



SHARING CONFIDENTIAL MICRODATA FOR STATISTICAL PURPOSES – OPPORTUNITIES AND CHALLENGES

Julien Dupont, OECD Statistics and Data Directorate

UNECE/EFTA Workshop on Modernising Statistical Legislation

29-30 May 2024, Tirana, Albania



Outline

1. Why is microdata important for statistical production?
2. Legal environment
3. Technical and institutional arrangements
4. Conclusions

A night cityscape with a digital overlay of glowing lines and dots, suggesting data and technology. The background shows a city skyline with illuminated buildings and streets. Overlaid on this are numerous vertical lines of varying heights and colors (blue, purple, pink) that terminate in small glowing circles, resembling data points or network nodes. The lines are more densely packed in the foreground and become sparser towards the horizon. The overall color palette is dark blue and black, with vibrant highlights from the city lights and the digital overlay.

WHY IS MICRODATA IMPORTANT FOR STATISTICS?



Microdata is an **increasingly important** and valuable resource (1)

1. Demand for **granular and timely** information has increased, including
 - **Great Financial Crisis 2007-2009: G20 DGI** stressed the importance of complementing macro-level aggregates with data on financial linkages between sectors and countries and distributional information
 - **Pandemic:** microdata needed for monitoring granular effects:
 - Income group
 - Region
 - Gender
 - Firm size
2. **Declining efficiency of established statistical instruments** (surveys, censuses) require complementing with microdata



Microdata is an **increasingly important** and valuable resource (2)

3. Policy questions increasingly cross-cutting, for instance

- Environmental and social policies
- Digital transition, green transition, mobility transition
- Government finances and income and wealth distribution
- Enterprise demography, productivity and employment
-

➤ Microdata unlocks **new possibilities for cross-cutting evidence**

4. **Key statistical infrastructure** increasingly using microdata, including:

- Register-based **censuses**
- Economic census, statistical business registers, population registers, etc.
- Car registrations, platform data, supervisory data (banks, insurance corporations, pension funds, investment funds,...), credit registers and securities databases for **national accounts**



Microdata is an **increasingly important** and valuable resource (3)

5. **Reducing response burden** for businesses and individuals: *Once-only principle* is impossible without microdata use

6. Efficiency: administrative microdata can be collected for **multiple purposes** (monitoring financial stability, compiling macroeconomic statistics)



Discussions in international fora

Discussions in international fora (UN, UNECE, OECD, etc.) are predominately about:

- Right to access microdata from private companies (e.g., large digital players) for statistical purposes
- Safe microdata access for research purposes
- *Practicalities* of the use of administrative data for statistics rather than *principle* of access
- Ensuring inter-operability of administrative micro-data
- Meta data catalogues
- Ex-ante harmonisation of administrative data collections for multiple purpose use



Benefits and costs

- **Costs** to using microdata for statistical purposes
 - **Direct costs** of creating and handling microdata for statistics
 - Societal concern is often **privacy protection**
 - Assessment from perspective of overall societal good
- Note that the **increasing trend of administrative microdata sharing** in OECD member countries is not in general considered to conflict with privacy protection and safeguarding confidentiality
- **Legal, institutional and technical safeguards** are key

A globe of the Earth is centered in the image, overlaid with a complex network of white lines and nodes, representing a global network or digital infrastructure. The globe is set against a dark, starry space background with streaks of light. A large, solid blue diagonal shape covers the bottom-left portion of the image, containing the text.

LEGAL ENVIRONMENT



Different legal frameworks with different purposes

	Personal data protection	Statistical confidentiality
Laws	<ul style="list-style-type: none"> • Privacy Act • EU GDPR and equivalent 	Statistics Act, laws regulating central banks
Objectives of laws	<ul style="list-style-type: none"> • Data protection • Regulate rights and obligations as regards personal data protection • Consent based • Data security • Data traceability • Data access 	<ul style="list-style-type: none"> • Statistical confidentiality, protection of data collected for statistical purpose • Professional independence • Access to administrative sources • Legal authority to collect data for statistical purpose
Institutions	<ul style="list-style-type: none"> • National Data Protection Authority, Data Protection Officer 	National Statistics Office Official producers of official statistics
Purpose	<ul style="list-style-type: none"> • Data collected for different purposes 	Data collected for statistical purposes only
Unit	<ul style="list-style-type: none"> • Person 	Statistical unit
Examples of practices	<ul style="list-style-type: none"> • Privacy Impact Assessment • Data Sharing Agreement 	<ul style="list-style-type: none"> • Privacy Enhancing Techniques (PETs) • Guidelines • Penalties in case of breach



Weighing societal goals

- **Privacy protection** (right to privacy in an ever-growing data environment)
- **Production of official statistics** (right to live in an informed society governed by evidence-based policies), in a context of growing user need for granular data, need to be balanced
- **Could be seen as potential conflict**
- But much attenuated through:
 - clear **legal basis for data sharing** based on **accountability, transparency**
 - **alignment** of laws and their implementation with international standards
 - secured **technical data environment** and secured data sharing mechanisms
 - **formal and detailed institutional arrangements** for sharing and cooperation



Alignment with international standards – examples (1)

OECD Recommendation on Good Statistical Practice

Legal mandate to collect data for statistical purposes

Professional independence

Statistical confidentiality

Access to administrative sources for statistical purposes

Quality management

Coordination

List of good practices in implementation



Alignment with international standards – examples (2)

OECD
Recommendation
on Enhancing
Access to and
Sharing of Data
(beyond statistical
purposes)

Reinforcing **trust** across the data system

Adopt a strategic **whole-of-government approach** to data access and sharing

Support **long-term investments in data access and sharing arrangements** to ensure their sustainability

Foster awareness about the **benefits and risks** of data access, sharing, and use to encourage responsible data governance throughout the data value cycle.



Alignment with international standards – examples (3)

OECD

Recommendation
concerning
guidelines
governing the
protection of
privacy and trans-
border flows of
personal data

Collection limitation

Data quality

Purpose specification

Use limitation

Security safeguards

Openness

Accountability

Individual participation



**TECHNICAL AND
INSTITUTIONAL
ARRANGEMENTS FOR
DATA SHARING**



Important practical questions

- How to establish an **efficient** and **sustainable** data collection to avoid inconsistent data requirements, overlaps and duplication, and to ensure data quality?
- How to develop a **secured data environment**?
- What does **anonymisation** mean in practice?
- How to ensure that information is of sufficient **quality** for data-driven decision making?
- How are **safeguards** implemented?
- What are the most efficient **institutional arrangements**?



Various situations in OECD Members

- **Internal or external data sharing** between administrative authorities, central banks, national statistics offices, and international organisations
- For example, data sharing between **regulatory supervisor and central bank** can be internal or external, depending whether data collection is performed by different authorities or by different departments within the same authority
- Note: internal data sharing **not necessarily less stringent** than external
- **Data sharing arrangements** may take **different forms** from ad hoc access on a need-to-know basis to fully-fledged regular sharing of individual micro-data

Various approaches to data integration regarding Central Banks



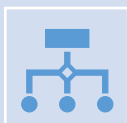
Highly integrated/cooperation approach: **common secured database** and a **single centralised data collection system** either managed by the Central bank (Italy) or based on a cooperation framework between authorities for technical and statistical requirements (Canada)



Partially integrated system (France, Finland) based either on a common database for multiple purposes with a clear **separation between statistical and supervisory reporting**, each institutions having access only to the information relevant for its work (France), or a joint data collection process



Decentralised approach (ex. Germany) based on separate legislations and different collection systems and sharing agreements



Coordination may cover one or several steps of the data collection process



Good practices: examples of institutional arrangements

- Data lifecycle perspective for **integrity** and **confidentiality**
- **Harmonised domestic standards** for sharing microdata
- **Joint data collection processes**, based on agreed purpose specification and efficient coordination mechanisms
- Committees/programmes to foster data sharing
- **Dynamic design of** institutional and technical **arrangements** with regular review and amendments of arrangements given the speed of IT and mathematical developments (encryption techniques), or emerging new personal and confidential data



Good practices **IT and anonymisation** - examples (2)

- Regular review of adequate **level of aggregation** (threshold) and relating anonymisation techniques
- **Develop common secured IT platforms and databases**, where it is possible to access all available data, even if those data are stored in different datasets
- Use of emerging **privacy-by-design principles** (e.g., Privacy Enhancing Technologies) and integration into business processes and **data governance frameworks**
- **Limit duration of data storage** (data kept in a form which permits identification of data subjects for no longer than necessary)
- The **collateral management of a unique identification** (database specific) of entities and instruments is key to avoid duplication of work.



Good practices: examples from **communication and engagement**

- Implementation of a public **confidentiality policy**
- **Privacy Impact Assessments**
- Commitment to **quality** and **quality assurance framework**.
- Communication on mechanisms in place to safeguard confidentiality and on **official statistics as a public good**
- Promotion of **culture** of privacy protection, openness to innovation and information sharing at the operational level
- **Staff secondments** to improve knowledge and share experiences in various entities
- **Institutionalized consultation** and communication



CONCLUSIONS



Conclusions

- Administrative microdata are of **increasing importance** for a functioning statistical system and for evidence-based policy making
- Societal goals of **privacy protection** are **also increasingly important** and there is a structural tension
- Tension can be eased or eliminated with the right **legal, institutional and technical arrangements** (no single format for this arrangements)
- **Clear legal basis is recommended** to support the institutional arrangements and implementation of secured technical data environment and data sharing mechanisms



THANK YOU

 @OECD_Stat  OECD Statistics  www.oecd.org/sdd  www.stats.oecd.org