a number of financial indicators, measured over time. The first use for the EFIGE dataset was to explore the correlation patterns between the various international activities of firms (imports, exports, foreign direct investment, international outsourcing) and firms' competitiveness, as measured by various proxies of productivity, in the countries surveyed. The information from the survey allows firms to be classified into seven non-mutually exclusive internationalisation categories. Firms are considered exporters if they reply 'yes, directly from the home country' to a question asking if the firm sold abroad some or all of its own products/services in 2008. The project followed the same procedure with imports, distinguishing between imports of materials and services. With respect to foreign direct investment (FDI) and international outsourcing (IO), EFIGE asked if firms were carrying out at least part of their production activity in another country. Firms replying 'yes, through direct investment (ie foreign affiliates/controlled firms) are considered to be undertaking FDI, while firms replying 'yes, through contracts and arm's length agreements with local firms', are considered to be pursuing an active international outsourcing strategy. Furthermore, EFIGE allows the identification of firms involved in global value chains, although not actively pursuing an internationalisation strategy, based on responses to a question asking if part of the firm's turnover was made up of sales generated by a specific order coming from a customer (produced-to-order goods). Firms replying positively, and indicating that their main customers for the production-to-order activity are other firms located abroad, are considered to be pursuing a passive outsourcing strategy. Hence, a passive

outsourcer is the counterpart to an active outsourcer in an arm's length transaction. Finally, on the basis of responses to a question that allows the identification of the main geographical areas of the exporting activity, EFIGE identified 'global exporters', ie firms that export to countries outside the EU. For all these types of firms, and using also the information derived from Amadeus, EFIGE computed various points of the distribution of an array of productivity measures, as well as unit labour cost and measures of intangible assets intensity. The project also assessed innovation strategies and innovative output and other aspects of price and non-price competitiveness.

Unlike some publicly available micro-based datasets developed at the European level (eg the European Union Labour Force Survey, the Community Innovation Statistics or the European Community Household Panel), which focus on one specific dimension of economic activity, EFIGE focused on international operations, but also contained a broad range of other different sets of firms' activities. With respect to commercially available cross-European datasets (eg Amadeus from Bureau van Dijk), EFIGE assembled not only balance-sheet data, but also both qualitative and quantitative information on firms' characteristics and activities which are typically not observable, but are crucial for competitiveness analysis. Finally the survey design enabled reliable comparisons of countries. Conversely, for example, official micro-based national statistics are not always harmonised across countries and cannot be used effectively for consistent cross-country analysis.

Consequently, EFIGE data can be uniquely used to identify and compare firms in different countries in terms of their different modes of internationalisation, and to analyse how these outcomes are related to other firm-specific variables and broader indicators of competitiveness.

3.3.2.2 The Competitiveness Research Network (CompNet)

CompNet is a network set up by the European Central Bank (ECB) in March 2012 that includes all national central banks within the EU. International organisations also participate. In addition, international scholars specialising in competitiveness issues support the Network.

CompNet is meant to improve the existing frameworks and indicators of competitiveness in all dimensions (macro, micro and cross-border). Additionally, the Network is trying to establish a better connection between identified competitiveness drivers and resulting outcomes (trade, aggregate productivity, employment, growth

and welfare) also by building a bridge between micro and macro analysis, in order to support the design of adequate policies.

On the micro level, the research conducted within the Network has confirmed the importance of firm-level factors (such as size, ownership and technological capacity) in understanding the drivers of aggregate performance. It has also developed a centralised project to compute cross-country homogenous indicators of labour and total factor productivity, and analyse the role of resource reallocation in increasing aggregate productivity.

CompNet is organised in three work streams related to:

1. Aggregate measures of competitiveness;
2. Firm-level studies;
3. Global value chains (GVCs).

One of the main policy questions addressed by CompNet is how aggregate productivity can be enhanced. As discussed earlier, a thorough analysis of competitiveness in different countries is best done by using firm-level data because firms are very heterogeneous. Therefore, information on firm-level drivers of competitiveness is being lost when working with country- or sector-level aggregates. However, because of confidentiality restrictions, the necessary firm-level datasets are not readily available in different countries. Nevertheless, in many European countries the micro-level data can be accessed from within the respective countries. Exploiting this fact, CompNet has opted to employ the Distributed Micro-data Approach (DMD) (see section 3.1.6) in order to compute different indicators of competitiveness at the micro level.

As such, CompNet has created an active network of country teams that independently run a common algorithm to compute a large number of competitiveness indicators. The CompNet firm-level indicator database is superior to others available because of: (i) coverage (58 2-digit, NACE Rev. 2, manufacturing and non-manufacturing sectors in 13 EU countries); (ii) time horizon (2002-2010), since it includes the recent boom-bust cycle and (iii) cross-country comparability. The first round of the so-called Do-File exercise has been completed and the second round is underway. Research output of the network can be accessed via:
http://www.ecb.europa.eu/home/html/researcher_compnet.en.html.

3.3.2.3 Combined firm data in Germany (KombiFiD)

The German KombiFiD project was a feasibility study to assess the potential, the obstacles and the benefits of matching official micro-level data from different institutions in Germany, also with regard to a future replication of such an effort on a larger scale or in different contexts. A unique business micro-dataset (also called KombiFiD) was created. This effort with the resulting unique new business micro-dataset was expected to provide *"enhanced information background for entrepreneurial decision-making"* and to reduce the *"respondent burden for businesses in official surveys and notification procedures"* (see http://fdz.iab.de/en/FDZ_Projects/kombifid.aspx). By matching data on firms from different sources, it was also expected to gain additional information, eg for scientific research or for policymakers, by combining information formerly only available separately. The project started in January 2008 and finished at the end of 2010, with the dataset for researchers released in early 2011 (see Biewen *et al*, 2012, for an overview).

The micro-data involved includes both survey and process-generated data. In particular, several Federal Statistical Office datasets were used such as the *Business Register*, the *Cost Structure Survey*, different tax statistics and the *Structure of Earnings Survey*. From the Federal Employment Office, the *Establishment History Panel (BHP)* has been added to the study and the Deutsche Bundesbank provided their firm-level database on 'Foreign Direct Investment Stock Statistics and Financial Statements'. For a complete list of datasets and for more detailed information on these datasets see http://fdz.iab.de/en/FDZ_Projects/kombifid.aspx.

A major challenge of the KombiFiD project was that German legislation (ie the Federal Data Protection Act) in principle does not allow the linking of the micro-level data of businesses or individuals without the explicit written consent of the affected firms or individuals. Thus, although the technical process of matching the data (ie linking the information contained in the different datasets by using common identifiers) has been quite straightforward, the requirement to obtain consent of the firms involved generated a high level of complexity. As it was not possible to include all businesses in Germany, a sample of 54,960 firms was selected. For a detailed description of the selection of the sample see Gruhl *et al* (2012, p. 7f).

These firms were asked for their consent to matching the available information in the respective databases. From that sample, nearly 31,000 firms responded, and 16,571 responses were positive, corresponding to an acceptance rate of 30.7 percent (see Vogel and Wagner, 2012, p. 3). The information from the different datasets on these

firms was then matched using the available common identifiers, and is used as the KombiFiD dataset.

Technically, the linking of the information from the different datasets was realised via common identifiers jointly available across the different sources and via record linkage techniques. The basic dataset for linking data from the Statistical Offices and the Federal Employment Office is the Business Register, which has been constructed since the 1990s in Germany (and in other European countries due to EU legislation[68]). The Business Register contains several firm identifiers: a unique Business Register ID, the establishment numbers of all corresponding establishments and tax numbers (see Gruhl *et al*, 2012, pp. 10-15, for a detailed assessment of this matching process).

Matching data from the Deutsche Bundesbank was less straightforward. As no common identifiers are available between the datasets described above and the data to be used from the Bundesbank, record linkage techniques based on the firms' names and addresses were used (see Koch and Neugebauer, 2014, for a more thorough description).

The resulting KombiFiD dataset contains all the information from its constituent datasets for the firms which agreed to the matching of their data. A detailed description and lists of variables are available in Gruhl *et al* (2012, pp. 21-85). The data is accessible to external researchers in a weakly anonymised version[69].

In general, a broad range of issues can be examined using the KombiFiD data. Up to now, however, the dataset has been only sparsely used in economic and statistical analyses. Exceptions are the papers by Wagner (2012 and 2012a) and Vogel and Wagner (2012), whereas only Wagner (2012a) goes beyond methodological aspects. This relatively scant utilisation of the potentially very rich KombiFiD data can first be attributed to the fact that the data has been made available to the public only quite recently. With regard to the analysis of competitiveness, the dataset contains a comprehensive set of variables from the different sources allowing evidence to be generated on, inter alia, growth, productivity, trade or employment.

It may, however, also be attributed to the fact that the data has some major drawbacks: first and foremost, it has to be pointed out that the use of the KombiFiD data was

68. Council Regulation No. 2186/93.
69. This type of anonymisation means that some variables, eg regional and sectoral identifiers, are only available in an aggregate form.

restricted until 31 December 2014 which made the serious utilisation of the data very difficult. To our knowledge, the data has to be erased completely from the servers of the data providers after that date, thus making research projects or even working papers nearly impossible as results cannot be verified after that date. Another serious drawback of the data itself is that no information is available about the firms from the original sample that refused consent for their data to be matched for the project. This results in no information on a potential selection bias, making thorough analyses hard to realise.

Wagner (2012) and Wagner and Vogel (2012) performed tests on the quality of the KombiFiD sample for the manufacturing and the service industries on the basis of data from the Statistical Offices. They come to the conclusion that the quality of the KombiFiD sample can be regarded as high only for the former West Germany, whereas for the former East Germany an assessment of quality is not possible because of the small sample size.

Ultimately, the KombiFiD project was a huge and ambitious effort with very meaningful objectives, ie creating a 'new' dataset building on existing information and thus sparing firms from participating in further surveys. The expectation was also to evaluate the future potential of similar projects.

The expectations have only partially been met, and the main drawbacks can be traced back to existing legal regulations preventing deeper cooperation or even exchange of data between data providers. Although a relatively large sample was used for the survey, even taking into account the need to obtain consent from the selected firms, there was a relatively high response rate and a high acceptance rate of more than 30 percent. Nevertheless strict regulations prevent reasonable use of the data: first, the limited time window of opportunity for using the data is a problem, and, second, the unknown nature of the potential selection bias.

In summary, the KombiFiD project generated much new knowledge on the technical aspects of data matching, experience with regard to firm behaviour and practical knowledge about cooperation between different data-providing institutions. Hopefully, future projects will be set up in order to proceed in this promising direction.

3.3.2.4 Data without Boundaries

A very promising, large-scale programme, which is connected to the MAPCOMPETE project in many ways, is 'Data without Boundaries' (DwB). DwB is another European

FP7 project, which aims to enhance transnational access to official micro-data for researchers[70]. The project will be finished in 2015. The motivation behind the project is that "*currently OS micro-data repositories are underutilised resources within research, eg within the social science research area, both nationally in many countries and internationally*"[71]. Programme participants cooperate with NSIs and European data archives to create an integrated model of transnational micro-data access. As part of the project, a comprehensive, structured meta-database providing information on official micro-data available for research purposes in Europe and on the procedures for requesting access to these data, is being built[72].

3.3.2.5 The Global Value Chain project and the Eurostat International Sourcing Survey

The Global Value Chain[73] project was coordinated by Statistics Denmark and carried out from 2011-13 within the ESSnet by Statistics Finland, Statistics Norway, CBS Netherlands, Instituto Nacional de Estatística (Portugal), National Institute of Statistics (Romania), National Institute of Statistics and Economic Studies (France). The aim of the project was to strengthen ESS capacity (conceptually and methodologically) to measure economic globalisation and the globalisation of business, and to concretely establish statistical evidence about the increasingly globalised ways of doing business and organisation of companies. The objectives were to help policymakers to make better informed decisions and to monitor the globalisation of economies by developing and providing indicators on economic globalisation.

The GVC project is intertwined with Eurostat's International Sourcing Surveys (ISS)[74], which were carried out in 2007 and in 2012. The latest survey gathered data on the international organisation and sourcing of business functions in 15 European countries, while in 2007, the coverage was 11 EU countries plus Norway. The surveys cover nearly 40,000 businesses with more than 100 employees.

70. http://www.dwbproject.org/
71. Data without Boundaries, DELIVERABLE D7.1, Metadata Standards – usage and needs in NSIs and Data Archives, 2013, http://www.dwbproject.org/export/sites/default/about/public_deliveraples/dwb_d7-1_metadata-standards-usage_report.pdf
72. Data without Boundaries, DELIVERABLE D5.2, Report and Databank Documenting OS Micro-data, 2013, http://www.dwbproject.org/export/sites/default/about/public_deliveraples/dwb_d5-2_databank-national-survey_report_final2.pdf
73. http://www.cros-portal.eu/content/global-value-chains-0
74. See http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/International_sourcing_of_business_functions (for the 2013 survey) and http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/International_sourcing_statistics (for the 2007 survey)

More specifically, the GVC project:

- Identified and developed a set of standardised indicators on economic globalisation to be collected and published as reference indicators within the European Statistical System, subject to political approval.
- Identified a set of supplementary indicators which could be collected to measure more industry-specific elements of the globalisation process utilising existing statistical sources.
- Identified possible experimental indicators based on micro-data linking. The project further developed the methodology for micro-data linking and identifying different types of statistical registers relevant for measuring globalisation.
- Supported the set up and implementation of the methodology to carry out the micro-data linking between different types of statistical registers in participating countries.
- Fine-tuned the survey methodology including finalisation of the survey contents and establishment of the required set of harmonised definitions to be used in the survey.
- Supported NSIs to set up and implement the survey on global value chains and international sourcing in participating countries.
- Produced statistical analyses of the global value chains and international sourcing survey and micro-data linking results to be published by Eurostat.
- Tested possible methods of improving the quality of the foreign affiliate statistics by utilising information available within the European Statistical System related to the population of foreign affiliates.

In summary, the GVC project and the ISS are interesting examples of how the ESS can leverage existing data from business registers, trade or foreign affiliates, by linking such micro-data with a new harmonised cross-country survey, to provide a rich information base, which can allow researchers to produce new knowledge useful to inform appropriate policy decisions.

4. Barriers to data access and matching in Europe: concluding remarks

This Blueprint so far has investigated the extent to which a wide range of competitiveness indicators, especially those that are built from micro-data and that we have defined as bottom-up indicators, can be computed for EU countries and what data is actually accessible for researchers. In chapter 2, we highlighted issues at the level of individual countries, while in chapter 3, we focused on the challenges of using micro-data to construct indicators of competitiveness across countries. In this chapter, we pick up on the main conclusions emerging from chapters 2 and 3 (in sections 4.1 and 4.2, respectively). Building on these considerations, in the next chapter we offer some policy recommendations.

4.1 Issues regarding the availability of data at country level

The availability of an indicator of competitiveness depends on different factors. In the MAPCOMPETE data mapping exercise (see chapter 2), we distinguish between factors that determine the computability of an indicator and factors that influence accessibility. By computability we mean the quality of data and the length of time coverage. Computability of an indicator relies mainly on data existence and the possibility to merge data from different sources, if necessary. The accessibility of data depends on the rules of access and their clarity. As part of the MAPCOMPETE data mapping exercise, statistical institutes of EU member states were approached to collect information on micro-data availability. Project participants surveyed several bottom-up competitiveness indicators – firms' productivity, dynamics, international activities, R&D activities and some other features – with respect to computability and accessibility.

4.1.1 Availability of data for statistical/research purposes

MAPCOMPETE participants surveyed several bottom-up competitiveness indicators, which are based on basic information about enterprises, such as year of establishment, number of employees or financial statement and balance sheet items. Although such information is usually collected by national authorities for administrative purposes, our findings on the availability of this data present a mixed picture.

We find that those indicators that require the use of basic balance sheet data (eg labour productivity, TFP) – along with trade indicators – are the most computable among the bottom-up indicators we surveyed, but there are country-specific problems. Also, bottom-up indicators on firm dynamics, which are based on data about company entries and exits, are poorly computable for several member states. In some cases the information needed is available, but only for a subset of enterprises or for a limited time period.

Much of this heterogeneity can be explained by the fact that countries report various databases as the best possible source of information on firm dynamics, balance sheet and financial statement items. There are NSIs that report survey data as the best possible source of information, while others indicate that administrative databases are available for statistical use.

Our findings are consistent with the findings of a recent ESSnet project. The ESSnet Admin Data project[75] examined the use of administrative and accounts data for producing national statistics. The project outcomes show that both legislation and existing practices regarding the use of administrative data differ in different EU member states[76]. They highlight the possibility to improve the quality of business statistics and to reduce the administrative burden on enterprises by finding common ways for using administrative data. It is also stated that relevant administrative data is available to a greater extent than is actually used. In some countries, administrative data is only used as a sampling framework, or for imputation and validation, while NSIs compute national statistics using survey data.

In most member states, national legislation supports the use of administrative data

75. http://essnet.admindata.eu/.
76. Costanzo, L. (2013) Report to Eurostat on the "Overview of Existing Practices", Admin Data, Work Package 1. http://essnet.admindata.eu/Document/GetFile?objectId=5995.

for statistical purposes – under different confidentiality restrictions – and provides special rights for the NSIs to access these sources. However, the ESSnet Admin Data project identified several factors that hamper the effective use of administrative sources. First, legislation that requires the use of administrative data whenever possible is rare (exceptions are Finland and the Netherlands). As a consequence, NSIs are not motivated to make investments in order to fully exploit administrative data. They use such data, but only if it can be used with minor adjustments as part of existing practices.

Second, most countries lack a coherent and comprehensive framework for collecting, storing and providing access to collected data. Different production units of NSIs perform admin-data related tasks separately, thus the use of administrative data is based on *ad-hoc* agreements with limited scope between the NSIs' production units and the data holders. There are, however, positive examples: Portugal replaced all surveys of Structural Business Statistics with one new data-collection system for administrative and statistical use, while Bulgaria introduced a single entry point for reporting fiscal and statistical information.

Third, cooperation between admin-data holders and NSIs is weak or difficult in several countries, partly because of the lack of legislation establishing the corresponding duties of data holders. In most countries, NSIs have no impact on the design of administrative data collection and authorities do not have to consult NSIs when introducing changes to data collection practices.

These aspects have been addressed in a amendment to Regulation (EC) No 223/2009 – being finalised at the time of writing[77] – which aims at establishing a legal framework for more extensive use of administrative data sources for the production of European statistics without increasing the burden on respondents, NSIs and other national authorities. NSIs should be involved, to the extent necessary, in decisions about the design, development and discontinuation of administrative records that could be used in the production of statistical data. NSIs should also coordinate relevant standardisation activities and receive metadata on administrative data extracted for statistical purposes. Free and timely access to administrative records should be granted to NSIs, other national authorities and Eurostat, but only within their own respective public administrative system and to the extent necessary for the development, production and dissemination of European statistics.

77. See footnote 57.

4.1.2 Legal and administrative constraints of access to micro-level data

The MAPCOMPETE data mapping exercise revealed substantial differences between EU member states in terms of the accessibility of micro-level information needed to compute the surveyed competitiveness indicators. We observe that there are countries for which many bottom-up indicators have a relatively high level of computability, meaning that the required information exists in some meaningful format at the local statistical authorities, but micro-data access is not allowed for outside users.

Legal barriers related to confidentiality

While the rules of micro-data access are not clearly specified in several countries, it is clear that confidentiality restrictions substantially differ in different member states. The common feature of national laws is that they oblige institutions collecting personal or firm-level data to guarantee the anonymity of respondents. However, various definitions of confidential data and different approaches to data protection are present. Research entities have the option to access personal data in the majority of countries, but there are significant differences in national confidentiality restrictions regarding the transmission of data from the collecting institution to other entities[78]. Some member states do not allow the transmission of certain confidential data, or the implementation is problematic.

Importantly, regulations concerning Eurostat itself also differ in different member states: Eurostat can't access confidential data from some countries.

The new EU statistical law[79] emphasises the importance of the availability of confidential data within the ESS network. It states that the transmission of confidential data between ESS partners may take place *"provided that this transmission is necessary for the efficient development, production, and dissemination of European Statistics or for increasing the quality of European statistics"*. The access to confidential data for scientific purposes also requires the approval of the national authorities which provide the data. However, our experience suggests that despite the legislative underpinning, there are several factors that hinder the research use of micro-data, and the exact methods, rules and conditions of access are still to be developed in many member states.

78. Ichim D., Franconi L. Strategies to achieve SDC harmonisation at European level: multiple countries, multiple files, multiple surveys, http://neon.vb.cbs.nl/casc/..%5Ccasc%5CESSnet%5Ccomparable%20dissemination%20v-1.pdf
79. Eurostat, Legal Framework for European Statistics - The Statistical Law, 2010 Edition, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-31-09-254/EN/KS-31-09-254-EN.PDF

The mapping of micro-level information also highlights the fact that different types of data are treated differently. In some EU member states, different regulations apply to different databases. Databases with the full population compiled by National Statistics Institute of Italy are not accessible to researchers, who can only access descriptive statistics upon request, but micro-data stemming from surveys is available. In the Czech Republic, business register data can be accessed relatively easily, while for other types of data, such as custom data and FATS data, conditions are more stringent. Malta allows access to firm-level information for research purposes, except for data on foreign ownership and capital. In Latvia, data is available upon request, except for data on trade by destination and product, which is confidential.

Our results show that in general there are stricter regulations on registry-type data and on databases that have full coverage over the observed population. Survey type data, especially data from harmonised surveys like CIS, is usually easier to access. Our findings on individual-level trade data are mixed, since these databases include information both from administrative sources (ExtraStat) and from a harmonised survey (IntraStat).

A distinction in confidentiality restrictions is particularly important when we consider the potential use of bottom-up indicators that are based on information obtained from different sources in different countries. For instance, firm entry and exit information and balance sheet data are obtained from administrative sources in some countries, while others conduct surveys to collect the information. Consequently, the computability and accessibility of bottom-up indicators based on these data is likely to differ in different countries and a harmonised approach to confidentiality protection is hard to achieve.

It is worth mentioning that Eurostat provides access for scientific purposes to certain European survey data[80] including the Labour Force Survey and the Community Innovation Survey. Recognised research entities conditional on the approval of their research proposal might access micro-data anonymised by Eurostat on electronic devices or non-anonymised data in Eurostat's 'safe centre'. Currently, Eurostat negotiates on the possible dissemination of the micro-data on a case-by-case basis and proposes a unique anonymisation methodology to all member states. Member states might refuse Eurostat's proposal if it conflicts with national legislation, and thus micro-data will not be available for all member states[81].

80. Comission Regulation 831/2002 specifies the surveys and the rules of access.
81. Ichim D., Franconi L. Strategies to achieve SDC harmonisation at European level: multiple countries, multiple files, multiple surveys, http://neon.vb.cbs.nl/casc/..%5Ccasc%5CESSnet%5Ccomparable%20dissemination%20v-1.pdf

Practical (technical) constraints on accessibility

We observe that in addition to national legislation, the internal regulations of data-collecting institutions and practical constraints also affect the accessibility of micro-data. In Romania, practical barriers hinder the accessibility of the databases compiled by the NSO: a safe environment for data security is at the time of writing not yet in place. Part of the variation in these matters can be explained by the fact that the increased demand for micro-data is a relatively new phenomenon. The resources available to NSIs for disclosure control, and their prior experience in the field, might influence the speed and direction of adaption. The development of new statistical disclosure methods needed to provide access to micro-data might be hindered by organisational, methodological and software problems.

Our results show that currently, at the national level, the most commonly used method to provide access to micro-data is the release of scientific use files. In case of research use files, statistical disclosure methods and restrictions on access and use – eg license or access agreements – are applied simultaneously[82]. Our data mapping exercise shows that several NSIs provide access to micro-data in data laboratories. Data laboratories allow researchers to use more identifiable data under strict conditions. In most cases, users are legally obliged to keep the data confidential, and are subject to close supervision and output checking. Since setting up a data laboratory takes time and resources, there are countries where this form of micro-data access is not yet available. Remote execution is also possible in a few member states. Note that the cost of operating a data laboratory or remote access services significantly increases with the number of users, mostly because output checking is completely manual in almost all of the member states. Consequently, even in the countries where the NSI already provides access to micro-data, revision of data protection practices will be inevitable in the near future.

4.1.3 Non-legal barriers

Issues with metadata

Having basic information about datasets in advance is a very important factor that might affect the success of a research project. Researchers need to have detailed

82. Eurostat, Handbook on statistical disclosure control (January 2010) http://unstats.un.org/unsd/EconStatKB/Attachment474.aspx

information on the available datasets including the identity of the owner of the data, the exact content, the quality of data and the rules of access. These pieces of information are necessary to decide whether the dataset is suitable to their needs and whether they apply for access.

International standards already exist for the international exchange of metadata. Statistical Data and Metadata Exchange (SDMX), an initiative sponsored by the Bank for International Settlements, ECB, Eurostat, International Monetary Fund, OECD, United Nations and the World Bank, aims to provide standards for the exchange of statistical information (eg formats for data and metadata, content guidelines, IT standards)[83]. Particularly for Europe, the European Commission has set up a recommendation 'on reference metadata for the European Statistical System'[84], which refers to the European Statistics Code of Practice[85] and is based on the SDMX framework.

While ESMS Metadata files for all of the statistics published by Eurostat are provided – and other international organisations also provide structured metadata on their statistics – our experience shows that there is still a big hole in the information on data. ESMS metadata files present useful information on methodologies, quality and the statistical production processes in general, but usually provide very little information on the link between the aggregate indicator and micro-data used to compute the given indicator. Also, country-specific information on survey and sampling design is often sketchy. We made use of the information provided in ESMS Metadata files when mapping the readily-available aggregate indicators, but we found that in order to be able to assess the strengths and weaknesses of these indicators to improve their quality or to propose new ones, much more information on the available national micro-data would be needed.

Gathering comprehensive information on micro-data available in EU member states proved to be a challenging and time-consuming task. The amount and structure of information available on the websites of NSIs and other national data providers is very different in different countries. It is usually insufficient to fill the MAPCOMPETE MetaDatabase and it is definitely insufficient to plan a research project. In many cases, researchers obtain information on given datasets from scientific publications or

83. SDMX (2009), Content-Oriented Guidelines, Statistical Data and Metadata eXchange. Vale, S. (2009), Generic Statistical Business Process Model, Version 4.0 – April 2009, UNECE Secretariat.
84. See European Commission (2009), Commission recommendation of 23 June 2009 on reference metadata for the European Statistical System, Official Journal of the European Union L 168/50, 50-55.
85. Eurostat (2011), European Statistics Code of Practice for the National and Community Statistical Authorities, Eurostat, European Statistical System, Luxembourg.

through informal channels, which are burdensome and usually result in incomplete information. Also, when conducting cross-country comparative research or research that requires the use of information from more than one source, researchers have to search through several websites and publications, each with different metadata structure and information content.

Since in MAPCOMPETE we collected a huge amount of information in a systematic manner, we tried to directly contact staff within the NSIs in all the EU28 countries to gather the relevant information. After a few months of the project, it became apparent that this was highly complicated, so we decided to gather information by exploiting existing contacts built up in another international project (CompNet) and from other personal contacts. In some cases, these contact persons were able to help us fill in the MAPCOMPETE MetaDatabase and in other cases they referred us to people within the NSI. The fact that in most countries economic databases are collected and handled by more than one institution – the NSI and the national central bank (and sometimes other institutions) both collect data in most cases – made it even harder to obtain the required information. Also, smaller countries and newer EU members tend to have less experience in handling requests for micro-data access, and consequently are usually less prepared to provide systematic information on existing data.

The experience we gained during the data-gathering process shows that the availability of information on the data is at least as important as the availability of data itself. Performing EU-wide research projects on competitiveness or designing new indicators is not feasible without easily available, comprehensive information on national micro-data. This is why the MAPCOMPETE MetaDatabase is especially useful for future research on measures of competitiveness. Furthermore, it serves as a basis for suggestions for possible improvements to data sources, treatment of data, conditions of access etc. It might promote quality research by providing detailed information on the accessibility and availability of data related to the measurement of competitiveness. However, the MAPCOMPETE MetaDatabase is only a snapshot of competitiveness-related data. A regularly updated, structured, easily available and comprehensive meta database on national micro-data – that might include the experience of other researchers working with the data – might substantially increase the efficiency of international research projects.

Issues related to the nationality of the data user

As part of establishing the European research space, conducting research and analysis on the basis of foreign data becomes important. Several specific problems arise in

terms of foreign access to datasets located in countries other than the nationality of the researcher. First, in some countries, such as Belgium, Denmark, Hungary and the United Kingdom, access to micro-data is allowed only to researchers who are citizens of the country of the data provider or affiliated with a national institution. Second, language barriers are obviously a serious burden, since in many countries information is provided only in the national language, but one that can be solved by simply offering data description and variables in English. Several NSIs have made a great deal of progress in this respect, including metadata provision in English. Third, the provision of data on site might not be a burden for locals, but can be very costly for foreign researchers. Hence, setting up secure remote access – such as is available in Finland, France, Germany and Sweden – would be an important step. Finally, making access by foreigners easier by appointing an English-speaking specialist could indeed facilitate European research integration.

Unclear rules of access

When mapping the accessibility of data, we faced the obstacle that it is often challenging to obtain precise information on the conditions of access to confidential data. Information on the accreditation process, statistical disclosure control methods applied and the practical details of access is usually not clearly specified on the website of the data provider or at any other publicly-available source. We found that one had to contact the data provider directly in order to clear up the details and to find out if access to the data is possible and under what conditions.

Our results show that there are substantial differences between countries in terms of the clarity of rules of access. In many countries there is some settled, formal procedure of applying for access (eg Denmark, Finland, France, Netherlands, Slovenia and Sweden) while other countries are less advanced in this respect and handle requests on a case-by-case basis. However, regardless of the sophistication of the application procedure, in most cases, it is required to present a research project which needs to be approved. This approval creates room for discretionary decision-making and informality which might differ from country to country, but is really difficult to assess.

The approval procedure might be more problematic when the data provider does not perform output checking itself, but it is the researcher's responsibility to protect the confidentiality of data. If data protection is delegated to the researchers then the cooperation strongly relies on trust between the data provider and the researcher, and it might be hard to define exact criteria.

Truncated data

In many cases, micro-data is provided in truncated form; that is it is made available with less information than the original source, in order to prevent the risk of disclosure (sensitivity) and for cost reasons. For the purposes of our discussion, this aspect is related to accessibility, but it can affect computability when it prevents the merging of different datasets.

Sensitivity truncation

Several statistical disclosure methods used to protect the confidentiality of data lead to a loss of information and might affect the quality of analysis carried out on the data. Let us first present key obstacles and make suggestions for their treatment (for details and a broad discussion, see Hundepool *et al*, 2010). According to statistical best practice, this implies *"first a definition of possible situations at risk (disclosure scenarios) and second, a proper definition of the 'risk' in order to quantify the phenomenon (risk assessment)"* (Hundepool *et al*, 2010, p. 30).

In this chapter, we identify four issues that matter for practitioners:

1. Sensitivity of information on selected firms;
2. Recoding data into broader categories;
3. Removing or modifying variables;
4. Other disclosure measures.

The first issue is related to the sensitivity issues of aggregated data. In some sectors, size categories or regions, there are only very few firms. Aggregating data on them would imply that in some categories only one or very few firms would feature and hence, their individual data would not be protected. To avoid this scenario, most statistics institutions and central banks or research outlets protect confidentiality by setting up compulsory aggregation rules. Typical rules include a minimum number of firms per aggregated band (this ranges between 4 and 9, in our experience) and maybe other controls such as market share of the top 5 firms in the aggregate.

The second topic is a more general solution to keep identification impossible. This entails aggregating some existing firm categories such as industry or location address to protect the identity of firms. This process is especially useful in smaller countries where some regions or industries might include only a few firms, even if they are not large. Examples include merging four-digit industry codes into two-digit codes, merging

municipalities or NUTS3 regions into NUTS2 regions, or replacing employment data with firm size brackets.

Third, authorities might remove or replace variables. This might include the deletion of variables that would allow identification – this happens when some activity occurs rarely or is carried out by only a few firms. This might include balance-sheet items, such as subsidies, or some research activities in an innovation survey.

Another option to prevent identification in general, and merging of datasets, in particular, is masking. This approach is divided into two categories depending on their effect on the original data: perturbative and non-perturbative masking methods. Perturbation implies the multiplication of all values by a random variable of unit expected value and a small but significant variance. This implies that say, sales values would be altered by a few percent without affecting any statistical relationship (given the unit expected value). Other options include rounding or truncation. In these cases identification or linking of the data to other data sources would be impossible or difficult because of the lack of exact matching (for more details, see Willenborg and de Waal, 2001).

Importantly, researchers can often access sensitive information in, for example, the research lab, but there are strict rules for the information available outside the safe environment. Apart from these more common issues, authorities might apply individual controls or ask for a list of descriptive statistics to control the process. Statistics offices will often ask researchers to submit all relevant documentation – including programme code files, and descriptive tables for output checking before releasing results.

Finally, note that in some cases an extreme application of this sensitivity approach is applied: individual data is aggregated right after data collection. In this scenario, firms are clustered by industry, location, size and only aggregate information is released. While this may indeed provide security, it washes out important features of observations that may be important for research.

Dataset reduction for cost saving

Another factor that might reduce the scope of available datasets is cost saving. Every aspect of a dataset – number of variables, dimensionality and frequency of observations – will generate additional costs, mainly in terms of attention. Supervisors need to spend time on organisation of dataset management, cleaning and provision,

and the costs of these will depend on the size and complexity of the data at hand. Saving resources and reducing administrative burdens are important in an era when NSI budgets are often being cut. As a result, aggregation and truncation of raw data are often carried out not for sensitivity but for cost purposes.

One such practice is aggregation of some part of the dataset. Transaction-level data might be aggregated into annual aggregates. For instance, foreign trade is often registered at a very fine transaction level, but available data is mostly at annual aggregate level. Several variables might be deleted in order to avoid spending the time that would be required for consideration of sensitivity issues.

Finally, another approach is exclusion of small firms. Dropping firms with fewer than five employees could reduce the size of a dataset by 80-90 percent, while retaining 95 percent of value added. However, such an exercise will limit analysis and understanding of important issues, such as entrepreneurship and firm dynamics.

An important aspect of dataset reduction for cost saving reasons is European/international harmonisation. Comparing statistics computed on the whole dataset or on firms with more than 10 employees might yield rather different results (for an application for exporters, see Békés *et al*, 2011).

4.2 Accessibility and matching of data from different countries

As we argued in chapter 3, data matching opens up rich and novel research opportunities, especially when micro-level datasets are concerned. Existing micro-level data in European countries has significant potential in terms of record linkage and matching, including also commercial data and Big Data. Data matching and issues of matchability have considerably gained in importance in recent years. One reason for this lies in the increased accessibility of micro-level datasets and in the desire of researchers to merge these datasets within and between countries in order to increase the research potential of the data. There has also been significant progress on technical issues, not least driven by the rapid development of computer technology and data storage.

The issue of data matching and matchability is of course not confined to the social sciences, but the recent economic crisis has made clear that economists require high-quality data, especially at the micro level, that is comparable across countries, in order to examine cross-country differences in competitiveness. However, comparable micro-data at the firm level in different EU countries is so far only available for some topics,

most of which are not directly relevant for competitiveness (notable exceptions are the Community Innovation Survey, the International Sourcing Survey or the EFIGE survey). These comparable micro-level datasets are, however, all based on sample surveys.

The huge potential of administrative data, which is already leveraged in many countries, is still waiting to be fully realised (see Agafitei and Vaju, 2013, for instance). There are, however, some serious endeavours in this direction, mainly based on the ESSnet projects and on the Framework Regulation for Integrating Business Statistics (FRIBS, see section 3.2). These projects are of special importance because they are concerned with administrative data within the EU, which is of high quality. Any step towards making these data more comparable and accessible is more than welcome by researchers and policymakers. Therefore, ensuring the availability of such data should be a priority for the European Commission because this would ensure vastly improved analysis of cross-country differences in competitiveness, and of labour market issues and related fields.

The most serious obstacles to matching micro-level data from different countries are still legal restrictions preventing data from being matched, because privacy and confidentiality are at stake. However, there is some activity in this area, namely within projects to evaluate the potential of analysing micro-level data without directly accessing the data.

There are also obstacles to data matching within countries (see the KombiFiD example from Germany). This holds especially true if the datasets to be matched are held by different data providers, eg statistical offices, central banks, employment agencies or private data providers. However, progress has been made in this regard in recent years.

Important steps to overcome the problem of data comparability between countries, particularly with regard to cross-country analyses of competitiveness, have been taken, for instance by the EFIGE project providing comparable firm-level data for 15,000 firms from seven EU countries. The ECB's CompNet project is following suit. However, these two projects can only be regarded as first tentative steps towards data that can be used for cross-country analyses in the field of competitiveness, and that is highly useful for policymakers.

Overall, much has been achieved in the field of data matching within Europe in recent years, but the universe of cross-country and matched datasets is still sparsely populated and quite heterogeneous, with potential for improvement. Because of the

ever-increasing need for high-quality datasets that can be used to inform policymakers, much more needs to be done. Cooperation between data providers within and in different countries is key, as is the reduction of red tape. Comparative analysis of competitiveness in different countries is ultimately only possible if comparable (micro) data exists in different countries or if data can be harmonised and made accessible to researchers. Ensuring the availability of such data should be a priority for the European Commission, because it would enable vastly improved analysis of policy-relevant issues.

5 Policy recommendations: towards better access, computability and matchability of micro-level data

This Blueprint has shown that the information currently available to researchers on comparable measures of competitiveness for different countries is insufficient. Aggregate data, which is easily accessible and widely available, does not allow researchers to provide the answers that policymakers need. Micro-data on individual countries is mostly inaccessible to external researchers, and the situation is even worse when one tries to compare figures based on micro-data which are comparable for different countries. Only a few firm-level surveys are available, mostly only for one or a few years; there are few examples of matched data from different countries, and internationally comparable figures can be gathered only from a few micro-distributed data exercises. This is very different from, for example, the United States, where micro-level data from different states has been matchable and comparable since at least the mid-2000s. This implies that Europe lacks proper information to assess of the state of competitiveness at European level, compared to the situation in the United States.

The first-best solution to overcome these bottlenecks would be to change the national and EU-level rules of data content, data availability, data matching and data access. The efforts undertaken by the ESS, with programmes such as MEETS, FRIBS, FATS, SIMSTAT and ESS.VIP (see section 3.2.2), towards greater harmonisation of data and the construction of pan-European data sets are useful initial steps in this direction. In particular, these initiatives can contribute to:

- The reduction of the burden on enterprises in collecting and providing internal data;
- The provision of a common ESS infrastructure framework for the production and compilation of business statistics with an appropriate legal background and new administrative mechanisms allowing for the sharing of information, services and costs among all ESS partners;
- The definition of consistent data requirements and a common data quality framework, which will enable the linking and matching of statistics obtained as part of the regular collection of global business statistics.

However, the timeline to complete this process, and for its effects to be felt by researchers, is far too long and in the end might even prove almost useless, since it might well be that when this time comes, the next generation of researchers might highlight a different set of needs.

Therefore, such long-term actions to change regulations need to be complemented with more short-term workarounds.

The first workaround is to exploit the availability of improved methods and techniques, such as matching after separate processing (eg the Distributed Micro-Data Approach) or imputation. Projects such as CompNet (see Table 3.1 and section 3.3.2.2) or ESSLait (see Tables 3.1 and 3.2) provided important insights into new aspects of competitiveness by producing micro-aggregated statistics going beyond the first moment of the distribution of firms' competitiveness indicators. However, if not properly supported by policy, these initiatives might remain one-shot exercises, whereas they need to be refined, constantly updated and carried out in a timely way in order to provide the more up-to-date figures for policy decisions. Two examples we have already mentioned clearly highlight these risks: the ESSLait exercise provided figures up to 2010 (see http://www.cros-portal.eu/content/metadata-work), while the more recent CompNet figures refer to the year 2011. Since these initiatives require researchers within data-providing institutions to run the codes prepared by the researchers, proper policy support is needed to enforce in as many countries as possible the requests to run micro-distributed exercises.

The second workaround would be to improve techniques for matching and accessing micro-level data, either by improving architectures for data matching (eg by involving 'matching institutions') or for access to data by researchers (eg by improving techniques of data anonymisation). Many NSIs have already developed or adopted elaborate methods and organisational arrangements in these areas. For example, in Germany, there is a well-established system of research data centres at several official

data providers. Other countries like the Netherlands or France have established techniques of remote-data access. From a theoretical perspective there are several additional ideas which could be rather easily adopted or, if necessary, adapted to national systems and legislation (see section 3.1.5, and Koch and Neugebauer, 2014, for an overview).

It is worth mentioning that – after speaking to officials in NSIs, national central banks and other official data providers in many EU countries – we are quite persuaded that in most countries access to micro-data would be feasible for external researchers, but it is easier for the data providers to restrict access. While the official reason is often linked to legal issues about confidentiality, it seems that other factors might play a role. We have described several approaches to allow researchers access to data while maintaining confidentiality (such various forms of anonymisation, or the creation of 'matching institutions'), but these solutions have costs, and require the data provider to take some responsibility for the release of the data. Restricting access is cost- and responsibility-efficient for the data providers, although very inefficient from the researcher's perspective. To some extent, it is also a way to protect the monopoly of the data provider in terms of use of the data. But if these are the real issues behind the restrictions on data access, there are readily-available solutions.

Data access does not need to be free for all researchers. Instead, researchers can contribute to cover the costs of setting-up the infrastructure for data access using their research funds. Since there are mainly fixed costs, related to setting up the facilities for safe access (including remote connections) and to the anonymisation of the data, while the marginal costs for an additional user are relatively low, data providers could use a sort of average incremental cost to establish access. This pricing structure is not new to economists, and it is similar to what happens in network industries. On top of this, since data providers are multi-product monopolies, they would obtain an advantage from allowing access to the greatest number of data sources, in order to increase the number of users[86].

Furthermore, when contacting national statistics institutes and national central banks, we found a generally high level of competence. However, in order to foster co-operation and build a truly European infrastructure for accessing micro-data, it is very important that there is also investment in developing capabilities such as language skills and economics knowledge. In this respect, EU support is crucial, especially for smaller

86. We thank Jan Hagemeier for an illuminating discussion on this point.

member states, which might not be able to afford to bear the fixed costs of setting up new infrastructures and developing the necessary capabilities.

The third workaround is to support multiscope cross-country surveys, which allow researchers to gather information on a wide range of firms' activities and performance indicators, in order to enable them to assess their contribution to overall competitiveness. The Community Innovation Surveys and the International Sourcing Surveys (see Table 3.1 and section 3.3.2.5) are interesting examples of this, although they both focus on specific aspects of competitiveness. The EFIGE survey (section 3.3.2.1) is another example, which takes into consideration more aspects of competitiveness. However, in order for this solution to be effective, there is a need for greater harmonisation and coordination. Concentrating resources on fewer surveys could be more effective in covering many aspects of competitiveness and basing results on a larger number of firms followed constantly over time. Thereby, the dynamics of firm competitiveness could also be accurately assessed. Such multiscope cross-country surveys could then be linked to administrative and registry data, and trade and foreign affiliate data, exploiting protocols for micro-data linking, as tested, for example, within the GVC project (section 3.3.2.5).

In summary, developing national capabilities in order to better service micro-level data is the most cost-effective and sustainable way to generate new indicators of competitiveness. Once these permanent structures are in place, access by individual researchers to micro-level data or projects based on the distributed micro-data approach could be more feasible. At the same time, given that setting up these capabilities for all EU28 countries will take time and, in some cases, legislation, we also recommend unification and extension of corporate surveys piloted under various projects funded by the European Commission's Seventh Framework and Horizon 2020 programmes. Carefully crafted annual surveys will allow new measures of competitiveness to be constructed and of greater understanding of its dynamics even in the short term.

6 Annex

6.1 Assessment of the indicators of competiveness

The annex provides more detailed information on the concepts of competitiveness, and also provides a technical assessment of the main indicators introduced in section 2.1. We tackle all different aspects highlighted in section 2.1 (fitness, reliability of the statistical techniques, complementarity, micro vs. micro dimensions) within each category. Finally, we provide the shortlist of selected indicators.

6.1.1 Productivity

Productivity measures how efficiently resources are employed. As made clear in our running definition, productivity is the quintessence of competitiveness and indeed the indicators collected here are among the most widely used proxies for competitiveness. Productivity is commonly defined as a ratio of a volume measure of output to a measure of input use. The micro–macro distinction of productivity is crucial.

A common measure, typically used for country-level analysis, is represented by *labour productivity.*

Labour productivity:

- *Description*: Productivity is commonly defined as a ratio of a volume measure of output to a measure of input use:

  $$\text{volume measure of output/measure of labour input use}$$

  Output measures to be used: GDP (Region) or Gross Value Added (country, sector) per hours worked. Labour input measures: number of hours worked and number of people in employment.
- *Rationale:* This indicator measures final production per person of final production per hour worked. Labour productivity offers a dynamic measure of economic growth and competitiveness within an economy. Growing labour productivity depends on

three main factors: investment and saving in physical capital, new technology and human capital.

- *Problems*: The comparability of output measures can be negatively affected by the use of different valuations (inclusion of taxes, different deflation indexes). Labour input can be biased by different methods used to estimate average hours or to estimate employed persons[87,88].

Multi-factor productivity:

- *Description* (1): Multi-factor productivity (MFP) relates output to a combined set of inputs. KLEMS MFP is a productivity measure that relates gross output to primary (capital (K) and labour (L)) and intermediate inputs (energy (E), other intermediate goods (M), services (S)):

$$MFP = \frac{Output}{KLEMS}$$

- *Description (2)*: the OECD MFP growth indicator is computed as the difference between the rate of change of output and the rate of change of total inputs.

$$MFP_i^t = \Delta ln(Q_i^t) - \alpha_i^t \Delta ln(L_i^t) - (1 - \alpha_i^t)\Delta ln(K_i^t)$$

Where α_i^t is the share of labour in total costs in industry *i*, $(1 - \alpha_i^t)$ is the share of capital in total costs, Q_i^t is value-added at constant prices, L_i^t and K_i^t are the labour and capital inputs respectively.

- *Rationale:* In theory, it's a more comprehensive measure than labour productivity. MFP shows the time profile of how productively combined inputs are used to generate gross output. Conceptually, the KLEMS productivity measure captures disembodied technical change. In practice, it reflects also efficiency change, economies of scale, variations in capacity utilisation and measurement errors[89].

The *OECD Multi-factor Productivity index* is a harmonised index that allows for country and sectoral comparisons.

87. "International comparisons of manufacturing productivity and unit labor costs trends". International Labor Comparisons Program. Bureau of Labor Statistics. U.S. Department of Labor.
88. Fleck, S. E. "International comparisons of hours worked: an assessment of the statistics". Monthly Labor Review, May 2009.
89. OECD Manual "Measuring Productivity: measurement of aggregate and industry-level productivity growth".

- *Problems*: Significant data requirements, in particular timely availability of input-output tables that are consistent with national accounts.

Total factor productivity growth:

- *Description*: Total factor productivity (TFP) growth accounts for the changes in output not caused by changes in labour and capital inputs. It is estimated as the residual by subtracting the sum of two-period average compensation share weighted input growth rates from the output growth rate. Log differences of level are used for growth rates, and hence TFP growth rates are Tornqvist indexes (definition from The Conference Board). As such, the output measure is gross value added. In the EUKLEMS database, TFP growth is identically defined.
- *Rationale:* TFP growth represents the effect of technological change, efficiency improvements, and our inability to measure the contribution of all other inputs. It is the closest approximation of productivity growth, which is the ultimate source of growth.
- *Problems*: As it is technically computed as a residual of the growth rates that is not accounted for by inputs growth, TFP growth measures the contribution of all other possible factors.

Total factor productivity (using micro-data):

- *Description*: TFP is calculated from the residual of a production function, where the output variable is production value and the input variables are capital, labour and materials costs. For firm-level productivity, the employed technique is borrowed from Levinshon and Petrin (2003) who employ intermediate inputs to control for correlation between input levels and the unobserved firm-specific productivity process.
- *Rationale:* Accounts for all effects in total output not caused by traditional inputs (labour, capital, materials etc.). Ready for cross-country and/or cross-sector comparison. Overcomes the simultaneity bias that affects standard estimates of firm-level productivity. Better measure of competitiveness than unit labour cost. Change in TFP captures technology catch-up, dynamism.
- *Problems*: Computationally intensive to calculate, and suffers from potential aggregation biases when calculated at the industry or country level.

Olley and Pakes productivity decomposition[90].

- *Description*: Productivity, defined at the industry level and computed as a weighted

90. Olley, S. and Pakes, A. (1996) "The Dynamics of Productivity in the Telecommunications Industry." Econometrica, 64(6), pp. 1263-1298.

average of firm-level productivity, can be decomposed into an unweighted industry average of the firm-level productivity and a covariance term between size and productivity:

$$\Omega_t = \bar{\Omega}_t + \Sigma_i \Delta s_{it} \Delta \omega_{it}$$

Where $\bar{\Omega}_t = \frac{1}{N} \Sigma_i \omega_i$ is the unweighted average of firms productivities, $\Delta s_{it} = s_{it} - \bar{s}_{it}$ and $\Delta \omega_{it} = \omega_{it} - \bar{\omega}_{it}$

- *Rationale:* the covariance term is a cross country comparable measure of the extent to which firms with higher than average productivity, have a higher than average share of activity and indicate the degree of resource misallocation. In fact, if $\Delta s_{it} \Delta \omega_{it}$ is positive, it implies that firms with above average productivity compared to other display above average market shares in a given year. It is a bottom-up approach for a cross-country comparable measure.
- *Problems*: OP decomposition compares productivity allocation across firms in a given year, and hence it does not give a comparison over time.

Foster decomposition of TFP growth[91]

- *Description:*

$$\Delta\Omega_t = \sum_{i \epsilon C} s_{it-k} \Delta \omega_{it} + \sum_{i \epsilon C} \Delta s_{it} (\omega_{it-k} - \Omega_{t-k}) + \sum_{i \epsilon C} \Delta s_{it} \Delta \Omega_{it} +$$

$$+ \sum_{i \epsilon E} s_{it} (\omega_{it-k} - \Omega_{t-k}) - \sum_{i \epsilon X} s_{it-k} (\omega_{it-k} - \Omega_{t-k})$$

Where C = plants that continue their business over time; E = plants that enter at a given time and X = plants that exit; while Ω_{t-k} is the weighted average productivity at the beginning of the period

- *Rationale:* The first three terms of the decomposition are known as the 'within', 'between' and 'covariance' component of firms' contribution in productivity, while the last two terms account for the net entry effects. This decomposition method has two advantages: an integrated treatment of entry/exit and continuing plants (measure of firm dynamics); separating-out within effect (based on plant-level changes) and between effect (that reflects changing shares) from cross/covariance effects. Focusing on the covariance term $\Delta s_{it} \Delta \Omega_{it}$: if this is positive, it means that

91. Foster, L., Haltiwanger, J. and C. J. Krizan (2001), 'Aggregate Productivity Growth. Lessons from Microeconomic Evidence', in: New Developments in Productivity Analysis, 303 372 National Bureau of Economic Research.

firms who are becoming more (less) productive over time are also able to attract more (less) workers; if it is negative or non-significant, then the functioning of the labour market (wage-setting mechanism) contributes negatively to productivity growth.

- *Problems*: While OP decomposition compares productivity allocation across firms in a given year; Foster-type decompositions compare productivity growth within firms over time.

BOX 6.1: OTHER MICRO-FOUNDED PRODUCTIVITY INDICATORS

The *bottom-up approach* is particularly useful while assessing productivity. Productivity measures are, commonly, Pareto distributed and then the average is not a sufficiently significant measure. Thanks to micro (firm-level) data, it is possible to retrieve the *medians and the distribution of productivity measures as TFP, labour productivity, ULC, and mark-ups*. These measures could be combined with the international status of the firm (domestic, exporter, importer, foreign direct investor or owned by a foreign firm) and can be computed for the total economy or by sector.

On top of that, another indicator that comes from a micro-level analysis is the *productivity threshold by international status of the firm*. Since at the micro level a self-selection occurs, an analysis based on productivity cut-offs is helpful to better understand the international status decision of firms. On the other side it could be interesting to analyse, also, the specificities of existing firms that are below the productivity threshold.

6.1.2 Trade competitiveness

Export **market shares** aim at capturing structural gains or losses in competitiveness. At macro level, indicators in this category track the export performance of a country/sector and are often used to check for international imbalances. In fact this broad concept masks different effects: economic growth in destination countries, product differentiation, price vs. non-price competitiveness, imports of intermediates and so on (see for instance a recent decomposition by Guaglier, Taglioni and Zignago, 2013).

Micro-founded indicators in this category are based on the intensive and extensive margin of trade, ie how much each firm exports (imports) and how many firms export (import).

5-year change in export market shares:

- *Description*: percentage change of export market shares over five years, based on balance of payments (Eurostat data)
- *Rationale:* This measure, used also by the Macroeconomic Imbalance Procedure (MIP)[92], aims at capturing structural losses in competitiveness. Export market shares can be driven by the increase/decrease of a country's export volume (numerator effect) but also by the growth of total world exports in goods and services (denominator effect). The five years span allows to measure long-term competitiveness development (non-idiosyncratic trade shocks).
- *Problems*: The main problem of market shares measures is that they are unrelated to competitiveness in a world characterised by global value chains.

Relative trade balance (RTB):

Description: The RTB indicator for product *i* is defined as follows:[93]

$$RTB_i = \frac{(X_i - M_i)}{(X_i + M_i)}$$

Where X= value of exports and M= value of imports is.

- *Rationale:* The relative trade balance (RTB), measures the trade balance relative to total trade in the sector. It is used to rank sectors according to their competitiveness vis à vis the rest of the world and to measure gains and losses in competitiveness over time.
- *Problems*: A negative trade balance is not necessarily a bad sign. Imports can contribute to a country's economy and might stimulate production in other sectors. Also, trade balances are dependent on domestic and foreign demand. This means that this indicator does not exclusively reflect external competitive strength; it also indicates a difference between domestic and international demand[94].

Dieppe *et al* (2012) propose a *decomposition of the trade balance into price and non price competitiveness.* This measure, build on Aiginger (1997), decomposes trade disentangling the respective roles of price and non-price factors allowing to take into account of, among others, quality, product reputation and variety, consumer preferences, etc.

92. Macroeconomic Imbalances Procedure Scoreboard Headline Indicators, 1 November 2012 Statistical information.
93. European Commission Enterprise and Industry: "EU industrial structure 2011 Trends and Performance", chapter iv – international competitiveness of EU industry.
94. Industrial competitiveness "European Competitiveness Report" 15th edition (2012).

- *Description*: Price and non-price determinants of the trade balance are identified at sector/country level through the relative unit values of imports and exports, which are computed out of imports and exports values and quantities. The technique is described in Dieppe *et al* (2012) which builds on Aiginger (1998), where X= value of exports and M= value of imports.
- *Rationale:* This decomposition analysis helps to disentangle the respective roles of price and non-price factors into sectorial/country competitiveness, as identified by the trade balance.

Revealed Comparative Advantages (RCA):

- *Description*: The Revealed Comparative Advantage based on trade is obtained as the fraction of the sector-country export shares over the sector-EU export shares. Other country groups can be used as reference. Formally, for sector i, country j, it is calculated as

$$RCA_i = \frac{X_{j,i}/\Sigma_i X_{j,i}}{X_{world,i}/\Sigma_i X_{world,i}}$$

where X is the value of exports.

- *Rationale:* Compares the share of a given sector's exports in the EU's total manufacturing exports with the share of the same sector's exports in the total manufacturing exports of a group of reference countries. Values higher (lower) than 1 mean that a given industry performs better (worse) than the reference group, and are interpreted as a sign of comparative advantage. The RCA indicator is thus used to rank EU products by comparative advantage. (From International competitiveness of EU industry - DG ENTR[95]).

Current Account as % of GDP:

- *Description*: The Current Account as Percentage of GDP is defined as the sum of the net income from abroad, the net current transfers and the difference between nationwide exports and imports, over GDP.
- *Rationale: "The current account balance determines the exposure of an economy to the rest of the world, whereas the capital and financial account explains how it is financed"* (Eurostat Balance of payment statistics). The indicator tracks imbalances in the nationwide Import/Export and measures the realised competitiveness of an economy.
- *Problems*: The indicator carries endogeneity problems. It also includes non-trade related components.

95. European Commission Enterprise and Industry: "EU industrial structure 2011 Trends and Performance", chapter iv – international competitiveness of EU industry.

BOX 6.2: OTHER MICRO-FOUNDED INDICATORS

Market shares and the international exposure of firms could be investigated also with a bottom-up approach that, starting from micro-level data on firms, outlines the country or sectoral macro outlook. For example, on top of measures of export or import market shares, from micro-data is possible to retrieve the *median or the variance of export (import) share*, and the *distribution of exporting (importing) firms by export (import) share*.

A similar analysis could be conducted by focusing on the *average, the median, and the distribution of the value of exports, value of imports or value of foreign production*. Since firms, normally, are Pareto distributed, these measures provide a good insight into the 'happy few' firms involved in foreign activities or production.

Moreover, to better underline the complexity of foreign operations, other useful measures are the *average, the median and the variance of the number of product exported*.

6.1.3 Price and cost competitiveness

Price and cost competitiveness reflects the ability of firms to sell cheaply in international markets. Among these indicators, we distinguish four main subgroups:

- Real Effective Exchange Rates (REERs), which reflect relative changes in the prices of a country's exports goods due to changes in nominal exchange rates and inflation differentials.
- Unit Labour Costs, which reflect cost competitiveness in an important share of value added.
- Price Cost Margins, which measure the intensity of price competition.

REERs reflect relative changes in the prices of a country's export goods due to changes in nominal exchange rates and inflation differentials. The REER is computed by deflating the Nominal Effective Exchange Rate (NEER), the unadjusted weighted average value of a country's currency relative to all major currencies being traded within a pool of currencies. The NEER can be deflated by selected relative price or cost

deflators, leading to different measures of real exchange rate[96,97,98]. The two suggested ones are the PPI-based REER and the UCLM-based REER.

The PPI-based REER index uses as deflator the producer prices index:

- *Rationale*: is closer to the production side of the economy (includes industrial products and intermediate goods that can be traded internationally) than the CPI; in fact CPI-based index shows the dynamics of relative consumer prices, and hence it can be a rather poor approximation of the dynamics in relative export prices.
- Even though PPI-based REER still includes production for the domestic market, PPIs are viewed as a reasonable proxy for tradable goods prices.
- *Problems*: data on export-oriented PPI are usually very scarce and their composition and compilation varies considerably across countries. It is important to collect comparable measure of PPI at the European level.

The ULCM-based REER index:

- *Rationale*: Unit labour costs in the manufacturing sector (ULCM) are often used as a proxy for unit labour costs in the tradable goods sector. ULCM-based REER is considered a better measure compared to the ULC-based index that usually refers to the total economy, including also the services sector.
- *Problems*: Unit labour costs do not cover all of the costs incurred by firms; factor substitution may affect these indicators without necessarily resulting in a change in productivity. Moreover, as for ULC-based index, cost measures are typically more affected by data quality issues than price measures. The last problem is related to the fact that this popular measure of competitiveness may, however, be too narrow a concept as it only focuses on a certain sector of the economy.

The percentage change over three years of the real effective exchange rate (REER) based on consumer price index deflators:

- *Rationale:* This measure captures the drivers of persistent changes in price and cost competitiveness of each member state relative to its major trading partners, and thus illustrates the magnitude of developments in price and cost competitiveness. The three years span casts a more comprehensive picture of global 'price' pressure on domestic producers in a medium-term perspective

96. Turner P and Van't Dack J. (1993) 'Measuring International Price and Cost Competitiveness', BIS Economic Paper No. 39.
97. Benkovskis Konstantins & Worz Julia (2012) 'Evaluation of Non-Price Competitiveness of Exports from CESEE Countries in the EU Market', Bank of Latvia WP 1/2012.
98. Schmitz, M., De Clercq, M., Fidora, M., Lauro, B. and Pinheiro C., (2012) 'Revisiting the effective exchange rates of the euro', ECB Occasional paper series N. 134, June 2012.

- Other commonly used deflators are: the Consumer Price Index, the Gross Domestic Product, export prices and Unit Labour Costs.

The Unit Labour Cost (ULC):

- *Description*: ULC is calculated as the ratio of total labour costs to real output, or equivalently, as the ratio of mean labour costs per hour to labour productivity (output per hour).
- *Rationale:* ULC represents a link between productivity and the cost of labour in producing output. Unit Labour Costs are seen as one of the most relevant measures of efficiency and aggregate competitiveness. Any increase in added value will translate into a higher level of firm competitiveness, while an increase in the cost of employees would reduce firm's competitiveness. They are easy to compute and are typically used for country level analysis.
- *Problems*: This measure, per contra, presents shortcomings both at the macro and the micro level. At the macro level ULC are not considered to be a comprehensive measure of competitiveness (labour earnings represent just one component of total value added). Moreover, the high heterogeneity across firms induces an aggregation bias. The effect of the aggregation bias on the adequacy of standard aggregate cost measures in capturing export capability can be shown with reference to the so-called Spanish paradox[99]. At the micro level the bias could derive from the fact that 'high-quality' firms might be associated with a higher total cost of employees and thus, if not perfectly reflected in higher added value, in a higher (rather than lower) ULC.

6.1.4 Innovation & technology

The **Innovation & technology** category is fundamental to assess *non-price competitiveness*. Through non-price competitiveness firms try to distinguish their products or services from competitors on the basis of attributes like quality, design or any other sustainable competitive advantage than price. Several indicators are used to determine the rate of firm's innovation.

Innovation & technology, on the other hand, could affect also prices: for example a process innovation could result in a reduction of the production costs, both fixed and variable, of a given good.

99. Altomonte, C., di Mauro, F. and Osbat, C. (2013) 'Going beyond labour costs: How and why "structural" and micro-based factors can help explaining export performance?' CompNet Policy Brief no.1, 15 January 2013.

R&D as percentage of GDP:

Other similar measures take into account public R&D expenditures or business expenditure on R&D.

- *Description*: R&D investments as percentage of total GDP
- *Rationale:* one of the targets of EU2020 is that 3 percent of the EU's GDP should be invested in R&D, since R&D investments foster quality and competitiveness.
- *Problems*: although R&D is related with technical change, it does not measure it. It does not encompass all the efforts of firms and governments in this area, as there are other sources of technical change, such as learning by doing, which are not covered by this narrow definition.

R&D expenditure:

- *Rationale*: Spending more on innovation-enhancing activities enables firms to improve their quality and hence increase their competitiveness. It is also a measure of internal and external knowledge spillovers.
- *Problems*: Although it is obviously related to technical change, it does not measure it. Moreover, R&D does not encompass all the efforts of firms and governments in this area, as there are other sources of technical change, such as learning by doing, which are not covered by this narrow definition.

Patent applications to the European Patent Office (EPO):

- *Description*: Number of patents applied for at the European Patent Office (EPO) per million population, by country and region
- *Rationale*: Aggregate measure of patent applications. Patents are strictly connected to innovation and hence to competitiveness. Spending more on innovation-enhancing activities enables firms to improve their quality and hence increase their competiteness. Is also a measure of internal and external knowledge spillovers. The number of patents granted to a given firm may reflect its technological dynamism; examination of the growth of patent classes can give some indication of the direction of technological change.

EPO patent application per billion GDP (in PPP€):

- *Description*: Number of patents applied for at the European Patent Office (EPO) by year of filing, over Regional GDP in PPP euros. The national distribution of the patent applications is assigned according to the address of the inventor.
- *Rationale:* The capacity of firms to develop new products will determine their competitive advantage. One indicator of the rate of new product innovation is the number of patents. This indicator measures the number of patent applications at the European Patent Office.

License and patent revenues from abroad as % of GDP:

- *Description*: License and patent revenues from abroad as % GDP
- *Rationale*: This indicator reflects a broader definition of innovation. License and patent revenues from abroad capture disembodied technology acquisition. Technology exports reflect the successful commercialisation of close-to-the-frontier technological activities. The number of patents granted to a given country may reflect its technological dynamism; examination of the growth of patent classes can give some indication of the direction of technological change.

On the other side, even though patents and R&D expenditures are good proxies, they are not at all comprehensive measures for innovation and technology. The *Community Innovation Survey* (*CIS*) allows to better investigate, at the micro level, many other aspects of this topic. The Community innovation survey is conducted in every European Union member state to collect data on innovation activities in enterprises, ie on product innovation (goods or services) and process innovation (organisational and marketing aspects)[100,101,102,103].

Non-R&D innovation expenditures (% of turnover):

- *Description*: Sum of total innovation expenditure for enterprises, in thousand Euros and current prices excluding intramural and extramural R&D expenditures over total turnover for all enterprises
- *Rationale:* is an important indicator that targets non-R&D innovation expenditure such as investment in equipment and machinery and the acquisition of patents and licenses. It measures the diffusion of new production technology and ideas.

Enterprises introducing product and/or process innovation (%):

- Description: Number of enterprises who introduced a new product and/ or a new process to one of their markets, over total number of enterprises.

SMEs introducing product or process innovations (% of SMEs):

- *Description*: number of SMEs who introduced a new product or a new process, to one of their markets, over total number of SMEs.
- *Rationale:* Technological innovation, as measured by the introduction of new products (goods or services) and processes, is a key ingredient to innovation in

100. European Commission / Research & Innovation "Innovation Union Scoreboard" (2013 and previous).).
101. Hollanders, H. and Tarantola, S. (2011) Innovation Union Scoreboard – Methodology report.
102. Directorate-General for Enterprise and Industry (DG ENTR) Regional Innovation Scoreboard (2012 and previous).
103. Derbyshire, J., Hollanders, H., Lewney R., Rivera Leon, L., Tarantola, S. and Tijssen R. (2012) Regional Innovation Scoreboard 2012 – Methodology report.

manufacturing activities. The rationale is that higher shares of technological innovators should reflect a higher level of innovation activities and hence higher competitiveness.

Enterprises introducing marketing and/or organisational innovation

- *Description*: Number of enterprises who introduced a new product and/or a new process to one of their markets over total number of enterprises.
- *Rationale*: Many firms, in particular in the service sector, innovate through other non-technological forms of innovation (ie marketing and organisational innovations).

SMEs introducing marketing or organisational innovations (% of SMEs):

- *Description*: Number of SMEs who introduced a new marketing innovation and/or organisational innovation to one of their markets over total number of SMEs.
- *Rationale:* many firms, in particular in the services sectors, innovate through other non-technological forms of innovation (ie marketing and organisational innovations). This indicator tries to capture the extent that SMEs innovate through non-technological innovation.

Intangible investments as percentage of GDP:

- *Description*: Intangible investments are made up of expenditures in the market sector in computerised information (eg software and database), innovative property (R&D, new product/systems in financial services, design etc.) and economic competencies (brand equity or firm-specific resources). The indicator is computed as

$$\frac{Intangible\ Investment}{GDP}$$

(The GDP used in this indicator is corrected for the presence of intangibles[104]).

- *Rationale:* Intangible investments are crucial drivers of knowledge creation. Recent research has shown that these spendings boost productivity and growth and foster a sustainable comparative advantage on knowledge-intensive tasks/products. As part of long-term strategies, these spendings are therefore considered as investments. In addition, high-wage economies are gradually increasing their investments in intangibles with respect to tangibles like buildings or machinery.

104. Corrado, Carol; Jonathan Haskel, Cecilia Jona-Lasinio and Massimiliano Iommi, (2012) 'Intangible Capital and Growth in Advanced Economies: Measurement Methods and Comparative Results', Working Paper, June, available at http://www.intan-invest.net.

Firm level estimates of quality

- *Description*: Firm-level quality indicator estimated using the KSW methodology: the quality for each firm-product-country-year observation can be estimated from a demand function using information on exported product quantities, the value of exports and assuming an elasticity of substitution.
- *Rationale*: The estimated quality for each firm-product-country and year depends on the residual and the elasticity of substitution and the quality-adjusted prices are computed as the difference between unit values and the estimated quality measure. This methodology assigns a higher quality to varieties with higher quantity conditional on prices.

BOX 6.3: OTHER CIS INDICATORS

CIS indicators measure also the effects of innovation and technology; *sales of new-to-market and new-to-firm innovations* is the sum of total turnover of new (either new to the firm or new to the market) or significantly improved products, over total turnover. The indicator captures both the creation of state-of-the-art technologies (new to market products) and the diffusion of these technologies (new to firm products).

Moreover, both *enterprises innovating in-house* and *innovative enterprises collaborating with others* examine the level of cooperation between firms in the innovation process. Complex innovations, in particular in ICT, often depend on the ability to draw on diverse sources of information and knowledge, or to collaborate on the development of an innovation. These indicators measure the flow of knowledge between firms and between public research institutions and firms.

Another important aspect to be investigated is the technological competitiveness of European countries. Creating, exploiting and commercialising new technologies is vital for the competitiveness of a country. *Medium and high-tech product exports as % of total product exports* and *knowledge-intensive services exports as % of total services exports* measure the technological competitiveness of the EU and reflect product specialisation by country.

BOX 6.4: OTHER SUGGESTED INDICATORS

Information and communication technologies (ICT) have fast become integral to EU enterprises. The extensive and intensive use of ICT, combined with new ways of accessing and using the internet efficiently have an important role in the competitive advantage and competitiveness of firms. The Community survey on ICT usage and e-commerce in enterprises is the better source to assess this topic. It is an annual survey conducted since 2002, collecting data on business use of ICT, the internet, e-government, e-business and e-commerce.

Human capital also has a leading role in enhancing innovation; numbers of researchers or employment in knowledge-intensive activities and investments in knowledge are, among others, good proxies for it.

BOX 6.5: SUMMARY INNOVATION INDEX (SII)

The innovation policy initiative PRO INNO Europe has also computed a Summary Innovation Index (SII) that is a composite indicator obtained by an appropriate aggregation of the 25 Innovation Union Scoreboard (IUS) indicators used for measuring innovation performance. The biggest advantage of this indicator is that it gives a composite, harmonised and comparable measure of overall innovation performance for each European country. The drawback is that, being a composite indicator, it does not represent an objective measure for innovation.

BOX 6.6: PROJECTS ON INTANGIBLE ASSETS

INTAN-Invest is the source of the indicator 'Intangible investments as percentage of GDP'. The dataset offers the latest and most comprehensive estimates of intangible investments. It was created in 2011 in a joint effort by Imperial College London, The Conference Board and LUISS Lab of European Economics. The project builds on previous research and estimation conducted by two EU-funded projects (Innodrive and Coinvest) and work done at The Conference Board. INTAN-Invest's contribution is to the harmonisation of different methodologies and the construction of a fully-comparable set of estimates in the cross-country analysis.

Extensive research projects in this field have been financed by the European Commission:

- INNODRIVE – Intangible capital and innovations: drivers of growth and location in the EU (2008-2011): the project tackles the intangible questions from the viewpoint of the firms.
- COINVEST– Competitiveness, innovation and intangible investment in Europe (2008-10): the project contributes to the understanding of intangible investments as drivers of innovation, competitiveness and growth and on supporting the view that they should be treated as investments instead of inputs.
- IAREG – Intangible assets and regional economic growth (2008-10): while developing new indicators, the special focus of this project was on a) the environment affecting firms' location and b) regional externality affecting the accumulation of intangibles.
- MERITUM – Intellectual capital guidelines for firms (1998-2001): the project elaborated a classification of intangibles and contributed in understanding how companies manage and control intangibles and whether these are relevant for equity valuation.

6.1.5 Firm dynamics

Measures for **firm dynamics** cover a crucial aspect in the analysis of competitiveness. The birth of enterprises is thought to enhance the competitiveness of enterprises, by forcing the incumbents to become more efficient. Indeed new entrants stimulate innovation and facilitate the adoption of new technologies, and hence contribute to the increase of overall productivity within an economy[105]. The survival and development over time of firms are important proxies of the dynamism of an economy. Exit is also important as the least-productive firms exit the market freeing up resources for the most productive.

105. Eurostat: Business Demography Statistics.

BOX 6.7: EUROSTAT BUSINESS DEMOGRAPHY

Eurostat business demography collects statistics on the entry rate (or birth rate) of enterprises. This useful measure could be disaggregated at the sector and size level. From a theoretical point of view, enterprise birth is related to the expectation of making profits. If the main objective of newly born enterprises is to make a profit, enterprise births are most likely to occur where profits are consistently high.

The counterpart is represented by the exit rate (or death rate) of enterprises; this measure focuses on less competitive firms in the market that are unable to outlive their competitors.

The analysis should be focused also on firms' survival rate that specifies the proportion of firms from a cohort of entrants that still exist at a given age. The rationale is to understand the post-entry performance and the market selection process that separates successful entrant firms that survive and prosper from others that stagnate and eventually exit. Since it measure the life cycle of newly born enterprises is a good measure of market selection.

Bartelsman *et al* (2004, 2005)[106,107] propose other useful indicators of firm dynamics:

Average firm size relative to entry, by age:

- *Description*: the evolution in average firm size of survivors as they age, corrected for possible changes in entry size of the actual survivors
- *Rationale:* it gives an insight to the gap between the firm size at entry and the average firm size of incumbents. The smaller relative size of entrants can be taken to indicate a greater degree of experimentation, with firms starting smalls and, if successful, expanding rapidly to approach the minimum efficient scale.

Dispersion of firms by size:

- *Description*: Coefficient of variation of firm size, normalised by the overall cross-country coefficient of variation.

106. Bartelsman E., Haltiwanger J. and Scarpetta S. (2004) 'Microeconomic evidence of creative destruction in industrial and developing countries', Policy Research Working Paper Series 3464, The World Bank.
107. Bartelsman E., Haltiwanger J. and Scarpetta S. (2005) 'Measuring and Analyzing Cross-country Differences in Firm Dynamics', paper prepared for NBER Conference on Research in Income and Wealth, Producer Dynamics: New Evidence from Micro-data, 8-9 April, 2005.

- *Rationale:* this indicator helps to see whether cross-country differences in the dispersion differ across sectors of the economy. If technological factors were predominant in determining the heterogeneity of firm size across countries, the values should be concentrated around one. If, on the contrary, the size differences were explained mainly by national factors inducing a consistent bias within sectors, then it would be expected that countries with an overall value above (below) the average are characterised by values generally above (below) one in the sub-sectors.

Another suggested indicator is constructed to measure *Shift and Share Decomposition of average firm size*. Both the size structure and the sectoral composition should be controlled for when analysing firms' dynamics and its effects on aggregate performance. This indicator assesses the role of sectoral specialisation versus within sector differences and is constructed such that: the first term accounts for differences in the sectoral composition of firms, the second for cross-country differences in firm size within each sector and the last represents an interaction term, which can be interpreted as an indicator of covariance: if it is positive, size and sectoral compositions deviate from the benchmark in the same direction.

Share of gazelles

- *Description:* Measured in terms of employment (or turnover), gazelles are enterprises which have been employers for a period of up to five years, with average annualised growth in employees (or in turnover) greater than 20 percent a year over a three-year period and with ten or more employees at the beginning of the observation period. The share of gazelles is expressed as a percentage of the population of enterprises with ten or more employees.
- *Rationale*: a high weight of Gazelles might signal that the most innovative and productive companies find it easy to employ resources and gain markets shares.

6.1.6 Global value chains

Global value chains (**GVCs**) have taken a predominant role in today's global economy and are a fundamental component of firms' competitiveness. Global value chains allow the international dimension and interconnectedness of production processes to be outlined. The growth in intermediate inputs, for example, is one way through which the fragmentation of production and the increasing importance of outsourcing can be tracked. Between 1995 and 2006, trade in intermediate inputs steadily grew at an

average annual growth rate of 6 percent (OECD, 2009)[108]. Moreover, participating in GVCs allows firms to benefit from highly fragmented production processes, complex outsourcing strategies and connections with foreign partners.

Intermediate import ratio:

- *Description*: ratio between the intermediate import amount and the total intermediate demand for each sector. The methodology that measures trade in intermediates is based on *Input-Output Tables*.
- *Rationale:* This indicator is a measure of the geographical fragmentation of production. The intermediate import ratio can be computed also from *OECD-STAN Input-Output dataset*. The advantage is that OECD Input-Output tables are harmonised and comparison among countries is more accessible.

Vertical specialisation (VS) share (import content of exports):

- *Description*: is measured as the share of total intermediate imports used in the production of a country's total exports. Import content of exports is measured using the domestic input coefficients and import matrices of the OECD's harmonised Input-Output Database.

$$Import\ content\ of\ exports = \frac{u\ Am(I - Ad)^{-1}\ Ex}{u\ Ex}$$

- Where Am and Ad are input coefficient matrices (n sectors by n sectors) of imported and domestic goods and services respectively; Ex is the export vector; and u is a (1 by n) vector with all elements equal to 1 [109].
- *Rationale:* VS indicator, proposed by Hummels *et al* (2001)[110], provides a good measure of the importance of the international fragmentation in the production processes. The OECD indicator '*import content of exports*', by using harmonised national input-output tables, computes the countries' degree of vertical specialisation. It measures the contribution that imports make in the production of exports of goods and services.

- *Problems*: one of the drawbacks is that the intensity in the use of imported inputs is assumed to be the same whether goods are produced for export or for domestic

108. Miroudot, S., Lanz, R., Ragoussis, A. (2009) 'Trade in Intermediate Goods and Services', Trade Policy Working Paper No. 93, OECD Publishing.
109. OECD (2011) 'Import content of exports', in OECD Science, Technology and Industry Scoreboard 2011, OECD Publishing.
110. Hummels, D., Ishii, J. and Yi, K. (2001) 'The nature and growth of vertical specialization in world trade', Journal of International Economics, Elsevier, vol. 54(1), pages 75-96, June.

final demand; the measure in fact is computed as the imported intermediate shares of gross production times exports

VS1 - Share of exports sent indirectly through third countries:

- *Description*: VS1 formula for a particular sector i and country k is:

$$\mathrm{VS1} = \sum_{j=1}^{n} (\textit{exported intermediates to country } j) \left[\frac{j\textit{'s exports}}{j\textit{'s gross production}} \right]$$

- *Rationale:* This indicator proposed by Hummels *et al* (1999)[111] is complementary to import content of exports since it captures the other half of the vertical specialisation transaction: VS1 measures the exported intermediates embodied in other countries' exports. The two indicators VS and VS1 together measure upward and downward participation to global value chains.
- *Problems:* VS1 is more difficult to measure than VS, because it requires matching bilateral trade flow data to the input-output relations.

Value added export ratio - domestic value added share of gross exports, % based on OECD_TiVA

- *Description*: EXGRDVA_EX: Value Added Export Ratio - total domestic value added share of gross exports in percent. From OECD TiVA dataset.
- *Rationale:* Measure of the international fragmentation of production, mapping trade flows in terms of value added and measuring the degree of participation in international production chain. Further decomposition of total gross export allows to more sophisticated indicators of participation in the global value chain: the domestic content of exports includes direct value added in export (ie, exported in final goods, exported in intermediate absorbed by final importers), indirect value added in export (ie, exported in intermediate re-exported in third countries) and exported in intermediate that returns in own imports (including double counting term). It complements the Hummels *et al* (2001) measure of global value chain participation from the export perspective.

Value added export ratio – domestic value added share of gross exports, % based on WIOTs

Description:

111. Hummels, Ishii and Yi (1999) 'The nature and growth of vertical specialization in world trade', Staff Reports 72, Federal Reserve Bank of New York.

(imported intermediates/gross output)

It is measured as the share of total domestic intermediate used in the production of a country's total exports. From WIOTs country tables. In order to derive the overall economy imports sum over industry imported inputs; in order to derive overall figures sum over output column; in order to derive overall figures sum over export column. All the measures are available at sector level.

- *Rationale:* The indicator measures the value of domestic inputs in the overall exports of a country, and can be computed on the basis of national input-output tables. It measures to what extent countries are involved in a vertically fragmented production. VS indicator, proposed by Hummels *et al* (2001), provides a good measure of the importance of the international fragmentation in the production processes. The OECD indicator 'import content of exports', by using harmonised national input-output tables, computes the countries' degree of vertical specialisation. It measures the contribution that imports make in the production of exports of goods and services. It is a measure of the international fragmentation of production, mapping trade flows in terms of value added and measuring the degree of participation in international production chain. By using international I-O tables it is possible to overcome the 'proportionality assumption' on which Hummels *et al* (2001) measure was based (ie using the same coefficients for the production sold in the domestic and in the foreign market).

Value added export ratio – foreign value added share of gross exports, %

- *Description*: EXGRDVA_EX): Value Added Export Ratio. Total foreign value added share of gross exports, %. From OECD TiVA dataset.
- *Rationale:* Measure of the international fragmentation of production, mapping trade flows in terms of value added and measuring the degree of participation in international production chain. Further decomposition of total gross export: Foreign content of Export, it includes other countries' domestic content in final goods, in intermediate goods and a double counting term. It corresponds to the VS measure in Hummels *et al* (2001).

Value added export ratio – foreign value added share of gross exports, %

- Description:

(imported intermediates/gross output)

It is measured as the share of total intermediate imports used in the production of a country's total exports. From WIOTs country tables. In order to derive the overall

economy imports sum over industry imported inputs; in order to derive overall figures sum over output column; in order to derive overall figures sum over export column. All the measures are available at sector level.

- *Rationale:* The indicator measures the value of imported inputs in the overall exports of a country, and can be computed on the basis of national input-output tables. It measures to what extent countries are involved in a vertically fragmented production. VS indicator, proposed by Hummels *et al* (2001), provides a good measure of the importance of the international fragmentation in the production processes. The OECD indicator 'import content of exports', by using harmonised national input-output tables, computes the countries' degree of vertical specialisation. It measures the contribution that imports make in the production of exports of goods and services. It is a measure of the international fragmentation of production, mapping trade flows in terms of value added and measuring the degree of participation in international production chain. By using international I-O tables it is possible to overcome the 'proportionality assumption' on which Hummels *et al* (2001) measure was based (ie using the same coefficients for the production sold in the domestic and in the foreign market).

BOX 6.8: OTHER SUGGESTED INDICATORS:

Some other measures of GVCs are based on value added, and hence are more computationally intensive. This is the case for the Ratio of Value Added to Gross Exports (VAX) and for the Domestic Value Added that Returns Home (VS1*). These two indicators summarise the amount of information of Hummels' indicators, but focus on value added, in contrast to many other indicators that use measures of intermediate goods trade or trade in parts and components, as a measure of fragmentation.

Two other useful indicators are the GVC participation index and the GVC position index, suggested by Koopman *et al* (2010). The participation index measures to what extent countries are involved in a vertically fragmented production: the higher the foreign value-added embodied in gross exports and the higher the value of inputs exported to third countries and used in their exports, the higher the participation of a given country in the value chain. In conjunction with this measure the position index define the country position in the GVC as the log ratio of a country's supply of intermediates used in other countries' exports to the use of imported intermediates in its own production. If the country lies upstream in a supply chain, the numerator tends to be large. On the other hand, if it lies downstream, then the denominator

tends to be large. Both these two indices can be computed, starting from input-output tables, for countries and sectors.

Fally (2011) proposes another indicator to measure the relative position in the GVC: the distance to final demand. The distance to final demand can be also interpreted as the length of the value chain when looking forward. The main drawback of this measure is that it comes from the solution of a system of linear equations for each industry i in country k, where the value of interest (D) is a function of D in all other industries and countries.

Bottom-up indicators of GVC:

At the micro level, some measures could help to better account for the inter-connectedness and geographical distribution of production. This is the case for the *distribution of exporting (importing) firms by country of destination (origin)* or for the *distribution of firms with production abroad (foreign affiliate and/or outsourcing) by country of location*. These measures allow also the GVC phenomena to be better depicted, outlining if the interconnectedness is at a global level, or more concentrated at a regional level.

On the other side, by focusing on the number of destination countries by firm, we can assess the complexity of foreign operations. The suggested measures are *the average, the median and the variance of the number of export destination countries per exporting firm* or the *distribution of exporting firms by number of export destination countries.* These indicators estimate the degree of involvement in the global economy.

Finally, firm-level data allows mapping of the ownership and affiliation of domestic registered firms. Thus, we can employ indicators that are normally not available from macro-level surveys such as *Share of foreign owned firms in total firms* and the *Share of domestic MNFs in total firms.*

Measures of foreign direct investment (FDI) should also be mentioned. *Inward* and *outward FDI* indicate the growing transnational ownership of production assets, and they capture different aspects: on the one side is a good proxy for globalisation and international interconnectedness, but on the other is also an indicator of international technological spillovers. We can capture FDI in stocks or flows. Additionally, we can look at the ownership/affiliation of firms from the Eurostat FATS database and retrieve

the following indicators: *Number of foreign-owned firms (affiliates of foreign multinationals), Number of affiliates abroad controlled by domestic firms, Number of domestic firms controlling affiliates abroad*.

6.2 List of sources for macro indicators

MACRO INDICATOR SOURCES

Dataset	Institution
OECD.StatExtracts	OECD
LFS-Strictness of EPL-(regular employment, temporary employment, collective dismissals)	OECD
OECD-WTO Trade in Value Added (TiVA)	OECD
Business Roundtable and PWC - Global Effective Tax Rates	PWC, Business Roundtable
American Enterprise Institute - Report card on effective corporate tax rates	American Enterprise Institute
The WIOD-database	WIOD
Bruegel - Effective Exchange Rate Database	Bruegel
EUKLEMS	EUKLEMS
World Bank - World Development Indicators	World Bank
The Conference Board - TED (Total Economy Database) Growth Accounting and Total Factor Productivity, 1990 - 2012	The Conference Board
UN ComTrade - Export	UN ComTrade
Institute for Fiscal Studies - Corporate tax rate data	Institute for Fiscal Studies
Statistical Data Warehouse	ECB
ESA95 National Accounts	ECB
Monetary and Financial Statistics - Bank Lending Survey - Supply - Enterprises - Q1-Q6	ECB
Monetary and Financial Statistics - Survey on the Access to Finance of SMEs	ECB
Eurostat Database	EUROSTAT
Eurostat Exchange Rates Database	EUROSTAT
Comext Database (Eurostat)	EUROSTAT
Eurostat Short-term Business Statistics Database	EUROSTAT
EU Labour Force Survey (LFS)Database	EUROSTAT
Eurostat Europe 2020 Indicators Database	EUROSTAT
Eurostat R&D expenditure at national and regional level	EUROSTAT
Eurostat Patent Statistics	EUROSTAT
Eurostat Annual National Accounts	EUROSTAT

Eurostat Community Innovation Survey	EUROSTAT
Balance of payments - International transactions	EUROSTAT
Structural business statistics	EUROSTAT
Eurostat Regional Transport Database	EUROSTAT
Eurostat - Tables on EU Policy - Macroeconomic Imbalance Procedure Scoreboard - Export market (tipsex)	EUROSTAT
Eurostat -Quarterly national account database	EUROSTAT
Eurostat - Regional Economic Accounts - ESA95	EUROSTAT
Annual Macro-Economic database	European Commission
Ameco Database	European Commission
Innovation Union Scoreboard-2013	European Commission
ZEW - Effective tax levels	ZEW
Amadeus	Bureau van Dijk
INTAN-Invest	INTAN-Invest consortium

6.3 The MAPCOMPETE meta-database

Table 6.1: Structure of the MAPCOMPETE Meta Database

Indicators description	Indicator split	Indicators computability	Variables	Sources	People
IndicatorID	IndicatorsID	IndicatorsID			Country
Reference	SubindicatorID	SubindicatorID			Title
Priority	Aggregation	ComputabilityID			Surname
Category		Country			Name
Indicator name		Degree of computability			Function
Type		Time	Variables		InstitutionID
Description		VarID_1	VarID		Institution
Rationale		VarID_2	Country		Existing contact?
Problems		VarID_3	Description		Email
Mapped		VarID_4	Time	**Sources**	Telephone
CompNet info		VarID_...	SourceID	SourceID	Skype
		Notes	Accessibility	Country	
			Disaggregation	Dataset	**Institutions**
			Notes	InstitutionID	InstitutionID
				Institution	Institution
				URL	Country
				Contact person	Main aim/function
				Topics covered	Street
				Temporal scope	Postal code
				Sectoral availability	City
				Regional availability	URL
				Publications	E-Mail
				Thresholds	Telephone
				Accessibility	Notes
				Statistical unit	
				Number of observations	
				Periodicity	
				Type of data	

The MAPCOMPETE meta-database (henceforth metaDB) have been defined in order to be able to organise information on availability and accessibility of data needed to construct the indicators of competitiveness, and with the goal of forming the basis for a web tool[112], which will visualise the information collected by MAPCOMPETE. The logic of the metaDB is as follows. Each of six tables contains a specific set of information and the tables are all connected to one another. Table 6.1 illustrates the contents of each table and the links between them.

IndicatorsDescription: is the table with the info on the indicators proposed. It contains:

- *IndicatorID* is a unique alphanumeric indicator, labelled as I_x (where the prefix I_ stands for indicator, and x is numerical identifier)
- Reference, which is the paper(s) where an indicator was mentioned
- *Category*, which indicates in which macro category of competitiveness an indicator falls into (eg price, productivity, GVC, firm dynamics, innovation, labour)
- *Indicator Name*
- *Type*, whether the indicator is macro, micro, sectoral, regional, a combination
- *Description*
- *Rationale* of the indicator as suggested by the literature

This table has **one row for each indicator**, and it is connected directly to 'IndicatorSplit' and 'IndicatorsComputability'.

IndicatorsSplit: is a table in which we split each indicator into the different levels of aggregation. Levels of aggregation can be: country, sector, region and bottom-up. The latter refers to indicators computed from firm-level data aggregated up at the sector, region and country level. For each 'Indicator_ID', there will be N 'Subindicator_ID'.

IndicatorsComputability: is a table where we check **to what extent an indicator is computable for a given country**. It contains C rows for each indicator (where C is the number of countries) and the following information:

- *Sub-IndicatorID*: same as above. They link IndicatorsComputability with IndicatorsDescription and IndicatorsSplit.
- *ComputabilityID*: alphanumeric indicator with the following structure C_XX_x, where C stands for Computability, XX is a two-letter code for a country, and x is the numeric identifier , as in IndicatorsID.
- *Country.*

112. The web tool can be reached at www.mapcompete.eu

- *Frequency*.
- *Disaggregation (country, sector, region)*.
- *Degree of computability*: a synthetic code for whether for a given country an indicator can be computed or not. We have opted for a three value scale: high, medium and low, to allow the fact that an indicator can be computed completely or partially.
- *Time*: the time span for which an indicator can be computed.
- *Notes*: any useful information on a given indicator-country pair.
- *VarID_1-VarID_20*: this is a key aspect of the structure of the dataset. Each indicator needs to be computed from some underlying variables, which may or may not come from the same source and for the same time period. We allow for the fact that each indicator can be computed from up to 20 underlying variables. For example, in order to compute the 'Unit Labour Cost of Manufacturing (UCLM)-based REER', we need bilateral trade flows, exchange rate, compensation per employee, value added per employee. In order to compute TFP, we need info on value added, tangible capital, number of employees. Variables are identified as V_XX_z, where V stands for variable, XX is the two-letter country code, and z is a numeric identifier for the variable. VarID_1-VarID_20 can be linked with VarID in the table variables.
- *One to one*: **in most cases an indicator uses information from two or more variables**. However, there are cases where an indicator is already available "as it is". This is the case of: TFP (total factor productivity) (Macro) (I_22), Number of hours worked (I_37), Participation of adults aged 25-64 in education and training by NUTS 2 regions % (I_40), Participation in lifelong learning of employed persons by sector (I_43), Training enterprises (%) (I_50), R&D as Percentage of GDP (I_58), EPO patent applications per billion GDP (in PPP€) (I_68), Non-R&D innovation expenditures (% of turnover) (I_70), SMEs introducing product or process innovations (% of SMEs) (I_73), SMEs introducing marketing or organisational innovations (% of SMEs) (I_75). In such cases there is a **one-to-one correspondence between indicators and variables**.

Variables: is a table providing info on sources and availability for each variable needed to construct a given indicator. Note that variables are country-specific. For example, compensation per employee (needed to construct the ULC) appears C times, one for each country.

VarID as above
Country
Description
Time: time coverage

SourceID: identifier of the Source for a given variable
Accessibility: indicates the extent of availability of the info on a given the variable
Disaggregation: indicates the level of aggregation (sectoral, regional, aggregate, firm)

Notes: contains any element that can be useful and does not fit the previous fields.

The dataset is complemented with information on the **Sources**, which are the specific data bases where one can get info on a specific variable, the **Institutions** which publish those sources and, eventually, **People** we may be in contact in the various institutions.

6.4 Detailed tables and comments for chapter 2.2

This section illustrates the results of the mapping of the degree of computability of indicators of competitiveness computed from aggregate data (section 2.2). As shown in Table 6.2, the degree of computability of indicators belonging to the exchange rate category is in general very good for most EU countries, since data is available since before the year 2000. Some relevant exceptions are nevertheless worth noting: Croatia, which shows good availability for only the change in CPI-based REER, while the availability for PPI-based REER is since only 2010 and no data are available for ULCM-based REER; Portugal, which lacks data on PPI-REER, since information on PPI is not available.

Information on PPI-based REER is available only since after 2000 for Bulgaria, Cyprus, Spain, Italy, Malta, Poland, Romania, 2001 for Belgium, 2003 for Slovakia, 2005 for Ireland.

ULCM based REER is available since after 2000 for Spain, Greece, Lithuania, Latvia, since 2004 and 2008 for Poland and Romania respectively, both at the country and sectoral level.

The information concerning Unit Labour Cost is available for all the EU28 members and with a good time span. For the EU15 countries, the availability starts from 1960/1970 while for all the other countries (Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Hungary, Lithuania, Latvia, Malta, Poland, Romania, Slovenia and Slovakia) the availability starts from 1990/1996.

The picture for the firms' dynamics indicators is mixed (see Table 6.3): a large time interval (since before 2000) is available for Spain, Finland, Italy, Luxembourg, Netherland, Portugal, Sweden and UK, while the availability is more restricted (since

after 2000) for the rest of the countries and no information at all is available for Croatia, Ireland and Malta.

As shown in Table 6.4, the indicators of productivity are generally available for all the EU countries for a large time span when the national level is considered.

An exception is Croatia, for which not only are all these indicators not available at the sectoral and the regional level, but data is missing at the country level for the aggregate labour productivity based on hours worked.

It is worth noting also that for some countries, ie Estonia, Luxembourg, Latvia, Greece, Malta and Slovenia, aggregate labour productivity based on hours worked is available for a shorter interval, since after 2000 (2002 for Luxembourg).

The availability is worse and less systematic when the sectoral and the regional levels are considered.

In particular, a large number of countries (Bulgaria, Croatia, Cyprus, Estonia, Greece, Lithuania, Luxembourg, Latvia, Malta, Poland, Portugal, Romania, Slovakia) lack data on the TFP growth rate at the sectoral level. A smaller, but relevant number of countries lacks information on the aggregate labour productivity based on hours worked at the sectoral level (Croatia, Cyprus, Ireland, Malta, Sweden), while data on aggregate labour productivity based on number of employees at the sectoral level is missing for Croatia, Cyprus, Hungary, Ireland, Malta, Poland and Sweden.

Data on the aggregate labour productivity based on hours worked at the regional level is missing for Croatia only. Data on the aggregate labour productivity based on number of employees at the regional level is missing for Italy, Estonia, Croatia, Belgium.

For some other countries, information on these indicators is available, but for a shorter time interval with respect to the rest of the countries: the aggregate labour productivity based on hours worked at the sectoral level is available since 2000 for Bulgaria, Greece, Lithuania, Latvia, Poland, since 2001, for Spain, since 2002, for Luxembourg, while at the regional level is available since 2000 for Austria, Finland, Greece, Hungary, Italy, Romania, and since 2001 for the Netherlands.

The aggregate labour productivity based on number of employees at the sectoral level is available since 2000 for Bulgaria, Estonia, Greece, Lithuania, Slovenia, since 2001 for Spain, since 2002 for Luxembourg, since 2003 for Latvia, since 2006 for Portugal,

while at the regional level is available since 2000 for Spain, Finland, Greece, Hungary, Latvia, Malta, Portugal, Romania, UK, since 2001 for the Netherlands, since 2002 for Luxembourg and 2008 for Poland and Slovenia. The indicators belonging to the Trade Competitiveness group are homogenously computable across the EU countries (see Table 6.5).

The 5-Year Change in Export Market Shares at the country level is provided for a large time span (at least 1997-2012) for all EU countries. As for the sectoral level, the same indicator is available for all countries since at least 1999, with the exceptions of Bulgaria (2001) and Luxembourg (2003). The Relative Trade Balance is available (4 digit level) for all the countries, monthly, since only 2002. The Decomposition of the Trade Balance in Price and non-Price Competitiveness is available at both country and sectoral level for a large time span (since before 2000) for all the EU countries. A similar availability applies to the Current Account as a Percentage of GDP (since before 1995 depending on the country). On the other side, the Revealed Comparative Advantage (RCA) at the sectoral level (4 digits) is available since 2002 only for all EU countries.

The Intangible Investments at the country level are available for a large time interval for all the EU countries with only some exceptions (see Table 6.6). Croatia has no information, while Greece, Luxembourg and Portugal provide information since 2000 instead of 1995 like most of the other countries.

The availability of the other two indicators considered, Loans to enterprises and Loan application success/failure, is not good in most of EU countries. For the former, there is no data available for Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Finland, Greece, Hungary, Lithuania, Latvia, Poland, Romania, Sweden, UK; data is available since 2003 for the rest of the countries.

The picture is worse when the loan application success/failure is considered, since there is no information for most countries, with the only exceptions being Germany, Spain, France, Italy, for which the indicator is available since only recently (2009-12).

Tables 6.7 annd 6.8 show that, as for the indicators providing comparable information across countries on inward and outward FDI, the coverage for EU countries is quite good.

This holds in particular with regard to the country-level indicators of both inward and outward FDI, both flows and stocks, which are available for a large time interval both annually and quarterly (since before the 1990s for most countries). The only exception

is Luxembourg for which both the flows are available since only 2002. The same indicators are available at the sectoral level with the only difference that for some countries the time interval is shorter; ie inward and outward FDI flows at the sectoral level for Belgium, Luxembourg, Slovenia, Slovakia, Malta and Romania are available since only after 2002, depending on the country (Ireland lacks data on outward FDI flows at the sectoral level before 2002); while inward and outward FDI stocks at the sectoral level are available since only after 2001, depending on the country, for Belgium, Cyprus, Ireland, Romania, Slovenia, Spain.

Information on both the number of foreign-owned firms (affiliates of foreign multinationals) and the number of affiliates abroad controlled by domestic firms is definitely worse in terms of time span for several countries. The number of foreign-owned firms at both the country and sectoral level is available since only 2001 for Austria, since 2007 for Belgium, since 2003 for Bulgaria, Estonia, Lithuania, Latvia, Romania, Slovenia, and Slovakia, since 2004 (and 2007 sectoral) for Cyprus, since 2008 for Malta, while there are no data for Greece and for Hungary data are available since 2003 at the sectoral level. The number of affiliates abroad controlled by domestic firms at both the country and sectoral level is good for only a few countries (Austria, Germany, Italy, Portugal and the Czech Republic), while it is available only for recent years in Belgium, Bulgaria, Cyprus, Denmark, Estonia, Finland, France, Hungary, Netherlands and Sweden (since 2007), Greece and Lithuania (since 2004), Slovakia (since 2005), Ireland (since 2010), Latvia (since 2006), Poland and Romania (since 2008), Spain and UK (since 2009), Slovenia in the time interval 2007-09. Data for Luxembourg is available in 2005 and then since 2009.

Tables 6.9 and 6.10 show the availability of information on EU countries' involvement in the global value chain as computed from the OECD-TiVA International Input-Output tables and from the WIOT tables. In both cases, the computability is high and comparable across all EU countries at both country and sectoral level. In particular, the domestic value added share of gross export and the foreign value added share of gross export are available for 1995-2000-2005-2008-2009 from the OECD-TiVA tables, while they are continuously available since 1995 to 2011 from the WIOT tables (Table 6.10).

The group of indicators on innovation activities for both all firms and SMEs are computable through the data provided by Eurostat based on the CIS survey, which is carried out in most EU countries. Nevertheless, both the number of waves available and what is publicly available through Eurostat varies across countries.

As for the innovation activity, indicators without distinguishing by firm size, the availability of comparable data is good when both country and sectoral level are considered, while data on the regional level is not available for any countries (see Table 6.11).

At the country and sectoral level, most indicators are available for at least four CIS waves, but there are nevertheless some exceptions. The most notable concern Finland, France, Greece, Croatia, Latvia, Sweden, Slovenia and the UK, for which data is available for three or less than three waves for the majority of the indicators at both sectoral and country level.

In particular, this applies to non-R&D innovation expenditures, on which information is available in 2008-2010 for Austria, Latvia and France; in 2006-2008-2010 for Croatia and Slovenia, in 2000 only for UK, in 2000-2008-2010 for Finland and Italy, in 2000-2004-2006 for Greece and in 2000-2008 for Germany, while for Sweden, 2004-2008-2010 are available. As for Enterprises introducing marketing and/or organisational innovations, data are available in 2004-2008-2010 for Belgium, France, Ireland, Italy, Slovakia and Spain, 2006-2008-2010 for Croatia, in 2008-2010 for Finland, Latvia, Slovenia, Sweden and UK and in 2004-2006 for Greece. A similar picture emerges for the Enterprises innovating in-house, for which data are available in 2006-2008-2010 for Croatia and Slovenia, in 2000-2004-2006 for Denmark and Greece, in 2000-2004-2008 for France, in 2004-2006-2008 for Spain, in 2004-2006 Ireland, in 2008-2010 Latvia and no data are available on this indicator for UK.

As for innovative enterprises collaborating with others and sales of new-to-market and new-to-firm innovations, the availability is good (four waves or more), with only a few exceptions. For the former, data is available in only 2006-2008-2010 for Croatia, in 2004-2008-2010 for France, in 2004 and 2006 only for Greece. As for the latter, Croatia and Finland show data in 2006-2008-2010, France and Sweden in 2004-2008-2010, while the UK in 2004-2006-2008; data is available in only 2004-2006 for Greece. As for innovative enterprises collaborating with others, it is worth underlining the availability at the sectoral level since does not always coincides with the one at the aggregate level; at the sectoral level information is available in 2004-2006-2008 for Cyprus and France, in 2006-2008-2010 for Ireland, in 2004-2008-2010 for Sweden.

For the enterprises introducing product and/or process innovations, there is a good availability in most countries, the only exceptions regarding Croatia and Greece for which data are available in only 2006-2008-2010 (Croatia) and 2000-2004-2006 (Greece).

As shown in Table 6.12, when the same indicators are considered for the subsample of small and medium sized firms (SMEs), the picture worsens significantly: not only is information not available at the regional level for all EU countries, like in the case of the all firms' sample, but information is also missing at the sectoral level for all countries. For some indicators there is information at 2 digit level only for 2000.

As for the country level, the availability of information on innovation indicators in SMEs is quite good for all countries (ie four waves or more are available) with only a few exceptions. Information on SMEs introducing product and/or process innovations is available in only 2006-2008-2010 for Croatia and in 2000-2004-2006 for Greece; as for SMEs introducing marketing and/or organisational innovations information is available in 2004-2008-2010 for Belgium, France, Ireland, Italy, Slovakia and Spain; in 2006-2008-2010 for Croatia, in 2008-2010 for Finland, Latvia, Slovenia, Sweden and UK; in 2004-2006 for Greece.

The availability of information on the share of SMEs innovating in-house is less homogeneous for EU countries: in 2004-2006-2008 for Cyprus, in 2006-2008-2010 for Croatia and Luxembourg, in 2000-2004-2006 for Denmark and Greece, In 2000-2004-2008 for France; in 2004 and 2006 only for Ireland, in 2008-2010 for Latvia, Malta and Slovenia; in 2000-2008 for Spain while no data are available for UK.

Innovative SMEs collaborating with others is widely and comparably computable across EU countries for a large time span (at least four waves) with the only exceptions of Croatia, for which data are available in 2006-2008-2010 only, France, in 2004-2008-2010, and Greece in 2004-2006.

The four indicators patent applications to the European Patent Office (EPO), EPO patent applications per billion GDP (in PPP€), License and patent revenues from abroad as % GDP, and EU Summary Innovation Index (SII) show a quite good degree of computability across EU countries, while the picture varies when we look at the other two indicators R&D as Percentage of GDP and R&D Expenditure.

As shown in Table 6.13, the availability of information is good for all the EU countries when looking at the country level for patent applications to the European Patent Office (EPO) and EPO patent applications per billion GDP (in PPP€), ie, patent applications to the European Patent Office (EPO) are computable for most of the countries since the late 1970s and for all in any case since before 2000, while EPO patent applications per billion GDP (in PPP€) since the second half of the 1990s.

At the regional level, again for all countries the time span is shorter, being in between 2000 and 2009 for all countries for the patent applications to the European Patent Office (EPO) and since 2000 to now for the EPO patent applications per billion GDP. It is worth noting that there are no data at the regional level for Croatia in both cases.

On the other side, license and patent revenues from abroad as % of GDP and EU Summary Innovation Index (SII) are computed and comparable across countries for all EU countries since 2004 (2006 for Spain and Greece), as for the former, and since 2008, the latter.

Turning the attention to R&D as percentage of GDP and R&D Expenditure at the country level, the availability of information is good for most countries since before 2000, with only some exceptions: Croatia and Malta (since 2002), and Greece, Luxembourg and Sweden showing a large discontinuity in the availability of data.

The sectoral level data on R&D as Percentage of GDP is continuously available for a time span since before 2000 for only Belgium, Bulgaria, Cyprus, Estonia, Ireland, Netherlands, Poland, the Czech Republic, Romania, Slovakia, Slovenia, Sweden and Hungary. For the rest of the countries the information is limited to a shorter time interval, ie Latvia, Lithuania and Portugal (since 2000), Finland, Germany, Greece, Italy and Spain (since 2001), Croatia (since 2002), Denmark, France and the UK (since 2007), and/or quite discontinuous in time (Austria and Malta), with several missing data.

At the regional level, the information is continuously available for a large time span (since before 2000) for only Cyprus, Estonia, Finland, Latvia, Lithuania, Portugal, Spain and Hungary; for the rest of the countries, information is given for a shorter time interval, ie Slovakia and Poland (since 2000), the Czech Republic and Romania (since 2001), Bulgaria, Ireland and Malta (since 2002), Slovenia (since 2003), the UK (since 2005), Belgium (since 2006), and/or quite discontinuous in time (Austria, Croatia, France, Germany, Greece, Italy, Luxembourg, Netherlands, Sweden).

The sectoral level data on R&D expenditures is continuously available for a time span since before 2000 for only Austria, Belgium, Cyprus, Estonia, Ireland, Netherland, Poland, the Czech Republic, Romania, Slovakia, Slovenia, Sweden and Hungary. For the rest of the countries the information is available for only a shorter time interval, ie Latvia, Lithuania and Portugal (since 2000), Finland, Germany, Greece, Italy and Spain (since 2001), Croatia (since 2002), Denmark, France and the UK (since 2007), Luxembourg, (since 2009) and/or quite discontinuous in time (Bulgaria and Malta).

At the regional level, information on R&D expenditures are continuously available since before 2000 for only Cyprus, Estonia, Finland, France, Latvia, Lithuania, Portugal, Spain and Hungary, while for the rest of the countries the information is limited to a shorter time interval, ie, Poland and Slovakia (since 2000), the Czech Republic and Romania (since 2001), Belgium, Ireland and Malta (since 2002), Slovenia (since 2003), Croatia (since 2008) and/or quite discontinuous in time (Austria, Bulgaria, Denmark, Germany, Greece, Italy, Luxembourg, Netherlands, Sweden, the UK).

Table 6.2: Macro-level indicators: price and cost – exchange rate and ULC

Index/level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_010_01 Country	2	1	1	1	2	2	2	2	2	1	2	2	0	2	1	1	2	2	2	1	2	1	0	1	2	2	1	2
I_011_01 Country	2	2	2	2	2	2	2	2	1	1	2	2	0	2	2	2	1	2	1	2	2	1	2	1	2	2	2	2
I_011_02 Sector	2	2	2	2	2	2	2	2	1	1	2	2	0	2	2	2	1	2	1	2	2	1	2	1	2	2	2	2
I_012_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_013_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

I_010: Producer Price Index (PPI)-based REER
I_011: Unit Labour Costs of Manufacturing (ULCM)-based REER
I_012: % change (3 years) in REER based on consumer price index (CPI) deflators
I_013: Unit Labour Cost (ULC)

Table 6.3: Macro-level indicators: firm dynamics

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_051_01 Country	1	1	1	1	1	1	1	1	1	2	2	1	0	1	0	2	1	2	1	0	2	1	2	1	2	1	1	2
I_051_02 Sector	1	1	1	1	1	1	1	1	1	2	2	2	0	1	0	2	1	2	1	0	2	1	2	1	2	1	1	2
I_056_01 Country	1	1	1	1	1	1	1	1	1	2	2	1	0	1	0	2	1	2	1	0	2	1	2	1	2	1	1	2
I_052_02 Sector	1	1	1	1	1	1	1	1	1	2	2	1	0	1	0	2	1	2	1	0	2	1	2	1	2	1	1	2

I_051: Entry rate (birth rate)
I_052: Exit rate (death rate)

Table 6.4: Macro-level indicators: labour productivity and Total Factor Productivity

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_001a_01 Country	2	2	2	2	2	2	2	1	1	2	2	2	0	2	2	2	2	1	1	1	2	2	2	2	2	1	2	2
I_001a_02 Sector	2	2	1	0	2	2	2	2	1	1	2	2	0	2	0	2	1	1	1	0	2	1	2	2	0	2	2	2
I_001a_03 Region	1	2	2	2	2	2	2	2	1	2	1	2	0	1	2	1	2	2	2	2	1	2	2	1	2	2	2	2
I_001b_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_001b_02 Sector	2	2	1	0	2	2	2	1	1	1	2	2	0	0	0	2	1	1	1	0	2	0	1	2	0	1	2	2
I_001b_03 Region	2	0	2	2	2	2	2	0	1	1	1	2	0	1	2	0	2	1	1	1	1	1	1	1	2	1	2	1
I_002_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_002_02 Sector	2	2	0	0	2	2	2	0	0	2	2	2	0	2	2	2	0	0	0	0	2	0	0	0	2	2	0	2

I_001a: Aggregate labour productivity based on hours worked
I_001b: Aggregate labour productivity based on number of employees
I_002: Aggregate TFP (total/multi factor productivity) growth

Table 6.5: Macro-level indicators: trade competitiveness

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_006_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_006_02 Sector	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2
I_007_02 Sector	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
I_008_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_008_02 Sector	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_064_02 Sector	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
I_065_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

I_006: 5 Year Change in Export Market Shares
I_007: Relative Trade Balance
I_008: Decomposition of the trade balance into price and non-price
I_064: Revealed Comparative Advantage (RCA)
I_065: Current Account as Percentage of GDP

Table 6.6: Macro-level indicators: intangible assets and financial activity

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_057_01 Country	1	0	0	1	0	1	0	1	0	1	0	1	0	0	1	1	0	1	0	1	1	0	1	0	0	1	1	0
I_058_01 Country	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_066_01 Country	2	2	2	2	2	2	2	1	2	2	2	0	2	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2

I_057: Loans to enterprises
I_058: Loan applications success/failure
I_066: Intangible Investments as Percentage of GDP

Table 6.7: Macro-level indicators: inward FDI

Index/level	AT	BE	BG	CY	CZ	DE	DK	EE	ES	FI	FR	GR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_041a_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2
I_041a_02 Sectoral	2	1	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	1	1	1	2	2	2	1	2	1	1	2
I_041b_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_041b_02 Sectoral	2	1	2	1	2	2	2	2	1	2	2	2	0	2	1	2	2	2	2	2	2	2	2	1	2	1	2	2
I_041c_01 Country	1	1	1	1	2	2	2	1	2	2	2	0	0	2	2	2	1	2	1	1	2	2	2	1	2	1	1	2
I_041c_02 Sectoral	1	1	1	1	2	2	2	1	2	2	2	0	0	1	2	2	1	2	1	1	2	2	2	1	2	1	1	2

I_041a: Inward FDI flows
I_041b: Inward FDI stock
I_041c: Number of foreign-owned firms (affiliates of foreign multinationals)

Table 6.8: Macro-level indicators: outward FDI

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_042a_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2
I_042a_02 Sector	2	1	2	2	2	2	2	2	1	2	2	2	0	2	1	2	2	1	2	1	2	2	2	1	2	1	1	2
I_042b_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_042b_02 Sector	2	1	2	1	2	2	2	2	2	1	2	2	0	2	1	2	2	2	2	1	2	2	2	1	2	1	2	2
I_042c_01 Country	2	1	1	1	2	2	1	1	1	0	1	1	0	1	0	2	1	1	1	1	1	1	2	1	1	0	1	0
I_042c_02 Sector	2	1	1	1	1	2	1	1	1	0	1	1	0	1	1	2	1	1	1	0	1	1	2	1	1	0	1	0

I_42a: Outward FDI flows
I_42b: Outward FDI stock
I_42c: Number of affiliates abroad controlled by domestic firms

Table 6.9: Macro-level indicators: global value chains

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_039a_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_039a_02 Sector	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_040a_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_040a_02 Sector	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

I_039a: Value Added Export Ratio - domestic value added share of gross exports, % - OECD TiVA
I_040a: Value Added Export Ratio - foreign value added share of gross exports, % - OECD TiVA

Table 6.10: Macro-level indicators: global value chains

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_039b_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_039b_02 Sector	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_040b_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_040b_02 Sector	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

I_039b: Value Added Export Ratio - domestic value added share of gross exports, % - WIOT
I_040b: Value Added Export Ratio - foreign value added share of gross exports, % - WIOT

Table 6.11: Macro-level indicators: innovation activity, all firms

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_027_01 Country	0	2	2	2	2	0	2	2	1	2	1	0	1	2	2	1	2	2	0	2	2	2	2	2	1	1	2	0
I_027_02 Sector	0	2	2	2	2	1	2	2	1	2	1	2	1	2	2	2	2	2	1	2	2	2	2	2	1	1	2	0
I_027_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_028_01 Country	2	2	2	2	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_028_02 Sector	2	2	2	2	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_028_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_030_01 Country	2	1	2	2	2	2	2	2	0	1	0	1	1	2	1	1	2	2	0	2	2	2	2	2	0	0	1	0
I_030_02 Sector	2	1	2	2	2	2	2	2	0	1	0	1	1	2	1	1	2	2	0	2	2	2	2	2	0	0	1	0
I_030_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_032_01 Country	2	2	2	2	2	2	1	2	1	1	2	1	1	2	0	2	2	2	0	2	2	2	2	2	2	1	2	0
I_032_02 Sector	2	2	2	2	2	2	0	2	1	2	2	1	1	2	0	2	2	2	0	1	2	2	2	2	2	2	2	0
I_032_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_035_01 Country	2	2	2	2	2	2	2	2	0	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_035_02 Sector	2	2	2	1	2	2	2	2	0	2	2	1	2	2	1	2	2	2	2	2	2	2	2	2	1	2	2	2
I_035_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_036_01 Country	2	2	2	2	2	2	2	2	0	2	1	1	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1
I_036_02 Sector	2	2	2	2	2	2	2	2	0	2	1	1	1	2	1	2	2	2	2	2	2	2	2	2	1	2	2	2
I_036_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

I_027: Non-R&D innovation expenditures (% of turnover)
I_028: Enterprises introducing product and/or process innovations (%)
I_030: Enterprises introducing marketing and/or organisational innovations (%)
I_032: Enterprises innovating in-house (%)
I_035: Innovative enterprises collaborating with others (%)
I_036: Sales of new-to-market and new-to-firm innovations as % of turnover

Table 6.12: Macro-level indicators: innovation activity, SMEs

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_029_01 Country	2	2	2	2	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_029_02 Sector	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_029_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_031_01 Country	2	1	2	2	2	2	2	2	0	1	0	1	1	2	1	1	2	2	0	2	2	2	2	2	0	0	1	0
I_031_02 Sector	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_031_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_033_01 Country	2	2	2	1	2	2	1	2	1	0	2	1	1	2	0	2	2	1	0	0	2	2	2	2	2	0	2	0
I_033_02 Sector	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_033_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_034_01 Country	2	2	2	2	2	2	2	2	0	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_034_02 Sector	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I_034_03 Region	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

I_029: SMEs introducing product and/or process innovations (% of SMEs)
I_031: SMEs introducing marketing and/or organisational innovations (% of SMEs)
I_033: SMEs innovating in-house (% of SMEs)
I_034: Innovative SMEs collaborating with others (% of SMEs)

Table 6.13: Macro-level indicators: R&D expenditure and output

Index/Level	AT	BE	BG	CY	CZ	DE	DK	EE	ES	FI	FR	GR	HR	HU	IE	IT	LT	LU	LV	MT	NL	PL	PT	RO	SE	SI	SK	UK
I_022_01 Country	2	2	2	2	2	2	2	2	2	2	2	1	1	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2
I_022_02 Sectoral	1	2	2	2	2	1	0	2	1	1	0	1	1	2	2	1	1	0	1	0	2	2	1	2	2	2	2	0
I_022_03 Regional	1	1	1	2	1	1	0	2	2	2	1	0	1	2	1	1	2	1	2	1	1	1	2	1	0	1	1	1
I_023_01 Country	2	2	2	2	2	2	2	2	2	2	2	1	1	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2
I_023_02 Sectoral	2	2	1	2	2	1	0	2	1	1	0	1	1	2	2	1	1	0	1	1	2	2	1	2	2	2	2	0
I_023_03 Regional	1	1	1	2	1	1	0	2	2	2	2	1	0	2	1	1	2	1	2	1	1	1	2	1	0	1	1	1
I_024_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_024_03 Regional	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
I_025_01 Country	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
I_025_03 Regional	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
I_026_01 Country	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
I_037_01 Country	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

I_022: R&D as Percentage of GDP
I_023: R&D Expenditure
I_024: Patent applications to the European Patent Office (EPO)
I_025: EPO patent applications per billion GDP (in PPP€)
I_026: License and patent revenues from abroad as % GDP
I_037: EU Summary Innovation Index (SII)

6.5 Detailed tables for Chapter 2.3

Table 6.14: Bottom-up indicators: labour productivity

	Computability									Accessibility							
	I_001_04	I_001_05	I_001_06	I_001_07	I_001_08	I_001_09	I_001_10	I_001_11	I_013_02	I_001_04	I_001_05	I_001_06	I_001_07	I_001_08	I_001_09	I_001_10	I_001_11	I_013_02
Austria	1	9	1	1	9	9	9	9	1	0	9	0	0	9	9	9	9	0
Belgium	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	2
Bulgaria	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1
Croatia	0	0	0	0	0	0	0	0	0	9	9	9	9	9	9	9	9	9
Czech Rep.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Denmark	2	1	2	2	0	0	1	1	2	1	1	1	1	9	9	1	1	1
Estonia	2	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1
Finland	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1
France	2	1	2	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1
Germany	1	1	1	0	0	0	1	1	1	1	1	1	9	9	9	1	1	1
Hungary	2	2	2	2	0	0	2	2	2	1	1	1	1	9	9	1	1	1
Ireland	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1
Italy	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0
Latvia	1	1	1	1	0	0	1	1	1	1	1	1	1	9	9	1	1	1
Lithuania	2	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
Malta	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1
Netherlands	2	9	9	9	9	9	9	9	2	1	9	9	9	9	9	9	9	1
Poland	2	2	2	2	0	0	2	2	2	1	1	1	1	9	9	1	1	1
Portugal	1	1	1	1	0	0	0	0	1	1	1	1	1	9	9	9	9	1
Romania	1	1	1	0	1	1	1	1	1	0	0	0	9	0	0	0	0	0
Slovakia	2	2	2	1	1	1	1	1	2	0	0	0	0	0	0	0	0	0
Slovenia	2	2	2	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1
Spain	2	2	2	2	9	2	2	2	2	0	0	0	0	9	0	0	0	0
Sweden	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1
UK	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1

I_001_04: Micro-aggregated labour productivity (av., median, other moments) - all firms
I_001_05: Micro-aggregated labour productivity (av., median, other moments) - domestic firms
I_001_06: Micro-aggregated labour productivity (av., median, other moments) - exporters
I_001_07: Micro-aggregated labour productivity (av., median, other moments) - importers
I_001_08: Micro-aggregated labour productivity (av., median, other moments) - domestic multinationals
I_001_09: Micro-aggregated labour productivity (av., median, other moments) - affiliates of foreign multinationals
I_001_10: Micro-aggregated labour productivity (av., median, other moments) - foreign owned exporter
I_001_11: Micro-aggregated labour productivity (av., median, other moments) - domestic owned exporters
I_013_02: Micro-aggregated ULC (av., median, other moments) - all firms

Table 6.15: Bottom-up indicators: Total Factor Productivity

	Computability										Accessibility									
	I_003_03	I_003_04	I_003_05	I_003_06	I_003_07	I_003_08	I_003_09	I_003_10	I_004_01	I_005_01	I_003_03	I_003_04	I_003_05	I_003_06	I_003_07	I_003_08	I_003_09	I_003_10	I_004_01	I_005_01
Austria	1	9	1	1	9	9	9	9	1	1	0	9	0	0	9	9	9	9	0	0
Belgium	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	2	2
Bulgaria	2	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1
Croatia	0	0	0	0	0	0	0	0	0	0	9	9	9	9	9	9	9	9	9	9
Czech Rep.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Denmark	2	1	2	2	0	0	1	1	2	2	1	1	1	1	9	9	1	1	1	1
Estonia	2	2	1	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1
Finland	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
France	2	1	2	2	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1
Germany	1	1	1	0	0	0	1	1	1	1	1	1	1	9	9	9	1	1	1	1
Hungary	2	2	2	2	2	0	0	2	2	2	1	1	1	1	1	9	9	1	1	1
Ireland	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Italy	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1
Latvia	1	1	1	1	0	0	1	1	1	1	1	1	1	1	9	9	1	1	1	1
Lithuania	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Malta	0	0	0	0	0	0	0	0	0	0	9	9	9	9	9	9	9	9	9	9
Netherlands	2	9	9	9	9	9	9	9	2	2	1	9	9	9	9	9	9	9	1	1
Poland	2	2	2	2	0	0	2	2	2	2	1	1	1	1	9	9	1	1	1	1
Portugal	1	1	1	1	0	0	0	0	1	1	1	1	1	1	9	9	9	9	1	1
Romania	1	1	1	0	1	1	1	1	1	1	0	0	0	9	0	0	0	0	0	0
Slovakia	2	2	2	1	1	1	1	1	2	2	0	0	0	0	0	0	0	0	0	0
Slovenia	2	2	2	2	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1
Spain	2	2	2	2	9	9	2	2	2	2	0	0	0	0	9	9	0	0	0	0
Sweden	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
UK	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1

I_003_03: Micro-aggregated TFP (average, median, other moments) - all firms
I_003_04: Micro-aggregated TFP (average, median, other moments) - domestic firms
I_003_05: Micro-aggregated TFP (average, median, other moments) - exporters
I_003_06: Micro-aggregated TFP (average, median, other moments) - importers
I_003_07: Micro-aggregated TFP (average, median, other moments) - domestic multinationals
I_003_08: Micro-aggregated TFP (average, median, other moments) - affiliates of foreign multinationals
I_003_09: Micro-aggregated TFP (average, median, other moments) - foreign owned exporters
I_003_10: Micro-aggregated TFP (average, median, other moments) - domestic owned exporters
I_004_01: Olley and Pakes TFP decomposition
I_005_01: Foster decomposition of TFP growth

Table 6.16: Bottom-up indicators: firm dynamics

	Computability						Accessibility					
	I_051_03	I_052_03	I_053_01	I_054_01	I_055_01	I_056_01	I_051_03	I_052_03	I_053_01	I_054_01	I_055_01	I_056_01
Austria	9	9	9	9	1	1	9	9	9	9	0	0
Belgium	9	9	9	2	2	2	0	0	0	0	2	0
Bulgaria	1	1	1	1	2	2	1	1	1	1	1	1
Croatia	1	1	0	0	0	0	0	0	9	9	9	9
Czech Republic	1	0	0	1	1	1	1	9	9	1	1	1
Denmark	2	2	2	2	2	2	1	1	1	1	1	1
Estonia	2	2	2	2	2	2	1	1	1	1	1	1
Finland	1	1	1	1	2	2	1	1	1	1	1	1
France	2	2	2	2	2	2	0	0	0	0	0	0
Germany	2	2	2	2	1	1	1	1	1	1	1	1
Hungary	2	2	2	2	2	2	1	1	1	1	1	1
Ireland	2	2	2	2	2	2	1	1	1	1	1	1
Italy	1	1	1	1	1	1	0	0	0	0	0	0
Latvia	1	0	0	0	1	1	1	9	9	9	1	1
Lithuania	1	1	1	1	1	1	0	0	0	0	0	0
Malta	1	1	1	1	1	1	1	1	1	1	1	1
Netherlands	2	2	2	2	2	2	1	1	1	1	1	1
Poland	1	1	1	1	2	2	1	1	1	1	1	1
Portugal	1	1	1	1	1	1	1	1	1	1	1	1
Romania	1	1	1	1	1	1	0	0	0	0	0	0
Slovakia	2	2	2	2	2	2	0	0	0	0	0	0
Slovenia	2	1	1	2	2	2	1	1	1	1	1	1
Spain	2	2	2	9	2	2	9	9	2	9	0	0
Sweden	2	2	2	2	2	2	1	1	1	1	1	1
UK	2	0	2	2	2	2	1	0	1	1	1	1

I_051_03: Entry rate (birth rate)
I_052_03: Exit rate (death rate)
I_053_01: Firm survival at different lifetimes
I_054_01: Average firm size relative to entry, by age
I_055_01: Dispersion of firm by size
I_056_01: Share of gazelles: firms with average growth of revenues (in euro) reaches 20% p.a. over 3 consecutive years. Small gazelles: start employment 10-49; medium gazelles: start employment 50-249 compared to reference population; at NACE2 level.

Table 6.17: Bottom-up indicators: international activities

	Computability										Accessibility									
	I_009_02	I_043_01	I_043_02	I_044_01	I_045_01	I_046_01	I_047_01	I_048_01	I_049_01	I_050_01	I_009_02	I_043_01	I_043_02	I_044_01	I_045_01	I_046_01	I_047_01	I_048_01	I_049_01	I_050_01
Austria	1	9	9	9	2	9	1	2	9	1	0	9	9	9	0	9	0	0	9	0
Belgium	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0
Bulgaria	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Croatia	0	2	2	2	2	1	0	2	1	0	9	0	0	0	0	0	9	0	0	9
Czech Rep.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Denmark	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
Estonia	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
Finland	1	1	1	1	2	2	1	2	2	1	1	1	1	1	1	1	1	1	1	1
France	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
Germany	2	0	0	0	2	2	2	0	0	0	1	9	9	9	1	1	1	9	9	9
Hungary	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
Ireland	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
Italy	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0
Latvia	1	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0
Lithuania	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	9
Malta	2	2	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1
Netherlands	2	2	2	2	2	9	9	2	9	9	1	1	1	1	1	9	9	1	9	9
Poland	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Portugal	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Romania	1	0	0	0	1	1	1	0	0	0	0	9	9	9	0	0	0	9	9	9
Slovakia	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Slovenia	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
Spain	2	9	9	9	2	2	2	2	2	2	0	9	9	9	9	9	0	9	9	0
Sweden	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
UK	1	1	1	1	2	2	1	2	2	1	1	1	1	1	1	1	1	1	1	1

I_009_02: Average, median and other moments of value of exports per exporting firm, total
I_043_01: Average, median, variance, other moments of number of export destination countries per exporting firm
I_043_02: Number of exporting firms by number of export destination countries.
I_044_01: Average, median, variance, other moments of number of export destination countries *number of products exported per exporting firm;
I_045_01: Number of exporting firms (extensive margin)
I_046_01: % of exporting firms in total number of firms (extensive margin)
I_047_01: Average, median, other moments of export sales as a share of total turnover (intensive margin)
I_048_01: Number of importing firms (extensive margin)
I_049_01: % of importing firms in total number of firms (extensive margin)
I_050_01: Average, median, other moments of imported intermediates as a share of total cost of materials (intensive margin)

Table 6.18: Bottom-up indicators: R&D and other activities

	Computability						Accessibility				
	I_023_04	I_023_05	I_041_03	I_042_03	I_059_03	I_070_01	I_023_04	I_023_05	I_041_03	I_042_03	I_059_03
Austria	1	1	9	9	1	9	0	0	9	9	0
Belgium	1	1	2	2	2	2	0	0	0	0	2
Bulgaria	2	2	1	1	2	1	1	1	1	1	1
Croatia	1	1	1	1	0	0	1	1	0	0	9
Czech Republic	1	1	1	1	1	1	1	1	1	1	1
Denmark	2	2	1	0	2	2	1	1	1	9	1
Estonia	2	2	2	2	2	2	1	1	1	1	1
Finland	2	2	2	2	1	1	1	1	1	1	1
France	2	2	1	1	2	2	1	1	1	1	1
Germany	1	1	1	0	1	0	1	1	1	9	1
Hungary	2	2	2	0	2	2	2	1	1	9	1
Ireland	1	1	2	2	2	1	1	1	1	1	1
Italy	1	1	1	1	1	0	1	0	0	0	1
Latvia	2	0	1	0	1	2	1	9	1	9	1
Lithuania	2	1	1	1	2	1	0	0	0	0	0
Malta	1	1	1	1	0	2	1	1	0	0	9
Netherlands	1	1	9	9	9	1	1	1	9	9	9
Poland	2	2	1	0	1	1	1	1	1	9	1
Portugal	1	1	0	1	1	1	1	1	9	1	1
Romania	1	1	1	1	1	0	0	0	0	0	0
Slovakia	2	2	1	1	2	1	0	0	0	0	0
Slovenia	2	2	1	1	2	2	1	1	1	1	1
Spain	2	2	2	9	2	9	2	2	9	9	9
Sweden	2	2	2	2	2	2	1	1	1	1	1
UK	2	2	2	2	2	1	1	1	1	1	1

I_023_04: R&D expenditure - mean
I_023_05: R&D expenditure (% of turnover) - mean
I_041_03: Share of foreign-owned firms in total firms (by country, sector, region)
I_042_03: Share of domestic MNFs in total firms (by country, sector, region)
I_059_03: Asset tangibility
I_070_01: Firm level estimates of quality

6.6 Synthesis of accessibility conditions for micro-data in EU

Country	Accessibility conditions
Austria	The sources are not publicly available.
Belgium	NBB data are confidential and restricted, and the use is allowed only to NBB members (or affiliated). NBB data on firms' balance sheet is the same data provided by Belfirst, and this source is available upon payment of a fee.
Bulgaria	All the sources mentioned above are restricted, and access is strictly regulated by the Protection of Secrecy (chapter 6, of Statistical Act). The micro-data from different statistical fields are accessible, if it does not conflict with existing regulations, and after the decision of the Commission appointed under Art.10 of the 'Rules for providing of anonymised data on scientific and research purposes'. These rules govern the relationship of providing by BNSI of micro-data and the procedure for obtaining them. The rules are based on, and in accordance with, requirements of national and relevant EU legislation. See https://unstats.un.org/unsd/dnss/docViewer.aspx?docID=2772. See also indicator 15.4 in http://www.nsi.bg/sites/default/files/files/pages/LegalBasis_e/BG_report_FINAL.pdf.
Croatia	Access to most data is restricted. Data collected for CIS (Turnover and R&D expenditure) can be accessed under certain conditions (for scientific purposes according to Ordinance on the methods of statistical data protection and Ordinance on Conditions and Terms of Using Confidential Data for Scientific Purposes).
Czech Republic	Business register data can be accessed both at NCB and CZSO. For the access, an external researcher has to provide a research project and to pay a fee. Data can be accessed both on-site and with CDs (depending on the agreement). According to NCB, custom data are available only for NCB employees, and the NCB does not report the conditions to use FDI, and outward FATS data. The access's conditions for the External Trade Database at CZSO are regulated by special contract of confidentiality, and the access is only for research purposes (upon payment of a fee). More details are available at http://www.czso.cz/eng/redakce.nsf/i/statistical_data_for_scientific_research_purposes
Denmark	Data are accessible for persons affiliated to Danish institutions which are recognised by Statistics Denmark, conditional to the approval of a project. In principle, foreign researchers can access to data if they have an affiliation with a Danish institution. Affiliation can only take place if the authorised environment is willing to take the responsibility for the foreign researcher making sure that all existing rules governing access to micro-data are observed. Data can be accessed on site or from a remote access. See more information at http://www.dst.dk/en/TilSalg/Forskningsservice.aspx
Estonia	Data are at SE, and the availability of micro-data for scientific purposes is regulated by legal

	acts and it can be used in the safe-centre (see http://www.stat.ee/legal-acts). In addition, all the sources mentioned above are highly confidential, so accessibility rules are quite restrictive.
Finland	Data are accessible at the Research Laboratory or via the remote access system conditional on a user license, access agreements and a fee payment. See more details at http://www.stat.fi/tup/mikroaineistot/index_en.html.
France	All the mentioned sources are highly confidential, but micro level data will be accessible with the new system by submitting a research proposal and conditional to a committee approval. Details on the accessibility can be find at http://www.casd.eu/
Germany	Most datasets are available under certain conditions at the respective institutions. Destatis, the Federal Employment Office (Bundesagentur für Arbeit, BA) and the Bundesbank all have dedicated Research Data Centres which offer on-site or remote access (or direct access via Scientific Use Files) to many of their micro-level datasets according to the German laws of privacy protection. Data is accessible to researchers, but only at the BA foreign researchers can get access to the data without cooperating with a partner from Germany. Data from the Deutsche Bundesbank are accessible only at the Research Centre (in Frankfurt am Main). The use of data from the Deutsche Bundesbank is subject to special confidentiality conditions. Due to legal requirements, individual data cannot be made generally available. However, these data are made available under strict conditions and for clearly defined academic research purposes. Bundesbank has visiting researcher programme at the Research Centre. In the case of BA, the FDZ offers three ways of data access for researchers. These three ways differ according the degree of anonymity of the data and the terms of data use: (i) on-site, (ii) remote data access, and (iii) Scientific Use File (rare). In all the three cases, the researchers have to present a research project that has to be approved by FDZ. In the case of on-site access, there is the possibility to apply for financial support. More details are at http://fdz.iab.de/en.aspx. The research data centre of the Destatis offers four different forms of access to selected micro-data of official statistics: (i) public use files, (ii) scientific use files, (iii) safe centres, and (iv) remote execution. They differ with regard to both the anonymity of the data, and the form of data provision. The scientific use files are well-suited for large part of the scientific data analyses. Foreign users not employed by German institutions may work with the data both at the research centre and via remote executions. More details are at http://www.forschungsdatenzentrum.de/en/datenzugang.asp
Hungary	The Hungarian matched data was created by the CSO by assigning an anonymised identifier to each company, which is consistent between years and databases. Data protection, required by the law, is a key element in the operations of the CSO. Therefore,

	variables that provide a direct possibility to reveal the identity of a company (eg name of the company, address of the headquarters or tax number) were deleted. Technically, the data is stored on a server in separate files according to topics. Merging the different databases using the id numbers assigned by the CSO is performed by the researcher. The matched database is accessible only to the researchers who have an agreement with CSO, such as the Hungarian Academy of Sciences or some ministries. Access is granted after registering the project at the CSO. The accessibility of the matched database is restricted to a safe research room inside the building of the CSO where researchers can work on the data on site, and save their results. Note that accessibility is still limited and burdened and occasionally quite slow. The researcher who works with the data has to be in the Research room in Budapest and needs be affiliated with a partner.
Ireland	The access to the data is in principle possible, but subject to stringent conditions. Firm-level data can be accessed on-site only, while the use and publication of results is subject to statistical office approval.
Italy	Firm-level data are confidential and restricted. Business Register (except for Business Demography) and micro-data stemming from surveys are available to the users at the ADELE Laboratory (Laboratory for Elementary Data Analysis). However, it should be stressed that identification code of single units are not available to external researchers; thus it is not possible to merge data stemming from different surveys without a specific agreement with Istat (research protocol). Databases with the full population are not accessible to researcher, but descriptive statistics from these databases are available upon request. See for example project Istat – Micro3. For further information about ADELE laboratory see http://www.istat.it/en/information/researchers/analysis-of-individual-data.
Latvia	Information on the value of export (import) by destination and product are not accessible because confidential. As for other data, in principle are available upon request, conditional to a fee payment.
Lithuania	Firm-level data are confidential. By the Law of Statistics, micro-level data could be used for research purposes. Confidential statistical data may be provided for scientific purposes to be used in a manner that it would be impossible to directly identify the respondents based on the data, where the research establishments ensure the protection of these data.
Malta	All the information is accessible upon request for research purposes, except data on foreign/domestic ownership.
Netherlands	In general, many issues of competitiveness are available to both domestic and foreign researchers. The accessibility to micro-level data follows explicit rules and specific charges apply. According to CBS “All datasets in the Centre for Policy Related Statistics’ micro-data catalogue are available for authorised external researchers to do their own research using these datasets. The catalogue does not contain all the datasets Statistics Netherlands uses to compile its statistics. CBS datasets not (yet) included in the catalogue may be made

	suitable for use by external researchers as custom-made datasets. The catalogue (classified by theme) includes documentation reports of the most recent version of datasets immediately available for use. This documentation contains a description of the contents and structure of the dataset. The enclosures referred to in this documentation are available only in Dutch and on request." More details can be found at http://www.cbs.nl/NR/rdonlyres/50625EDE-3274-4D7C-B19B-5E5D0F239E2F/0/131112dienstencatalogusosra2014eng.pdf
Poland	According to the information that we were able to gather, we can only state that the rules of statistical confidentiality are determined by the law on official statistics issued on 29 June 1995. In theory, access to micro-data is possible only under specific conditions, but the practice shows that access to individual data beyond CSO and NBP is nearly impossible.
Portugal	We are not in position to describe in details the accessibility conditions. However, in principle data seem accessible.
Romania	Data are not accessible since a safe environment for data security is not yet in place.
Slovakia	The firm-level databases are not available on-line, and the access is confidential: the rules of access have not been specified.
Slovenia	All the micro-data are accessible at the SURS and are restricted only for research purposes. See http://www.stat.si/eng/drz_stat_mikro.asp
Spain	In the case of the Industrial Economics Survey, only other statistical institutions (Statistical Institutes of Autonomous Communities) are provided with micro-data files. As for the CIS and the Pitec databases, it is possible to access to firm level data 'anonymised' on the INE web through a specific procedure. Researchers must submit a request by filling out the required fields in the tab 'Solicitud de descarga de BBDD'. Once the request is evaluated and approved, the researcher will receive within 72 hours an email providing a username and password, valid for three months. Except for anonymisation of a set of variables the files available on the web site correspond with the original files.
Sweden	All firm-level data are restricted but data can be accessed by European researchers on remote access, conditional on a confidentiality check and an administrative cost.
United Kingdom	All the sources are available via the submission of a research project to the correspondent institutions (UKDS, ONS, and HMRC Datalab). In addition, the HMRC Datalab requires a short training course, which includes legal issues as well as statistical disclosure control of output. At the moment the Datalab is only open to UK based institutions and by law HMRC is only allowed to share the data if it serves one of HMRC's functions. Data are available only on site

References

Agafitei, M. and Vaju, S. (2013) 'Addressing the Challenge of Producing European Comparable Data Using Administrative Data', *Working Paper* 10, UNECE, Geneva

Aleong, C. S. & Aleong, J. (2004) 'Harmonization of data: lessons to be learnt from major international organizations', manuscript, Delaware State University and University of Vermont

Altomonte, C. and Aquilante, T. (2012) 'The EU-EFIGE/Bruegel-Unicredit Dataset', *Working Paper* 2012/13, Bruegel

Altomonte, C., Barba Navaretti, G., Di Mauro, F. and Ottaviano, G.I.P. (2011) 'Assessing competitiveness: How firm-level data can help', *Policy Contribution* 2011/16, Bruegel

Altomonte, C., di Mauro, F. and Osbat, C. (2013) 'Going Beyond Labour Costs: How and why 'structural' and micro-based factors can help explaining export performance?' *CompNet Policy Brief* no.1, 15 January

Annoni, Paula and Dijkstra, Lewis (2013) 'EU Regional Competitiveness Index RCI 2013', JRC Scientific and Policy Report EUR 26060 EN, Luxembourg: EC

Arnaud, B., Dupont, J., Seung-Hee Koh and Schreyer, P. (2011) 'Measuring Multi-Factor Productivity by Industry: Methodology and First Results from the OECD Productivity Database', OECD Statistics Directorate

Bachteler, T. (2012) *Record Linkage Bibliography*, German Record Linkage Centre, Duisburg

Bakker, B. (2010) 'Micro-Integration: State of the art', in ESSnet on Data Integration (ed) *Report on WP1: State of the art on statistical methodologies for data integration*, pp77-107

Barnard, J. & Meng, X.-L. (1999) 'Applications of multiple imputation in medical studies: from AIDS to NHANES', *Statistical Methods in Medical Research* 8: 17-36

Bartelsman, E.J., Haltiwanger, J.C. and Scarpetta, S. (2004) 'Microeconomic evidence of creative destruction in industrial and developing countries', *IZA Discussion Paper* No. 1374, Institute for the Study of Labour (IZA), Bonn

Bartelsman E., Haltiwanger J. and Scarpetta S. (2005) 'Measuring and Analyzing Cross-country Differences in Firm Dynamics', paper prepared for NBER Conference on Research in Income and Wealth 'Producer Dynamics: New Evidence from Micro-data'

8-9 April
Bartelsman, E.J., Haltiwanger, J.C. and Scarpetta, S. (2009) 'Cross-Country Differences in Productivity: The Role of Allocation and Selection', NBER *Working Paper* No. 15490, National Bureau of Economic Research (NBER), Cambridge (MA)
Bartelsman, E.J., Haltiwanger, J.C. and Scarpetta, S. (2009) 'Measuring and analyzing cross-country differences in firm dynamics', in Timothy Dunne, J. B. Jensen & Roberts, M. J. (eds) *Producer Dynamics: New Evidence from Micro-data*, University of Chicago Press, pp15-76
Bartelsman, E.J. and Hamilton, A. (2004) *The analysis of micro-data from an international perspective*, OECD Statistics Directorate Committee on Statistics (STD/CSTAT(2004)12), OECD, Paris
Bartelsman, E.J., Haskel, J. & Martin, R. (2008) 'Distance to which frontier? Evidence on productivity convergence from international firm-level data', *Discussion Paper* No. 7032, CEPR, London
Bartelsman, E.J., Scarpetta, S. & Schivardi, F. (2005) 'Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries', *Industrial and Corporate Change* 14(3): 365-391
Békés, G., P. Harasztosi and B. Muraközy (2011) 'Firms and products in international trade: Evidence from Hungary', *Economic Systems* 35
Bender, S, Lane, J., Shaw, K.L., Andersson, F. & Wachter, T.v. (eds) (2008) *The Analysis of Firms and Employees. Quantitative and qualitative approaches.* University of Chicago Press
Benkovskis Konstantins & Worz Julia (2012) 'Evaluation of Non-Price Competitiveness of Exports from CESEE Countries in the EU Market', *Working Paper* 1/2012, Bank of Latvia
Biewen, Elena, Anja Gruhl, Christopher Gürke, Tanja Hethey-Maier and Emanuel Weiß (2012) 'Combined firm data for Germany – possibilities and consequences of merging firm data from different data producers', in *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften* 132(3): 361-377
Borgman, C.L. (2010) 'Research Data: Who will share what, with whom, when, and why?' *Working Paper* 161, Berlin: RatSWD
Boyd, D. & Crawford, K. (2011) 'Six provocations for big data', paper presented at Oxford Internet Institute's 'A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society,' 21 September
Brandt, M. (2012) 'Decentralised and Remote Access to Confidential Data in the ESS DARA', 4th Workshop on Data Access (WDA), Luxembourg
Broersma, L., Koch, A. & Rekveldt, B. (2010) 'Hiring by skill in innovative and non-innovative firms. An explorative comparison using German and Dutch matched employer-employee data bases', Micro-Dyn *Working Paper* No. 5/10, Vienna

Brook, E. L., Rosman, D. L. & Holman, C. J. (2008) 'Public good through data linkage: measuring research outputs from the Western Australian Data Linkage System', *Australian and New Zealand Journal of Public Health* 32(1): 19-23

Brown, B., Chui, M. & Manyika, J. (2011) *Are you ready for the era of 'big data'*, McKinsey Global Institute, San Francisco

Burkhauser, R. V. & Lillard, D. R. (2005) 'The contribution and potential of data harmonization for cross-national comparative research', *Journal of Comparative Policy Analysis 7*(4): 313-330

Büttner, T. & Rässler, S. (2008) 'Multiple Imputation of Right-Censored Wages in the German IAB Employment Sample Considering Heteroscedasticity', *Discussion Paper* 44/2008, Nuremberg, IAB

Cheser, A. & Nesheim, L. (2006) 'Review of the Literature on the Statistical Properties of Linked Datasets', *DTI Occasional Paper* 3, Department of Trade and Industry, London

CHINTEX (2001) *The change from input harmonization to ex-post harmonisation in national samples of the European Community Household Panel. Implications on data quality* (Synopsis), technical report, CHINTEX

Christen, P. (2012) *Data Matching Concepts and Techniques for Record Linkage, Entity Resolution and Duplicate Detection*, Springer, Berlin, Heidelberg

Christen, P. (2012a) 'A survey of indexing techniques for scalable record linkage and deduplication', *IEEE Transactions on Knowledge and Data Engineering* 24(9): 1537-1555

Cihak, M., Demirgüç-Kunt, A., Feyen, E., & Levine, R. (2012) 'Benchmarking financial systems around the world', *World Bank Policy Research Working Paper*, 6175

Clark, D. E. (2004) 'Practical introduction to record linkage for injury research', *Injury Prevention* 10(3): 186-191

CompNet Task Force (2014) 'Micro-based evidence of EU competitiveness: The CompNet database', *Working Paper Research* 253, National Bank of Belgium

Crawford, K. (2013) 'The Hidden Biases in Big Data', *Harvard Business Review Blog Network*

Cukier, K. & Mayer-Schoenberger, V. (2013) 'The Rise of Big Data: How it's Changing the Way We Think about the World', *Foreign Affairs* 2013(May/June): 28-40

Daas, P. & van der Loo, M. (2013) 'Big Data (and official statistics)', UNECE/OECD/EUROSTAT/ESCAP *Working Paper*

Daas, P., Puts, M., Buelens, B. & van den Hurk, P. (2013) 'Big Data and Official Statistics', manuscript, Statistics Netherlands

Data without Boundaries (DWB, 2012) *Report on the State of the Art of Current SC in Europe*, European Community, Work Package 4, Improving Access to OS Micro-data

De Backer, K. and Yamano N. (2011) *International Comparative Evidence on Global*

Value Chains, OECD, Directorate for Science, Technology and Industry

Derbyshire, J., Hollanders, H., Lewney R., Rivera Leon, L., Tarantola, S. and Tijssen R. (2012) *Regional Innovation Scoreboard 2012 - Methodology report*, European Commission

Desai, T. (2008) *A Guide to Linked Employer-Employee Data Sources in the EU and Beyond*, Research Laboratory, London School of Economics and Political Science, London

Dieppe, A., Dees, S., Jacquinot, P., Karlsson, Osbat, C., Özyurt, S., Vetlov, I., Jochem, A., Bragoudakis, Z., Sideris, D., Tello, P., Bricongne, J. C., Gaulier, G., Pisani, M., Papadopoulou, N., Micallef, B., Ajevskis, V., Brzoza-Brzezina, M., Gomes, S., Krekó, J. and Vyskrabka, M. (2012) 'Competitiveness and external imbalances within the euro area', *Occasional Paper Series* No. 139, European Central Bank

Doyle, P., Lane, J., Theeuwes, J.M. & Zayatz, L.V. (eds) (2001) *Confidentiality, disclosure, and data access: Theory and practical applications for Statistical Agencies*, Amsterdam: Elsevier

Drechsler, J. (2010) 'Multiple imputation of missing values in the wave 2007 of the IAB Establishment Panel', *IAB-Discussion Paper* 6/2010, Nuremberg, IAB

Drechsler, J., Dundler, A., Bender, S., Rässler, S. & Zwick, T. (2007) 'A New Approach for Disclosure Control in the IAB Establishment Panel', *IAB-Discussion Paper* 11/2007, Nuremberg, IAB

Dunn, H. L. (1946) 'Record Linkage', *American Journal of Public Health and the Nation's Health* 36(12): 1412-1416

Elliot, M. (undated) *What is Data Linkage?*, presentation

Elmagarmid, A. K., Ipeirotis, P. G. & Verykios, V. S. (2007) 'Duplicate record detection: A survey', *IEEE Transactions on Knowledge and Data Engineering* 19(1): 1-16

ESSnet Statistical Methodology Project on Integration of Survey and Administrative Data (ESSnet-ISAD, 2007) *Report of WP1: State of the art on statistical methodologies for integration of surveys and administrative data*, mimeo, European Statistical System

European Central Bank (ECB, 2012) 'Competitiveness and External Imbalances within the Euro Area', *Occasional Paper Series* No. 139, ECB, Frankfurt

European Central Bank (ECB, 2013) *Competitiveness Research Network: First Year Results*, Frankfurt

European Commission (1991) 'Council Regulation (EEC) No 3330/91 of 7 November', *Official Journal of the European Communities* L 316/1

European Commission (2001) *European Governance, A White Paper*, Commission of the European Communities, Brussels

European Commission (2002) 'Commission Regulation (EC) No 831/2002 of 17 May 2002 implementing Council Regulation (EC) No 322/97 on Community Statistics,

concerning access to confidential data for scientific purposes', *Official Journal of the European Communities* L133/7

European Commission (2006) *Communication from the Commission to the European Parliament and the Council on 'The Reduction of the Response Burden, Simplification and Priority-setting in the Field of Community Statistics'*, Commission of the European Communities, Brussels

European Commission (2009) *The production method of EU statistics: a vision for the next decade, Communication from the Commission to the European Parliament and the Council*, COM(2009) 404 final, Commission of the European Communities, Brussels

European Commission (2009a) 'Commission recommendation of 23 June 2009 on reference metadata for the European Statistical System', *Official Journal of the European Union* L 168/50: 50-55

European Commission (2010) *Fifth report on economic, social and territorial cohesion*

European Commission (2010) 'Investing in Europe's future', *Fifth report on economic, social and territorial cohesion*

European Commission (2011) *EU industrial structure 2011 Trends and Performance*, chapter iv – international competitiveness of EU industry

European Commission (2011) *Regulation of the European Parliament and of the Council on the European Statistical programme 2013-2017*, European Commission, Brussels

European Commission (2011) *EU industrial structure 2011 Trends and Performance*

European Commission (2011a) *Programme for the Modernisation of European Enterprise and Trade Statistics* (MEETS), European Commission, Brussels

European Commission (2012) *Roadmap: Framework Regulation Integrating Business Statistics* (FRIBS), European Commission, Brussels

European Commission (2012) *Macroeconomic Imbalances Procedure Scoreboard Headline Indicators, 1 November 2012 Statistical information*

European Commission (2012) *European Competitiveness Report*, 15th edition

European Commission (2012) *Regional Innovation Scoreboard* (2012 and previous editions).

European Commission (2013) *Innovation Union Scoreboard* (2013 and previous editions)

European Commission (2013) 'Towards knowledge driven reindustrialisation. European Competitiveness Report 2013', *Commission Staff Working Document* SWD (2013) 347 final, Luxembourg: EC

European Commission (2014) *Report from the Commission to the European Parliament and the Council on the implementation of Decision No 1297/2008/EC of the European Parliament and of the Council of 16 December 2008 on a Programme for*

the Modernisation of European Enterprise and Trade Statistics (MEETS), COM (2014) 444 final, Brussels

European Economic Community (2007) 'Regulation (EC) No 716/2007 of the European Parliament and of the Council of 20 June 2007 on Community Statistics on the Structure and Activity of Foreign Affiliates', *Official Journal of the European Union* L171/17

European Economic Community (2008) 'Decision No. 1297/2008/EC of the European Parliament and of the Council on 16 December 2008 on a Programme for the Modernisation of European Enterprise and Trade Statistics (MEETS)', *Official Journal of the European Union* L 340/76

European Economic Community (2009) 'Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities', *Official Journal of the European Union* L 87/164 (223/2009)

European Statistical Advisory Committee (2012) 'Opinion on the further development of statistics on international trade in goods and services in the European Union (SIMSTAT)', ESAC Doc. 2012/1115, European Statistical Advisory Committee

European Statistical System Committee (ESSC, 2013) '17th Meeting of the European Statistical System Committee', European Statistical System Committee.

Eurostat (2003) *Definition of Quality in Statistics*, Eurostat, Luxembourg

Eurostat (2009) *Foreign AffiliaTes Statistics (FATS) Recommendations Manual (1977-0375)*, Eurostat, Luxembourg

Eurostat (2011) *European Statistics Code of Practice for the National and Community Statistical Authorities*, Eurostat, European Statistical System, Luxembourg

Eurostat (2013) *ESSnet projects 2013 assessment report*, Eurostat

Fally, T. (2011) 'On the Fragmentation of Production in the US', University of Colorado-Boulder, July

Fellegi, I. P. & Sunter, A. B. (1969) 'A Theory for Record Linkage', *Journal of the American Statistical Association* 64(328): 1183-1210

Figueira, M.H. (2013) 'FRIBS – The EU framework for business statistics', presentation, Statistiktag Austria, 22/10/2013, Vienna

Fleck, S. E. (2009) 'International comparisons of hours worked: an assessment of the statistics', *Monthly Labor Review*, May

Foster, L., Haltiwanger, J. and C. J. Krizan (2001) 'Aggregate Productivity Growth. Lessons from Microeconomic Evidence', in *New Developments in Productivity*

Analysis: 303 372, National Bureau of Economic Research
Foster, R. (2012) 'How to Harness Big Data for improving Public Health', *Government Health IT*
Fursova, N. (2013) 'Data linking aspects of combining data (survey/administrative) including options for various hierarchies (S-DWH context)', *Working Document*, ESS-Net on Micro-data Linking and Data Warehousing in Production of Business Statistics, Centraal Bureau voor de Statistiek
Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. & Brilliant, L. (2008) 'Detecting influenza epidemics using search engine query data', *Nature* 457 (7232): 1012-1015
Granda, P. & Blasczyk, E. (2010) *Data Harmonization*, manuscript
Granner, F. (2013) *A new way of measuring foreign trade – SIMSTAT*, Statistics Austria, Vienna
Grünewald W. (2001) *Strengths and Weaknesses of the European Statistical System (ESS)*, http://www.scb.se/Grupp/Produkter_Tjanster/Kurser/Tidigare_kurser/q2001/Session_23.pdf
Gruhl, Anja, Guerke, Christopher, Hethey-Maier, Tanja, Oberschachtsiek, Dirk, Seitz, Julia, and Biewen, Elena (2012) 'Kombinierte Firmendaten für Deutschland – (KombiFiD)', Aktualisiert am 1. März, *FDZ-Datenreport*, 02/2011, Nürnberg, IAB
Günther, R. (2003) 'Report on compiled information', *Working Paper* 19, CHINTEX
Hall, B.H. (2010) 'The Financing of Innovative Firms', *Review of Economics and Institutions*, 1 (1), Article 4
Haltiwanger, J., Scarpetta, S. & Schweiger, H. (2008) 'Assessing job flows across countries: the role of industry, firm size and regulations', *NBER Working Paper* 13920, NBER, Cambridge (MA)
Haskel, J., Jona-Lasinio, C., & Iommi, M. (2012) 'Intangible capital and growth in advanced economies: measurement methods and comparative results', IZA
Heining, J., Jacobebbinghaus, P., Scholz, T. & Seth, S. (2012) 'Linked-Employer-Employee Data from the IAB: LIAB-Mover-Model 1993-2008 (LIAB MM 9308)', *FDZ-Datenreport* 01/2012, Nuremberg, IAB
Henrekson, M. & Johansson, D. (2010) 'Gazelles as job creators: a survey and interpretation of the evidence', *Small Business Economics* 35: 227-244
Herzog, T. H., Scheuren, F. & Winkler, W. E. (2010) 'Record linkage', *Wiley Interdisciplinary Reviews: Computational Statistics* 2(5), 535-543
Höhne, J. & Höninger, J. (2012) 'Das Verfahren Morpheus - Auf dem Weg zu Remote Access', *RatSWD Working Paper Series* 205, RatSWD, Berlin
Hollanders, H. and Tarantola, S. (2011) *Innovation Union Scoreboard – Methodology report*

Hollanders, H., Tarantola, S., Garda Porras, B. (2013) *Innovation Union Scoreboard 2013*, Directorate General for Enterprise and Industry, European Commission

Horvath, S. (2013), 'Big Data', *Aktueller Begriff* 37/13, Deutscher Bundestag, Wissenschaftliche Dienste, Berlin

Hummels, D., Ishii, J. and Yi, K. (2001) 'The nature and growth of vertical specialization in world trade', *Journal of International Economics*, Elsevier, vol. 54(1): 75-96

Hundepool, A. (2007) *CENEX summary report*, manuscript

Hundepool, A., J. Domingo-Ferrer, L. Franconi, S. Giessing, R. Lenz, J. Naylor, E. Schulte Nordholt, G. Seri, P.-P. De Wolf (2010) *Handbook on Statistical Disclosure Control*, ESSNet SDC, Luxembourg

Kallas, J. & Linardis, A. (2008) 'A Documentation Model for Comparative Research Based on Harmonization Strategies', *IASSIST Quarterly* 2008: 12-25

Kang, L., O'Mahony, M. & Peng, F. (2012) 'New Measures of Workforce Skills in the EU', *National Institute Economic Review* 220(1), R17-R28

Karlberg, M. and Skaliotis M. (2013) 'Big Data for Official Statistics – Strategies and some Initial European Applications', *Working Paper* 30, UNECE, Geneva

Karmel, R. (2005) 'Data linkage protocols using a statistical linkage key', *Data Linkage Series* 1, Australian Institute of Health and Welfare, Canberra

Khandelwal, A., Schott, P., and S.,Wei, forthcoming, 'Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters', *American Economic Review*

Kim, J.K. and Fuller, W. (2004) 'Fractional hot deck imputation', *Biometrika* 91(3), 559-578

Koch, A. (2008) 'How to analyse firm dynamics in European countries? Methodology and results of a comparative study', *Micro-Dyn Working Paper* 15/08, MicroDyn, Vienna

Koch, Andreas and Neugebauer, Katja (2014) *Technical report on the general considerations of the matchability of datasets within and across countries and regions*, Technical Report, Tübingen

Koopman, Robert, Powers, William M., Wang, Zhi and Wei, Shang-Jin, (2010) 'Give Credit Where Credit is Due: Tracing Value Added in Global Production Chains', *NBER Working Paper* No. w16426

Lamel, J. (2002) *The Future of the European Statistical System. Theme 2: New ideas for ESS development*, Wirtschaftskammer Österreich, Vienna

Levinsohn, J. and Petrin, A. (2003) 'Estimating Production Functions Using Inputs to Control for Unobservables,' *Review of Economic Studies*, Wiley Blackwell, vol. 70(2): 317-341, 04

Liotti, A. (2013) *Interoperability of business registers in the European Statistical System: the Eurostat VIP.ESBRs project*, manuscript, Eurostat

Little, R. & Rubin, D. (1987, 2002) *Statistical Analysis with Missing Data*, Wiley

Lohr, S. (2012) 'The Age of Big Data', *The New York Times*
Mayer, T. & Ottaviano G. (2008) 'The Happy Few: The Internationalisation of European Firms', *Intereconomics: Review of European Economic Policy* 43(3), 135-148
Mills, S. (2013) *Demystifying Big Data. A practical Guide to Transforming the Business of Government*, Tech America Foundation, Washington DC
Miroudot, S., Lanz, R., Ragoussis, A. (Nov 2009) 'Trade in Intermediate Goods and Services', *OECD Trade Policy Working Paper* No. 93, OECD Publishing
Museux, J.-M., Hilbert, N. & Barcellan, R. (2013) *Architecture the ESS.VIP Programme*, Eurostat
Narayanan, A. & Shmatikov, V. (2010) 'Myths and fallacies of personally identifiable information', *Communications of the ACM* 53(6): 24-26
Neff, G. (2013) 'Why Big Data Won't Cure Us', *Big Data* 1(3): 117-123
Newcombe, H. B. & Kennedy, J. M. (1962) 'Record linkage: making maximum use of the discriminating power of identifying information', *Communications of the ACM* 5(11): 563-566
Newcombe, H. B., Kennedy, J. M., Axford, S. & James, A. (1959) 'Automatic linkage of vital records', *Science*, New Series 130(3381), 954-959
O'Mahony, M. & Timmer, M.P. (2009) 'Output, Input and Productivity Measures at the Industry Level: The EU KLEMS Database', *The Economic Journal* 119(538), F374-F403
O'Mahony, M., Castaldi, C., Los, B., Bartelsman, E., Maimaiti, Y. & Peng, F. (2008) *EUKLEMS – Linked Data: Sources and Methods*, manuscript, University of Birmingham
OECD (2001) 'Measuring Productivity: measurement of aggregate and industry-level productivity growth', *OECD Manual*
OECD (2011) 'Import content of exports', in *OECD Science, Technology and Industry Scoreboard 2011*, OECD Publishing
OECD (2013) 'Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by Big Data', *OECD Digital Economy Papers* 222, OECD, Paris
OECD (2013) 'Calculating Summary Indicators of Employment Protection Strictness: Methodology' http://www.oecd.org/els/emp/EPL-Methodology.pdf
OECD (2013a) *OECD Health Data 2013*,
http://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT
OECD, STAN Input-Output Database:
http://stats.oecd.org/Index.aspx?DatasetCode=STAN_IO_M_X
OECD, STAN Input-Output Database:
http://stats.oecd.org/Index.aspx?DataSetCode=STAN_IO_INTERM_M
Okner, B. (1972) 'Constructing a New Data Base from Existing Micro-data Sets: The 1966 Merge File', *Annals of Economic and Social Measurement* 1(3): 326-361

Olley, S. and Pakes, A. (1996) 'The Dynamics of Productivity in the Telecommunications Industry.' *Econometrica*, 64(6): 1263-1298
Radermacher, W. (2013) *From Intrastat to SIMSTAT and ESS.VIP Programme*, Eurostat
Rässler, S. (2008) *Multiple Imputation*, manuscript, Otto-Friedrich-Universität Bamberg
Rubin, D. B. (1996) 'Multiple imputation after 18+ years', *Journal of the American Statistical Association* 91(434): 473-489
Rubin, D.B. (1993) 'Discussion: Statistical Disclosure Limitation', *Journal of Official Statistics* 9: 462-468
Ruggles, N. and Ruggles, R. (1974) 'A strategy for merging and matching micro-data sets', *Annals of Economic and Social Measurement* 3(2): 353-371
Santos, M. & Museux, J.-M. (2005) *Legal, political and methodological issues in confidentiality in the ESS. Presentation*, Eurostat
Schafer, J. L. (1997) *Analysis of Incomplete Multivariate Data*, Chapman and Hall, New York
Schafer, J. L. (1999) 'Multiple imputation: a primer', *Statistical Methods in Medical Research* 8(1), 3-15
Schafer, J. L. (2001) 'Thou Shalt not Impute only Once', introductory overview lecture, Joint Statistical Meetings, Atlanta, GA
Schiller, D. (2013) 'Proposal for a European Remote Access Network (EU-RAN) - Main Components', Joint UNECE/Eurostat work session on statistical data confidentiality, (Ottawa, Canada, 28-30 October, UNECE, European Commission
Schiller, D. and Welpton, R. (2013) 'Providing Remote Access to European Micro-data', mimeo, Institute for Employment Research and UK Data Archive
Schmitz, M., De Clercq, M., Fidora, M., Lauro, B. and Pinheiro C. (2012) 'Revisiting the effective exchange rates of the euro', *ECB Occasional paper series* N. 134, June, ECB
Schnell, R. (2009) 'Record-Linkage from a Technical Point of View', *RatSWD Working Paper Series* 124, Berlin
SDMX (2009) *Content-Oriented Guidelines*, Statistical Data and Metadata eXchange
Sharma, S. (2007) 'Financial development and innovation in small firms', *World Bank Policy Research Working Paper* 4350
Sharp, S. (undated) 'Deterministic and probabilistic record linkage', National Records of Scotland, presentation, available at http://www.scotland.gov.uk/Resource/0039/00390031.pdf
Shearing M. (2013) 'An Introduction to the Global Statistical System', available at http://www.statisticsauthority.gov.uk/about-the-authority/uk-statistical-system/an-introduction-to-the-global-statistical-system.pdf
Snorrason, H. (2002) *The Future of the European Statistical System: Towards a European Statistical Identity*, Statistics Iceland

Statistikrat der Bundesanstalt Statistik Österreich (2013) 'Positionspapier zur Rahmenverordnung für eine integrierte Unternehmensstatistik (FRIBS)', Statistikrat der Bundesanstalt Statistik Österreich, Wien

Sverdrup, U. (2005) 'Administering information: Eurostat and statistical integration', *Working Paper*, available at http://www.arena.uio.no.

Sweeney, L., Abu, A. & Winn, J. (2013) 'Identifying Participants in the Personal Genome Project by Name', available at SSRN 2257732

Taleb, N. (2013) 'Beware the Big Errors of "Big Data"', available at http://www.wired.com/opinion/2013/02/big-data-means-big-errors-people/

Teufel, H. (2008) *Privacy Policy Guidance Memorandum*, Homeland Security

Thoma, G., Torrisi, S., Gambardella, A., Guellec, D., Hall, B. H. & Harhoff, D. (2010) 'Harmonizing and Combining Large Datasets – An Application to Firm-Level Patent and Accounting Data', *NBER Working Paper* 15851, National Bureau of Economic Research, Cambridge, MA

Trade & Investment Division (TID, 2012) *Data Harmonization and Modelling Guide for Single Window Environment*, Trade and Investment

Turner P and Van't Dack J. (1993) 'Measuring International Price and Cost Competitiveness', *BIS Economic Paper* No. 39.

U.S. Department of Labor (2012) 'International comparisons of manufacturing productivity and unit labor costs trends', International Labor Comparisons Program, Bureau of Labor Statistics

UNECE (2007) 'UNECE Quality Improvement Strategy', Document SA/2007/14/Add.1, Committee for the Coordination of Statistical Activities, United Nations Economic Commission for Europe (UNECE), Madrid

UNECE (2007a) *Managing Statistical Confidentiality & Micro-data Access. Principles and Guidelines of Good Practice*, New York and Geneva: United Nations

Vale, S. (2006) *The international comparability of business start-up rates*, Final report, OECD, Paris

Vale, S. (2009) *Generic Statistical Business Process Model, Version 4.0 – April 2009*, UNECE Secretariat

van Ark, B., Hao, J. X., Corrado, C., & Hulten, C. (2009) 'Measuring intangible capital and its contribution to economic growth in Europe', *EIB papers*, 14(1): 62-93

Vogel, Alexander, Wagner, Joachim (2012) 'The Quality of the KombiFiD-Sample of Business Services Enterprises: Evidence from a Replication Study', in *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften* 132(3): 393-403

Wagner, Joachim (2012) 'The Quality of the KombiFiD-Sample of Enterprises from Manufacturing Industries: Evidence from a Replication Study', in *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften* 132(3): 379-392

Wagner, Joachim (2012a) 'Average wage, qualification of the workforce and export

performance in German enterprises: Evidence from KombiFiD data', *Journal of Labour Market Research* 45(2): 161-170

Wahlstrom, K., Roddick, J.F., Sarre, R., Estivill-Castro, V. & de Vries, D. (2009) 'Legal and Technical Issues of Privacy Preservation in Data Mining', mimeo

Willenborg, L. and de Waal, T. (2001) 'Elements of statistical disclosure control', *Lecture Notes in Statistics*, 115, New York: Springer

Winkler, W. E. (1990) 'String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage', *Proceedings of the Section on Survey Research Methods* 1990, 354-359

Winkler, W. E. (2006) 'Overview of Record Linkage and Current Research Directions', *Research Report Series* 2006-2, Statistical Research Division US Census Bureau, Washington DC

World Economic Forum (2012) *Big Data, Big Impact: New Possibilities for International Development*, World Economic Forum

World Economic Forum (2013) *The Global Competitiveness Report 2013-2014. Full Data Edition*, World Economic Forum

Yuan, Y. C. (2010) *Multiple imputation for missing data: Concepts and new development (Version 9.0)*, SAS Institute Inc, Rockville, MD

Zio, M. d. (2012) *Microintegration – description in the Memobust Handbook, Department of Integration, Quality, Research and Production Networks Development*, Istat – Italian National Institute of Statistics, Rome

Mapping competitiveness with European data

Davide Castellani and Andreas Koch

Europe needs improved competitiveness to escape the current economic malaise, so it might seem surprising that there is no common European definition of competitiveness, and no consensus on how to consistently measure it. There is no single and/or harmonised dataset allowing the different facets of competitiveness to be captured in an internationally comparative perspective.

In particular, there is a lack of clarity about competitiveness at the firm level. The international operations of firms are not adequately represented by standard trade statistics, even though a thorough understanding of firm-level competitiveness should be a central component of Europe's response to economic difficulties. To help address this situation, this Blueprint provides an inventory and an assessment of the data related to the measurement of competitiveness in Europe. It is intended as a handbook for researchers interested in measuring competeveness, and for policymakers interested in new and better measures of competitiveness. Policymakers have an important role to play to improve data accessibility for the economic analysis of competitiveness in Europe.

Bruegel is a European think tank devoted to international economics. It is supported by European governments and international corporations. Bruegel's aim is to contribute to the quality of economic policymaking in Europe through open, fact-based and policy-relevant research, analysis and discussion.

MAPCOMPETE is a project, supported by the European Union, to provide an assessment of data opportunities and requirements for the comparative analysis of competitiveness in European countries. Further information is available at www.mapcompete.eu.

bruegel
33, rue de la Charité, Box 4, 1210 Brussels, Belgium
www.bruegel.org

ISBN 978-90-78910-36-7

9 789078 910367 > €15