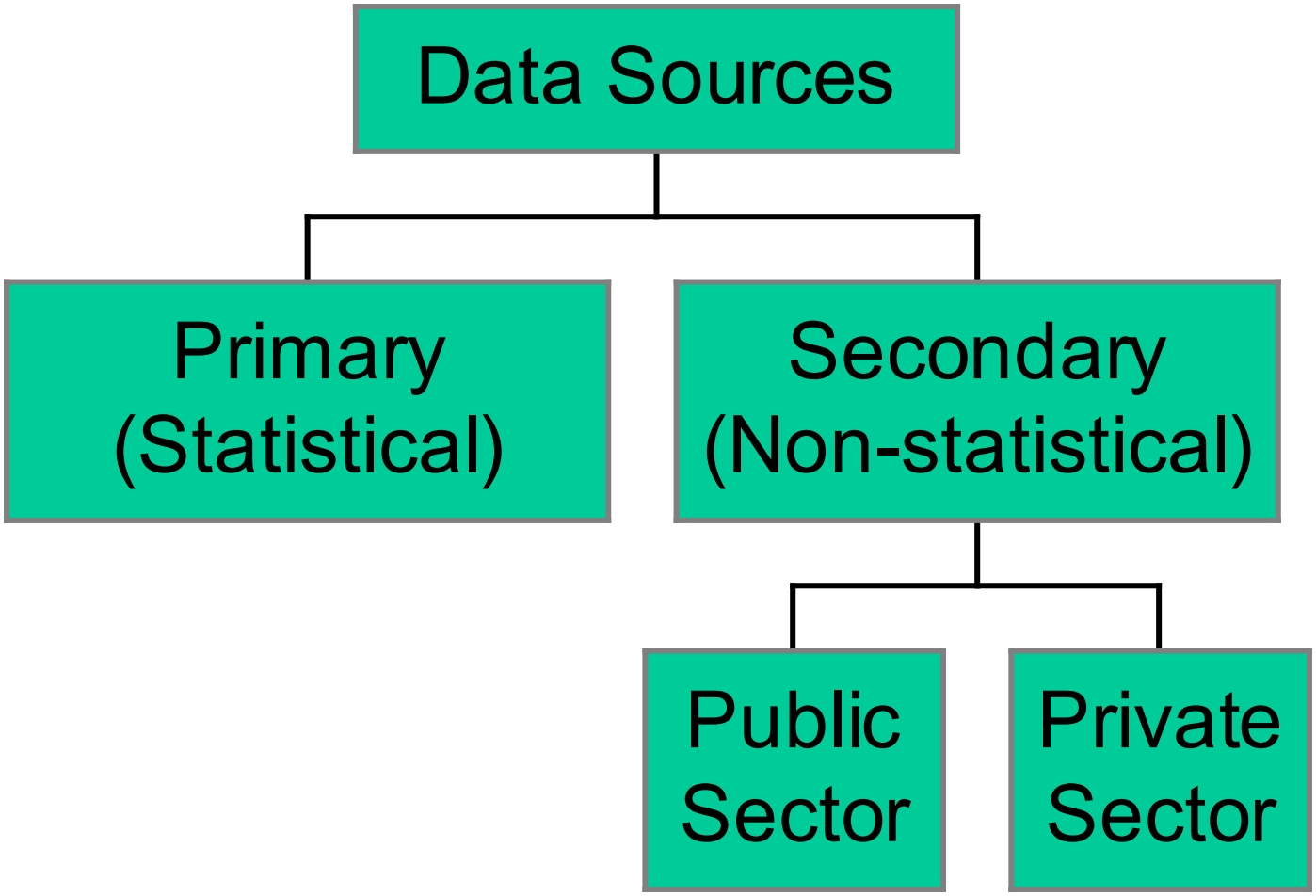
United Nations Economic Commission for Europe
Statistical Division

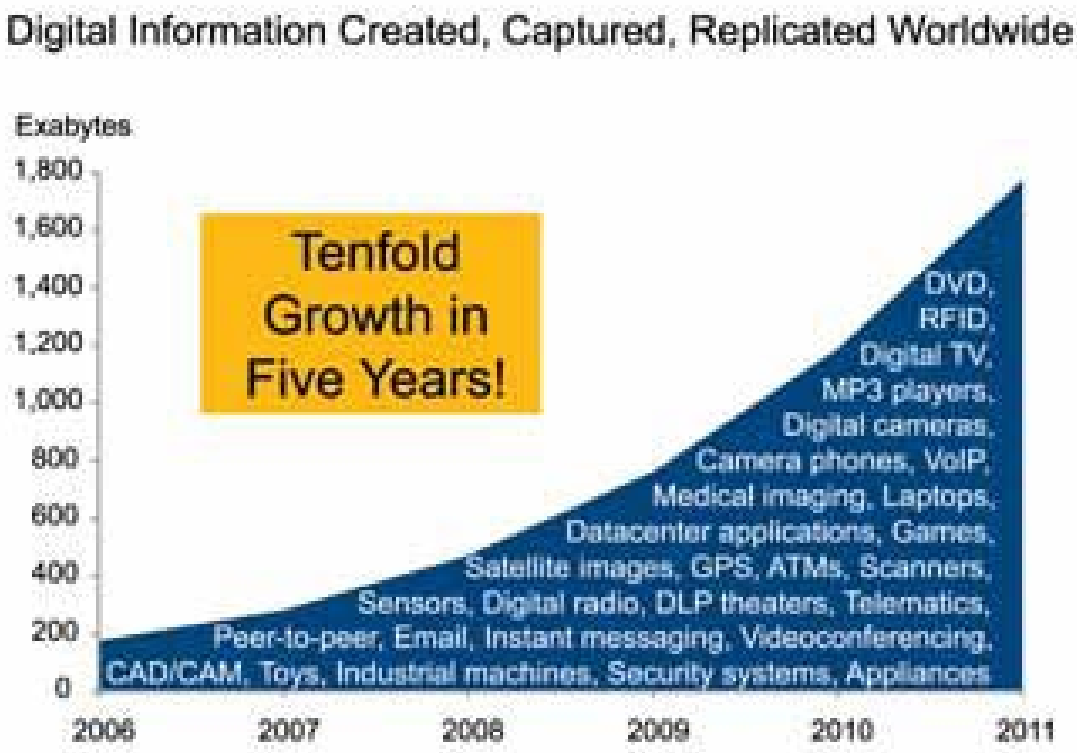
Use of Administrative Data

Steven Vale
Statistical Division, UNECE

Types of Data Source



Data availability



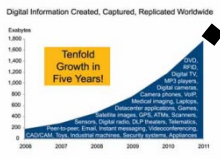
The internet had 1800 exabytes of data in 2011

exa = 10^18

50,000 Exabytes in 2020

UNECE

27x growth in 9 years



We live in exponential times!

Are these data interesting?

- Probably 99.9% are videos of cats, photos, audio files and other nonsense
- But that still leaves

50,000,000,000,000,000,000

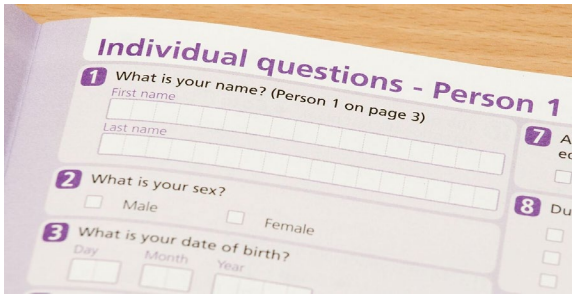
bytes of potentially relevant data

Why use Administrative Data

Population census costs

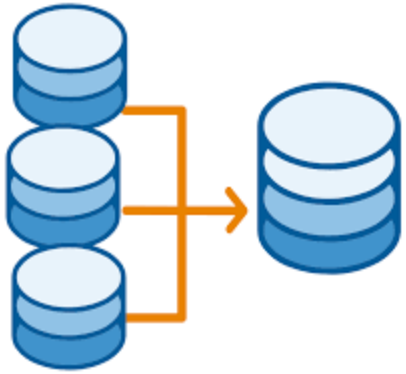
- UK, €367m

€6.20 per person



- Finland, €0.8m

€0.20 per person



Source: Eurostat – Documentation of the 2000 round of population and Housing censuses in the EU, EFTA and Candidate Countries; Table 22

Access to Administrative Data

4 Frameworks

- Legal
- Policy
- Organisational
- Technical

Legal Framework

Law on Official Statistics no. 93 of 26 May 2017

Article 16. Access to administrative data

(1) …. **the central statistical authority shall have the right to access and use, free of charge, administrative data sources, including sources containing individual data**.

(2) The owners of administrative data sources, and other legal entities managing databases on natural and legal persons, shall present, free of charge, to the central statistical authority aggregated and individual data, including metadata related to them. The **special confidentiality provisions from other laws cannot be invoked** in this case, except the cases when that legislation excludes explicitly the use of data for statistical purposes.

(3) The owners of administrative data shall coordinate with the central statistical authority the content and volume of data collected by them, including possible changes to the data, and, at the request of the central statistical authority, shall include additional indicators.

(4) The central statistical authority shall use the data obtained … for statistical purposes only.

EU Law: Regulation 322/97

Article 17a

1. … NSIs, other national authorities … shall have the right to access and use, promptly and free of charge, all administrative records … for the development, production and dissemination of European statistics

2. The NSIs … shall be consulted on, and involved in, the initial design, subsequent development and discontinuation of administrative records …

5. The NSIs and owners of administrative records shall establish the necessary cooperation mechanisms

Policy Framework

- National government data / digitalization policies
- European Union policies
  - General Data Protection Regulation (GDPR)
- Fundamental Principles of Official Statistics
- European Statistics Code of Practice

ESS Code of Practice

Principle 2: Statistical authorities have a clear legal mandate to collect and access information from multiple data sources for European statistical purposes. Administrations, enterprises and households, and the public at large may be compelled by law to allow access to or deliver data for European statistical purposes at the request of statistical authorities

Organisational Frameworks

The use of administrative data is enabled by a legal framework, in the context of a policy framework

But - these are not detailed enough to cover all the administrative arrangements for access and use

A Formal Agreement

- Sets out terms and responsibilities for the provision and use of data
- Signed by all parties
- Can cover one or more data flows

Agreement Contents (1)

- Legal basis
- Names of persons transferring and receiving data
- Detailed description of data covered
- Frequency of data supply
- Quality standards
- Confidentiality rules

Agreement Contents (2)

- Technical standards
- Provision of metadata
- Provisions for payment for supply data
- Period of agreement
- Contingencies for changes in circumstances
- Procedure for resolving disputes

Technical Frameworks

- Mechanisms for data transfer
  - Paper
  - Computer files (off-line)
  - Secure on-line connection
- File formats and structures
- Data / metadata standards

Quality of Administrative Data

- Are data from administrative sources as good as data from surveys?
- Who should judge this?
- How can we measure quality?
- How should we report and communicate quality?

Definition of Quality

ISO 9000/2015 defines quality as;

'The degree to which a set of inherent characteristics of an object fulfils requirements.'

What does this mean?

- Whose requirements?
  - The user of the goods or services
- A set of inherent characteristics?
  - Users judge quality against a set of criteria concerning different characteristics of the goods or services
- Therefore, quality is all about providing goods and services that meet the needs of users (customers) – **"Fit for purpose"**

Three Aspects of Quality

- To understand the quality of administrative sources we need to consider:
  - Quality of incoming data
  - Quality of processing
  - Quality of outputs

Incoming Data

- Timeliness
- Completeness – are there any missing units or variables?
- Comparability with other sources
- Quality check survey?

Tip: Knowledge of the source is vital!

Processing

- Quality of matching / linking
- Outlier detection and treatment
- Quality of data editing
- Quality of imputation

Tip: Keep raw data / metadata to refer back to if necessary

Outputs

- Are the users satisfied?
- Are the outputs comparable with data from other sources?
- What is the impact on time series?
- Are the outputs cost-effective?
- Quality reports to measure and communicate differences?

Tip: Be ready to explain discontinuities!

Metadata / Quality

- Knowledge and documentation of the data source is vital if you want to understand data quality:
  - How the data are collected
  - Why they are collected: priorities of collector
  - How they are processed
  - Concepts and definitions used
  - etc…

Quality Frameworks - Global

National Quality Assurance Framework

- Chapter 7 “Quality assurance for statistics compiled from different data sources”
  - Statistical
  - Administrative
  - Other
- Some indicators in Table 7.1

https://unstats.un.org/unsd/methodology/dataquality/un-nqaf-manual/

GSBPM Quality Indicators

- Developed by task team
  - Canada, Hungary, Italy, Turkey, Eurostat, UNECE
- Indicators for each GSBPM sub-process
  - Processes based on surveys: 2016
  - Expanded to include administrative data: 2017
- Consistent with existing frameworks:
  - UN NQAF, ESS Code of Practice / QAF

https://statswiki.unece.org/display/GSBPM/Quality+Indicators

Quality Frameworks - Regional

Quality Assurance Framework of the ESS

- Based on the ESS Code of Practice
- Provisions for administrative data throughout the framework
  - Indicators 2.1, 2.2, 8.1, 8.6, 8.7, 9.4, 10.3, 14.4, 15.6

https://ec.europa.eu/eurostat/web/quality/european-quality-standards/quality-assurance-framework

Quality Frameworks - Other

- Guidelines for assessing the quality of administrative sources for use in censuses
  https://unece.org/statistics/publications/CensusAdminQuality
- Quality assurance of administrative data (UK)
  https://gss.civilservice.gov.uk/policy-store/quality-assurance-of-administrative-data/
- Checklist for the quality evaluation of administrative data sources (Netherlands)
  https://ec.europa.eu/eurostat/documents/64157/4374310/45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf
- Guide to reporting on administrative data quality (New Zealand)
  https://www.stats.govt.nz/methods/guide-to-reporting-on-administrative-data-quality

Role of NSOs in Data Ecosystem

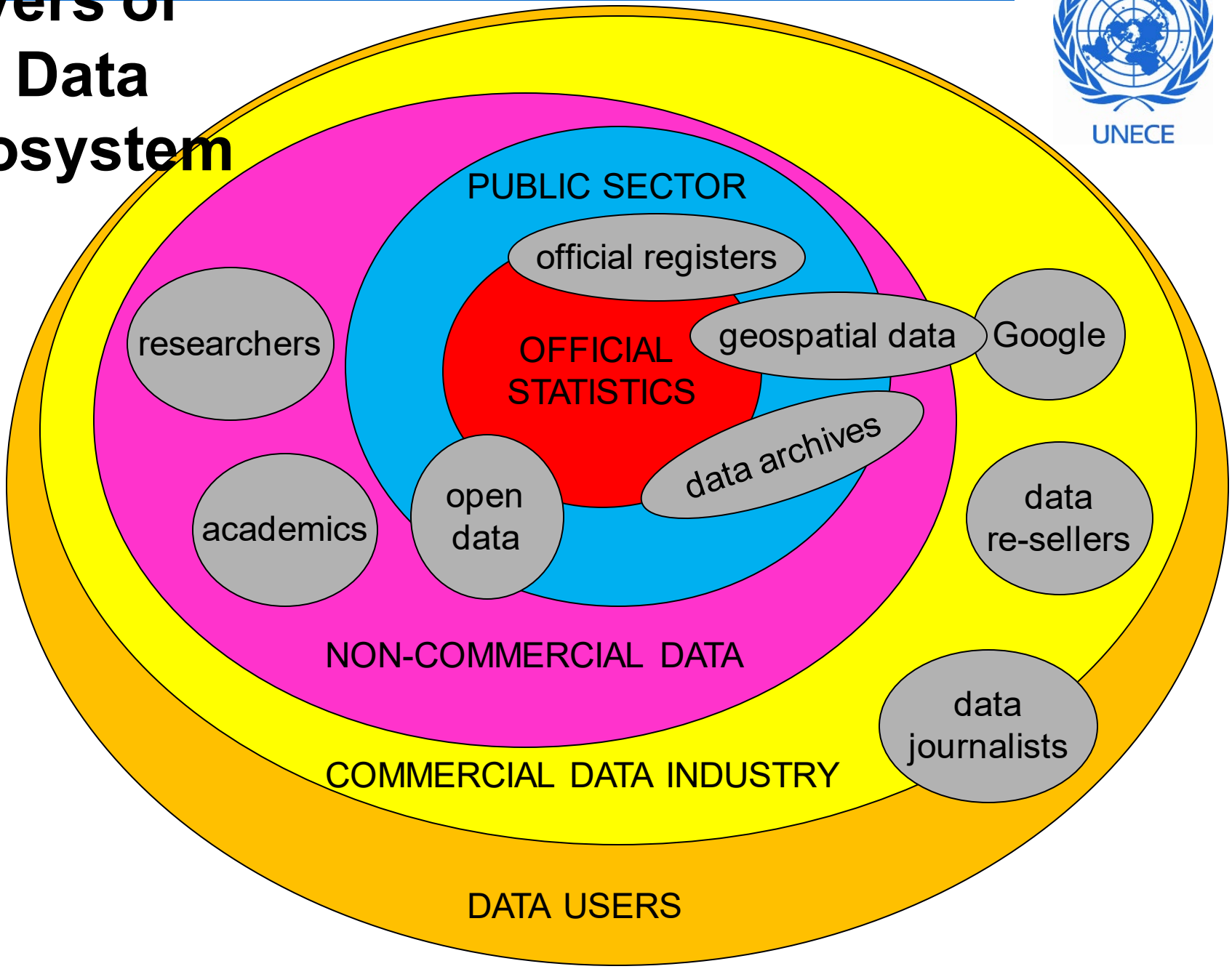
- Paper by: Estonia, Albania, Canada, Ireland, Italy, Netherlands, New Zealand, Poland and UNECE for the Conference of European Statisticians, June 2020
- Objective: To identify opportunities and address challenges for national statistical systems regarding the development of national data ecosystems

Some Key Points

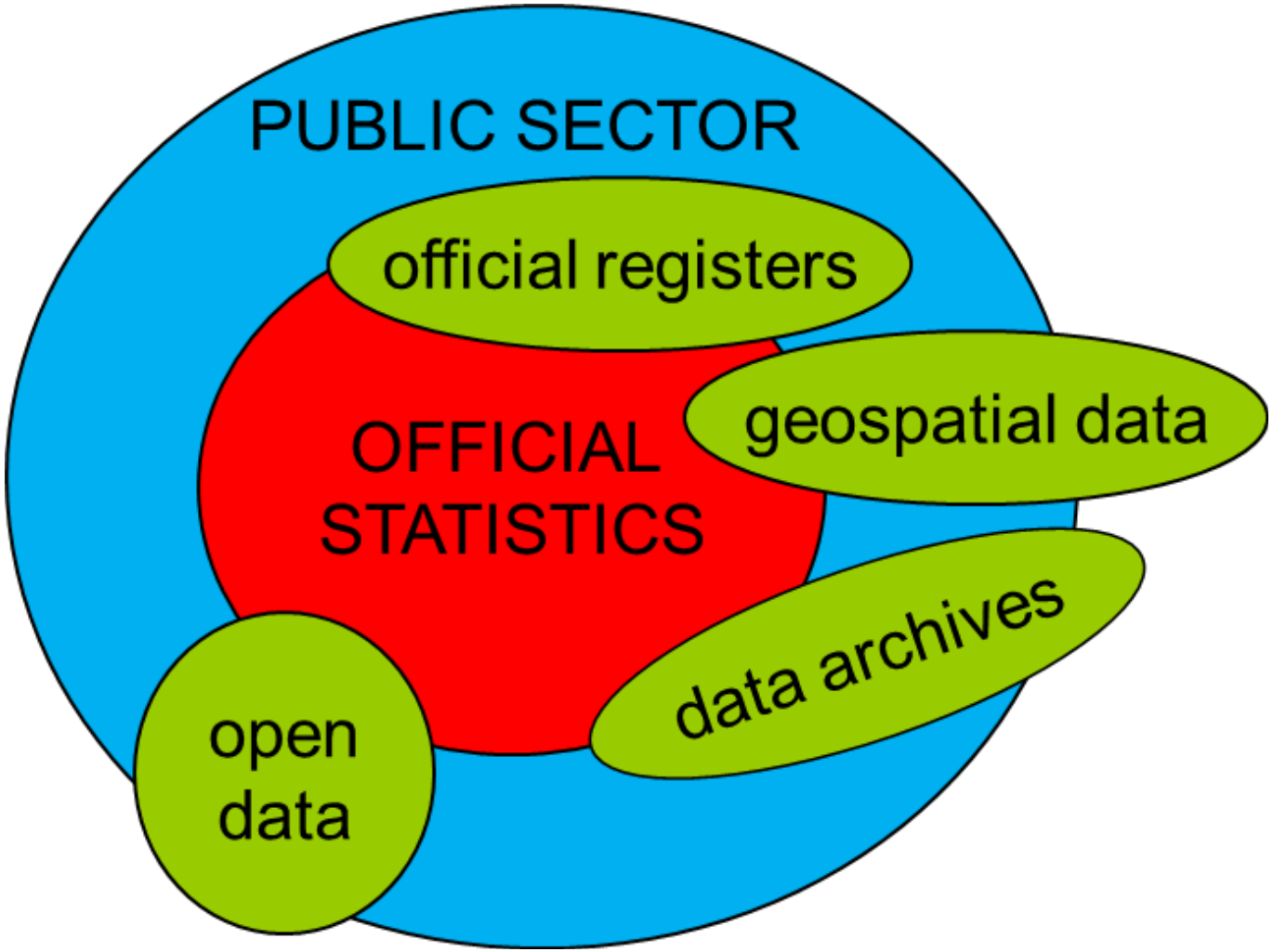
- Data ecosystems comprise many types of data: statistical, administrative, geospatial and other new sources including big data
- Data ecosystems provide new possibilities for linking data sources to create valuable information for policy makers and the public
- NSOs have data management skills
- What role should they play?

http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2019/ECE_CES_2019_16-1906490E.pdf

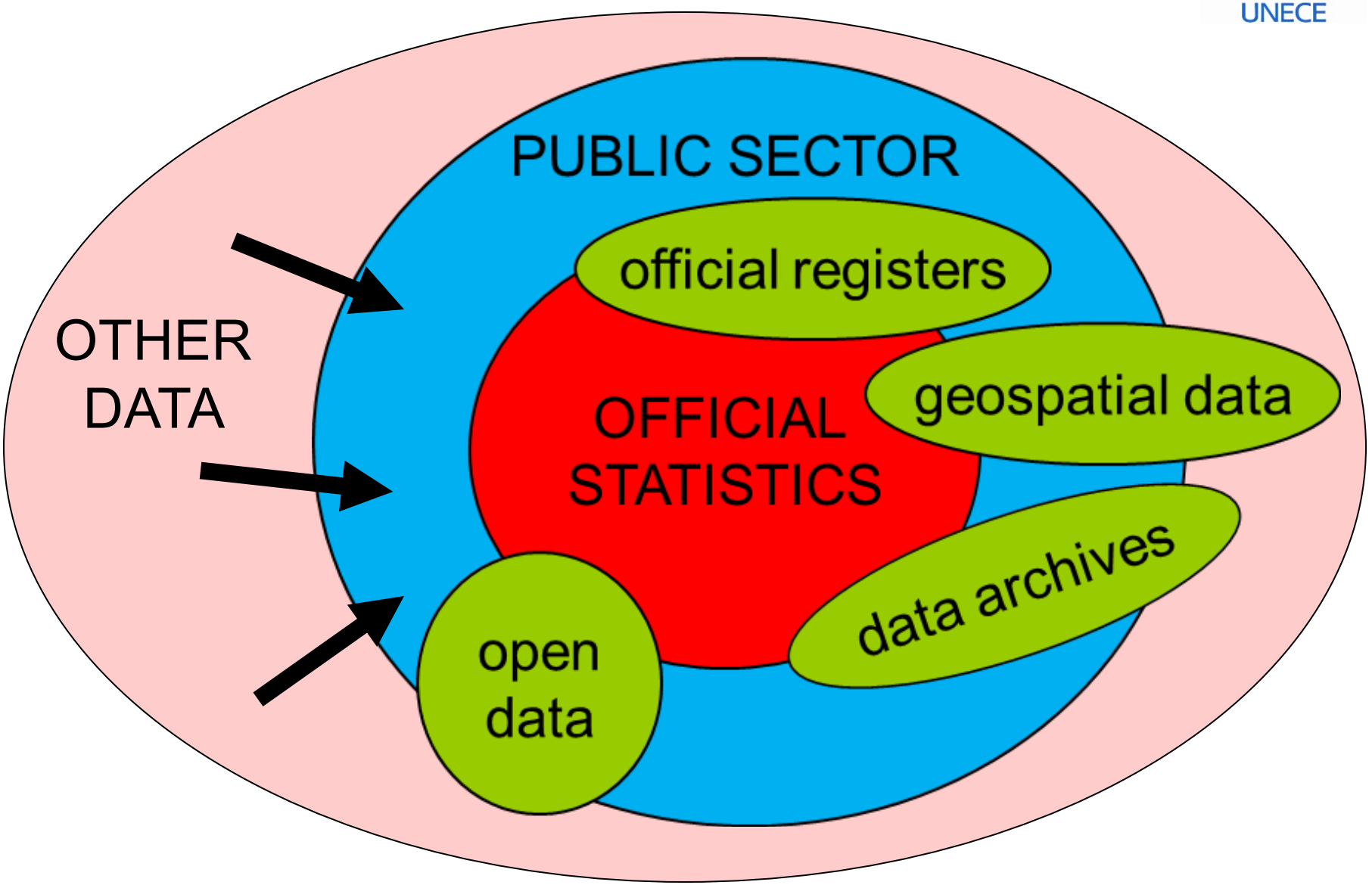
Layers of the Data Ecosystem



National Data Infrastructure



National Data Infrastructure



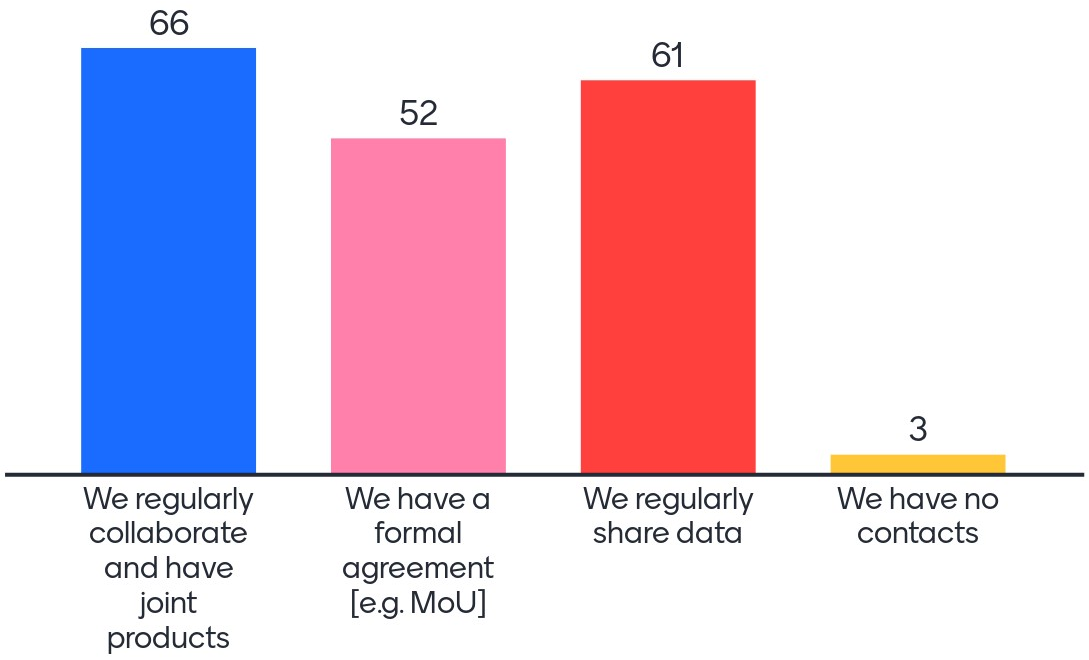
Next Steps

- Task Force on Data Stewardship
  - Estonia (Chair), Canada, Finland, Hungary, Ireland, Mexico, Netherlands, New Zealand, Poland, UK, Eurostat and OECD.
  - Draft report
    - Defining data stewardship
    - Roles for national statistical offices
    - Case studies
  - Maturity model
  - Guidance

Geospatial Data

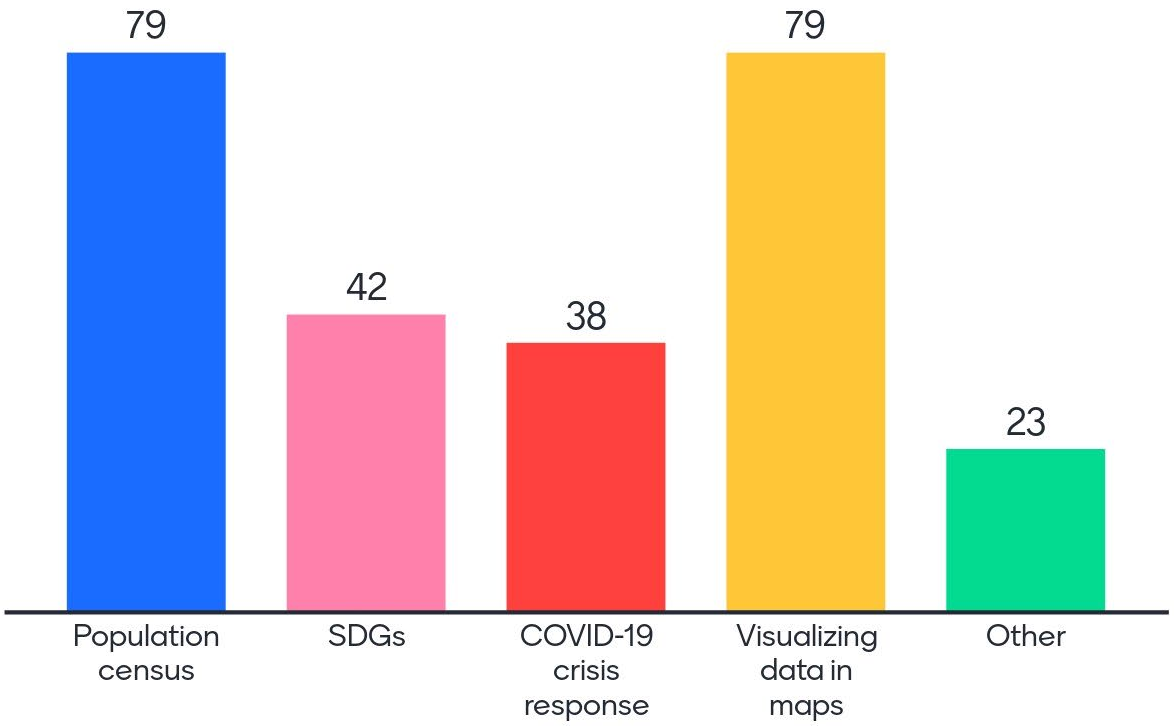
- Conference of European Statisticians / UN-GGIM: Europe Joint Plenary 2020:
  - Geospatial and statistical data are cornerstones of national data ecosystems. Statistical and geospatial agencies can support each other to enhance their roles
  - Create a joint task team to determine where geospatial and statistical standards need to be better aligned to ensure greater interoperability

How is the relationship between the statistical and geospatial communities in your country?



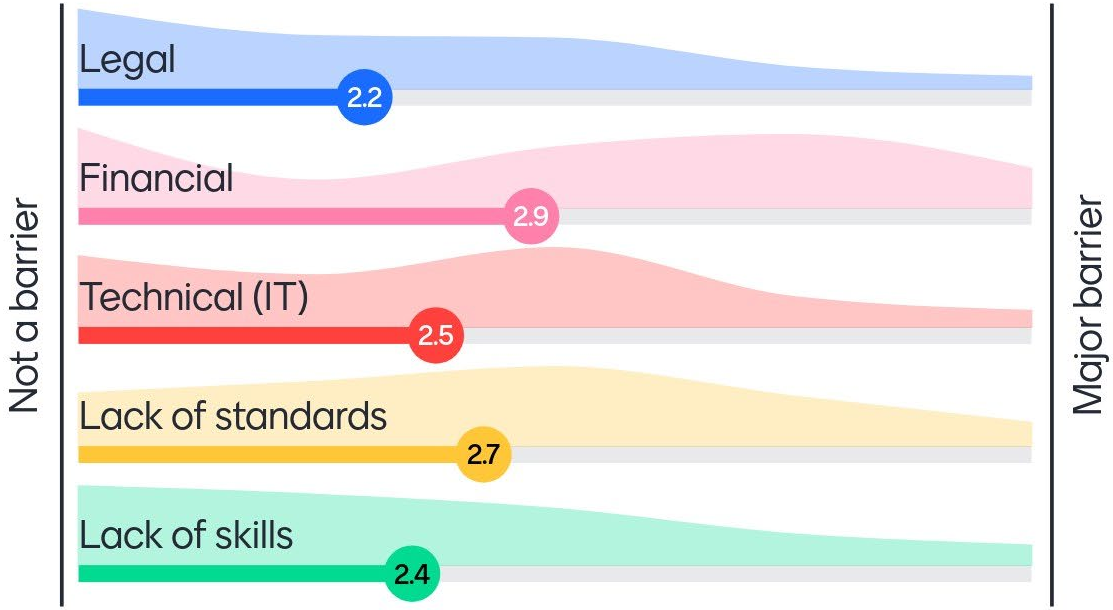
What are the main areas of collaboration?

Mentimeter



103

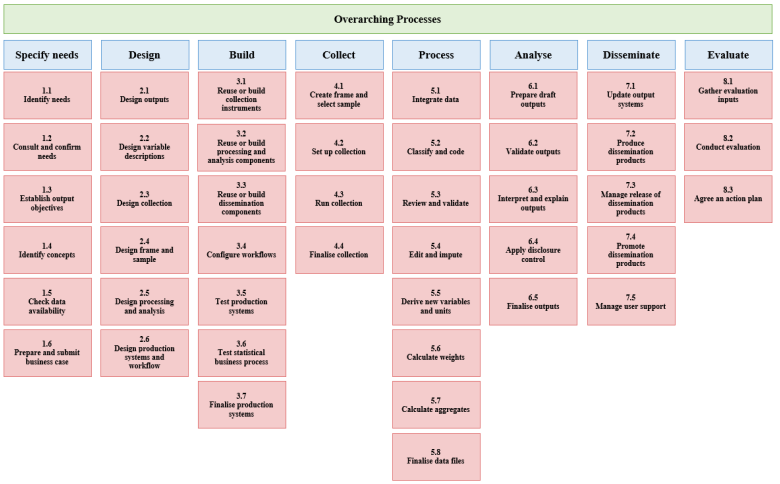
What are the main barriers to greater collaboration?



96

Geospatial view of GSBPM

- Description of how geospatial information can be used across all stages of the statistical production process



New Project

- Geospatial information
  - Project supported by a Eurostat grant
  - Developing standards for data integration
  - Support for NSOs to join international meetings on integrating statistical and geospatial information
  - 2022-2024

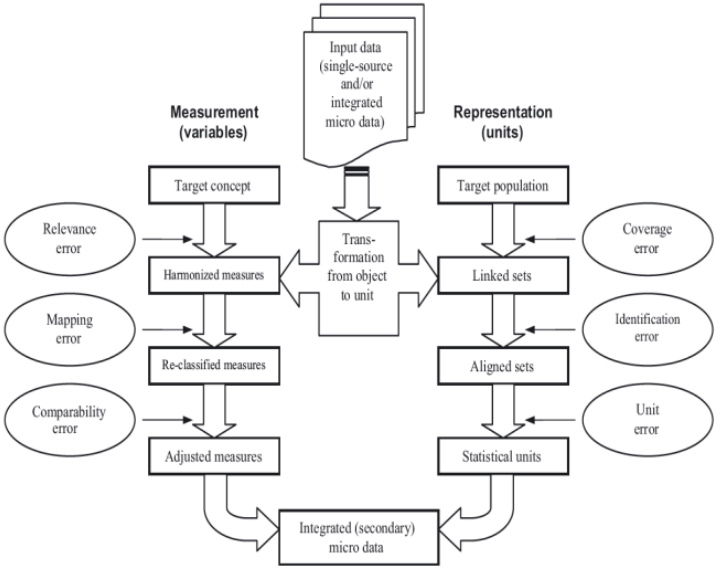
Quality of Big Data

- "A Suggested Framework for the Quality of Big Data" – December 2014
  - Australia, Canada, France, Italy, Mexico, Netherlands, Poland, Slovenia, UNSD, UNECE
- "extensions to existing statistical data quality frameworks were needed in order to encompass the quality of Big Data"

https://statswiki.unece.org/display/bigdata

Quality in data integration

- Part of UNECE guidelines prepared by international project
- Uses Zhang's two-phase life-cycle method model for integrated statistical microdata



https://statswiki.unece.org/display/DI/Quality

Any questions?