



# Trusted Smart Statistics: What it is Why it comes Where it brings us

**Fabio Ricciato** [fabio.ricciato@ec.europa.eu](mailto:fabio.ricciato@ec.europa.eu)

**EUROSTAT - Big Data Task Force**

**Smart Statistics 4 Smart Cities  
Kalamata, Greece, 6.10.2018**



# The new datafied world

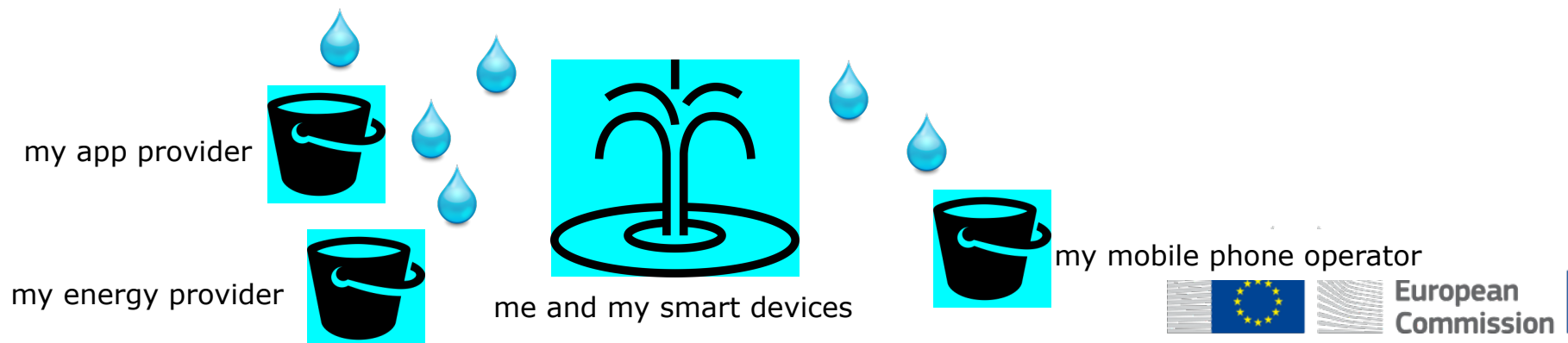
- *The cyber world is natively digital. And the physical world is being increasingly digitized (IoT, Smart Devices...)*
- **"Anything that goes digital, gets logged"**  
*(somewhere, by somebody) 1° fundamental law of datafication*  
**digitalization → datafication**

my mobile phone operator



# The new datafied world

- *The cyber world is natively digital. And the physical world is being increasingly digitized (IoT, Smart Devices...)*
- **"Anything that goes digital, gets logged"**  
*(somewhere, by somebody) 1° fundamental law of datafication*  
**digitalization → datafication**
- *Individuals, organizations, places ... become "data **fountains**"*
- *More and more business companies become "data **buckets**"*



# data and new data

- Features about the **individual**
- changing slowly or rarely
- recorded at coarse temporal aggregation (months, years).

**"micro-data"**

Name. Gender. Birth date.  
Marital Status. Residence address.  
Occupation. Household composition...  
Monthly income.  
Monthly expenditures per good category.  
Number of touristic trips in a year.

...

# data and new data

- Features about the **individual**
- changing slowly or rarely
- recorded at coarse temporal aggregation (months, years).

Name. Gender. Birth date.  
Marital Status. Residence address.  
Occupation. Household composition...

Monthly income.  
Monthly expenditures per good category.  
Number of touristic trips in a year.

**"micro-data"**

- Features about single **events, transactions**  
→ highly pervasive, *sub-individual* level
- changing continuously
- recorded at fine temporal aggregation (minutes, seconds)

...

Your exact location, every second.  
Every single heart-beat, blood pressure...  
Every single transaction, purchases, encounter, event involving you...  
Your current opinion on any single fact...

**"nano-data"**

# data and new data



**"micro-data"**

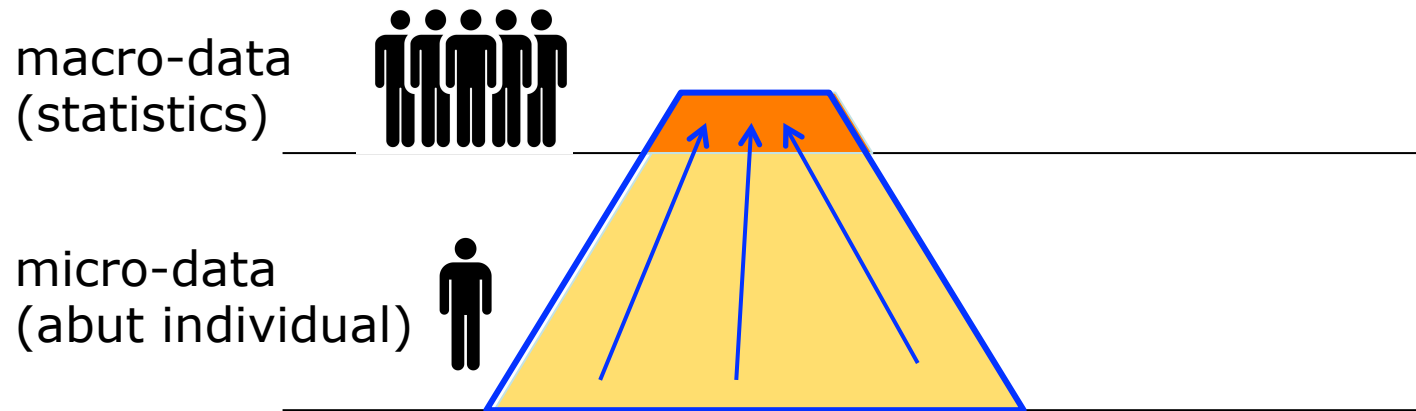
Name. Gender. Birth date.  
Marital Status. Residence address.  
Occupation. Household composition...  
...  
Monthly income.  
Monthly expenditures per good category.  
Number of touristic trips in a year

**"nano-data"**

...  
Your exact location, every second.  
Every single heart-beat, blood pressure...  
Every single transaction, purchases,  
encounter, event involving you...  
Your current opinion on any single fact...

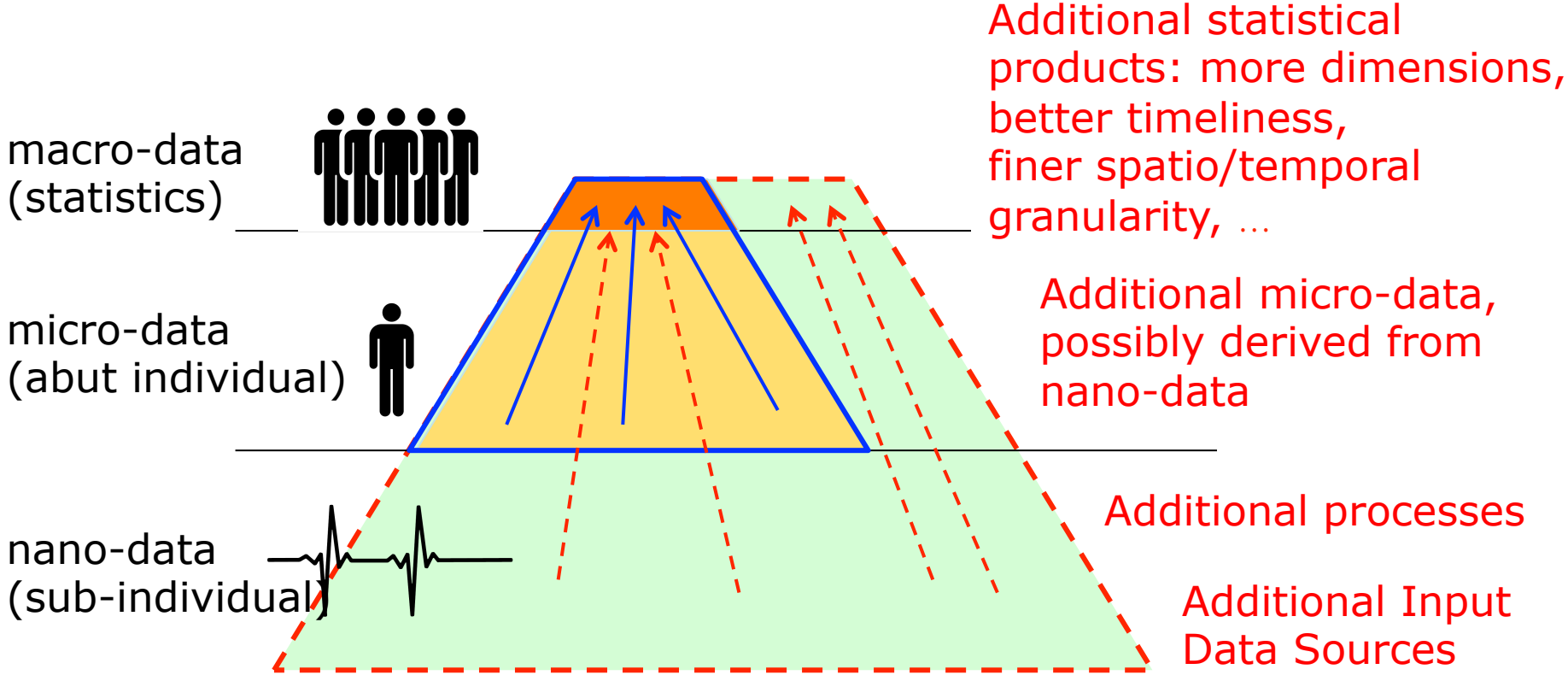
# Official Statistics.

- *The ultimate goal of Official Statistics is to produce **macro-data** (statistics) from input **micro-data***
  - Collection of micro-data as ancillary task



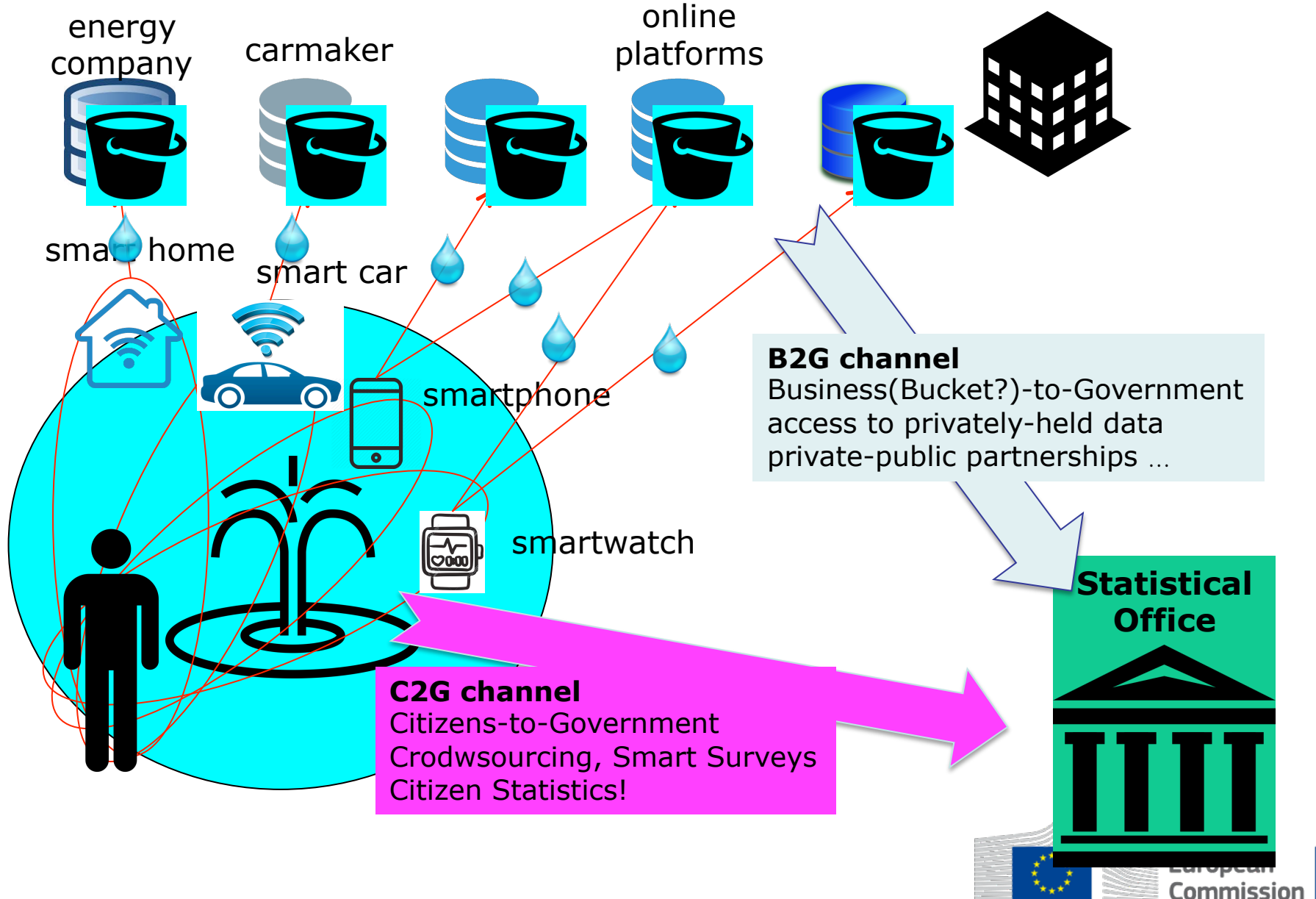
# Official Statistics. Augmented

- *Availability of new (deep, nano) data sources as opportunity to extend & empower Official Statistics*

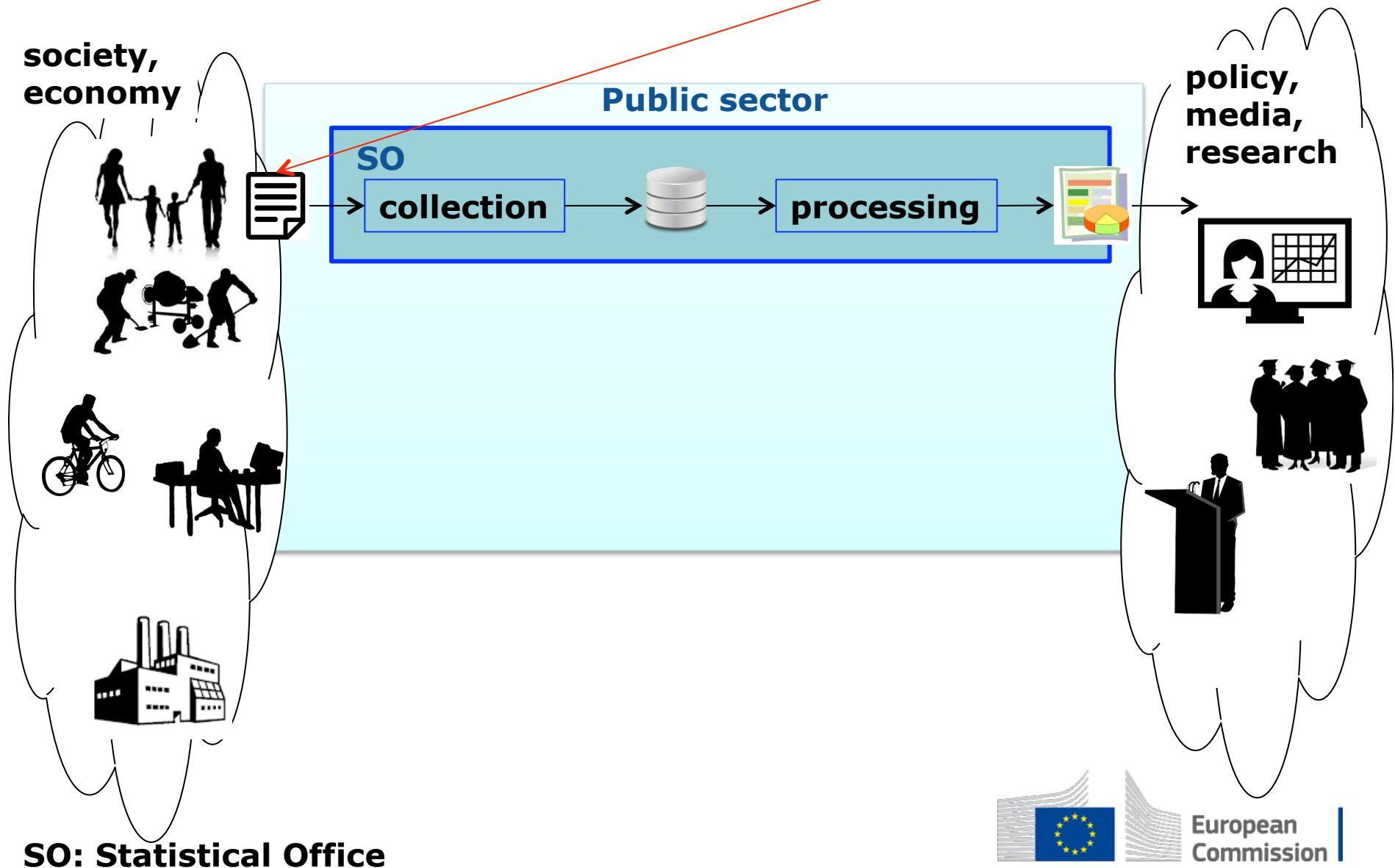




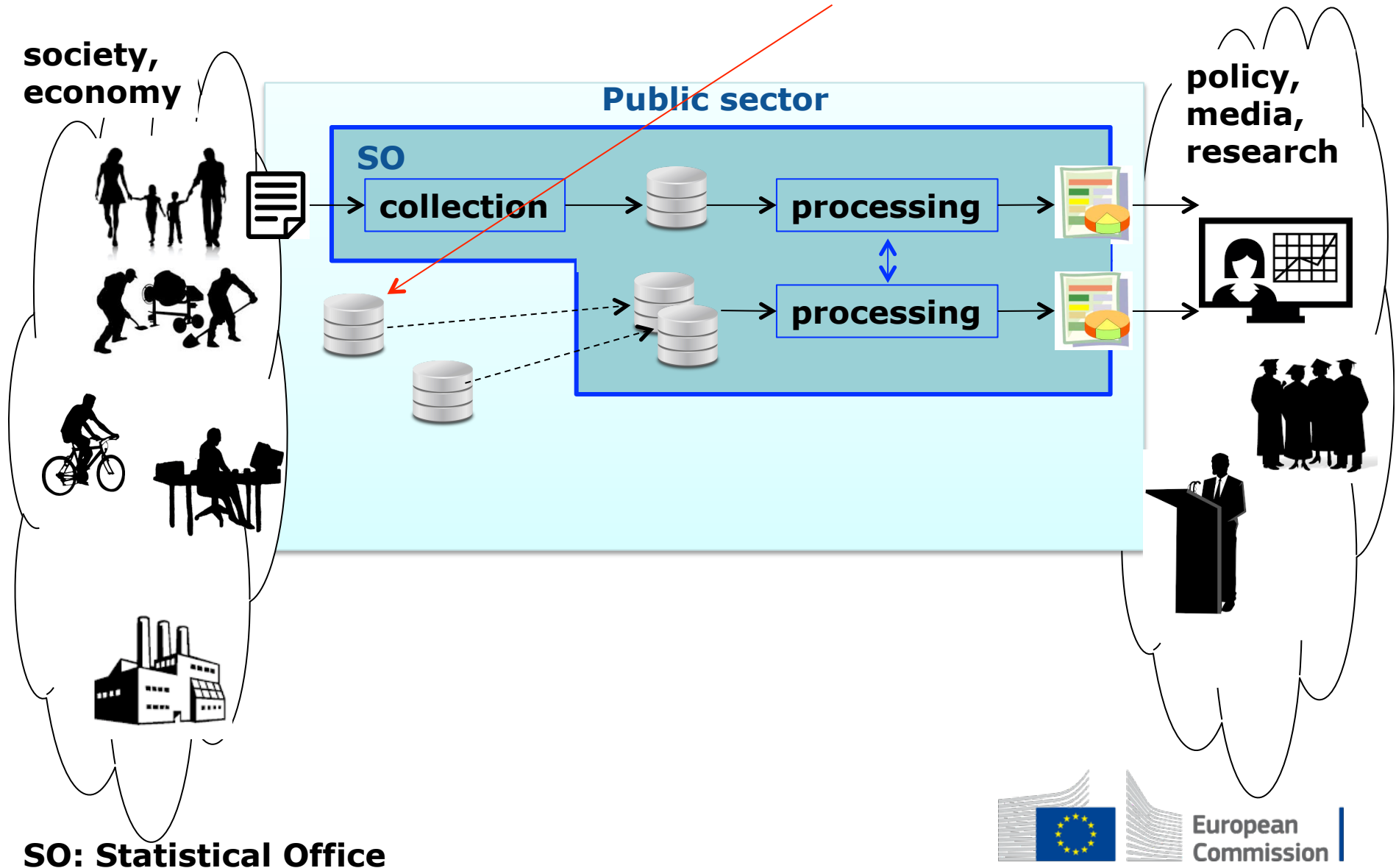
# Where the data can be accessed?



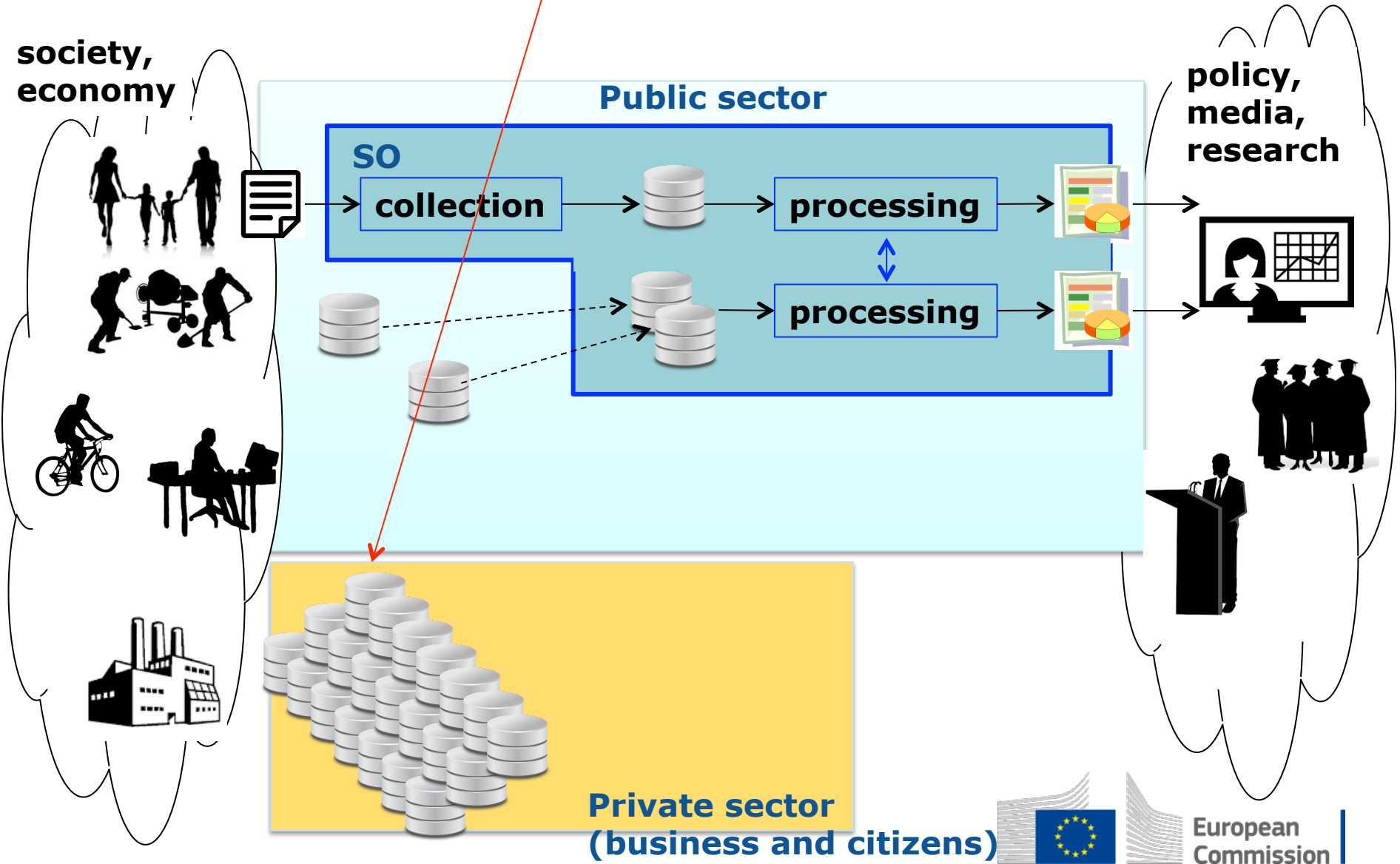
# Official Statistics based on **survey data**



# Official Statistics based on **survey data** and **administrative data**

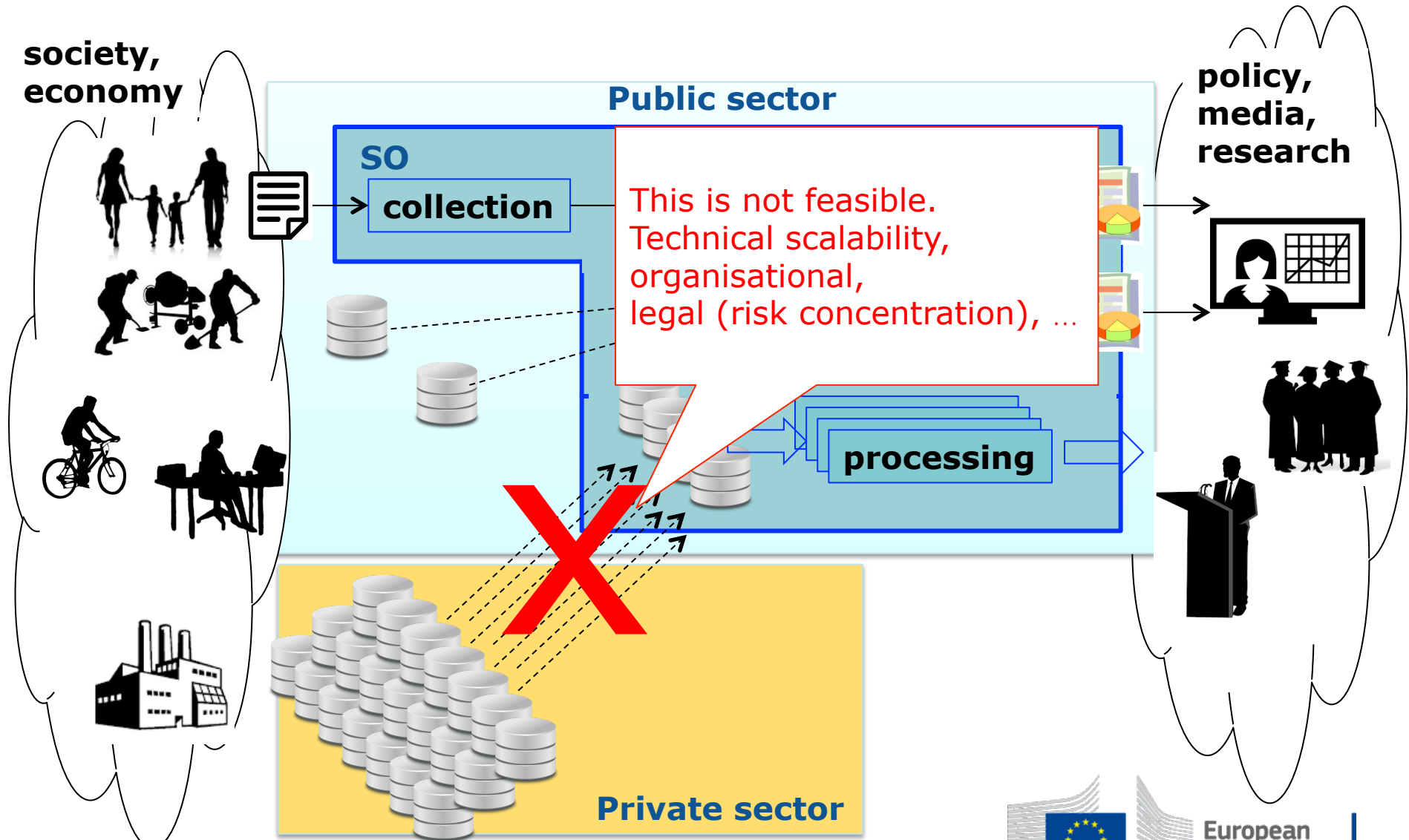


# and now **Big Data** come into play



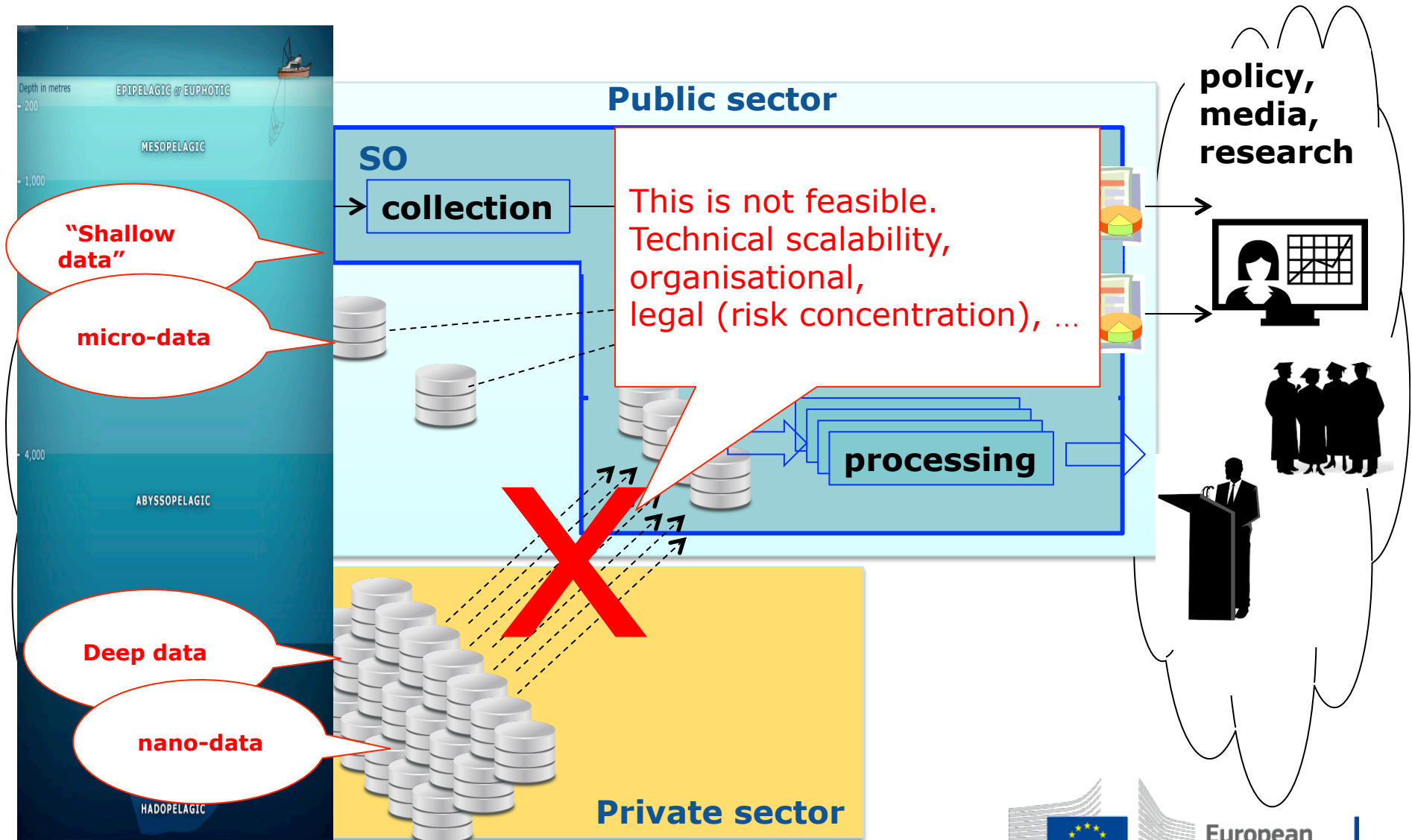
# Handling the new in old ways

*Pull data in*



# Handling the new in old ways

*Pull data in*



"Shallow data"

micro-data

Deep data

nano-data

**Public sector**

**SO**

**collection**

This is not feasible. Technical scalability, organisational, legal (risk concentration), ...

**processing**

**policy, media, research**

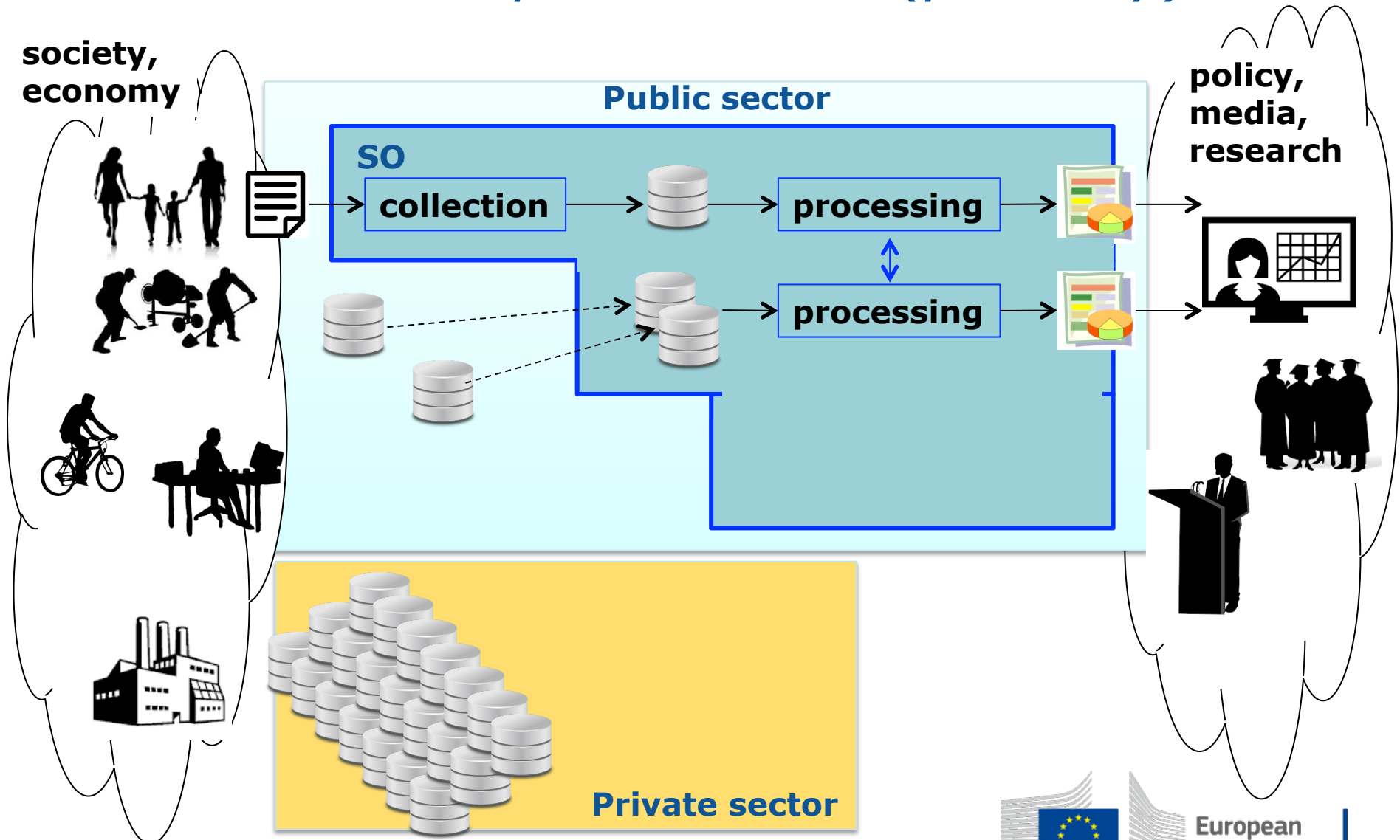
**Private sector**



European Commission

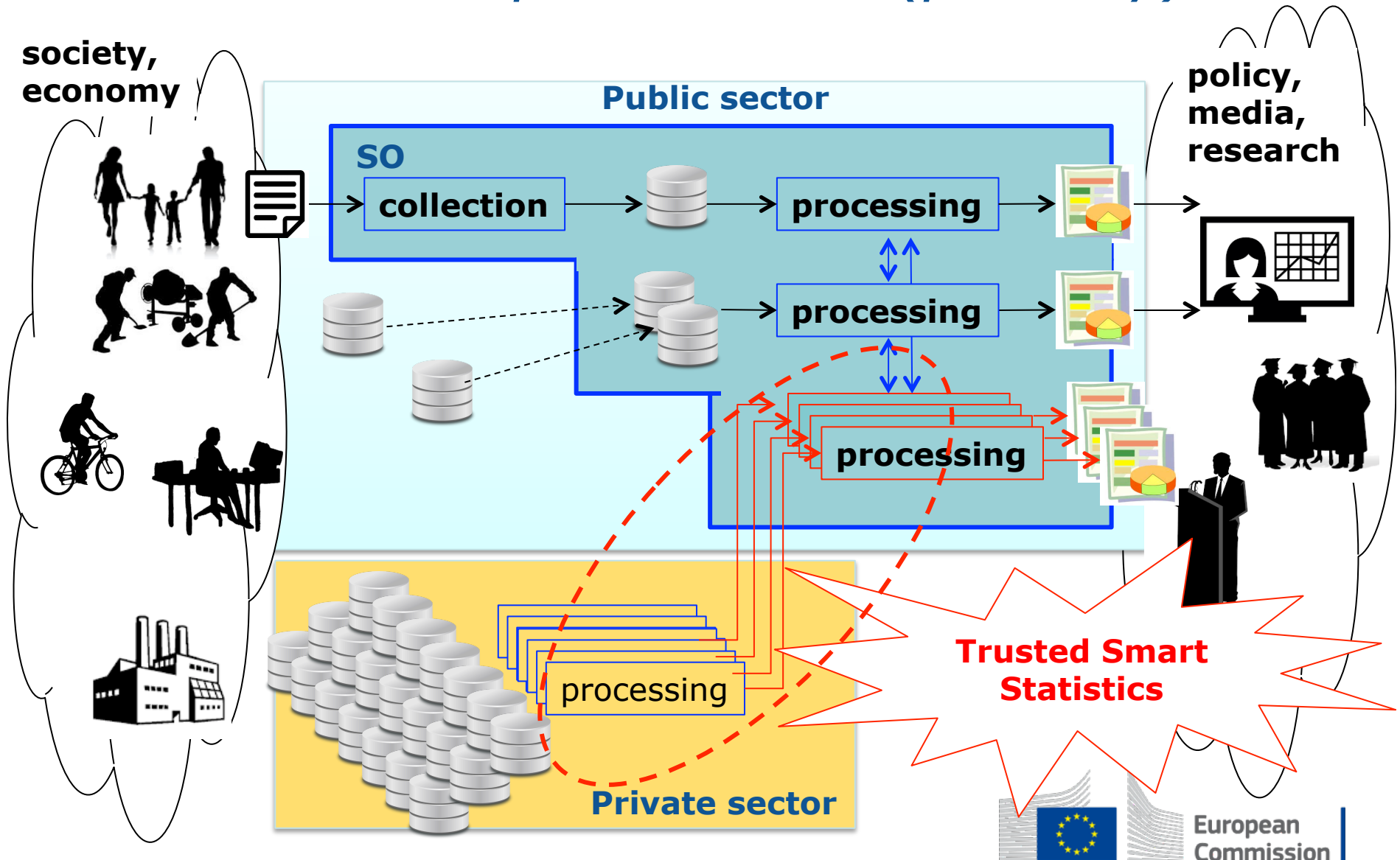
# Handle the new in new ways

*Push computation out (partially)*



# Handle the new in new ways

*Push computation out (partially)*





# Trusted Smart Statistics

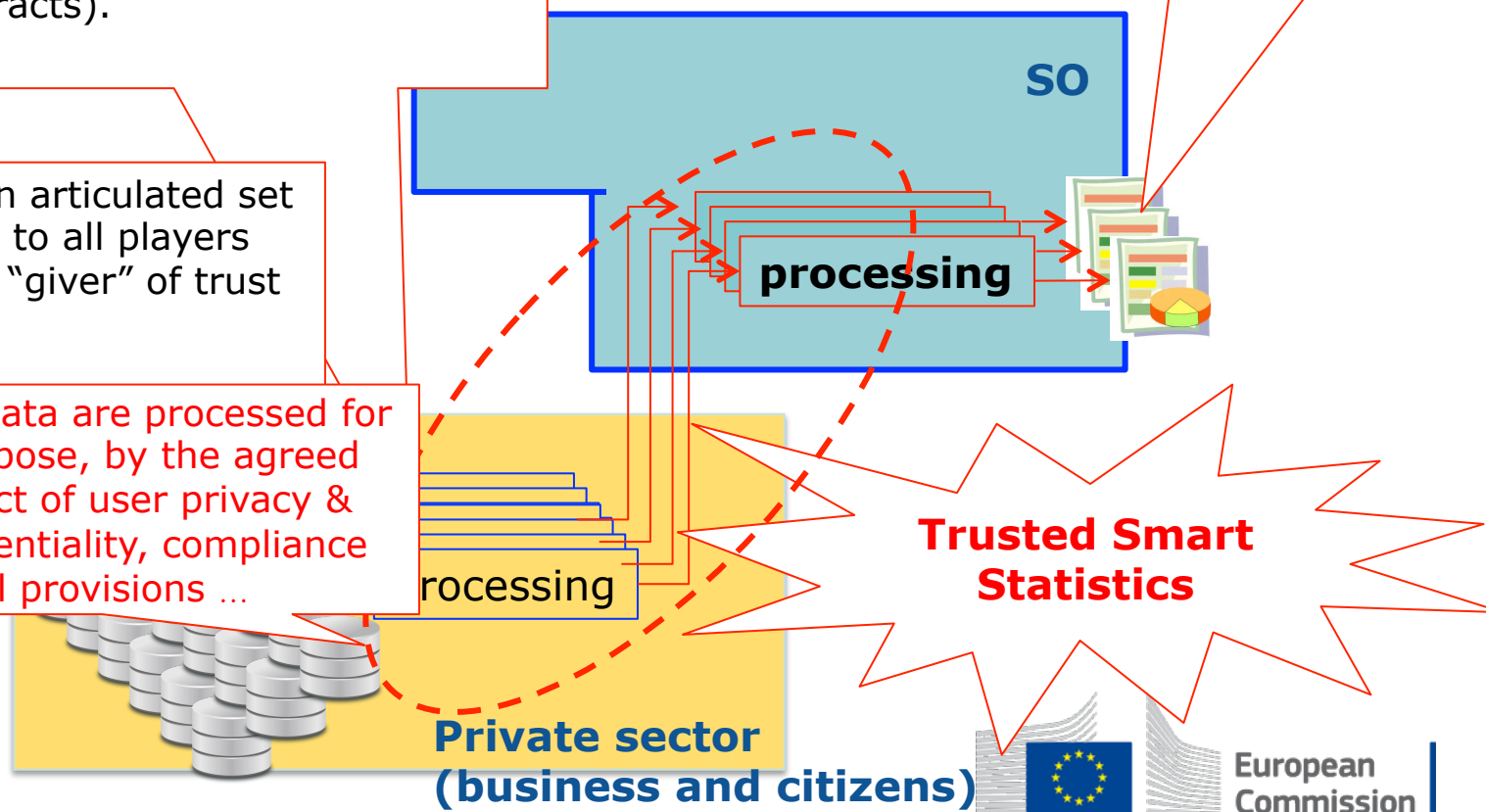
**Smart:** externalization towards **data sources** of the (initial) part of **processing execution**

Leveraging the “smart” features of the data sources (often Smart Systems, Smart Objects) and other “smart technologies” (e.g., Smart Contracts).

**Smart Statistics** as an opportunity to deliver more advanced statistical products, more timely (nowcasting), more targeted to specific user groups, through novel reporting and presentation ways ...

**Trusted:** ensure an articulated set of trust guarantees to all players (SO as “taker” and “giver” of trust guarantees)

Guarantee that data are processed for the agreed purpose, by the agreed method, respect of user privacy & business confidentiality, compliance with legal provisions ...

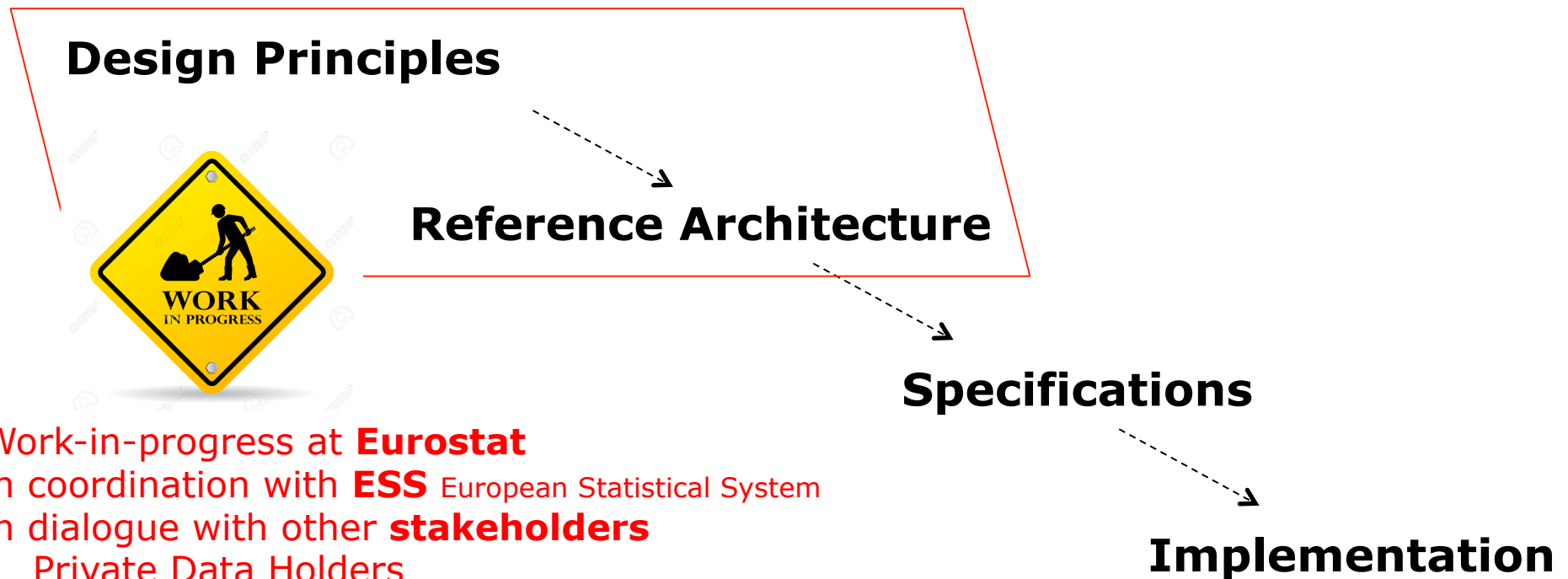


Private sector  
(business and citizens)

Trusted Smart  
Statistics



# Towards a Reference Architecture for Trusted Smart Statistics



Work-in-progress at **Eurostat**  
in coordination with **ESS** European Statistical System  
in dialogue with other **stakeholders**

- Private Data Holders
- Researchers, Academic communities
- Data Protection Authorities
- other arms of European Commission
- National and Local authorities
- ...

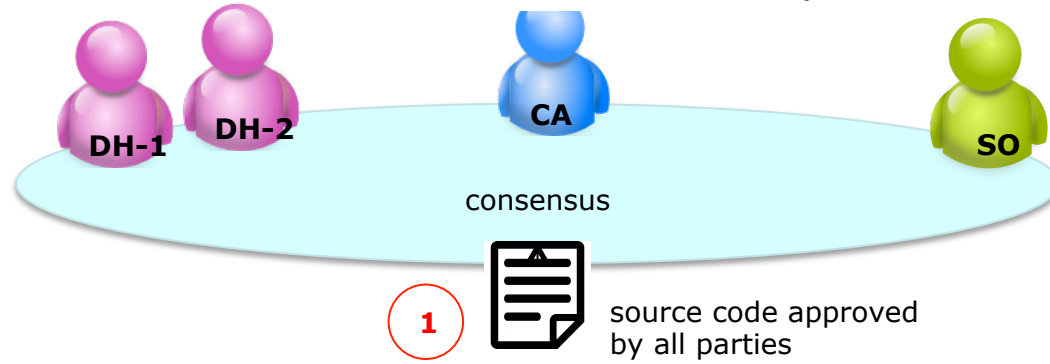
# Some design principles

1. *Processing method (algorithm) transparent to all involved parties*
  - co-designed or at least agreed-upon (consensus-based design)

Data Holders

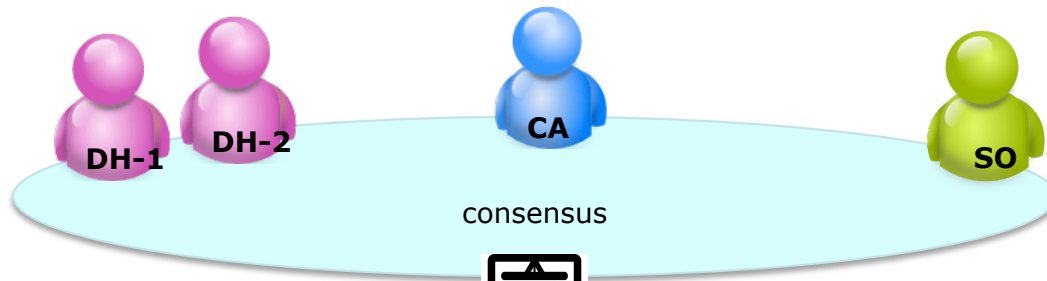
Certification Authority?

Statistical Office



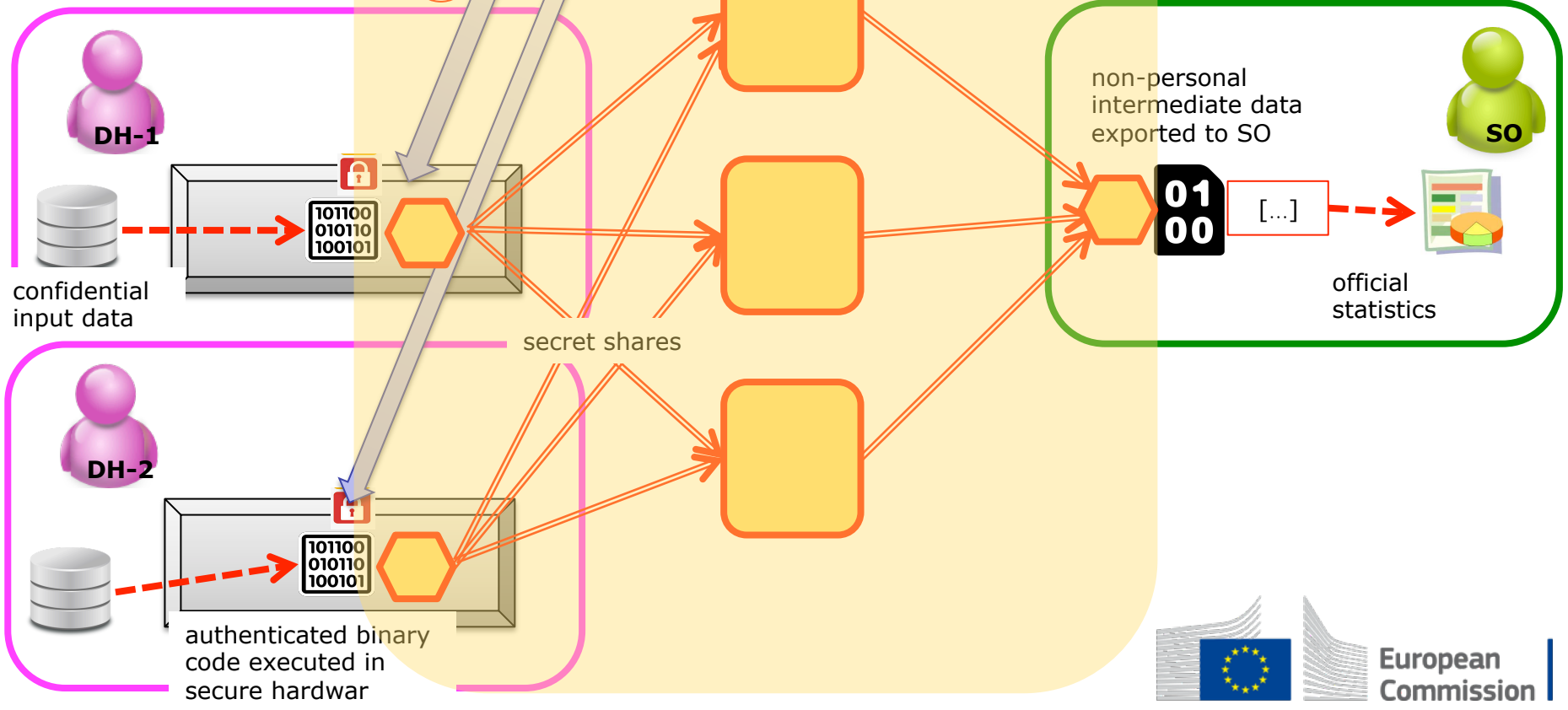
# Some design principles

2. *Data are not "moved to/shared with", but only "used by" the Statistical Office – goal is the output, not the input!*
  - Adopt technologies for Secure Private Computing technologies, e.g., Secure Multy-Party Computation



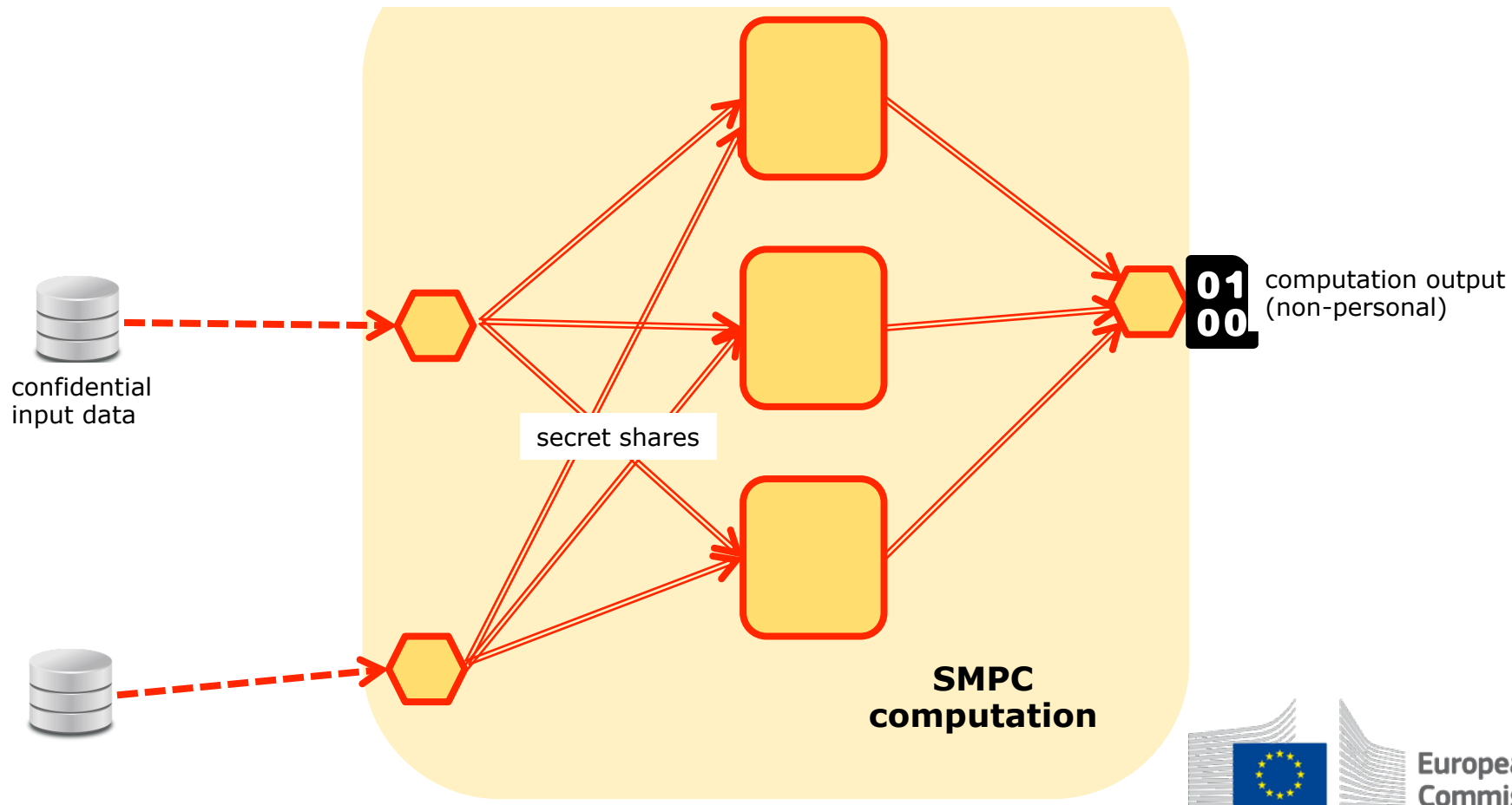
1 source code approved by all parties

2

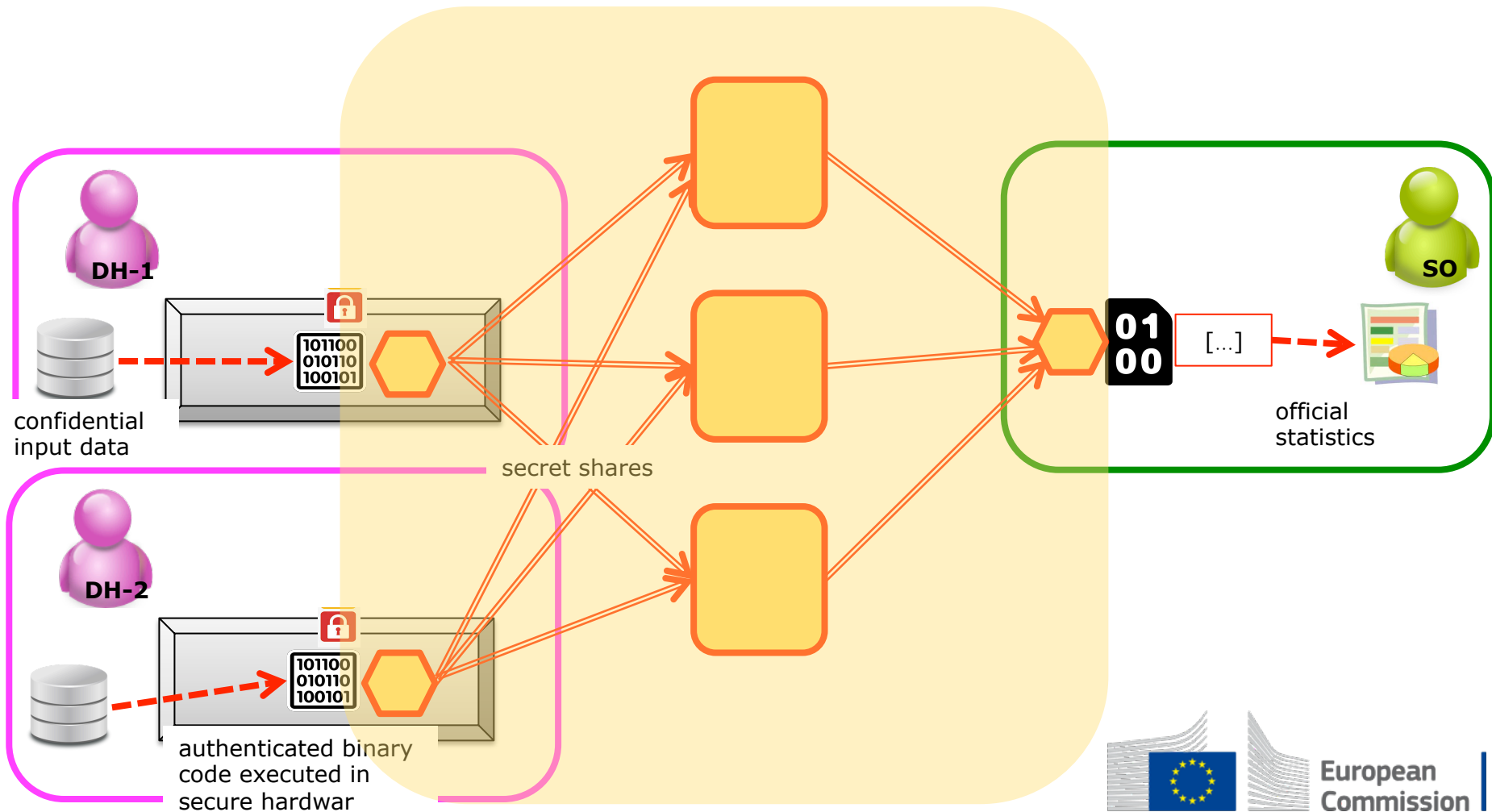


# Secure Multi-Party Computation (SMPC) infrastructure

An infrastructure (technology + organizational provisions) to let the output information be extracted without exchanging the input data

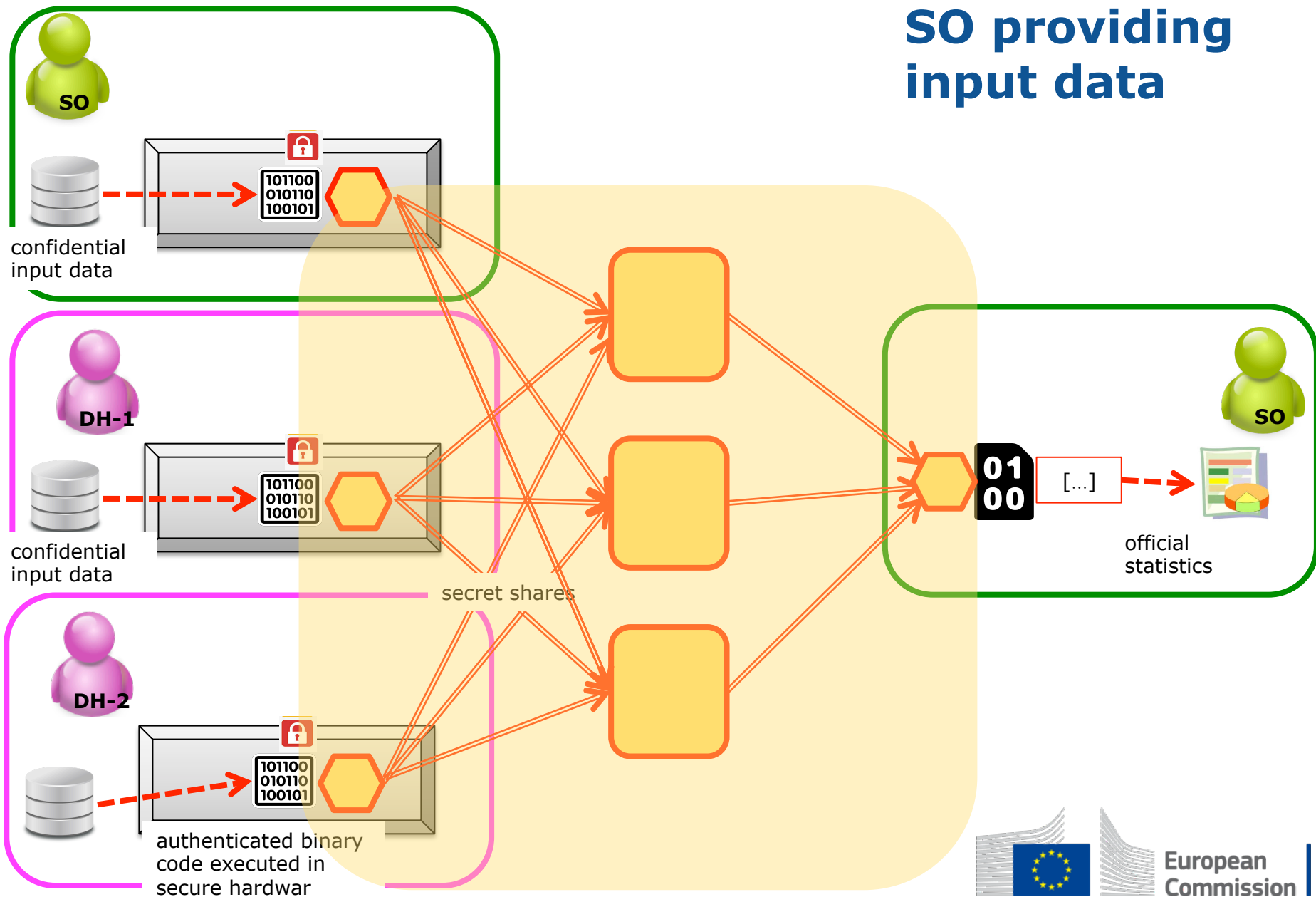


# B2G scenario with multiple DHs

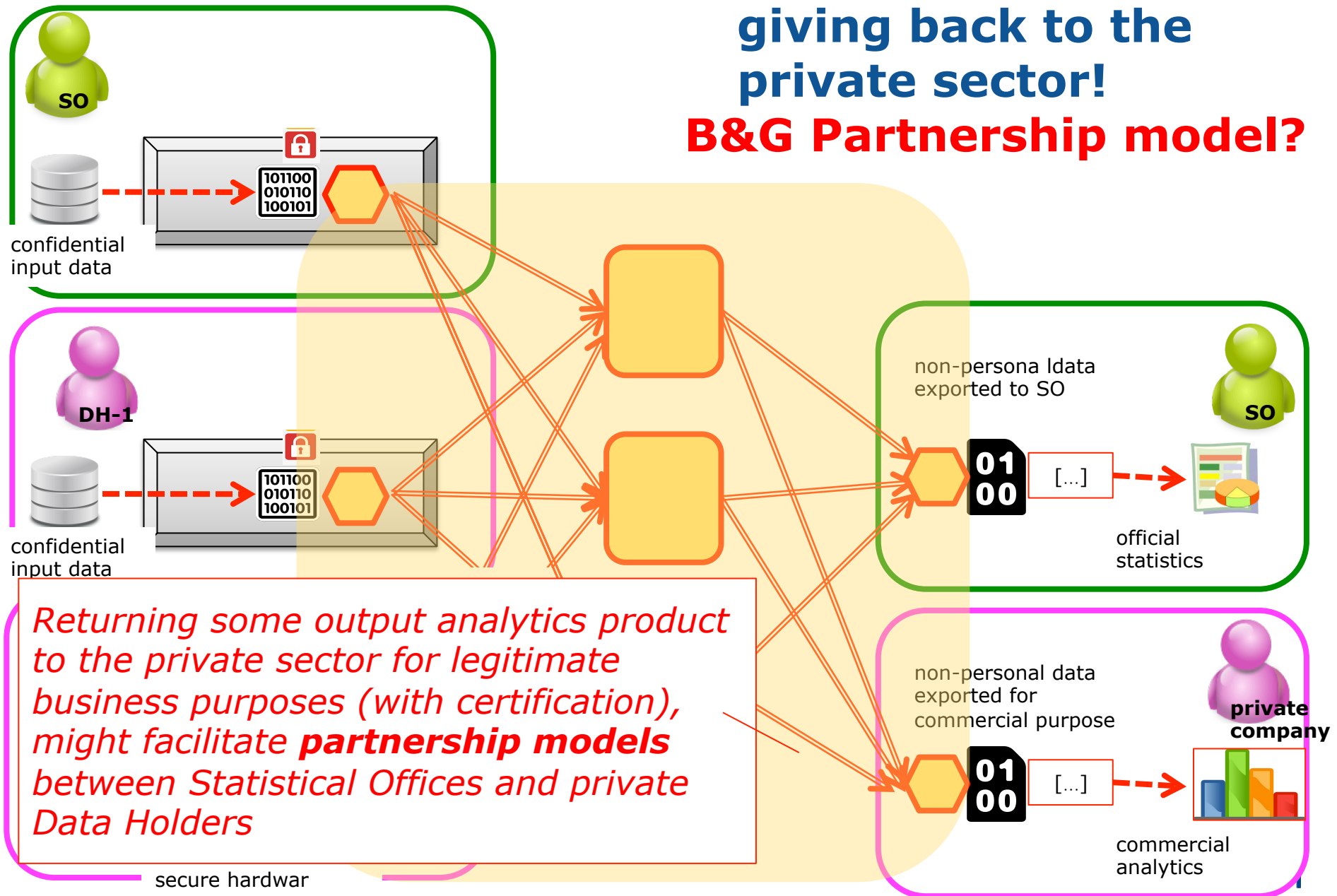




# BG2G scenario: SO providing input data

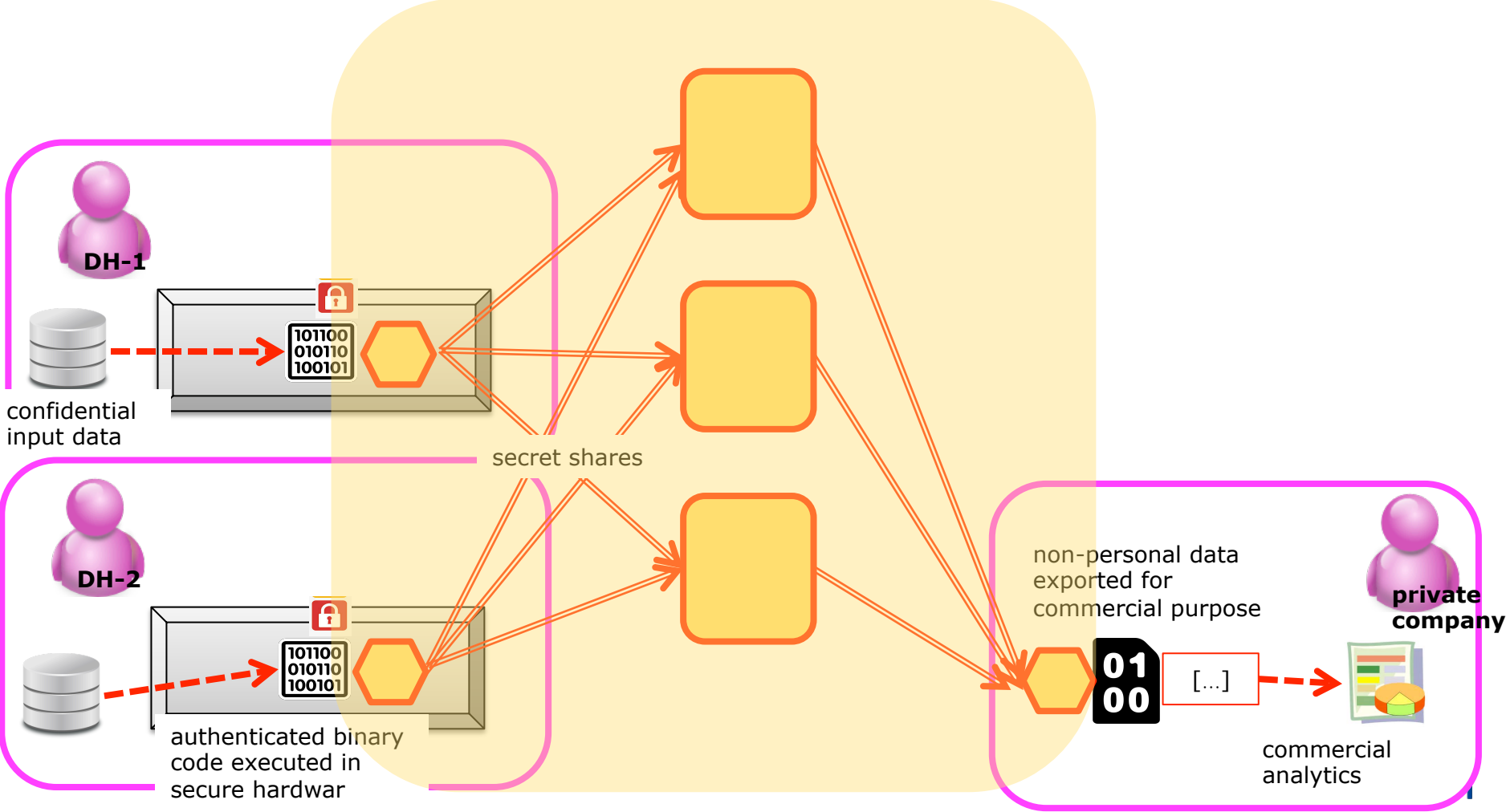


# B2G2B scenario: giving back to the private sector! B&G Partnership model?



secure hardware

# Reusing the infrastructure for B2B analytics?



# Some design principles

## 3. *Engage and partner with the input parties*

- Incentives might involve “giving back” computation output to them

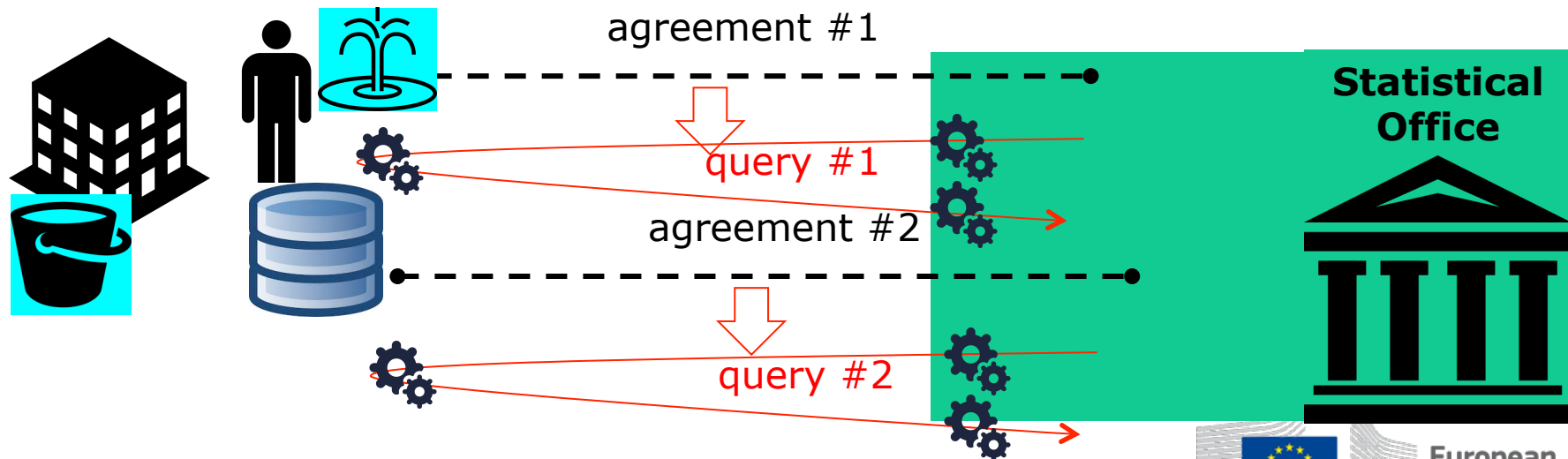
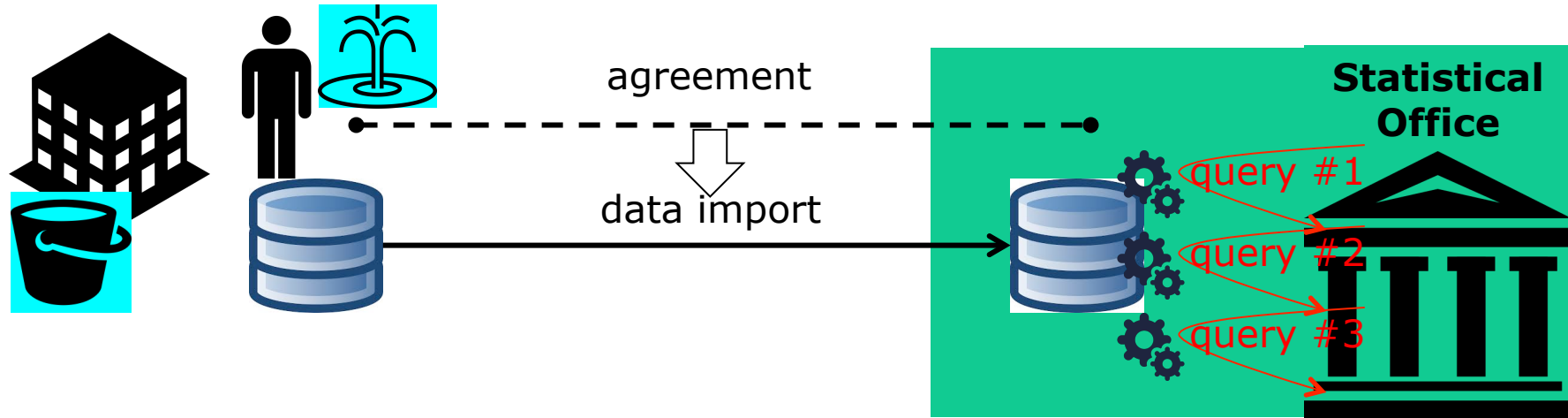
# Some design principles

## 4. *Agreement for data usage bound to computation instance.*

- Technological means guarantee that data cannot be used for other query/purpose other than the agreed one(s)

# Sharing input data

## Using input data on per-purpose basis



# Some design principles



## 5. *Purpose and algorithms **open** for public scrutiny*

- more public transparency → more public trust

# Some design principles

1. *Processing method (algorithm) transparent to all involved parties*
  - co-designed or at least agreed-upon (consensus-based design)
2. *Data are **not "moved to/shared with", but only "used by" the Statistical Office** – goal is the output, not the input!*
  - Adopt technologies for Secure Private Computing technologies, e.g., Secure Multy-Party Computation
3. *Engage and partner with the input parties*
  - Incentives might involve **"giving back"** computation output to them
4. ***Agreement for data usage bound to computation instance.***
  - Technological means guarantee that data cannot be used for other query/ purpose other than the agreed one(s)
5. *Purpose and algorithms **open** for public scrutiny*
  - more public transparency → more public trust



# Some slogans to shout loud

- *Let the **information** flow, not the **data**!*
- *Don't show your data to me, but let me use it!*
- *Share/distribute the **computation**  
share/distribute the **control**  
don't share/distribute the **data**!*
- ***Close the data, open the algorithms!***
- *Using more pervasive data calls for*
  - → more public transparency (open-source)
  - → more checks and balances  
(distributed control, consensus, certification authorities?)
  - → stronger *engagement* of sources (fountains and buckets)



# Take home message

- *Trusted Smart Statistics = the future of Official Statistics*
- *New sources of "big" data as input: more pervasive, timely, heterogeneous... and often privately held!*
- *Exploiting such data for Official Statistics requires a new architecture to build "trust" among all stakeholders  
→ ongoing work in Eurostat*
- *Key ingredients: SMPC and/or Trusted Hardware, open algorithms, source-code certification (?),...*
- *Once deployed, the same platform can be reused for other public interest purposes  
(and perhaps even for B2B applications)*





**Thanks for your attention**

**For follow-up contact**

*[fabio.ricciato@ec.europa.eu](mailto:fabio.ricciato@ec.europa.eu)*