DISCUSSION PAPER SERIES

DP18640

MICRODATA FOR ECONOMIC RESEARCH IN EUROPE: CHALLENGES AND PROPOSALS

Eric J Bartelsman, Filippo di Mauro, Sergio Inferrera, Marco Matani, Ugo Panizza and Michael Polder

OCCASIONAL PAPERS

CEPR

ISSN 0265-8003

MICRODATA FOR ECONOMIC RESEARCH IN EUROPE: CHALLENGES AND PROPOSALS

Eric J Bartelsman, Filippo di Mauro, Sergio Inferrera, Marco Matani, Ugo Panizza and Michael Polder

Discussion Paper DP18640
Published 24 November 2023
Submitted 24 November 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

MICRODATA FOR ECONOMIC RESEARCH IN EUROPE: CHALLENGES AND PROPOSALS

Abstract

While access to high-quality microdata is essential for economic research and policy evaluation, effective access to such data remains limited in Europe. It varies from country to country, with uneven information on access procedures. This is a major obstacle to social science research, including research on European competitiveness and the effects of climate change, inequality, globalization, and digitalization. The objective of this paper, which is based on a brainstorming exercise coordinated by CEPR and CompNet, is to assess the status quo and discuss a series of proposals for improving access to Microdata for economic research. We underline the need for developing the relevant tools for extended access to and use of European business statistics microdata. Building such tools entails both establishing the requested microdata and creating a body facilitating cross-country access to the established databases with harmonized content.

JEL Classification: C81, C82, C83, D22

Keywords: Microdata access, Data Harmonization, Policy evaluation

Eric J Bartelsman - e.j.bartelsman@vu.nl
Vrije Universiteit Amsterdam

Filippo di Mauro - filippodimauro1@me.com
IWH and CompNet

Sergio Inferrera - s.inferrera@hss22.qmul.ac.uk
Queen Mary University Of London

Marco Matani - marco.matani@studbocconi.it
IWH and CompNet

Ugo Panizza - ugo.panizza@graduateinstitute.ch
Geneva Graduate Institute and CEPR

Michael Polder - m.polder@mw.unimaas.nl
Statistics Netherlands

Acknowledgements
We would like to thank Giuseppe Berlingieri, Jo Van Biesebroeck, Patrick McGuire, Javier Miranda, Peter Boegh Nielsen, Regina Riphahn, Jean-Pierre Villetelle and participants to the CEPR-CompNet brainstorming exercise that took place in Paris on June 15-16, 2023 for helpful comments and suggestions. The views expressed in this paper do not necessarily reflect those of Statistics

Netherlands or any other institutions the authors are affiliated with.

Microdata for Economic Research in Europe: Challenges and Proposals

Eric Bartelsman
Vrije Universiteit

Filippo di Mauro
IWH and CompNet

Sergio Inferrera
Queen Mary University of London

Marco Matani
IWH and CompNet

Ugo Panizza
Geneva Graduate Institute and CEPR

Michael Polder
Statistics Netherlands

Abstract[*]

While access to high-quality microdata is essential for economic research and policy evaluation, effective access to such data remains limited in Europe. It varies from country to country, with uneven information on access procedures. This is a major obstacle to social science research, including research on European competitiveness and the effects of climate change, inequality, globalization, and digitalization. The objective of this paper, which is based on a brainstorming exercise coordinated by CEPR and CompNet, is to assess the status quo and discuss a series of proposals for improving access to Microdata for economic research. We underline the need for developing the relevant tools for extended access to and use of European business statistics microdata. Building such tools entails both establishing the requested microdata and creating a body facilitating cross-country access to the established databases with harmonized content.

Keywords: Microdata access, Data harmonization, Policy evaluation, European competitiveness
JEL codes: C81, C82, C83, D22

[*] We would like to thank Giuseppe Berlingieri, Jo Van Biesebroeck, Patrick McGuire, Javier Miranda, Peter Boegh Nielsen, Regina Riphahn, Jean-Pierre Villetelle and participants to the CEPR-CompNet brainstorming exercise that took place in Paris on June 15-16, 2023 for helpful comments and suggestions. The views expressed in this paper do not necessarily reflect those of Statistics Netherlands or any other institutions the authors are affiliated with.

EXECUTIVE SUMMARY

- Microdata are a public and essential good. Granular information broadens researchers' knowledge on firms, households, and public institutions across a vast range of critical economic, geographic, and social dimensions. As we live in a data-rich and data-driven society, Microdata are essential to build a solid foundation for research and policy advice.
- Drawing from a recent dedicated CEPR workshop in Paris (15-16 June 2023; see Appendix 3), this paper looks at Microdata access in Europe with the two main objectives of assessing: (1) where do we stand, and (2) how we can concretely improve it.
- Meeting participants agreed that effective Microdata access for economic research is not satisfactory in Europe and brainstormed on options for improving it. Lack of access to Microdata constitutes a major obstacle in all areas of economic and social science research, including research on European competitiveness and the effects of important developments such as climate change, inequality, globalisation, and digitalisation.
- Participants came out with some initial actionable ideas (e.g., building cross-institution consortia, establishing standards for data sources of published papers), which deserve a more detailed discussion focusing on their desirability and feasibility.
- All meeting participants agreed that the best arrangement for academic and policy researchers would ideally entail full access to the broadest possible range of datasets in each country, both in terms of being able to directly manipulate the data and to merge multiple datasets autonomously. This arrangement is far from being realistically implementable in the short run, due to a combination of national and EU legislation related to privacy and confidentiality concerns, as well as technical issues.
- Out of the Microdata sources indicated in Section 1 of this paper and discussed broadly in the CEPR workshop, we highlight current efforts around the business sector Microdata in the vaults of the National Statistical Institutes (NSIs) in the

EU.[1] and we present the Micro Data Infrastructure (MDI). This initiative, supported by the EU Commission and led by the Halle Institute of Economic Research (IWH), and CompNet represents a critical starting point for eventually achieving access to cross-country confidential micro data housed inside statistical offices. These data will not cover all needs discussed in the workshop; but the MDI represents an important example of an effort to solve a significant challenge in cross-country research and policy.

- The MDI is currently in development and is not yet accessible to the public. The raw data underlying the MDI is confidential and will always remain in the vault of the respective NSI. Such data have, however, been previously harmonised for a few countries and the same computer routine can run in any of the participating countries. The output from the MDI routines is in the form of micro-aggregated statistics that have gone through a disclosure process. This is in the tradition of NSI Research Data Centers as well as CompNet, but with potentially a much larger degree of access to the underlying raw data across countries, therefore allowing to develop even further the substantial academic literature already existent using that source.[2]
- The MDI is currently operational for two countries (FR and NL), but within a year is expected to include seven additional countries (Slovenia, Portugal, Finland, Germany, Austria, Italy, and Malta).
- In steady-state, the MDI will set up a mechanism to screen research proposals, 3rganize the creation and maintenance of open-source meta-data and tools, and interface with the institutional and technical situation at individual NSIs.
- A recently approved permanent line of funding from the German Government will allow the set-up of an operational unit within IWH-Halle; it will be composed and endowed with a team of experts and tasked to pursue the MDI project as well as continue the CompNet project.
- Support from the CEPR community will be key for this and other initiatives aimed at facilitating access to Microdata at the European level.

[1] Our initiative addresses more widely the NSIs of the European Statistical System (ESS), which also include EFTA countries (Iceland, Liechtenstein, Norway, and Switzerland) on top of the 27 EU Member States, as well as the UK. Nevertheless, we broadly refer to the EU as a more immediate shorthand for the ESS.

[2] The latest top publication using CompNet is Bighelli et al. (2023) in The Journal of the European Economic Association. Previous publications include Autor et al. (2014) in The Quarterly Journal of Economics.

1 INTRODUCTION

The significance of microdata lies in their ability to facilitate a comprehensive understanding of economic issues by examining the connections and interactions among various phenomena and agents.[3] Microdata also help to identify causal relationships and mechanisms which are difficult or even impossible to uncover with aggregate data. By studying these relationships, policymakers and researchers can effectively design projects, devise policies and target interventions where they are most needed, and closely monitor and assess the impact and outcomes of these initiatives, projects, and policies. This is necessary in a data-rich and data-driven society (Nagaraj and Tranchero, 2023). Microdata that are comparable across countries would also allow for investigating the role of different policy and institutional settings, going beyond the research possibilities of microdata for an individual country. Within Europe, for instance, the evaluation of the effectiveness of Next Generation (NG) EU investments as well as 'green transition' plans would be enriched by the availability of granular information.

While official statistics are crucial as the basis for evidence-based policy research, the statistical data are primarily available at an aggregate level. However, firms are heterogeneous even within narrowly defined sectors and may respond to exogenous shocks in different ways.[4] Similarly, households and individuals are heterogeneous. Consequently, there is a need for access to more granular data for research and policy purposes.

In most European countries, researchers face significant barriers to accessing official business statistics microdata due to legal, organizational, and technical reasons. Even though the European Statistical System (ESS) and many National Statistical Institutes

[3] The term 'micro-data' refers to unit-level data collected from sample surveys, censuses, or administrative systems. These data offer insights into the specific characteristics of individual people or entities, ranging from households and businesses to facilities, farms, and even geographical areas like villages or towns. For instance, they can reveal details such as the wages earned by individuals within a particular company during a specific time frame. The use of the term micro-data varies across fields. In finance, micro-data often refers to security or transaction-level information or information on specific traders or financial institutions; in industrial organization, micro-data refer to prices of certain products across different markets; in economic history, the term might refer to hand-collected data in various archives; and in political economy, it might refer to voting patterns or certain politicians or geographical area. In what follows, we will use the term micro-data to describe all types of granular data used in economic research.

[4] See Bartelsman and Doms (2000), Syverson (2011), and Blundell and Stoker (2005).

(NSIs) have in made progress in establishing firm-level databases (which improve researchers' access to firm level data at the national level), there is still a need to develop a method for cross-country analysis with firm-level data. Not only do legal, organizational, and technical details for access vary in each country, but so do the underlying data collection approaches, with differing statistical coverage and definitions of variables. The dearth of published research in the EU that exploits cross-country micro datasets is a consequence of these hurdles.

The objective of this paper, which is based on a brainstorming exercise coordinated by CEPR and CompNet, is to assess the status quo and discuss a series of proposals for improving access to microdata for economic research.[5]

The paper is organized as follows: Section 1 describes the status quo and the main challenges faced by researchers who work or would like to work with micro-data; Section 2 discusses a series of ideas that surfaced during the brainstorming exercise. Although none of these ideas is necessarily fully fleshed out, the points listed here can be considered as starting points for future discussion and action; Section 3 describes the Micro Data Infrastructure (MDI), a microdata access facilitator currently under experimentation by IWH and CompNet, which facilitates cross-country access to existing business-related firm-level databases with a harmonized content. Section 4 points to operational challenges faced by the MDI as a possible starting point for achieving – in a modular fashion – a fully satisfactory access to a much wider set of micro datasets for academic and policy use.

2. THE STATUS QUO

Microdata commonly used in economic research can be divided into five broad categories. These categories and their respective pros and cons are briefly described below.

[5] The brainstorming exercise took place in Paris on June 15-16, 2023, and included academic economists, policymakers, members of international organization and staff of statistical agencies. The program and list of participants is in Appendix 3.

2.1 NSI (OFFICIAL) MICRO-DATA

NSI microdata concern information from business registers, administrative sources, surveys, public registers, customs, et cetera. They are collected and held by the NSIs for the purpose of production of official aggregated statistics. According to the Code of Conduct of European NSIs, "statistical authorities should develop, produce and disseminate statistics in an impartial, objective, professional and transparent manner in which all users are treated equitably". The Code also contains indicators measuring whether "mechanisms are in place in the statistical authority to ensure equal access of all users to statistics".

These data are typically accessible through local research data centers. The data are linkable via unique identifiers for firms, households, persons, locations, and products. Some initiatives stand out for their ease of access outside the source country, or their cross country approach, including the French Secure Data Access Center (CASD, a consortium between INSEE, GENES, CNRS, École Polytechnique, HEC Paris and Banque de France), the International Data Access Network (IDAN, a network between six research data Centers from Germany, France, Netherlands and United Kingdom), and the Micro Data Linking (MDL) project. However, the MDL does not include access facilities for external researchers.

ESS (Eurostat and the NSIs), as producers of official statistics, need to invest in a European Microdata Infrastructure based on the business statistics governed by the newly adopted and implemented regulation on European Business Statistics.

2.2 COMMERCIAL MICRO-DATA

These datasets include publicly available, purpose-built data that have been privately collected, and data collected under a public mandate and then collated and re-sold by private firms. Examples are the Bureau van Dijck's (BvD) Orbis, Bloomberg, Dealscan, NL Analytics, CapitalIQ, Datastream, and Worldscope. These data sources are often expensive, provide incomplete information, or are biased towards large or listed firms. Moreover, these datasets are usually built with commercial operators (often in the financial industry) in mind and are distributed in a format that makes their use for

research particularly cumbersome (for instance it is often time consuming to download the data and use them in standard statistical packages).

2.3 ANONYMIZED MICRO-DATA

These are microdata where the unique identifier is typically removed. Two examples are the cross-country LIS income surveys and IPUMS. LIS harmonizes microdata on income and wealth collected from about 50 countries, while IPUMS collects nationally representative population samples from censuses in 103 countries. The advantage of these data is that they can be combined across countries and years (repeated cross sections). There are also some caveats. To start with, this type of data is not available for firms. Second, anonymization is not possible for 'thin tails' of household/individual data (top-coding, limited co-variates). Finally, it is not possible to link these data longitudinally or to other micro sources.

2.4 PRIVATE (TRANSACTIONS) MICRO-DATA

These are data collected by firms and cover transactions and traffic data, (e.g., bank accounts, e-commerce purchases, store loyalty cards, online b2c or b2b platforms, mobile phone tracking). In the EU, individuals have the right to 'be forgotten' and the right to 'get their own data' returned to them (GDPR). Arrangements have been put in place between researchers and companies (e.g. Google, Uber, etc.) to make these data available for scientific research. However, no systematic access procedures exist.

2.5 DATA AND SURVEYS COLLECTED BY INTERNATIONAL ORGANIZATIONS

International organizations collect microdata in many fields both for the purpose of their institutional mandates and for research interest. These include:

- the World Bank Enterprise Surveys (providing comparable firm-level data in emerging markets and developing economies) and Business Ready (the former Doing Business Survey),
- OECD surveys within the Local Employment and Economic Development (LEED) Programme, DBnomics (supported, among others, by Banque de France and pooling data from several international organizations),
- data on financial linkages collected from the BIS, the EU Labor Force Survey, the Eurobarometer,

- the ECB money and banking surveys like the Euro Area Bank Lending Survey (BLS), the Survey on the Access to Finance of Enterprises (SAFE), and the Euro Area Monetary Policy Event-Study Database (EA-MPD), as well as
- Microdata collected by the Single Supervisory Mechanism (SSM) and
- the European Stability Mechanism (ESM) during their operations. Also,
- the EBRD gathers microdata through ad-hoc randomized control trials and in collaboration with governments and National Central Banks, besides surveying businesses and households in countries of operations (Banking Environment and Performance Survey and Life in Transition Survey) and having designed a standardized legal framework (Development Data Partnership) that can be joined by private companies willing to share their data with researchers.

One caveat with these datasets is that they focus on specific fields and cover only subsets of firm population. Only rarely can researchers rely on systematic procedures to access the data.

Overall, however, at present, the microdata discussed in this section are under-utilized in academic and policy research, for two main interrelated reasons. First, as specified in Appendix 1, the possibility to access data varies by country from 'very easy' to 'restricted', or 'secluded'. In the case of the NSIs, either by statute or '*de facto*', tend to serve national audiences, sometimes even imposing a national (not even EU) affiliation requirement for research access to microdata (e.g., Denmark and Sweden). Secondly, many of these databases are costly making them outside the scope of many researchers.

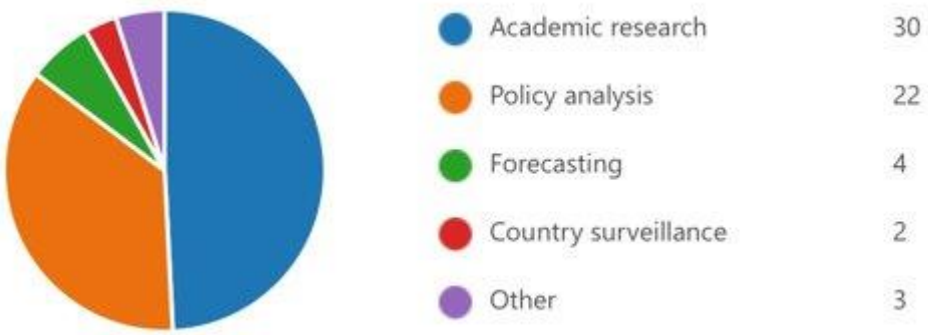
3. POTENTIALLY ACTIONABLE IDEAS

3.1 PARTICIPANTS PROFILE AND NEEDS

In preparation for the brainstorming exercise held in Paris on June 15-16, 2023 (see Agenda in Appendix 3), CEPR circulated a survey aimed at identifying the participants' profiles and needs in terms of accessing microdata.
Among the key results,

- Almost half of the respondents had an academic background and used (or are interested in using) microdata predominantly for academic or policy work (Fig.1).

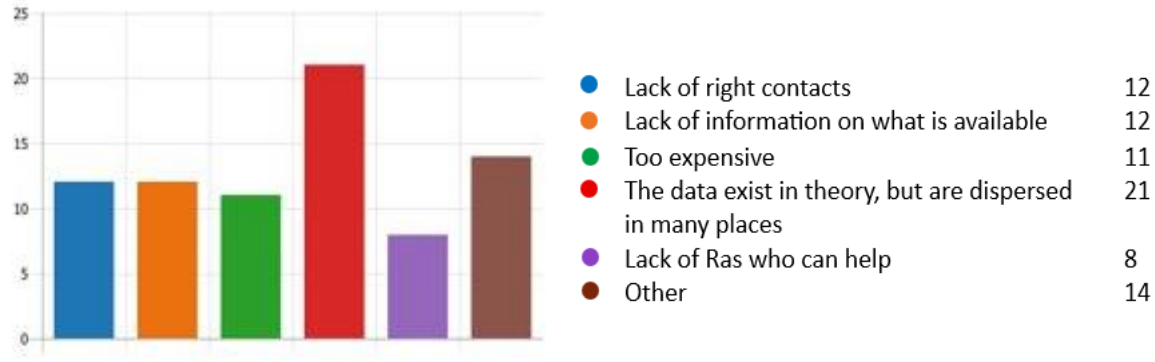
Figure 1. What kind of analysis do you use, or are you interested in using Microdata for?



Note: For the same respondent, answers are not mutually exclusive.

- When accessing microdata, most respondents faced obstacles that traced back to one among the following factors: contacts not being available, scarce information, expensiveness, or dispersion across different sources (Fig.2).

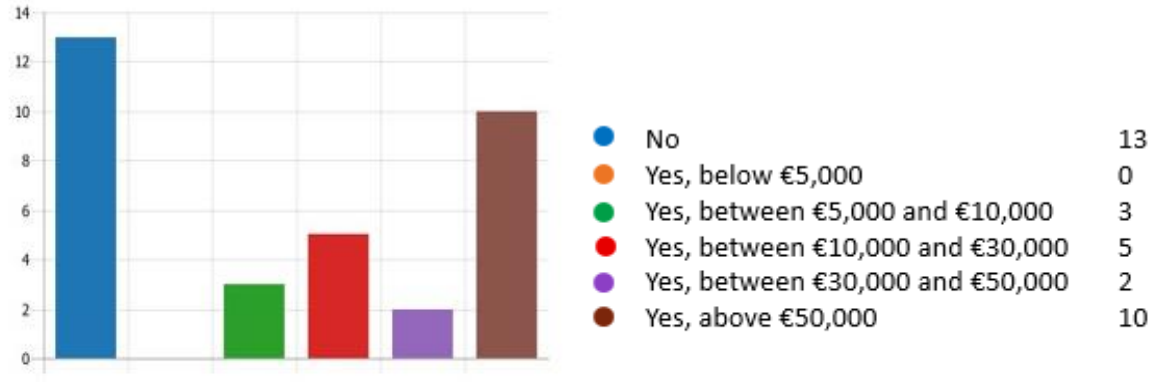
Figure 2. What are the main challenges that you or researchers in your field face in accessing microdata?



Note: For the same respondent, answers are not mutually exclusive.

- Notably, most respondents paid for accessing Microdata and for half of those who paid the price was above €50,000 (Fig. 3).

Figure 3. Did you pay for accessing such data? If so, what was the range of the cost?



➔ We interpret these results as reinforcing our claim about the urgent need for improved Microdata access in Europe.

3.2 ACTIONABLE IDEAS

Meeting participants brainstormed on options for improving access to microdata for economic research. Discussions were broad and are summarized here. This section lists **ideas** which, in our view, deserve a more detailed discussion focusing on their desirability and feasibility. Following this summary, we turn to a description of the MDI.

- **Build cross-institution access consortia.** Some researchers acquire commercial data with their own research funds while others have access to commercial data via the library of their institution. Several large universities or international organizations also have dedicated staff within the library that help researchers access and working with specific micro datasets. Access to microdata is thus easier for researchers who work in rich and large institutions which can afford to pay for such datasets. Researchers who work in smaller institutions tend to have limited access to such data because of scale problems (it is difficult for a library to justify purchasing an expensive dataset if this dataset is only used by one or two researchers). However, libraries in small universities could cooperate and build consortia to split the costs and grant access to data

to their researchers. In a sense, they already do this through inter-library loans, which could be extended to include similar arrangements for data sharing.

- **Build a repository of information about data access to and availability of the data used in existing pieces of economic research.** Data appendices or tables with data sources are a common feature of published research. However, not all papers include such information and, when they do, there are no established presentation standards. Working paper series and journals could ask authors to fill out a standardized web form with some basic information about the data used in the paper. Basic information (e.g., one paragraph describing the data, a weblink, data access policy of the data supplier, possible cost, and one contact person) could be provided in a few minutes at the time of submission. The information could then be used to form a global database of economic datasets.

- **Clear rules for access to confidential data maintained by international organisations, statistical offices, central banks, regulatory authorities, and other public institutions**. There are several organizations that have access to a wealth of confidential data that can be used for economic research. Certain institutions have clear and transparent rules for access to these data, e.g., via visiting programs or paid or free access (in some cases data can only be accessed in computers located within the institution, in some cases data can be remotely accessed with black boxes; see Table 1 in Appendix 1 for a first assessment related to business related micro-data). However, several institutions only grant ad hoc access to researchers who have personal connections with people within the institutions. It would be desirable if all publicly funded institutions (or institutions with a public mandate) had public and clear rules for access to data that do not favour “connected” researchers. For researchers interested in improving the data infrastructure in their own countries, the ‘best practice’ elsewhere could serve as a template.

- **Create political support for greater Microdata availability.** A coordinated campaign of op-eds by prominent scholars would make policymakers aware

that countries would obtain substantial benefits from better data access policies. Good access to microdata provides incentives for economic research. Researchers are happy to work for free to evaluate policies implemented by countries with good data. Better policy evaluation and analysis will eventually lead to better policymaking. This is especially important for projects/initiatives that involve the use of public funds, such as the those linked to the Next Generation EU program.

- **Coordinate interdisciplinary research on new methods of data collection and dissemination.** Some of the information contained in commercial databases could be scraped from the web and then made available for free to the research community. There are, however, three issues related to the implementation of such a strategy.

    - Incentive structure: What type of recognition would researchers obtain to produce this public good?
    - Technical coordination among different projects: How can projects be coordinated so that the various datasets are compatible and coherent?
    - Legal considerations: The fact that data are available on the web does not necessarily mean that it is legal to collect and disseminate them. How can data producers be sure that what they do is legal?

  These are complex interdisciplinary questions that could be an ideal match for a large interdisciplinary Synergia grant.

➔ **There was broad consensus among meeting participants on having the CEPR functioning as a hub for coordinating the deployment of concrete initiatives to ease microdata access in Europe**, also in view of similar initiatives promoted by the NBER, which were limited however to sharing commercial sources like BvD Orbis.[6]

[6] See Kalemli-Ozcan et al. (2015).

4. COORDINATING FIRM-LEVEL MICRO DATA ACCESS IN EUROPE: THE MDI

Drawing from the wish of **creating a comprehensive microdata hub**, we present here the concept and operations of the Micro Data Infrastructure (MDI), a microdata access facilitator currently under experimentation by IWH-CompNet. By streamlining cross-country access to business-related firm-level databases with a harmonised content, the MDI holds the promise of catalysing advances in the European microdata infrastructure. Such a scheme has the potential to be made fully operational as it concerns capacity of screening and run research codes. Moreover, it could be further widened in scope by increasing the number of variables as well as countries included, thus serving a broader range of economic fields.

4.1 BACKGROUND

Fragmented information, data unsureness, and bureaucratic hurdles jeopardise cross-country economic research in Europe. To conduct cross-country research on business-related NSI micro-data, a researcher must investigate access modalities and choose the datasets and variables that are available for a set of countries (see Appendix 1 for more details). Next, research must be set up separately for each country, even in the best case where all countries have remote access. One example is the OECD Multiprod, which is run mainly in cooperation with national researchers and statisticians. It requires separate access to the available NSI data in each country and uses remote execution and remote access. It is a valuable but time-consuming exercise. Although it yields a high-quality cross-country micro-aggregated database with adequate harmonisation across countries, non-OECD researchers cannot currently get access to it.

At present, three feasible alternatives exist for cross-country analysis on business firms by interested individual academic or policy researchers. However, all of them are suboptimal:[7]

[7] Another alternative is to setup and run one's own cross-country survey, such as the World Management Survey (www.worldmanagementsurvey.org). Clearly, unless one is fortunate to receive substantial funding, this does not qualify as feasible for the bulk of researchers, and access to such data may again be restricted for those not part of the project or network.

- **Commercial Microdata (e.g. Orbis)** - These include both commercial and non-commercial sources. The disadvantages of the former are that they are expensive, they have no external source linking and that have a limited number of variables being available. Moreover, it has been documented that, for some research purposes, these data are less suitable or of lesser quality than official sources (see, for instance, the extensive cleaning procedures needed in Bajgar et al (2019) due to Orbis' changing coverage, among other things).

- **Non-commercial sources** (IDAN). The International Data Access Network (IDAN) is a collaboration between six Research Data Centers to provide controlled access to microdata for these countries, an objective very close to what we have in mind. However, IDAN is rarely used for research, suggesting that effective access is hampered by bureaucratic access hurdles.

- **Public micro-aggregated cross-country Data -** ESSLait and CompNet are both examples of micro-aggregated databases.
    - The main task of ESSnet's ESSLait was to develop a common micro-aggregated data infrastructure to allow future analysis. Unfortunately, this initiative was discontinued. Data up to 2010 are available in the **Micro-Moments Dataset** at the Safe Centre at Eurostat. They include data on ICT usage, innovation, and economic performance in enterprises.
    - **CompNet** was founded by the European System of Central Banks in 2012 and since 2017 it is hosted at the Halle Institute for Economic Research (IWH). Over the years, CompNet has been producing a micro-aggregated dataset of harmonized indicators of productivity, competitiveness, trade, labor, and finance for more than 20 European countries. The dataset - updated on an annual basis – is now in its 9th vintage and is made available to accredited users. It covers the 1995 to 2020-21 period for most of the countries. CompNet is now being set up as an operational Unit within IWH.

4.2 THE MDI: CONCEPTS AND OPERATIONS

The MDI builds on existing European Statistical System (ESS) frameworks, the progress in microdata linking done by individual NSIs, as well as on many years of

applied research and data management experience by the authors of this paper, most lately within the IWH-CompNet/MDI project.

More specifically, the MDI draws on two IWH projects recently financed by the EU Commission:

1. MICROPROD (2019-2022) which expanded on the Micro Data Linking (MDL) lead by Statistics Denmark to create the Micro data Infrastructure (MDI) comprising the NSIs of Denmark, Finland, France, Netherlands, Norway, and Sweden.
2. MULTIMSPROD (2022-2024) which is expanding the set of countries included in the above project to Portugal and Slovenia, as well as Germany and Austria using a similar set up.

Like in the above projects, the MDI uses the Business Register as a 'backbone' (see Bartelsman et al, 2020), to link together several existing datasets and surveys. Figure 4 indicates the currently available data for the participating countries.

Figure 4. MDI dataset

The **MDI** is based on the following principles, set up in the formal Memoranda of Understating (MoU) between IWH and the individual NSIs:

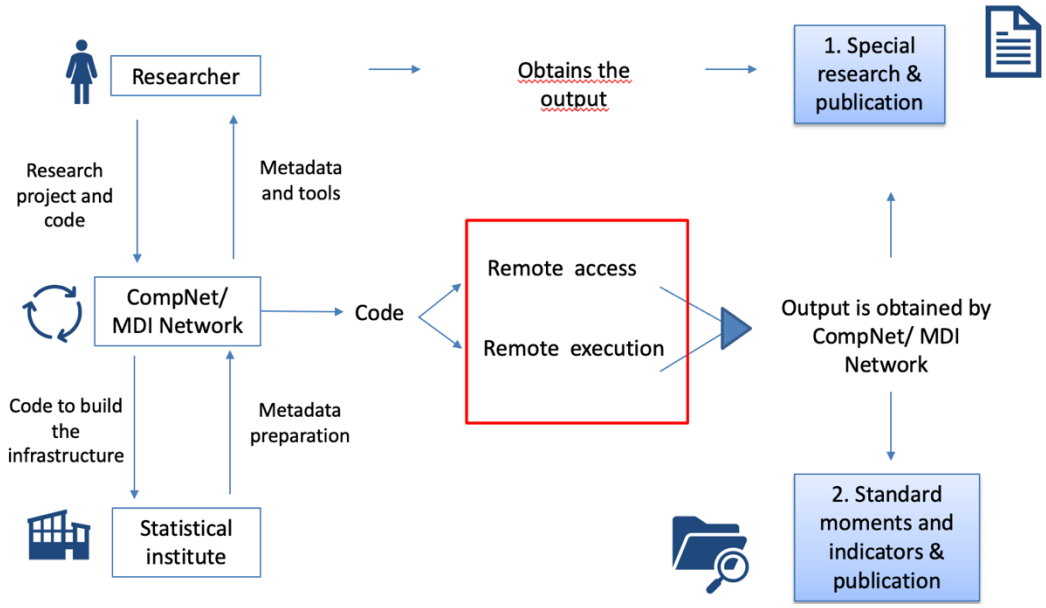
1- **Data location** – Microdata always remains strictly on the servers of the respective NSIs.

2- **Access** - The CompNet/MDI staff take care of arranging – on a country-to-country basis – the best available solution at the national level by facilitating

bilateral data access agreements with the respective NSI. This is a huge advantage for users, who would otherwise need to go through the lengthy, potentially opaque and quite heterogeneous procedures to gain access in each country. Access modalities are country-specific and involve remote execution by NSI staff (now the case for Portugal and Finland), or direct remote access to microdata by CompNet/MDI staff (for France, Netherlands and Slovenia), depending on country-specific arrangements.

3- Output – The bilateral agreement includes setting up adequate confidentiality arrangements for the disclosure of output produced in the different countries by the routines developed for research or policy purposes. Released output adheres to the regulations in place at the pertinent NSI. Disclosed output is eventually collected and 16armonizati by **CompNet/MDI** staff. Figure 5 below summarises the operational set-up.

Figure 5. The MDI operational set up



4- Operational costs – At present, in its experimental basis, IWH covers the operational costs of MDI via a two-year EU financing (**under grant agreement No 101101853**) to end in October 2024. The current MDI currently does not have in place a mechanism to interface with researchers. It will develop and test the operational procedures in the coming year.

5. CHALLENGES AND THE WAY FORWARD

5.1 FINANCING

Drawing from previous experience, ensuring permanent and adequate funding is critical to initiatives aimed at linking micro-data. Indeed, several of the earlier projects ceased as soon as funding ended. Financing needs relate to the coordinating organization, as well as to the NSIs providing the data. While on the former there are recent positive developments, on the latter some further discussions on best course of action is needed.

1. **Leading organization**. IWH is playing this role and expected to do so in the future, also in view of a permanent line of funding recently approved and awarded to IWH from the German Government for this purpose. The new line of funding will be fully disbursed in 2025. This is a positive development for the MDI initiative. It will allow the set-up of an operational CompNet unit within IWH-Halle composed and endowed with a team of permanent experts. This team will include Ph.D. economists, programmers, and statisticians tasked to coordinate the data 17armonization work by the associated NSIs, as well as the periodical execution of codes generated by users, while pursuing their research agenda.

2. **National Statistical offices (NSIs)**. Over and above the financing needs of the coordinating organization, there are two non-trivial costs at the NSIs, which we can distinguish in two main categories: upfront costs and operating costs.

    1) **Upfront costs**. These are the costs to create the infrastructure at the NSIs' data centres. These costs are country-specific depending on a) the NSI's current availability of Microdata linking infrastructure; and b) upfront financing requirements depending on the extent NSI's data access procedure. For some NSIs (e.g., the Nordic countries, Netherlands, France), a substantial set of enterprise data has already been harmonised in the above mentioned and other projects. Many NSIs consider this effort a part of their regular mandate. Other NSIs need to identify additional funding for

this initial investment. While these costs are hard to evaluate and are country specific, they range between 50-100,000 Euro per country.

2) **Operating costs**.

   These depend on the modality of data access.

   a. Costs for remote execution: Funds to finance working hours of statisticians and programmers at the NSIs who provide indirect data access (e.g., currently Portugal and Finland).

   b. Usage costs: These are the download costs charged by NSIs that provide direct (remote) data access to CompNet/MDI staff (at present France, Netherlands and Slovenia), who operates on behalf of the users. They include fees for access to data at each NSI, for disclosure of the output, and subscription costs for data use to cover ICT infrastructure and staffing costs of all the NSIs involved in the project. We estimate it to be, on average, 5,000€ per research project.

Optimally, those NSI related costs, should be taken up by national authorities (being interested anyway for complementary reasons to develop their micro data infrastructure). Going forward, it will be necessary to find sources of financing for NSIs that cannot be funded by national sources. Discussion might include the feasibility of recovering some of those costs via user fees.

5.2 INTERACTION WITH EUROSTAT

The MDI builds on past EU initiatives and intends to strictly coordinate with Eurostat to fully align its operations with the ultimate aims of Eurostat. This includes, for instance, eventually enlarging the infrastructure to include the entire EU, without excluding the possibility of expanding it further, even beyond EU borders.

5.3 THE MDI GOING FORWARD

The MDI is at present very much on an experimental basis. It is also rapidly expanding and developing its infrastructure both across countries and within IWH. Procedures for interactions with the larger community need to be developed. This means that it is not yet operational and able to engage with external researchers. Regardless, the MDI has a goal of becoming a fully operational tool open to researchers in steady-state. Its built-in “modularity” will allow to include additional data sources in the underlying

dataset in view of the needs and staff capacity both centrally and at the NSIs (e.g. Linked employer-employee data). This will allow an almost limitless capacity to encompass broader economic topics, increasing the number and types of potential users. Such expansion is also possible as far as country coverage is concerned, even beyond the EU.

CEPR researchers are invited to engage with MDI with the objective of expanding MDI coverage and devise the optimal sequencing of this expansion. Such dialogue should also involve international organisations engaged in parallel initiatives, such as OECD, IMF and World Bank.

6. CONCLUSIONS

This paper has detailed how access to cross-country granular data in Europe is hampered by a series of heterogeneous rules at the national level. Furthermore, the lack of homogeneous data access practices is costly for researchers. The outcome is a lack of cross-country granular analysis that would be of critical value for research and policy making.

As a remedy, the paper has presented several actionable ideas to be further discussed and developed, as well as a concrete microdata access facilitator: the Micro Data Infrastructure (MDI), handled by IWH-CompNet. Its goal is to ease the process of accessing cross-country micro-data, to fully exploit the potential of this public good. The MDI coordinates the required output provided by participating NSIs, building on existing national data access procedures and contacts already established with several NSIs. In steady-state, the MDI will enable researchers to access individual country microdata in a fair, fast, and transparent way and to link them as they please across data sources; with due respect of course of established confidentiality guarantees. In a “modular” fashion, the project could be expanded further to encompass a broader range of countries and economic fields.

The CEPR community can play a critical role in supporting initiatives aimed at improving microdata access in Europe. By informing the public and policymakers on how research that uses microdata can lead to better policies, CEPR researchers can

promote the funding of the National Statistical Institutes (NSIs) and make sure that data sharing is among the main objectives of these institutes.

REFERENCES

Bajgar, M., Berlingieri G., Calligaris S., Criscuolo C. and J. Timmis (2019), "Industry Concentration in Europe and North America", OECD Productivity Working Papers, No. 18, OECD Publishing, Paris, https://doi.org/10.1787/2ff98246-en.

Bartelsman, E. J. and M. Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature*, 38 (3): 569-594.

Bartelsman, E., Hagsten, E., & Polder, M. (2018). Micro Moments Database for cross-country analysis of ICT, innovation, and economic outcomes. *Journal of Economics & Management Strategy*, *27*(3), 626–648. https://doi.org/10.1111/jems.12256

Bartelsman, E. J., Haelbig, M. and F. di Mauro. 2020 – Micro Data Infrastructure (MDI)-Methodological Report on Cross-country Analysis of Newly Developed Firm-level Indicators, Microprod Mimeo.

Blundell, R., and T. M. Stoker. 2005. "Heterogeneity and Aggregation." *Journal of Economic Literature*, 43 (2): 347-391.

Bricongne, J.C., Carluccio J., Fontagné L., Gaulier G., and S. Stumpner (2022). From Macro to Micro: Large Exporters Coping with Common Shocks. Working Paper Bank of France No. 881.

Di Mauro, F. and P. Lopez Garcia – Assessing European competitiveness: the new CompNet microbased database, ECB Working Paper no 1764

Fauth, M., Jung, B. and W. K. Kohler, German Firms in International Trade: Evidence from Recent Microdata (2023). CESifo Working Paper No. 10523, Available at SSRN: https://ssrn.com/abstract=4499478 or http://dx.doi.org/10.2139/ssrn.4499478

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49 (2): 326-65.

APPENDIX 1 – NSI MICRODATA ACCESS IN EUROPE

Access to, and methods of access to, business-related NSI official microdata vary across NSIs and can be grouped as follows:

- No Access.
- Only accessible by NSI employees, for statistical purposes.
- Restricted access:
  - → NSI Research: an NSI employee can conduct analysis on behalf of clients (academic, government, private sector) against fees covering data access and analyst time.
  - → Remote execution: Researcher's analysis is run at NSI site, usually through intervention of NSI employee, with disclosure check of output. Also, in this case there are fees for data facilities and some analyst time.
  - → On-site access: Qualified researchers use facilities at NSI site with disclosure analysis of output. Fees are due for data access.

- Remote Access: Qualified researchers are provided with a facility to analyse data securely at their own site or even in any place. Fees are due for facility usage and data access (e.g., France, Netherlands, Portugal, Slovenia).

Even within the same category, NSIs' access methods vary by definition of qualified researcher, scope of available datasets, possibilities for data linking (longitudinally or across sources), quality of documentation (for instance, in native language only), and quality/cleanliness of data. Table 1 describes some of these heterogeneous practices.

Table 1. Heterogeneous business microdata availability across NSIs

EU Country	Access Method	Qualified Researcher[8]	Data Scope and Domain	Data Documentation
France	Remote through dedicated box (CASD)	Both national and international accredited institutions (Universities, Research Centers, etc.)	Business	
Germany	Physical access at research data center	Researchers affiliated with eligible institutions	Business related data, balance-sheet data, product-level data	
Italy	Remote (Scientific Use Files) and physical (Secure Use Files)	Research institutions recognized by Comstat or by Eurostat; alternatively, procedure for recognition of the relevant Entity as a matter of priority	Business and households	
Netherlands	Remote with your own PC	Both national and international accredited institutions (Universities, Research Centers, etc.)	Business and non-business sector; households and individuals; real estate; prices; trade; transport	Available (in Dutch)

[8] These can be, for instance, only researchers affiliated with national institutions.

APPENDIX 2 – PROGRAM AND LIST OF PARTICIPANTS

A3.1 PROGRAM

Thursday, June 15	
12:30-13:30	**Lunch**
13:30-13:45	**Welcome and Workshop's Objectives** Tessa Ogden (CEPR), Filippo di Mauro (CompNet and CEPR), Ugo Panizza (The Graduate Institute, Geneva and CEPR)
13:45-14:00	**Survey Results – brief overview** Sergio Inferrera (Queen Mary University)
14:00-14:45	**Session 1 - What We Have: LIS and CompNet** **Chair: Hibret Maemir (The World Bank)** **Panellists** Filippo di Mauro (CompNet and CEPR) Teresa Munzi (LIS) 10 minutes per presenter 15 mins discussion
14:45-16:05	**Session 2 - What We Have: International Organizations** **Chair: Christophe Benz (CEPREMAP)** **Panelists** Johannes Breckenfelder (ECB) (Online) Ralph de Haas (EBRD) Patrick McGuire (BIS) Norman Loayza (World Bank) (Online) Douglas Sutherland (OECD) Filiz Unsal (OECD) 10 minutes per presenter 40 mins discussion
16:05-16:25	**Coffee/Tea Break**

16:25-16:55	**Session 3 - What We Have: Microdatain France.** **Chair: Javier Miranda (IWH)** **Panelists** Michel Julliard (Banque de France) Rémy Marquier (CASD) Jean-Pierre Villetelle (Banque de France) 15 minutes presentation 15 mins discussion
16:55-18:45	**Session 4 - What we want: Development, Industrial Organization, Organizational Economics, Public Economics, Political Economics, labor.** **Chair: Regina Riphahn (FAU Erlangen-Nürnberg)** **Panelists** Marco Manacorda (Queen Mary University, London and CEPR) (Online) Javier Miranda (IWH) David Seim (Stockholm University and CEPR) Gabriel Ulyssea (UCL and CEPR) (TBC) Jo van Biesebroeck (KU Leuven and CEPR) (Online) 10 minutes per panelist (what is available and what is not but could be) 50 mins discussion.
18:45-19:00	**Wrap up – main takeaways**
20:00	**Dinner**
Friday, June 16 **Sciences Po, Rue de la Chaise (room 933)**	
08:30-08:40	**Coffee**
08:40-10:30	**Session 5 - What we want: Finance, International Macro, Macroeconomics and growth, Monetary Economics, Economic history, Trade** **Chair: Ugo Panizza (The Graduate Institute and CEPR)** **Panelists** Giuseppe Berlingieri (ESSEC) Giancarlo Corsetti (EUI and CEPR) (online) Miklos Koren (Central European University and CEPR) (online) 10 minutes per panelist (what is available and what is not but could be) 50 mins discussion

10:30-10:50	**Coffee/Tea Break**
10:50-12:30	**Session 6 - How do we get there?** **Chair: Filippo di Mauro (CompNet)** Open Discussion
12:30-13:00	**Next steps – Close of workshop** Tessa Ogden (CEPR), Filippo di Mauro (CompNet and CEPR), Ugo Panizza (The Graduate Institute, Geneva and CEPR)

A3.2 LIST OF PARTICIPANTS

Eric Bartelsman (Vrije Universiteit)
Christophe Benz (CEPREMAP)
Giuseppe Berlingieri (ESSEC)
Johannes Breckenfelder (ECB, Research Economist)
Giancarlo Corsetti (EUI and CEPR)
Filippo di Mauro (CompNet Chairperson and CEPR)
Ralph de Haas (EBRD Director of Research and CEPR)
Sergio Inferrera (PhD student, Queen Mary University of London)
Michel Juillard (Banque de France)
Miklos Koren (Central European University and CEPR)
Norman Loayza (World Bank)
Hibret Maemir (The World Bank)
Javier Miranda (Halle Institute for Economic Research and Friedrich-Schiller University Jena)
Marco Manacorda (Queen Mary University of London and CEPR)
Marco Matani (CompNet Consultant)
Patrick McGuire (International Banking and Financial Statistics)
Remy Marquier (CASD)
Teresa Munzi (Cross-national data centre in Luxembourg)
Tessa Ogden (CEPR Chief Executive Officer at Centre for Economic Policy Research)
Ugo Panizza (Graduate Institute of International and Development Studies)
Michael Polder (Centraal Bureau voor de Statistiek)
Regina Riphahn (FAU Erlangen-Nürnberg)
David Seim (Stockholm University and CEPR)
Douglas Sutherland (Head of Division at OECD)
Filiz Unsal (OECD)
Gabriel Ulyssea (UCL and CEPR)
Jean-Pierre Villetelle (Banque de France)
Jo van Biesebroeck (KU Leuven and CEPR)