



Creating Comprehensive Data Worlds using Standardization

Stephan Müller, Deutsche Bundesbank

Are post-crisis statistical initiatives completed? - The data universe is exploding -

Demand for new
statistical surveys

Exploding data
supply

Banking crisis

How heavily affected are investors in Europe?

Sovereign debt crisis

Who holds which government bonds?

Banking union

What is the scope of risk concentration?

Low-interest-rate environment

How healthy are euro area banks?

Data amount is growing constantly and rapidly

- Automatic recording of process data (sensors, Internet of Things)
- Social networks and search engines
- Mobile phones and tablets

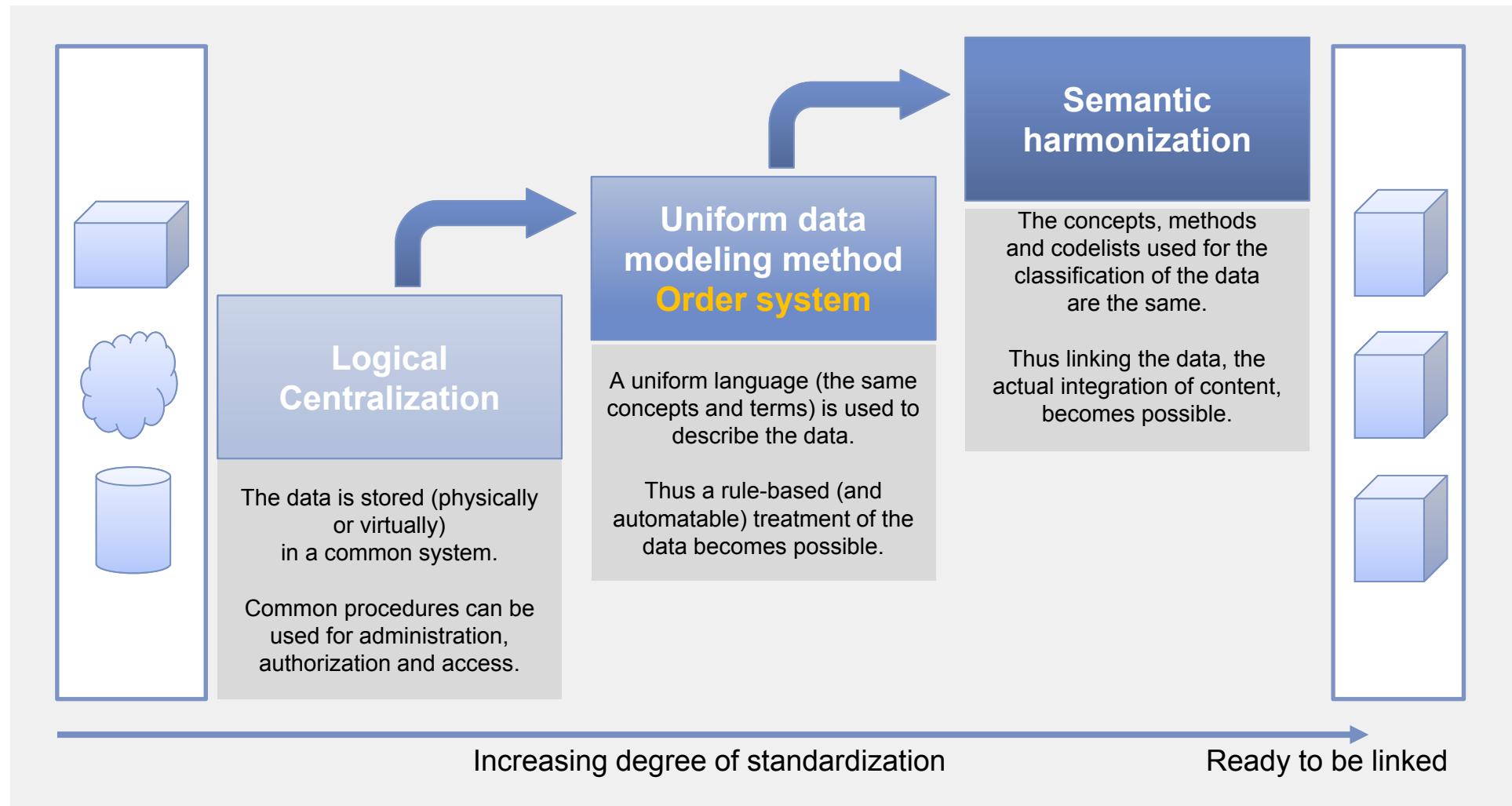
New technological developments

- More computing power: Big Data
- New analysis techniques: Machine Learning, AI

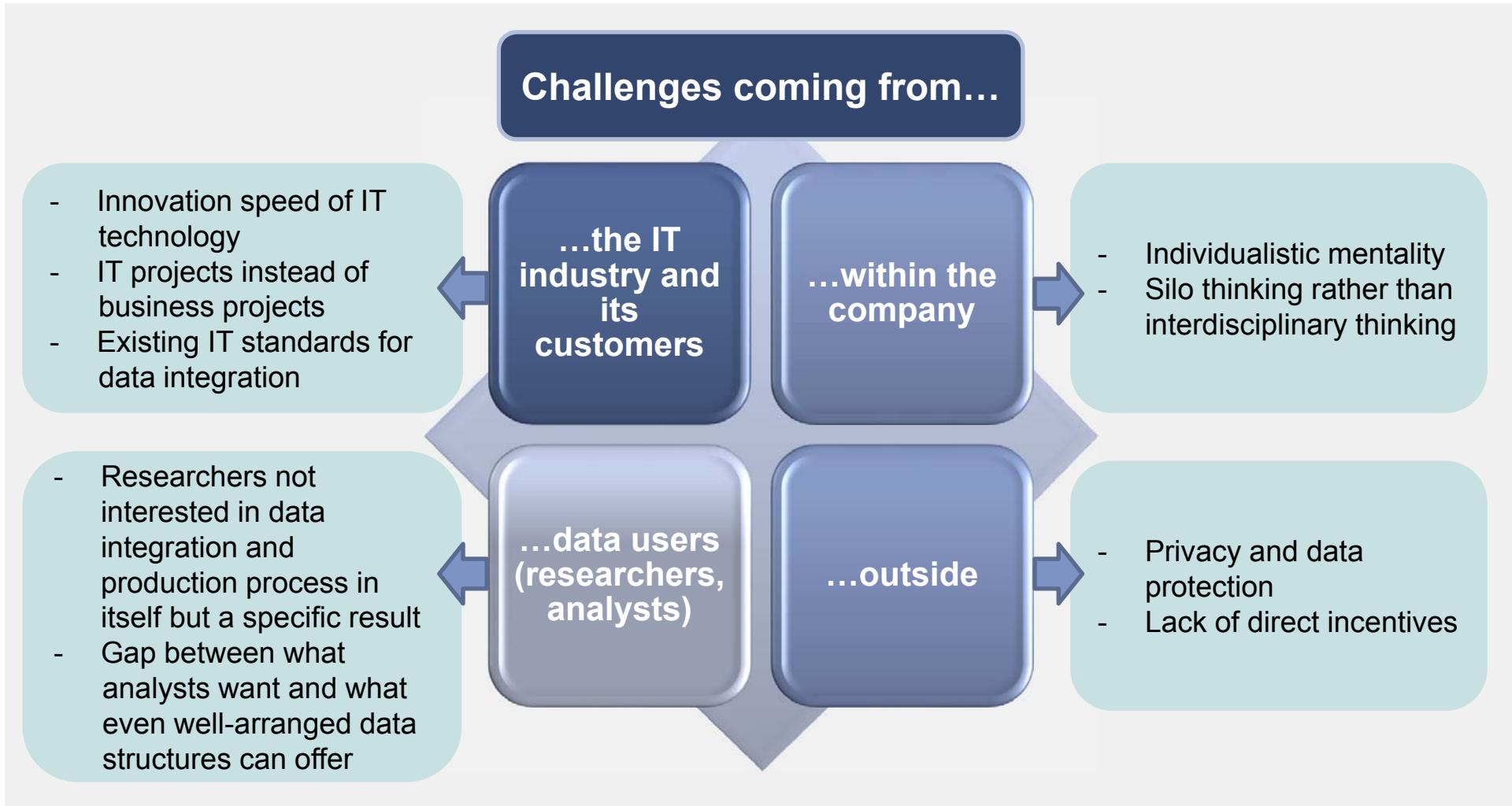
“Water, water, everywhere, but not a drop to drink.”

- **Yawning Data Gaps despite “Collectomania”**
 - Data is not collected where it's needed, but where it occurs. Still painful data gaps
- **The Data Universe lacks Order**
 - In IT: neither a system of order for data / information, nor a prominent standardization, nor a global identifier (“barcode for information”)
 - In companies: large part of the data stored in data silos; need for data integration / BI / DWH / Big Data projects / CIOs
 - In industry branches or countries: proprietary solutions
- **Using IT not Possible Without Content-Related Expertise**
 - No longer classical statistical production of prescribed indicators
 - Instead implementation of data analysis on demand
 - New style of data collections with hundreds of dimensions
 - Automation or lack of expertise could lead to comparing apples and oranges
 - Professional expertise crucial for evaluating and interpreting the results

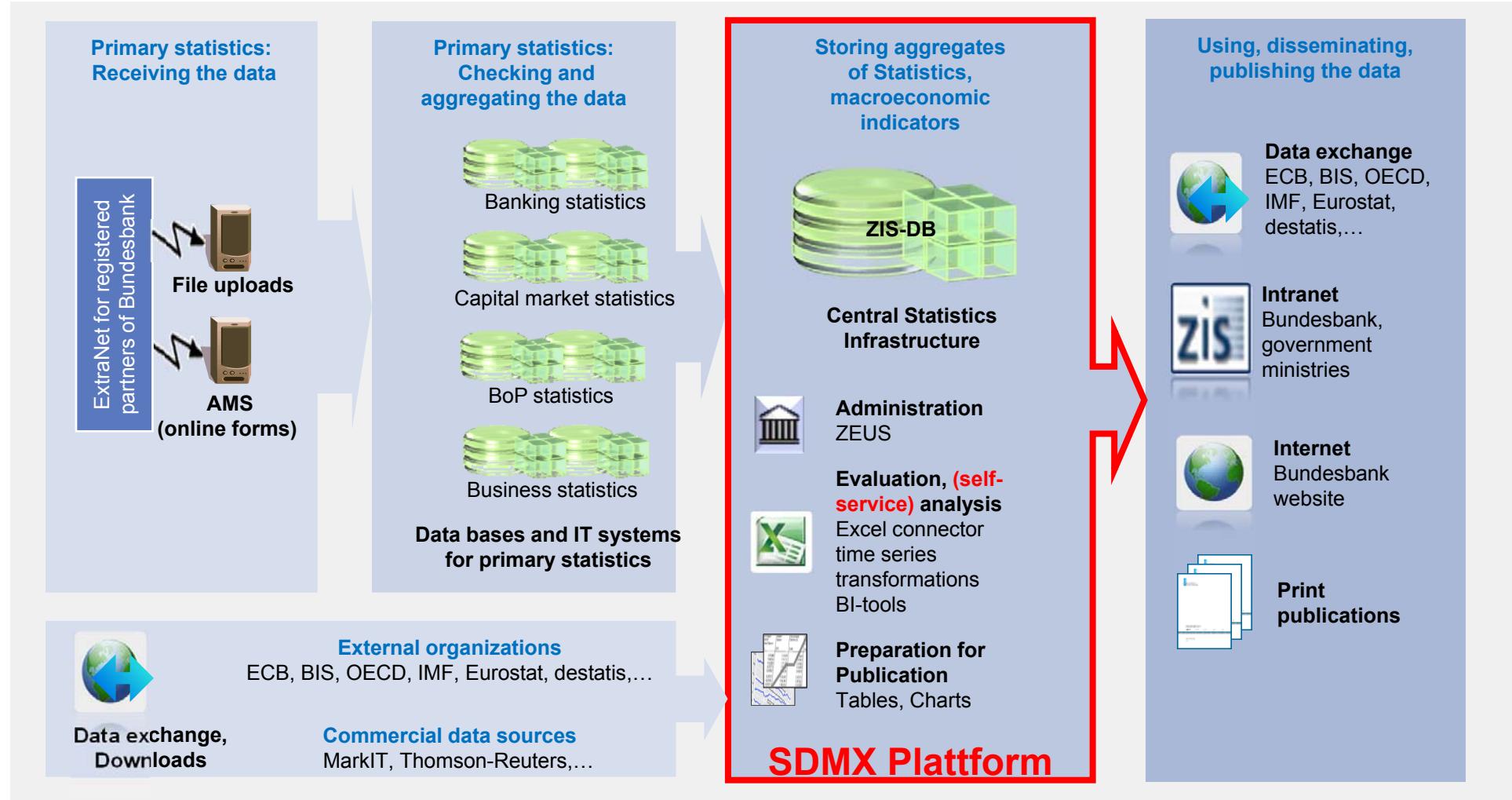
The three steps of data integration



Challenges for those who want to introduce data integration

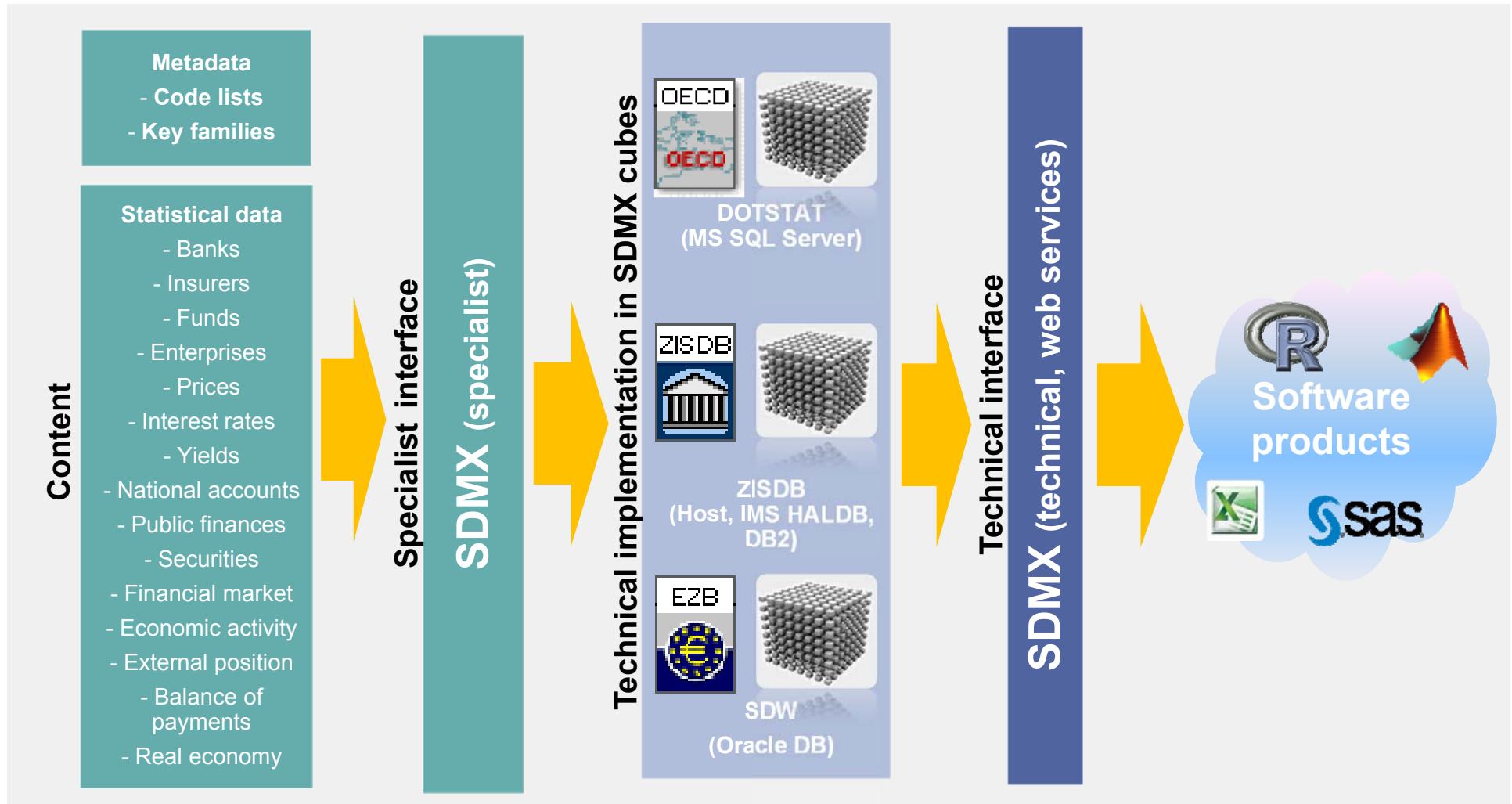


Directorate General Statistics Value Chain

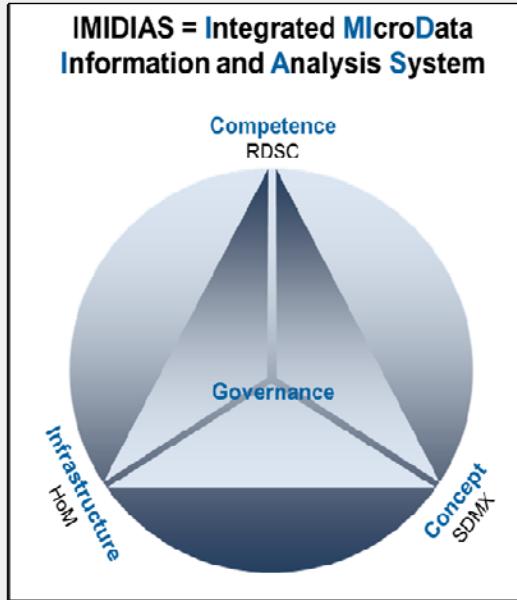


SDMX is used across domains and platforms

Decentralised data sinks on various technical platforms



SDMX and central statistics infrastructure Basis for House of Microdata (HoM)



- In 2013, the Statistics Department was mandated to establish an **integrated interdepartmental information system for analytical and research purposes based on microdata** for various user groups (financial stability, research, monetary policy, supervision)
- This should be achieved by developing a Research Data and Service Centre (RDSC) and a **microdatabase (HoM, “House of Microdata”)**
- This HoM is based on SDMX and the **Central Statistics Infrastructure**

- The SDMX model can be used without any problems for microdata.
- Data diversity requires standardization, SDMX provides a suitable framework
- Multidimensional approach, by using uniform code lists, offers an ideal means of linking and comparing data from different sources.

What is there to do?

- There are **no globally consistent code lists** so far
 - Is a **truly global standard** possible?
- Is the global approach and the open source idea in accordance with **confidentiality constraints and the legal framework?**
- **Open source** approach for future efforts?
- Do we invest enough time and effort in **Data Literacy?**