Micro Data Hubs for Central Banks and the Way into Big Data

Stefan Bender (Deutsche Bundesbank)

IMF Statistical Forum, Frankfurt, November, 19th 2015

The content of the paper represents the author‘s personal opinions and do not necessarily reflect the view of the Deutsche Bundesbank or its staff.

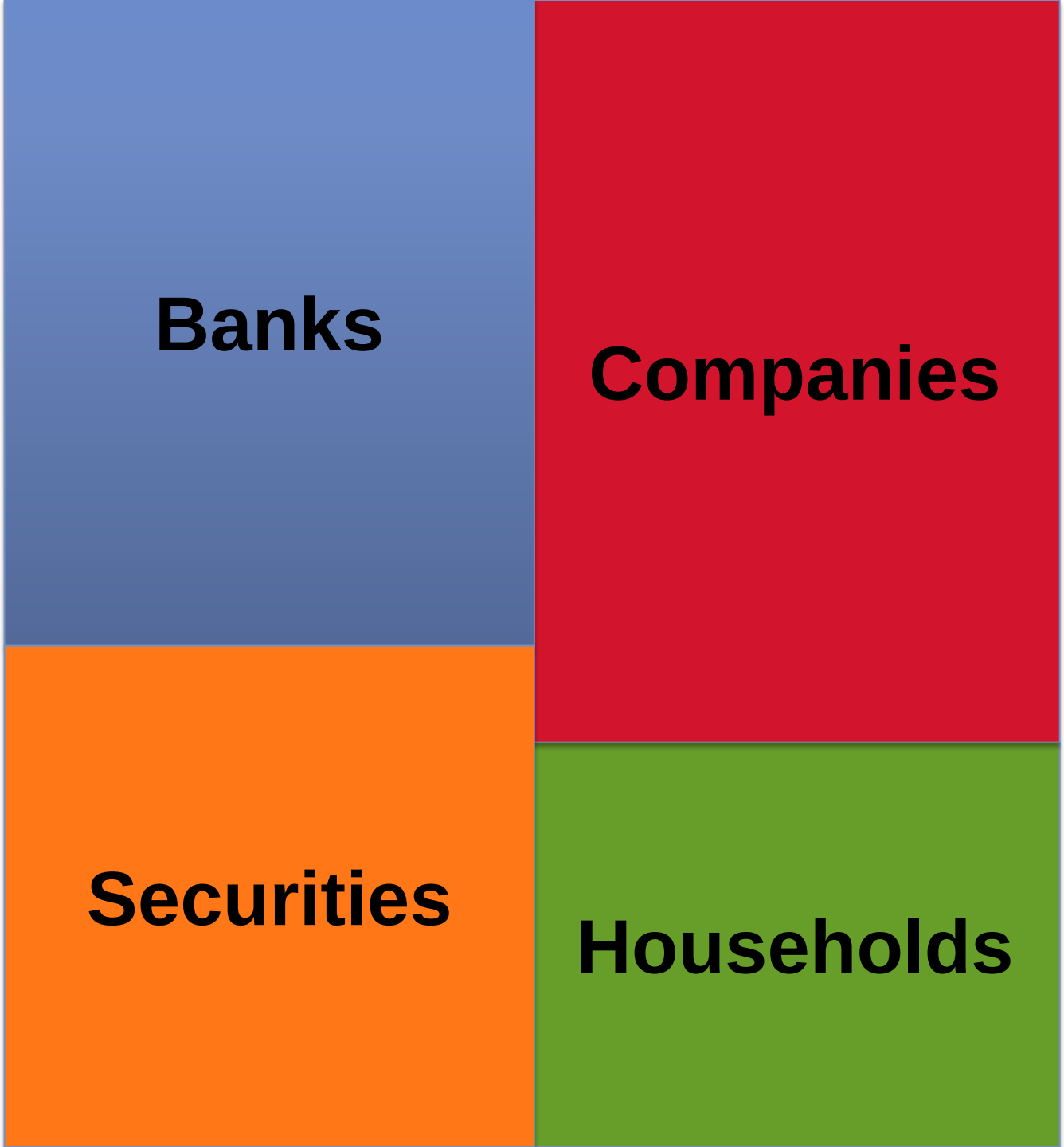
Policy evaluation can make better use of existing datasets

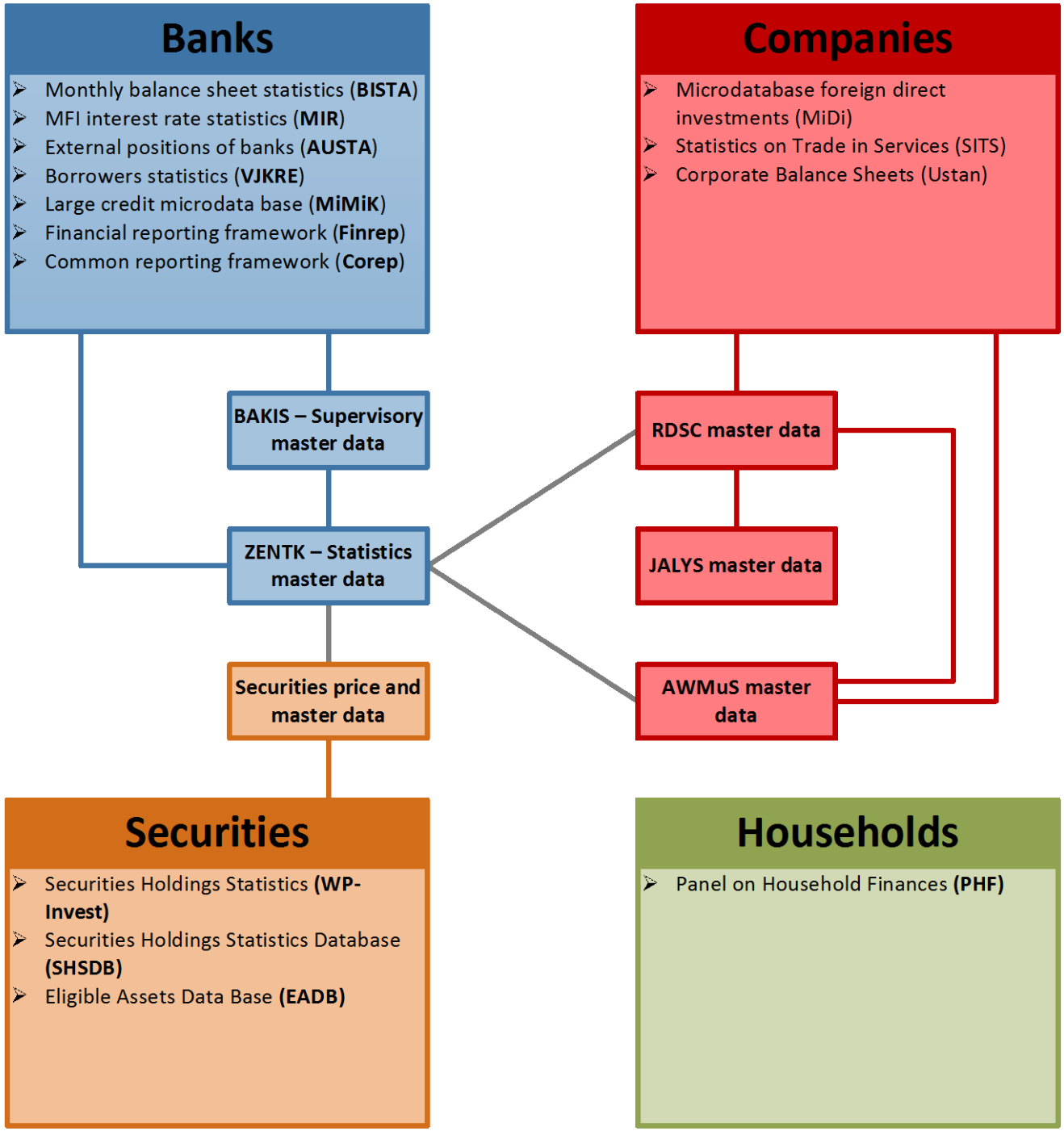
- The Bundesbank – like other central bank – produces datasets which are highly valuable for policy analysis and research.
- Systematic use of these data for policy analysis is often constrained by
  - Time
  - IT-resources
  - Legal restrictions

➢ The Bundesbank has launched a large-scale initiative aimed at making better use of existing data both, for policy analysis as well as internal and external researchers.

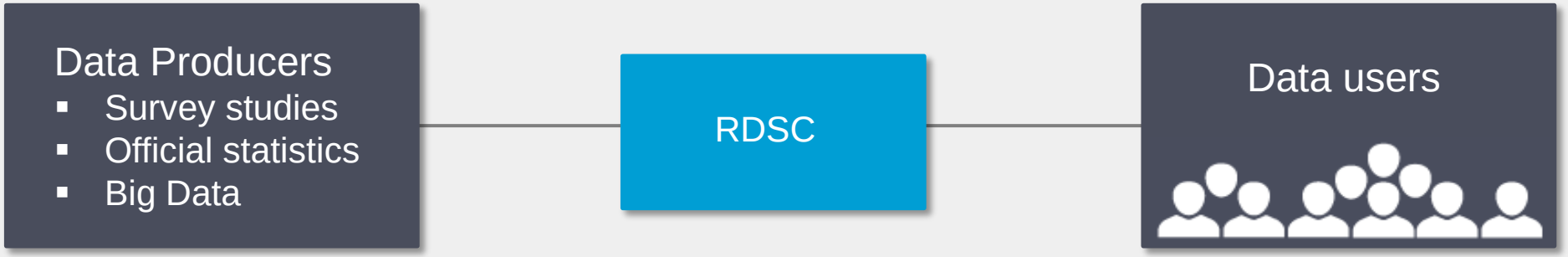
Scope of the Bundesbank‘s *Research Data and Service Center (RDSC)*

- **The RDSC is part of the Bundesbank internal project Integrated MicroData-based Information and Analysis System (IMIDIAS)**
- **Goals of IMIDIAS:**
  - Encourage cooperation with (external) researchers
  - Promote evidence-based policy-making
  - Support policymaking processes
- **Key principles:**
  - Data as a public good
  - Democratic data access
  - Data protection


Banks
Companies
Securities
Households


Banks
Monthly balance sheet statistics (BISTA)
MFI interest rate statistics (MIR)
External positions of banks (AUSTA)
Borrowers statistics (VJKRE)
Large credit microdata base (MiMiK)
Financial reporting framework (Finrep)
Common reporting framework (Corep)
Companies
Microdatabase foreign direct investments (MiDi)
Statistics on Trade in Services (SITS)
Corporate Balance Sheets (Ustan)
BAKIS – Supervisory master data
RDSC master data
ZENTK – Statistics master data
JALYS master data
Securities price and master data
AWMuS master data
Securities
Securities Holdings Statistics (WP-Invest)
Securities Holdings Statistics Database (SHSDB)
Eligible Assets Data Base (EADB)
Households
Panel on Household Finances (PHF)

Data Access in the RDSC



- RDSC mediates between data producers and external users.
- RDSC controls for compliance with data protection regulations.

Tasks of the RDSC

RDSC offers access for non-commercial research to the (highly sensitive) micro data.

- Generates micro data (linking data).
- Provides data access and data protection.
- Offers advisory service on data selection and data access (handling, potential, scope and validity of data).
- Documents data and methodological aspects of data.

Factsheet on the RDSC of the Bundesbank

- The RDSC has started in 2014 as part of the Statistics Department of the Bundesbank.
- Over 150 active projects.
- 12 employees (2016: 14).
- 12 working places for guest researchers.

Location of the RDSC

- 20th floor of the Trianon-Tower in Frankfurt.
- Near Frankfurt central station.

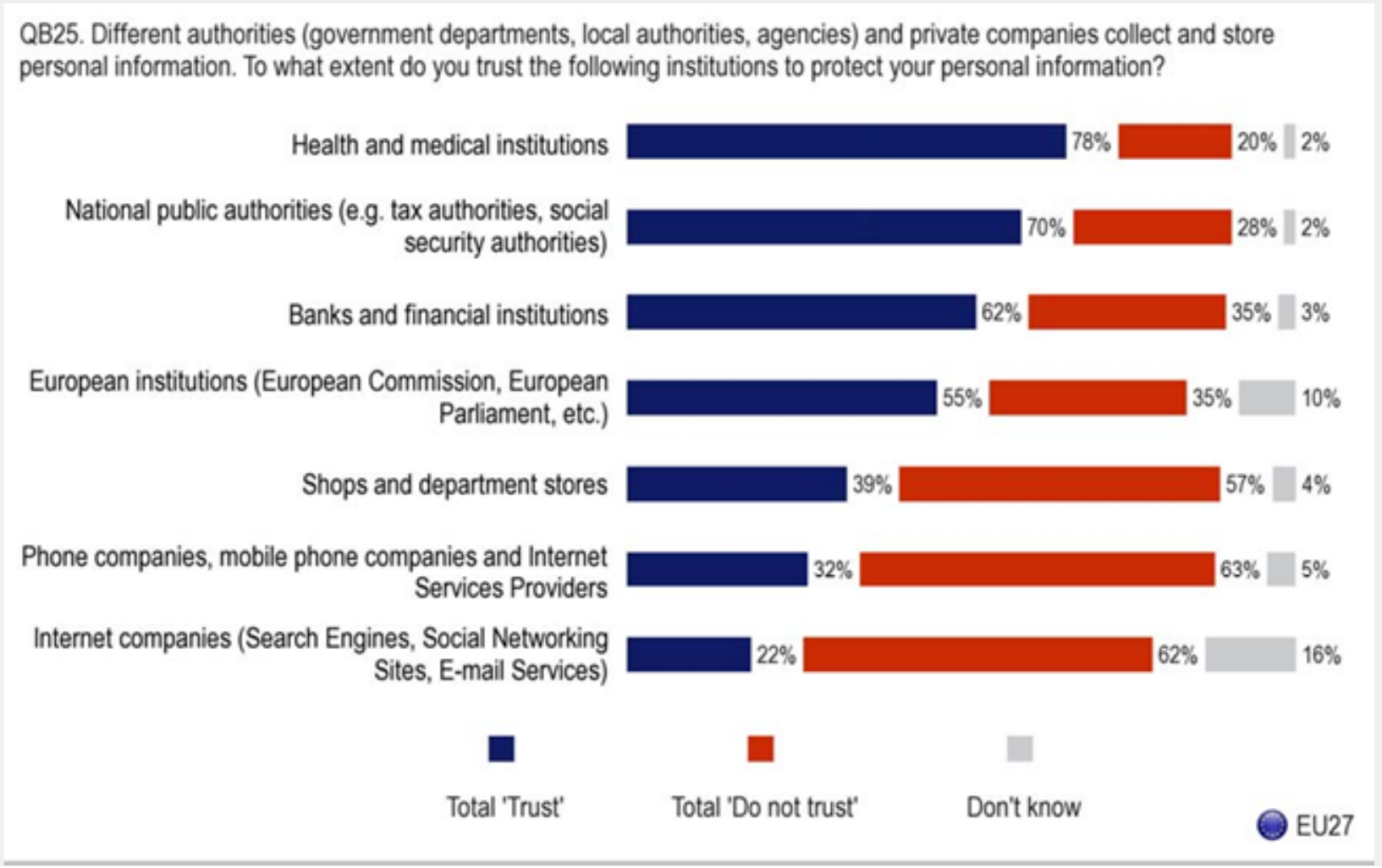
Working places for guest researchers

Trianon-Tower

Types of Information considered as Personal Information

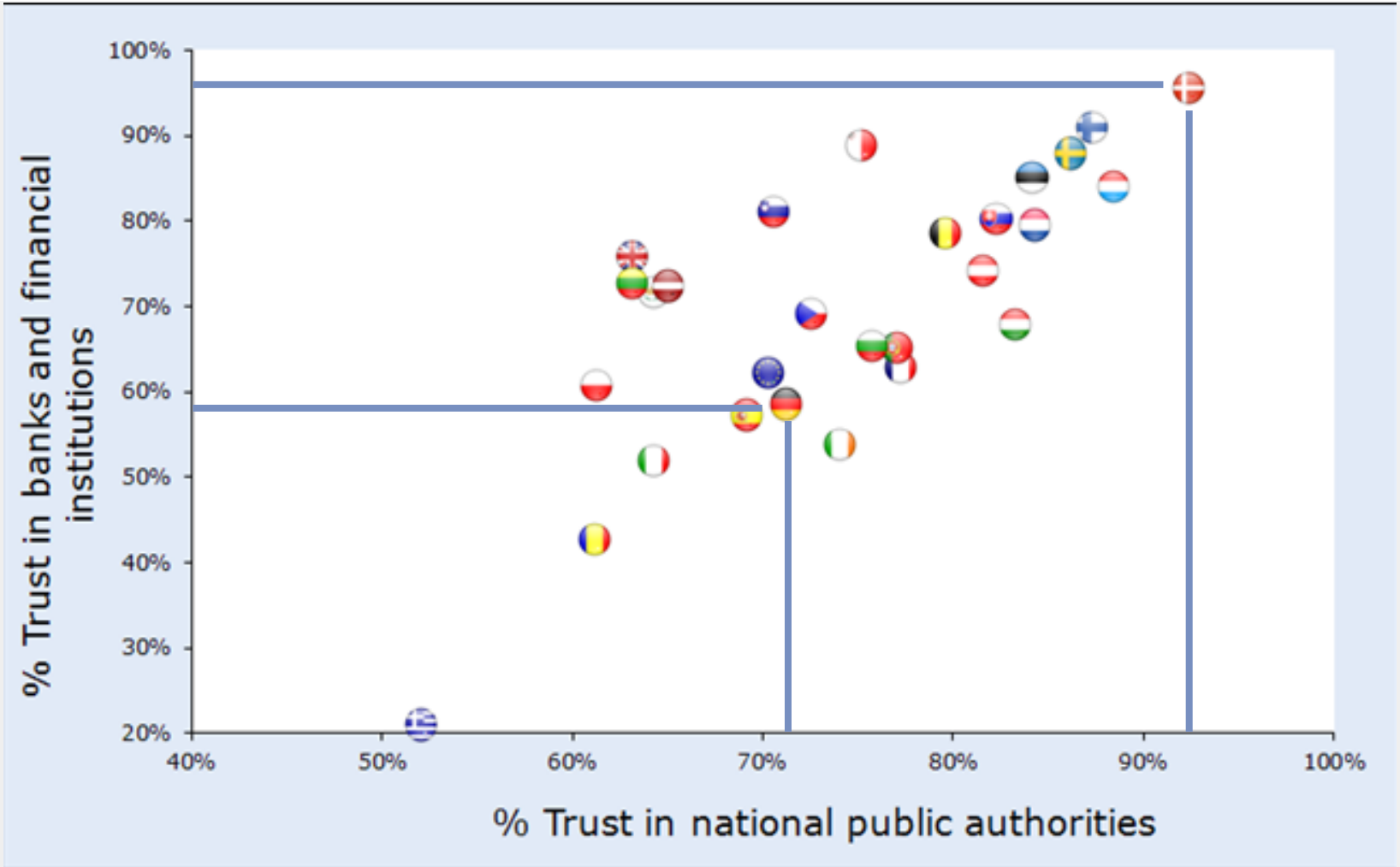
(Special Eurobarometer 359: Attitudes on Data Protection and, Electronic Identity in the European Union, Survey 11-12/2010)

Trust in Institutions to Protect Personal Information



(Special Eurobarometer 359: Attitudes on Data Protection and, Electronic Identity in the European Union, Survey 11-12/2010)

Correlation between Trust in Banks and National Public Authorities by Country

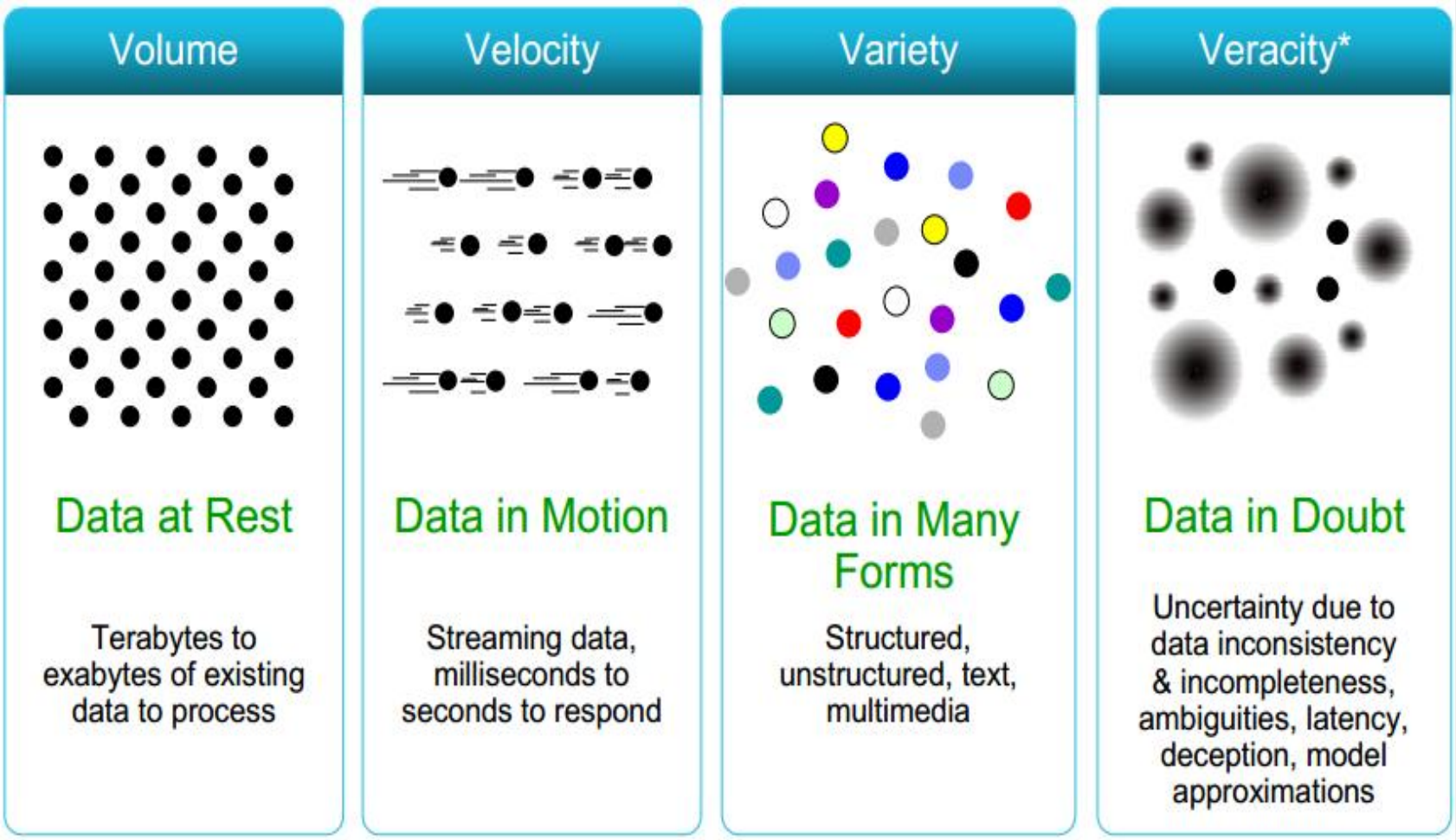


(Special Eurobarometer 359: Attitudes on Data Protection and, Electronic Identity in the European Union, Survey 11-12/2010)

Arguments for Moving into Big Data

- Use of Big Data for research and policy advice will increase.
- Big Data represent additional data sources we need.
- Additional arguments (topics of the presentation)
  - Nature of found data (Big Data).
  - Data generating process.
  - Paradigm shift in research.
  - Access to Big Data (found data).

Definition of Big Data Data – But (at least) one more V

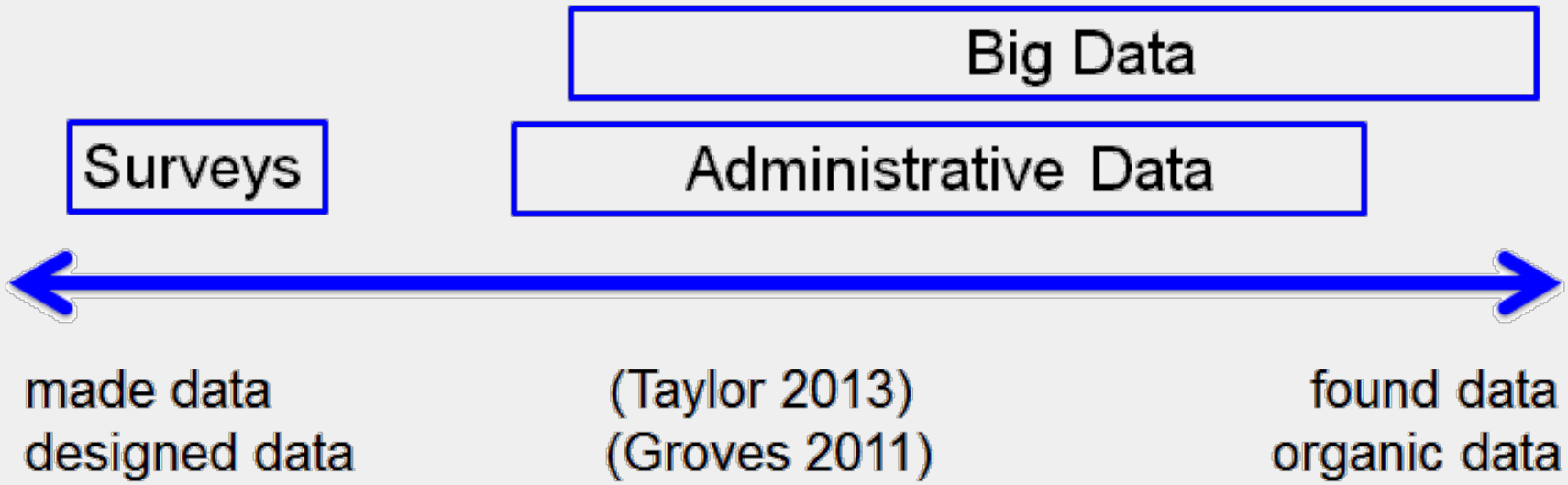


(http://www.rosebt.com/blog/data-veracity)

Characteristics of Big Data I

- Secondary data.
- Related to some non-research purpose and then reused by researchers.
- The amount of control a researcher has and the potential inferential power vary between the different types of Big Data sources.

Characteristics of Big Data II



Big Data are not just Data

Imprecise description of a rich and complicated set of

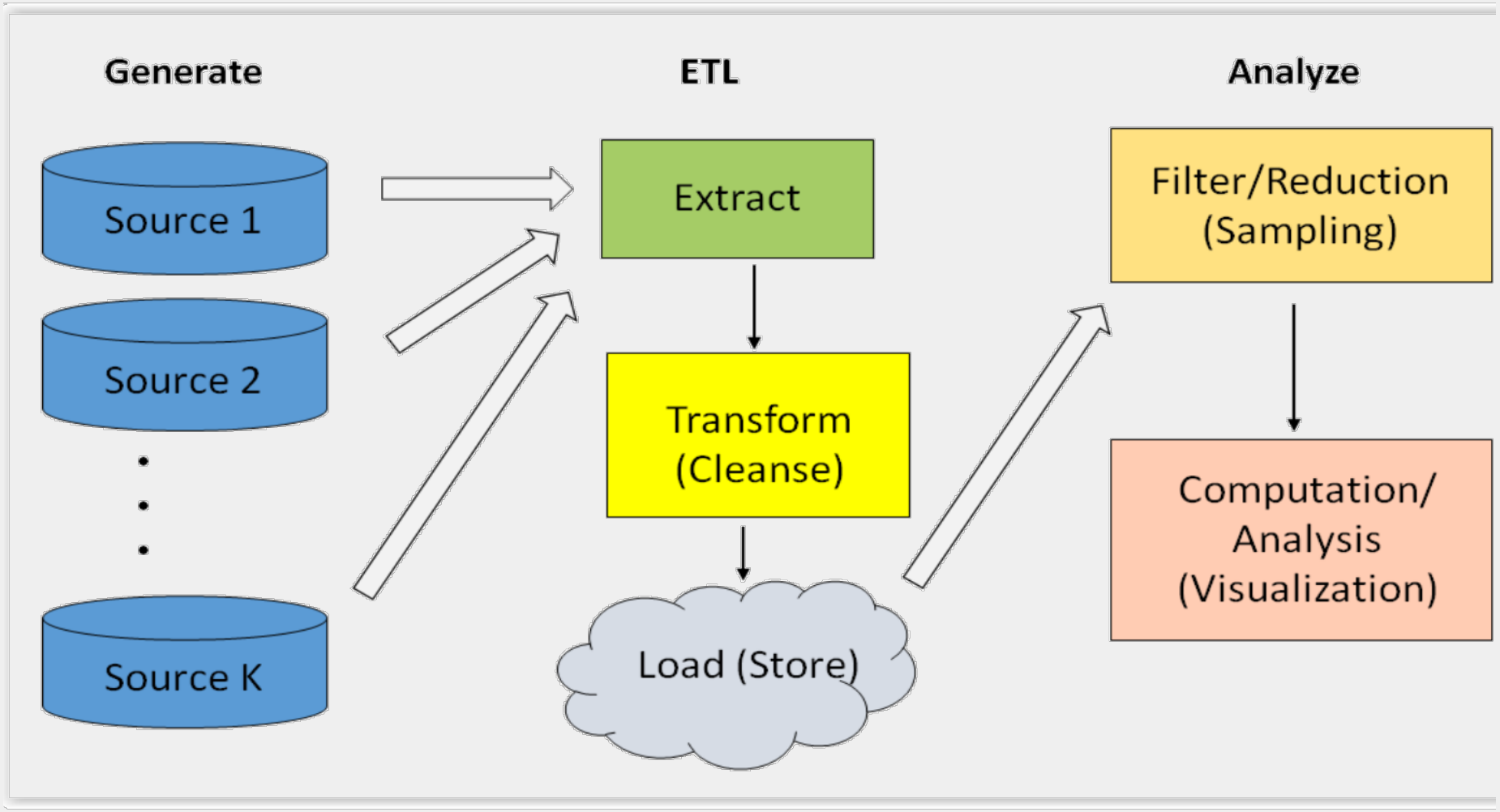
- characteristics,
- practices,
- techniques,
- ethical issues, and
- outcomes

all associated with data.

Data Generating Process

- Big Data are often selective, incomplete and erroneous.
- Big Data are typically aggregated from disparate sources at various points in time and integrated to form data sets.
- Thus in statistically valid ways, using Big Data is increasingly challenging.

Big Data Process Map: Big Data Total Error (BDTE)



Conclusion

- Policy advice and research is about answering questions.
- There is a strong need for granular data.
- Access to data is needed.
- Solutions are in place to fulfill privacy issues (RDSC).
- New possibilities of combining different data sources:
  a. Big Data, b. surveys, c. admin data.
- Big Data is not only data, it is a new thinking with data.

Cost-Benefit: Big Data (but not only Big Data)

“The mining of personal data can help increase welfare, lower search costs, and reduce economic inefficiencies;

at the same time,

it can be source of losses, economic inequalities, and power imbalances between those who hold the data and those whose data is controlled.“

(Acquisti 2014, p. 98)

Literature, Sources

- **AAPOR Report on Big Data**
  by AAPOR Big Data Task Force; February 12, 2015
  Lilli Japec, Frauke Kreuter, Marcus Berg,
  Paul Biemer, Paul Decker, Cliff Lampe, Julia Lane,
  Cathy O'Neil, Abe Usher.

- **Privacy, big data, and the public good** * frameworks for engagement. Cambridge: Cambridge University Press..Lane, Julia; Stodden, Victoria; Bender, Stefan, Nissenbaum, Helen (eds.) (2014)

- **Big Data Working Group of the Bundesbank**

Contact information

Research Data and Service Centre

Deutsche Bundesbank
Mainzer Landstraße 16
- TRIANON -
60325 Frankfurt am Main
Germany

+49 69 9566-7595

fdsz@bundesbank.de