CompNet The Competitiveness Research Network

Micro-Data Infrastructure CompNet

Filippo di Mauro (*)
Chairman of CompNet

8, 2022

(*) with input from Eric Bartelsman, Sergio Inferrera, Bernardo Mottironi and Alessandro Zona Mattioli

MDI ARCHITECTURE

- Introduction and motivation
- The players
  - Researchers (govt, academic), NSIs, MDI TEAM
- The micro data
  - Registers, surveys, administrative sources, other
- The “Modus Operandi”
  - General architecture
    - Remote execution
    - Remote access
- Legal status output
  - NSI owns the data,
  - MDI intellectual property tools

MDI – Conceptual framework

- In the EU, micro-level information is broadly available in the respective National statistical Institutes (NSIs), but only sparsely used for issues related to data access (confidentiality) and cross country comparability.
- In recent years there has been some progress in harmonizing micro-level data, for example by regulations on Business Registers (Regulation (EC) No 177/2008) and surveys on ICT usage in business (Regulation (EC) NO 808/2004), as well as by Eurostat model questionnaires, e.g. for the Community Innovation Survey (with voluntary participation).
- With the Business Register as a 'backbone', the MDI intends to assist NSIs (of a larger number of EU countries as possible) to link information from these datasets and other survey or register-based information at the individual enterprise-level (here loosely referred to as 'firm-level').
- The result is an incredibly rich set of information which will allow (academic and policy institutions) Researchers to understand for instance how a variety of factors affect productivity at the firm level.
- So far, six NSIs have been involved in the initiatives (FRA, NL, SWE, NOR, FIN, DEN)

MDI motivation- Current status with many hurdles

A Researcher to do do research goes through:

1. Research proposal
2. Data access procedure
3. NSI technical restrictions
4. Fixed costs of getting to know the data
5. Decisions on data handling (e.g. outliers treatment)
6. Programming the code

The NSI, for each research project has to:

1. Provide access procedures
2. Provide dedicated assistance
3. Provide Microdata research support
4. Do a disclosure analysis
5. Sometimes host researchers

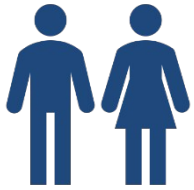
The resulting disclosure-free output will be used in a publication, **virtually impossible to replicate in other countries** after many hurdles!

Our set up

MDI – the players

NSIs

- Provide access to run code (remote access or remote execution)
- Provide metadata
- Authorize release of disclosure-free output



Researchers

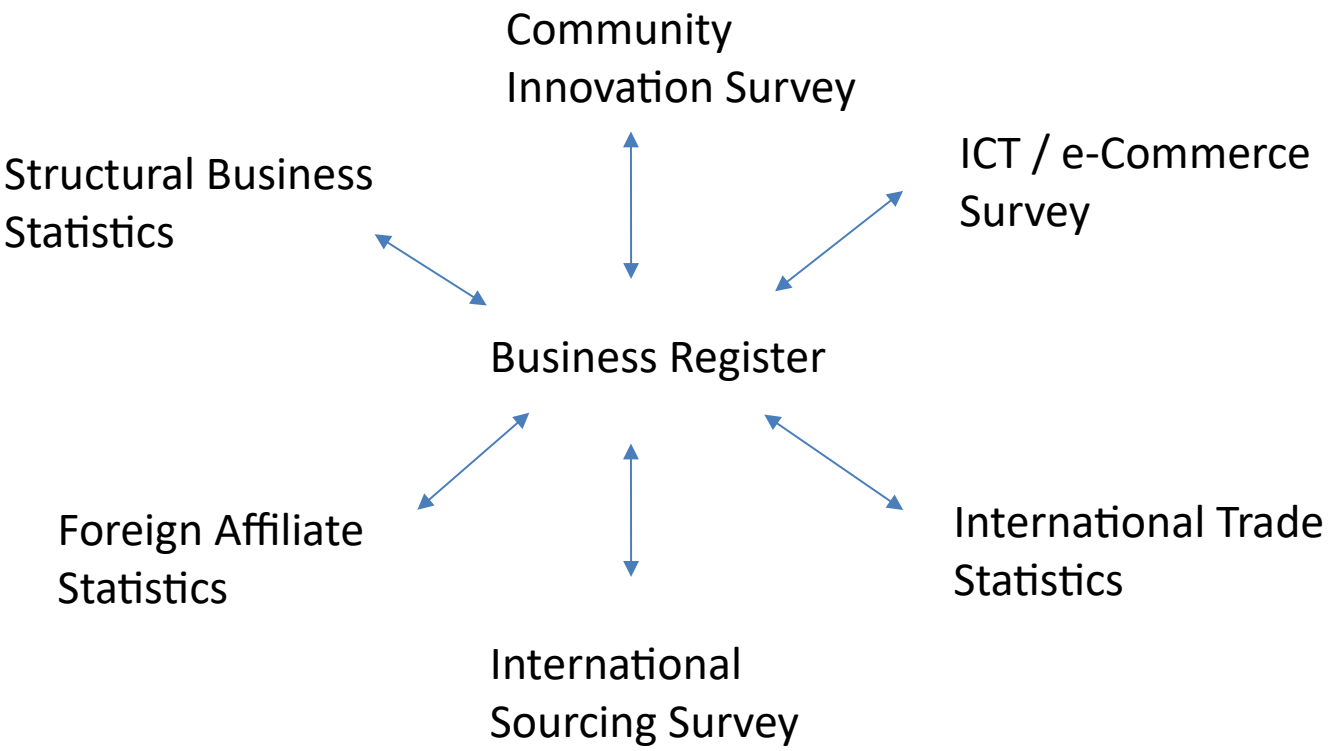
- Good research questions and analytical/programming expertise (Stata/SAS/R)
- Have funds

MDI - CompNet

- Legal entity with NSI access
- Expertise to evaluate proposals and support analysis/coding
- Coordination of 'open source' tool development
- Managing finance for fixed costs and for NSI subscriptions and fees

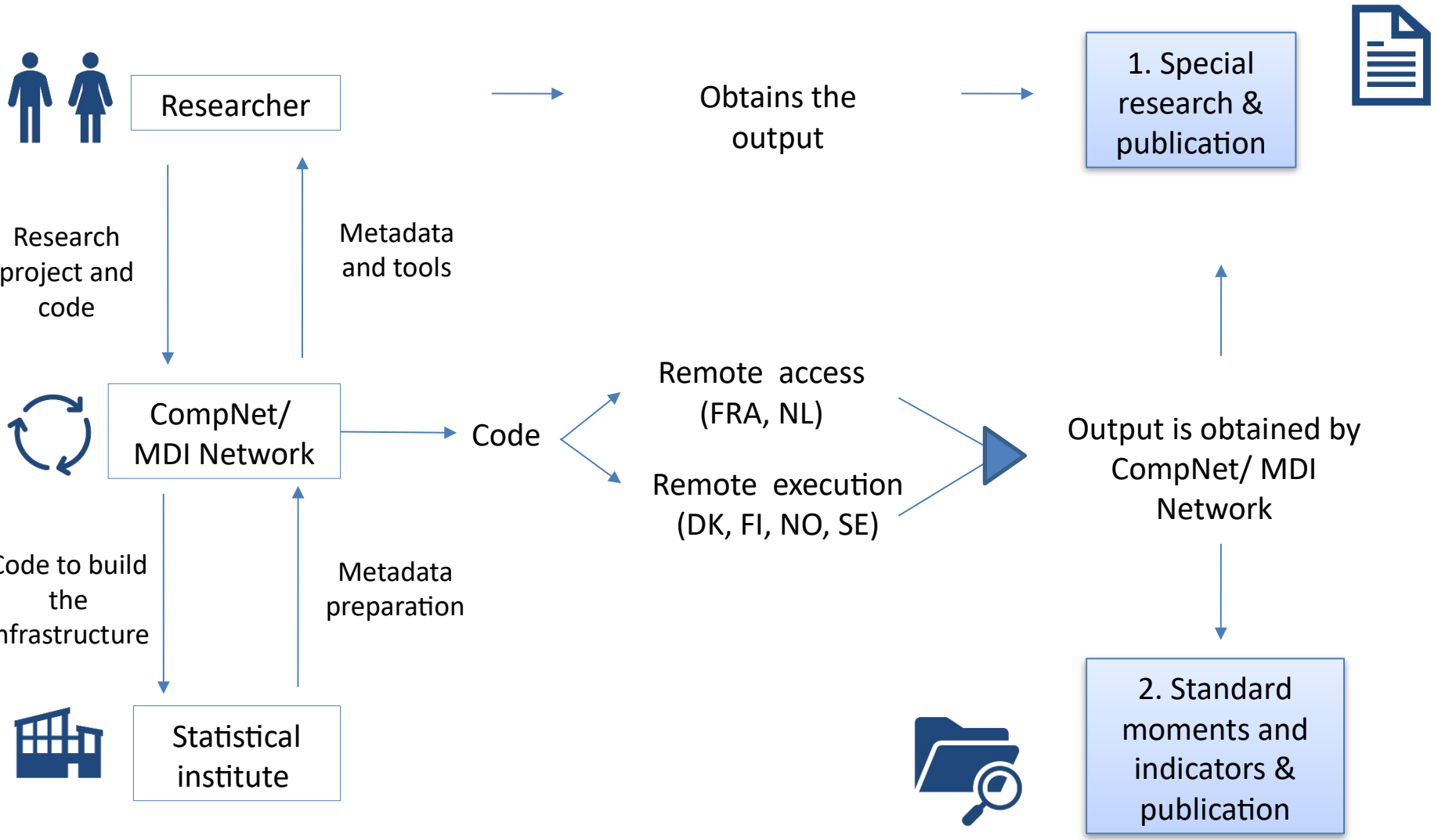
MDI – the micro data

Homogeneized set of **datasets** (sourced from Eurostat tables):



- For each dataset, **same list of indicators** (see excel file) with **Firm level** data.
- **Time coverage**: for now 2007-2017, expanding on a country basis.

MDI architecture



MDI- Micro-data access policy

- **MDI/CompNet team**:
  - has remote access to micro level raw data.
  - manages the running of the code written by users.
  - develops tools to increase data accessibility for the users.
- **NSIs**:
  - Update the data and prepare the infrastructure.
  - Ensure comparability cross country.

MDI- Legal Framework

- MDI/CompNet team **negotiates with NSIs long term contracts to access the raw data**:
  - Ownership of the data remains with the NSIs, who are active part of the project.
  - Intellectual property is of the data users/MDI-CompNet team.
- The Team has currently direct access to data from the **Netherlands** (CBS) and **France** (CASD).

Going Forward

- We are currently finalising with the Commission a program of assistance to EU national productivity Boards (PBs) which will allow to entend the MDI from 6 to 11 countries in a first stage (additional are GER, POR, SLO, SKO, LAT).
- And potentially to the whole EU going forward

APPENDIX 1
data sources

Business register

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

← BACK

Structural Business Statistics

Structural Business Statistics (SBS) The Structural Business Statistics (SBS) describe the economic activities within the business economy, including industry, construction, distributive trade and services. SBS indicators at the detailed sector level are transmitted to Eurostat and published by all European Statistical System (ESS) members (EU Member States, Norway and Switzerland, some candidate and potential candidate countries). Harmonization of the SBS has taken place regarding the detail and coverage of the sectors (now NACE 2.1) and the statistical definition of the transmitted indicators (Commission Regulation (EC) No 250/2009). Generally, the SBS indicators in each country are collected at the level of individual enterprises engaged in economic activity.

← BACK

Community Innovation Survey

Community Innovation Survey (CIS) The Community Innovation Survey (CIS) is part of the EU science and technology statistics and provides mostly qualitative information on firm innovative activity. Surveys are carried out every two years by EU member states and a number of ESS member countries on a voluntary basis. The harmonized survey contains information on the types of innovation and various aspects of the development of an innovation, such as the type of funding and innovation expenditures. The CIS covers both innovation outputs and the innovative process and inputs (type of funding, R&D expenditure) and distinguishes four innovation types: process, product, organizational, marketing, thus covering both innovative property as well as capabilities and organizational capital. Additionally, the CIS asks about the novelty of the innovation, i.e. whether it is new for the market, new to the country, developed by the firm or was adopted, and thus provides information about the innovative value.

← BACK

ICT usage/ E-commerce Survey

ICT usage/ E-Commerce Survey (ICTEC) The Community survey on ICT usage and e-commerce in enterprises is an annual survey conducted since 2002, which collects information on the use of information and communication technology, the internet, e-government, e-business and e-commerce in enterprises. Like the CIS, the EC survey contains mostly qualitative data. The ICT use survey measures various dimensions of firm technology use. Besides software and databases being considered as an integral part of intangibles, the adoption of certain technologies also provides information about firms' organizational capital and ICT capabilities both in the firms' internal operations and regarding the firms' supplier and buyer relationships. The qualitative information in the survey can be used to construct an ICT intensity index which allows for variation in the underlying source variables, thereby overcoming the issue with changing survey questions and the saturation of certain variables over time

← BACK

International Trade Statistics

International Trade Statistics Firm-level statistics concerning exports and imports are the International Trade in Goods Statistics (ITGS) and International Trade in Services Statistics (ITSS). International trade in goods statistics (ITGS) measure the value and quantity of goods traded between EU Member States (intra-EU trade) and goods traded by EU Member States with non-EU countries (extra-EU trade) broken down by types of goods (Combined Nomenclature) and by partner countries. The providers of statistical information differ between intra and extra EU-trade. In the first case, it corresponds to all taxable persons reporting transactions exceeding a certain threshold fixed by member states; in the second one, it corresponds to administrative data from the customs declarations lodged by natural or legal persons in the customs administration. International Trade in Services Statistics (ITSS) typically cover trade in services, i.e. transactions paid for the services that have taken place between the residents and non-residents.

← BACK

Foreign Affiliate Statistics

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

← BACK

International Sourcing Survey

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

← BACK