

Key Factors for Obtaining Access to Big Data

DGINS Conference, Bucharest (RO), 10-11 October 2018

Nico Waeyaert
Director general Statistics Belgium

Overview

- Official statistics and big data: the problem
- What is blocking us?
- CPI and scanner data: a model
- Approach to accessing private-sector big data
- Conclusions

Official statistics and big data: the problem

Five years of trying within ESS ...

... and still no regular official statistics using
privately owned big data?

What is blocking us?

NOT (although often far from easy)

- Data transfer, storage and processing
- IT infrastructure
- Software
- Methodology
- Data science capabilities

BUT

- Data access

Some factors blocking access

- Cost to create data
- Differences in perspective within private company
- Confidentiality concerns
- Privacy and reputation risk concerns
- Transfer models

CPI and scanner data: a model

Scanner data quite similar to big data:

- Detailed, almost-real-time
- Better statistics for lower cost and effort
- IT and methodological challenges
- Owned by profit-oriented private companies

Successful implementation in Belgium's CPI

Unblocking scanner data

Four key success factors

- Clear and detailed use case
- High-level engagement and active support
- Building trust
- Legislation

Unblocking private-sector big data: clear and detailed use case

Stating in detail

- which data needed
- which frequency
- which statistical output
- for what users and uses

Unblocking private-sector big data: high-level engagement and active support

- Assess feasibility at operational level
- Upper management decides
- Request from high up needed

Unblocking private-sector big data: building trust

- Explain statistical and privacy legislation
- Inform about all security aspects
- Contract to specify terms

Unblocking private-sector big data: legislation

- Supportive
- At EU and/or national level
- Generic and/or domain-specific
- Privacy, data protection or sector-specific barriers

Conclusions

- ❑ Solve storage and handling, IT and software, methodology
- ❑ Critical key problem is data access
- ❑ Systematic approach
 - Focus: identify a clear and detailed use case
 - Negotiate: ensure high-level support and build trust
 - Regulate: support by clear and specific legislation