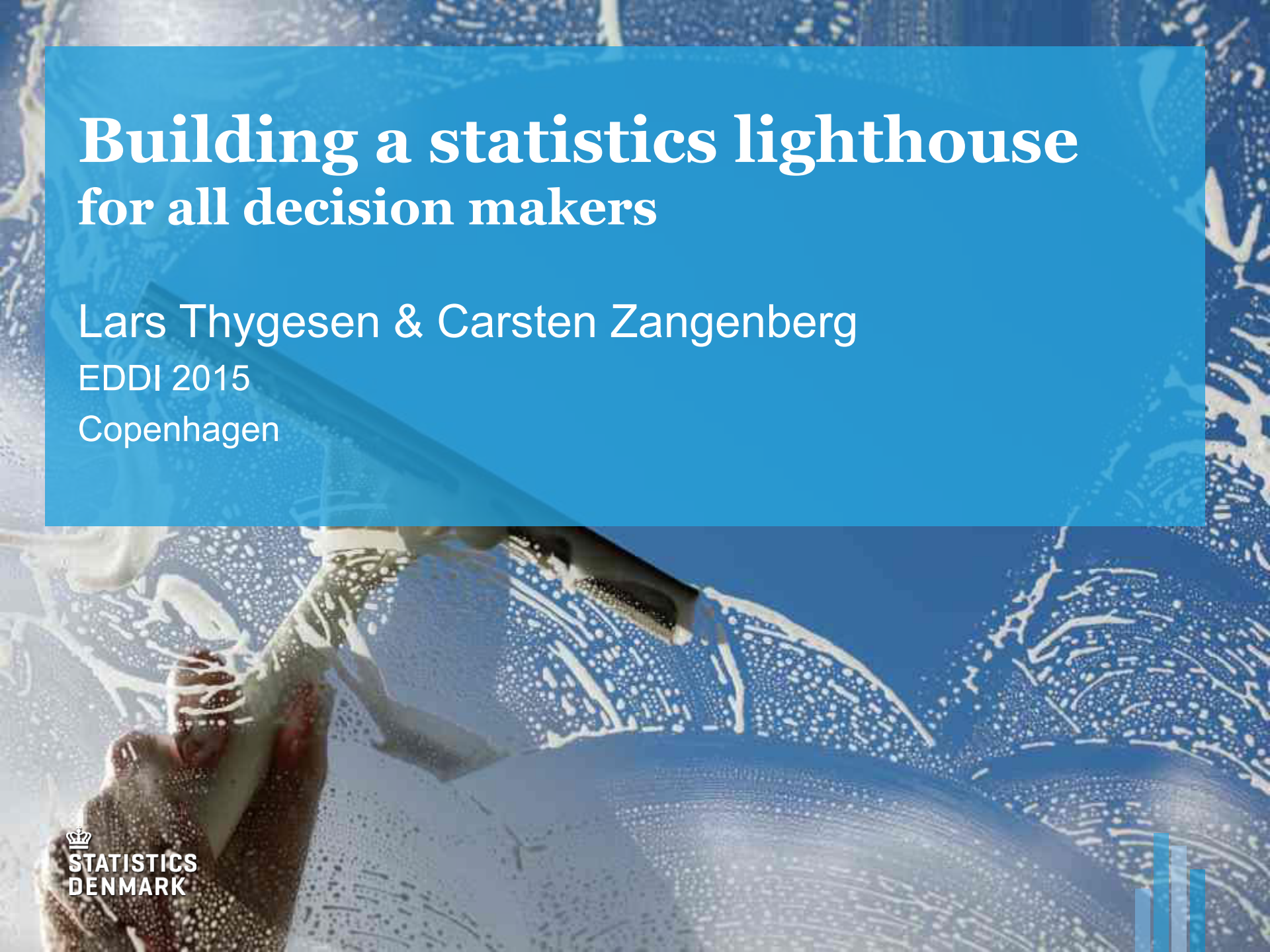
Building a statistics lighthouse
for all decision makers

Lars Thygesen & Carsten Zangenberg

EDDI 2015

Copenhagen



STATISTICS DENMARK

Why are we here?

- Because all good decisions rely on facts
- Provide the basic facts on society – a common reference point
- Basis for democracy and economy
- A lighthouse in the Sea of Information

Strategy of official statistics organisations

- Provide basis for informed decisions by:
  - Government
  - Research
  - Enterprises
  - Citizens
  - International organisations

What we have to do?

- Produce high quality statistics
- Make it available

But it is not enough!

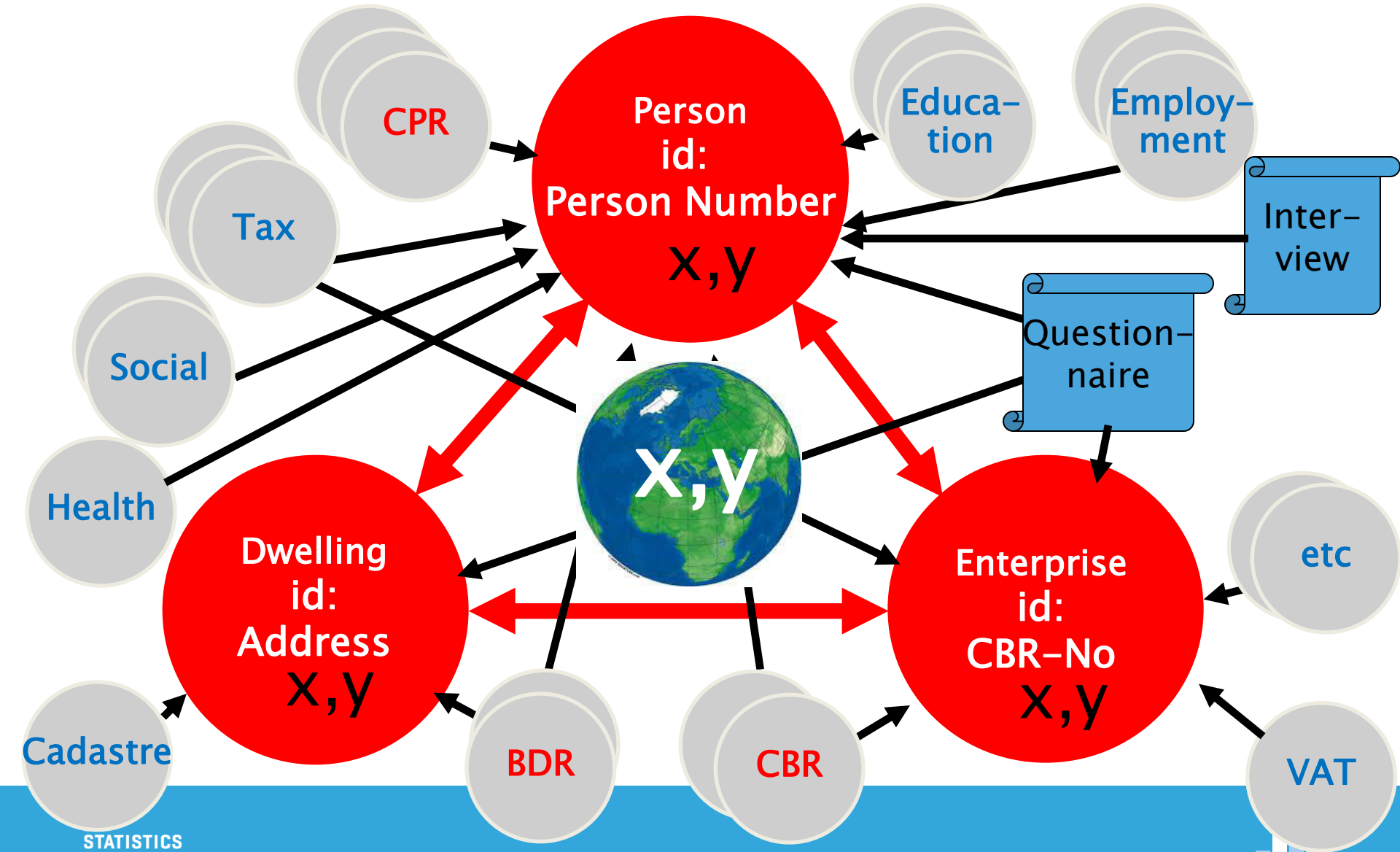
- Data ➡ information ➡ statistics ➡ knowledge
- Knowledge has to enter the heads of users

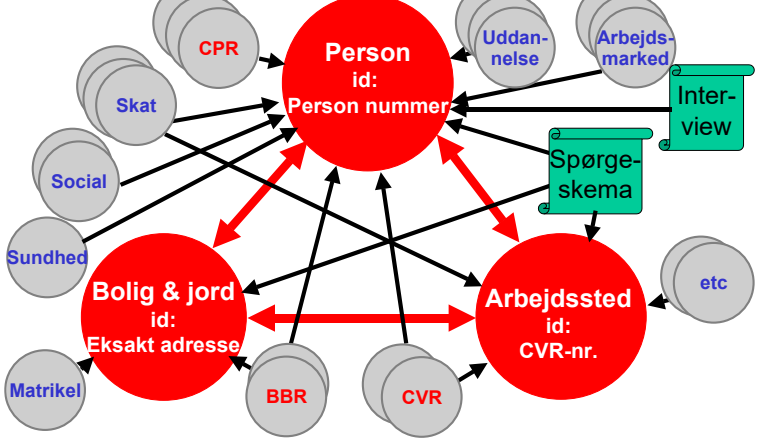
Simplified definition of statistical metadata

- Reference metadata:
  - Conceptual metadata (e.g. definition of income)
  - Methodological and processing metadata (e.g. description of data processing)
  - Quality metadata (e.g. Availability)

- Structural metadata:
  - Metadata act as identifiers and descriptors of the data (e.g. name on variables, dataset etc)

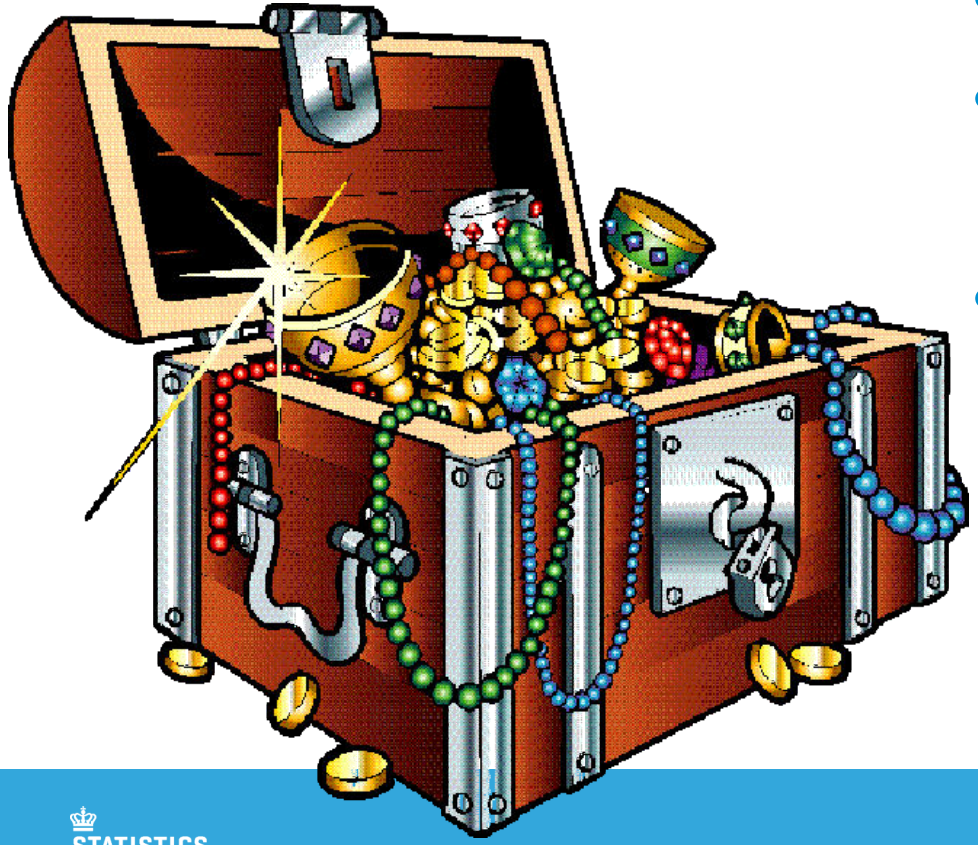
The Statistical Information System





A treasure!

- Immensely rich data
- High data quality
- Combining & linking
- Longitudinal studies



- But how can users benefit?

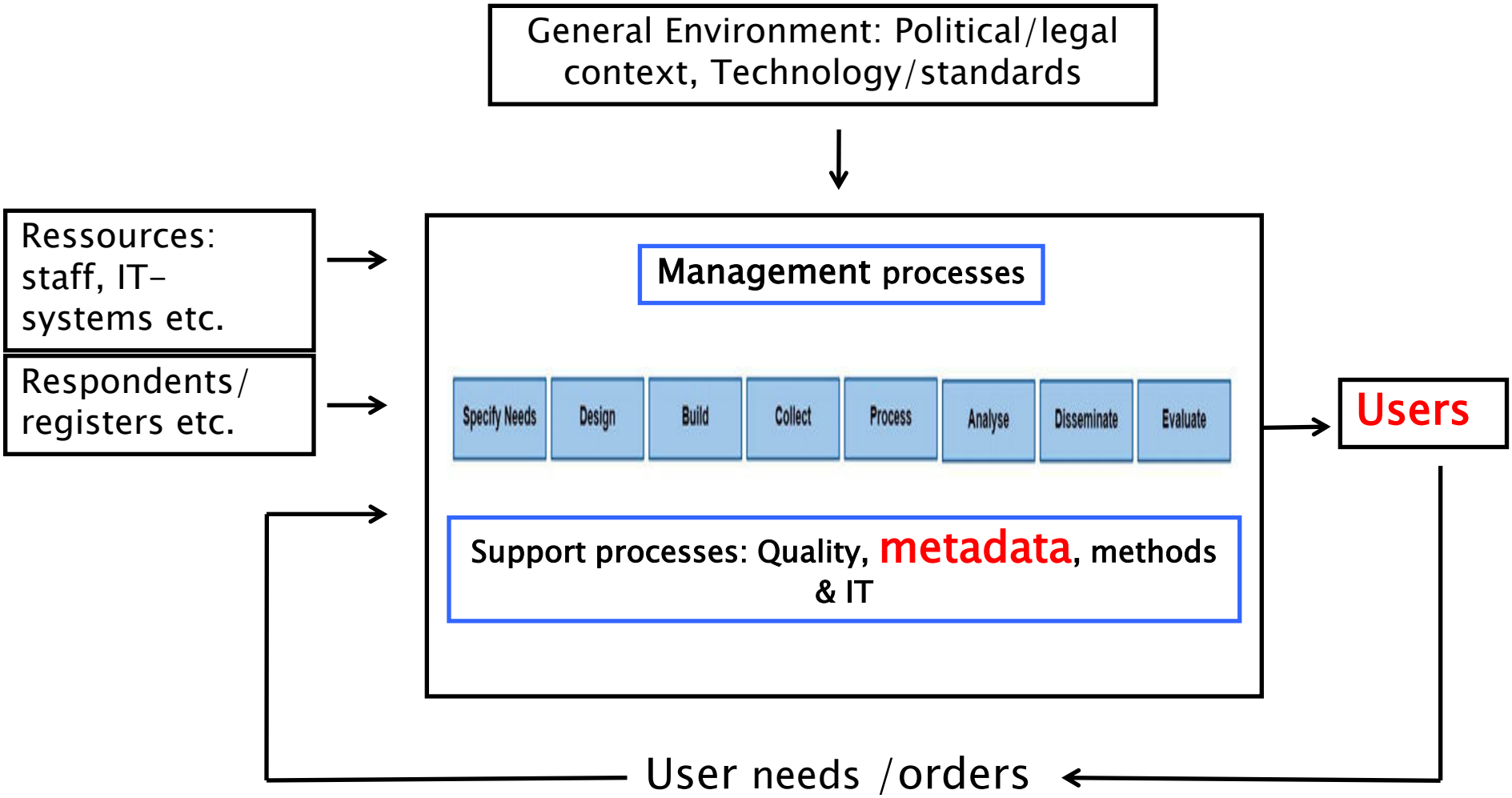
How metadata can help

- Support users to find the statistics most suitable for their purpose
- Users must be able to understand contents and quality, and thus fitness for their purpose
- Metadata must be very well structured and integrated with the data
- Easy to maneuver from one part to others

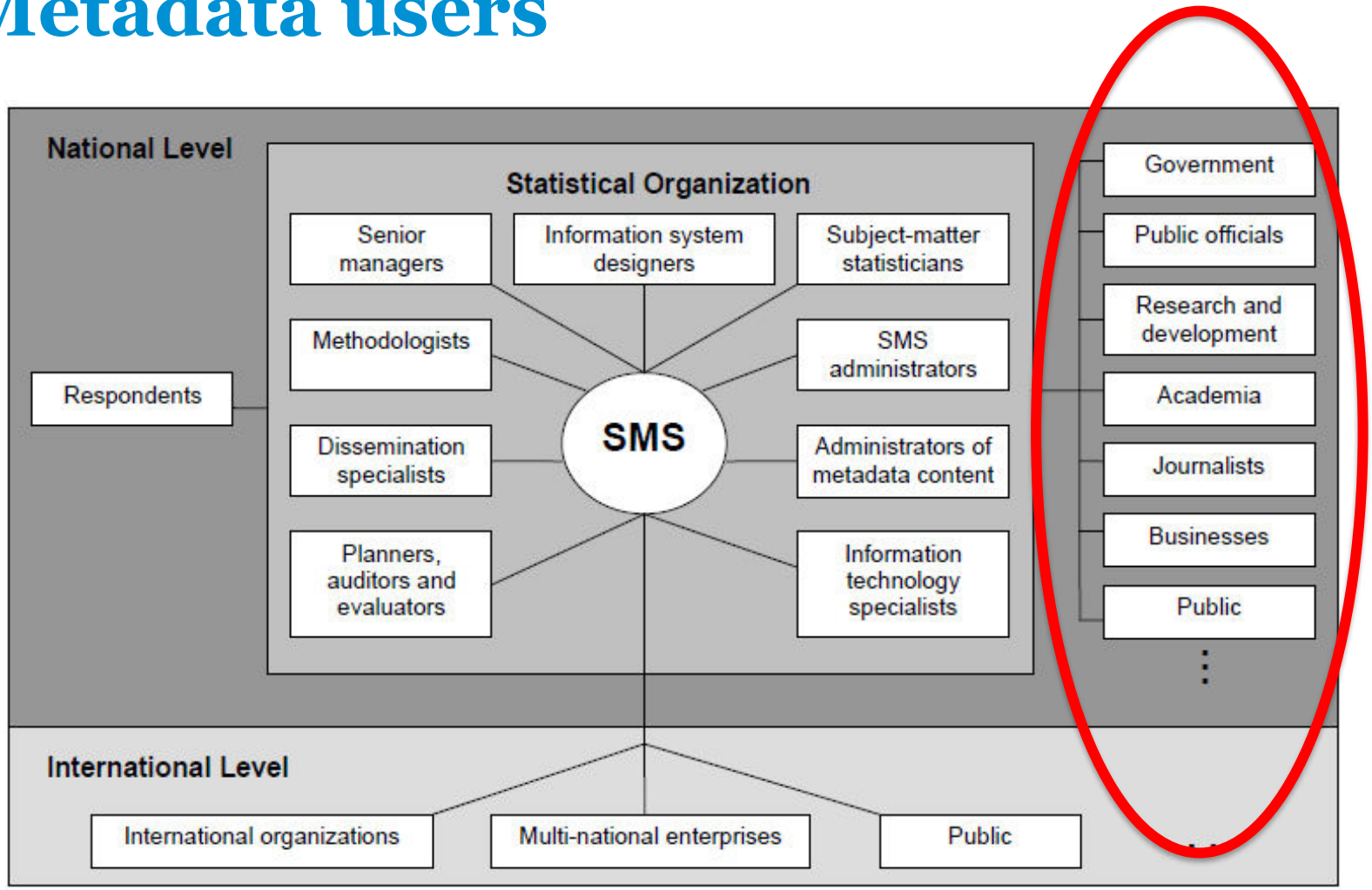
How can we serve users?
They should be helped to…

- find possibly useful statistics
  - Most users don't know what exists
  - Accessibility
  - Google Danmark or other search - needs metadata
- make sure if the statistics are suitable
- study exact characteristics
  - Contents
  - Quality
  - Methods
- compare with other statistics and metadata

The role of metadata



Metadata users



Metadata vital for end users of statistics

- Availability of statistical outputs
- Metadata related to the statistical outputs
  - Metadata on concepts and definitions, classifications, aggregations, statistical and evaluation methods, terminology, history, etc.
- Metadata about quality (e.g. explanatory notes, indicators)
- Access to microdata
- Time series
- Updating procedures
- Statistical revisions
- Responsibility for individual statistical outputs

Understanding statistics and metadata

- Language and Terminology
- Concepts versus data
- Populations and units
- Attributes
- Stocks versus flows / events

Statistics are not all in one place, nor completely coherent.

- Many organisations in each country produce official statistics
- Difficult to get an overview of what exists and how are the connections and differences between the concepts measured
- If you need data from several countries – which is increasingly needed – it becomes even worse

Where to go?

User

Coordinated metadata

- Within each data provider
- Between several data providers
- In a country
- Several countries
- Eventually across the world

Portals and standards

- Organisation-wide, nation-wide, Nordic, world-wide
- Metadata standards are required.
- Standards must be so intuitive that users, advanced as well as simple, can benefit.
- Must be easy to explain to producers of statistics
- Easy to implement in the production processes.

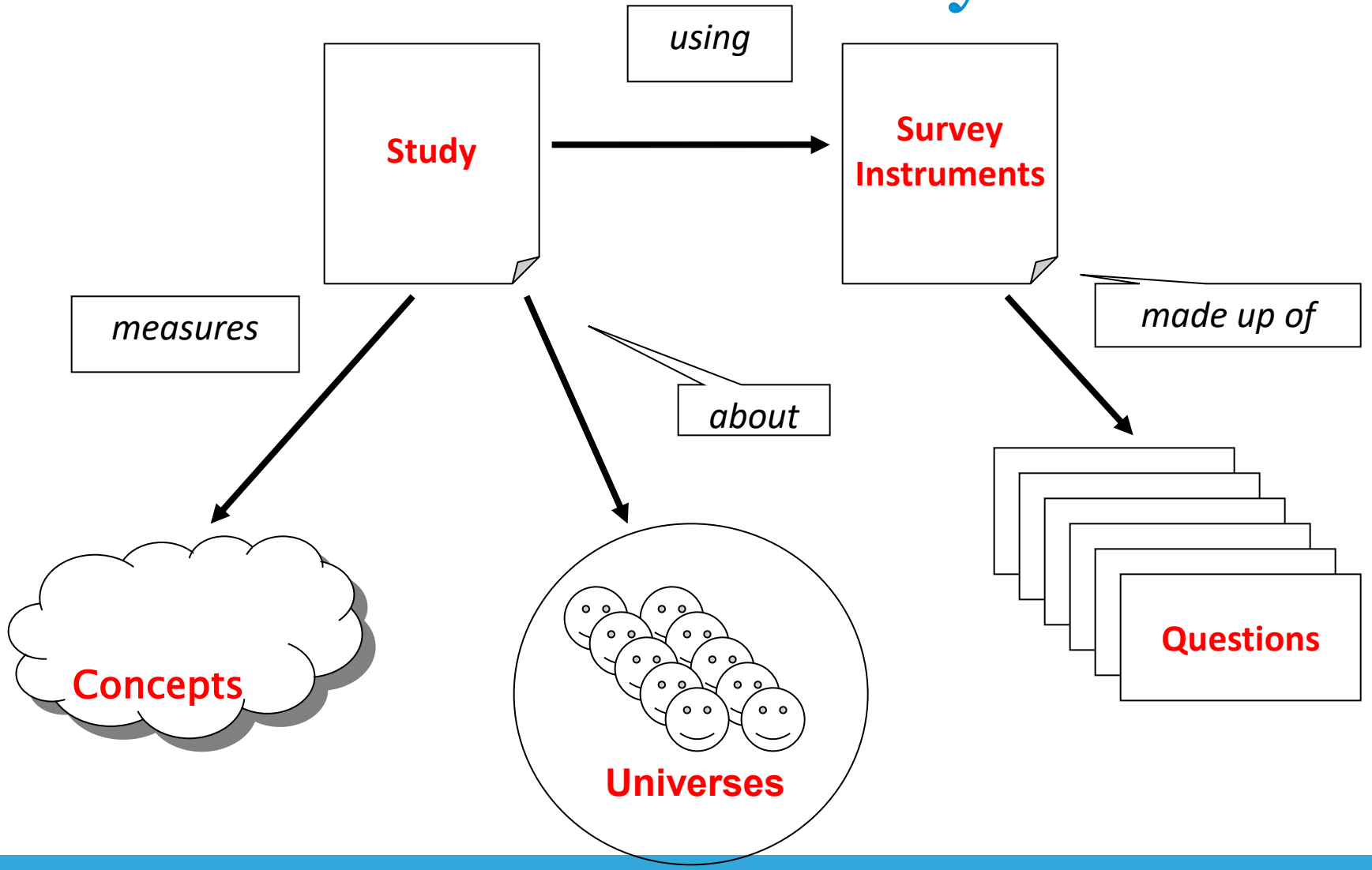
DDI

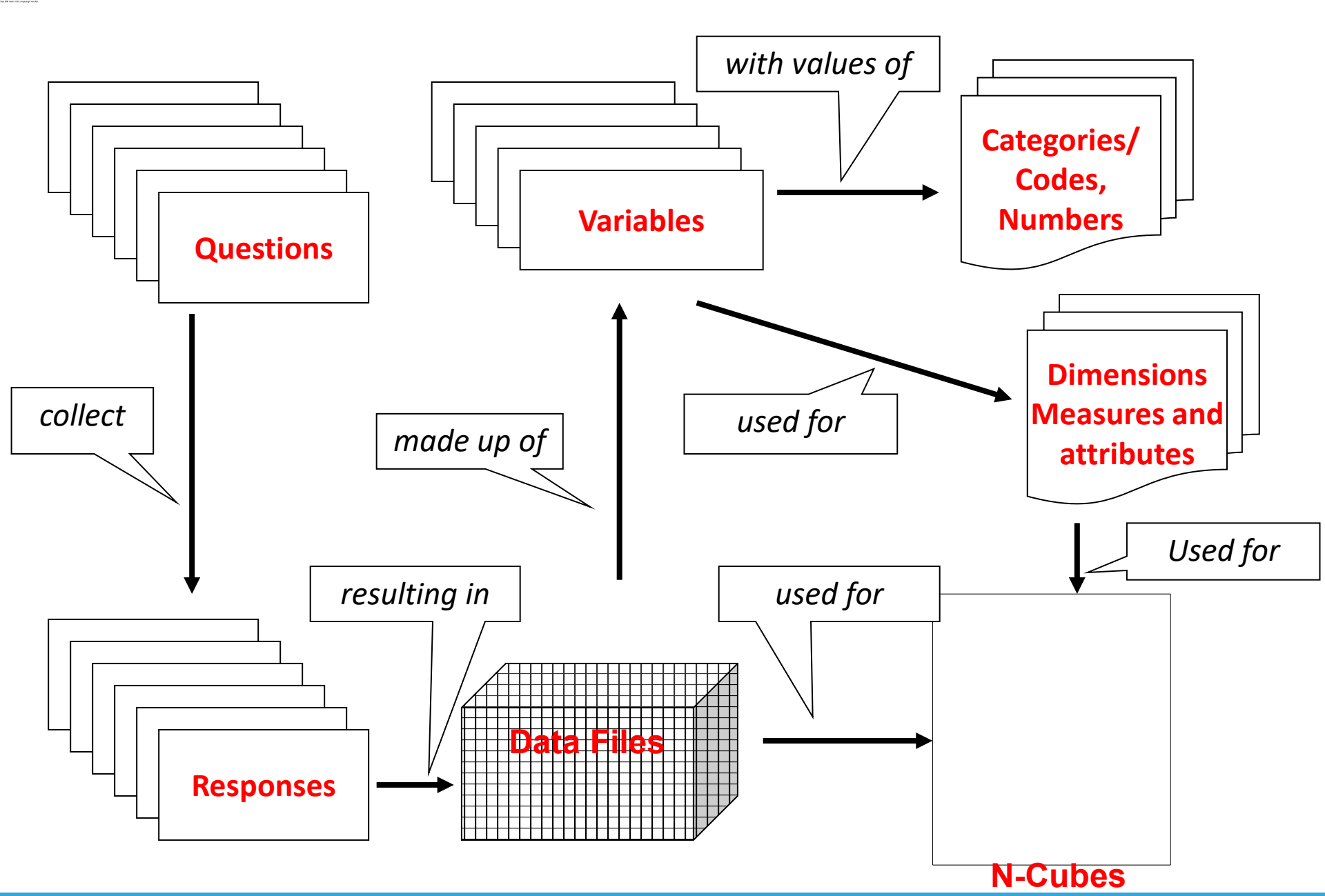
+ other standards specifying detailed contents

The chosen standards

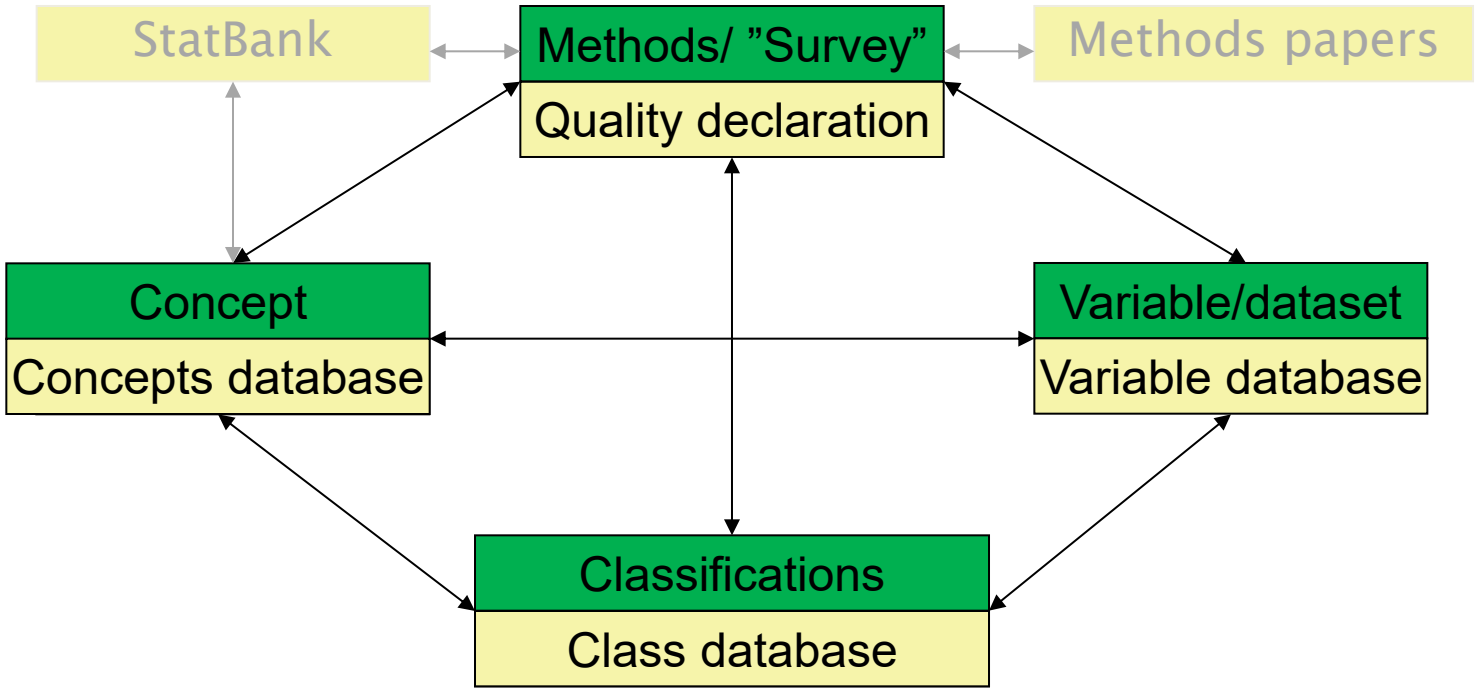
- DDI to achieve:
  - the right structure of metadata concepts,
  - links between metadata terms & concepts,
  - connection to business processes
  - independence of IT solution
- SDMX & SIMS to achieve:
  - Total coverage of metadata items
  - Inter-operability with other metadata systems
  - Flexibility in presentation
  - Easy exchange with Eurostat and others
- GSBPM to achieve:
  - Metadata to be produced in the right processes
  - Metadata guidelines integrated with other guidelines

Statistics and DDI in a hurry




Questions
Variables
with values of
Categories/
Codes,
Numbers
collect
made up of
used for
Dimensions
Measures and
attributes
Used for
resulting in
used for
Responses
Data Files
N-Cubes

Classical metadata elements: ”The Diamond model”




Dissemination aspects

Different users have different needs

- Pupils need comprehensible descriptions
- Analysts need for change management
- Researchers need specific and detailed information - also the historical

Metadata must provide a perceived added value for end-users

- Only one in four uses quality declarations

- For those who do not use quality declarations, two out of three indicates, they do not know what it is

Who uses documentation

• Municipalities	50 pct.
• Media and press	0 pct.
• Educational institutions	70 pct.
• Private companies	50 pct.
• Ministries and organizations	100 pct.

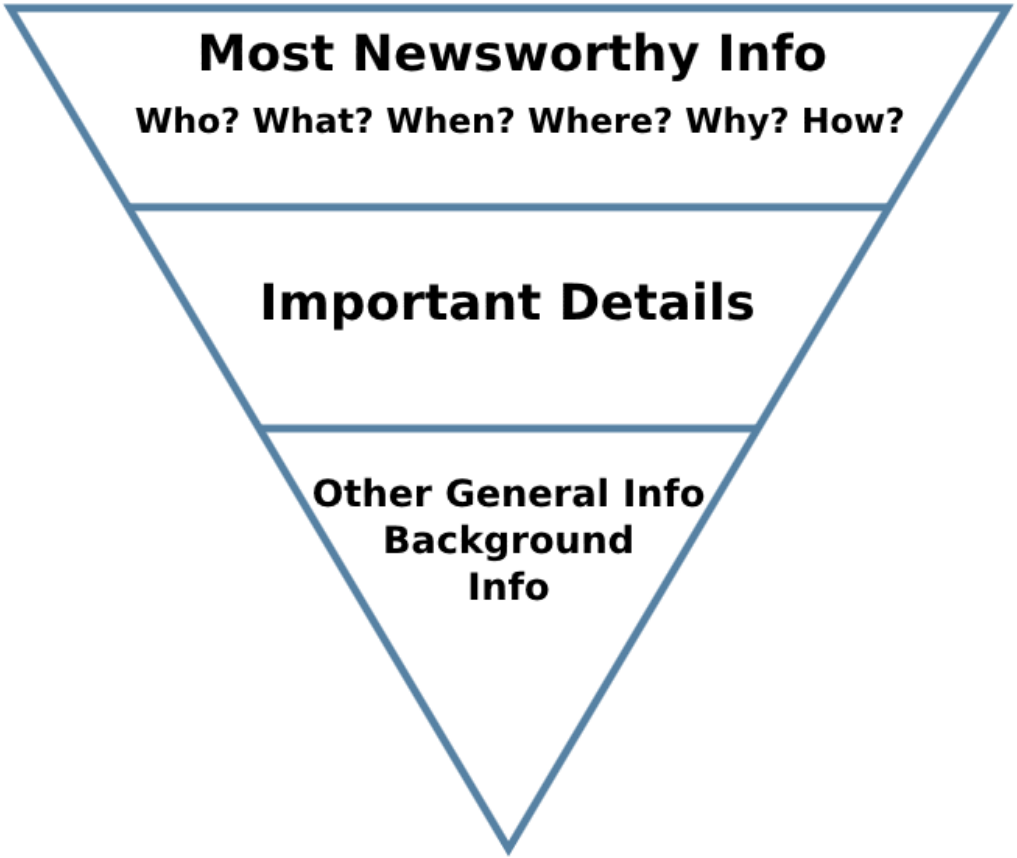
Who says that we need to improve

- Municipalities 50 pct. !
- Media and press 0 pct.
- Educational institutions 70 pct. !
- Private companies 50 pct. !
- Ministries and organizations 100 pct. !

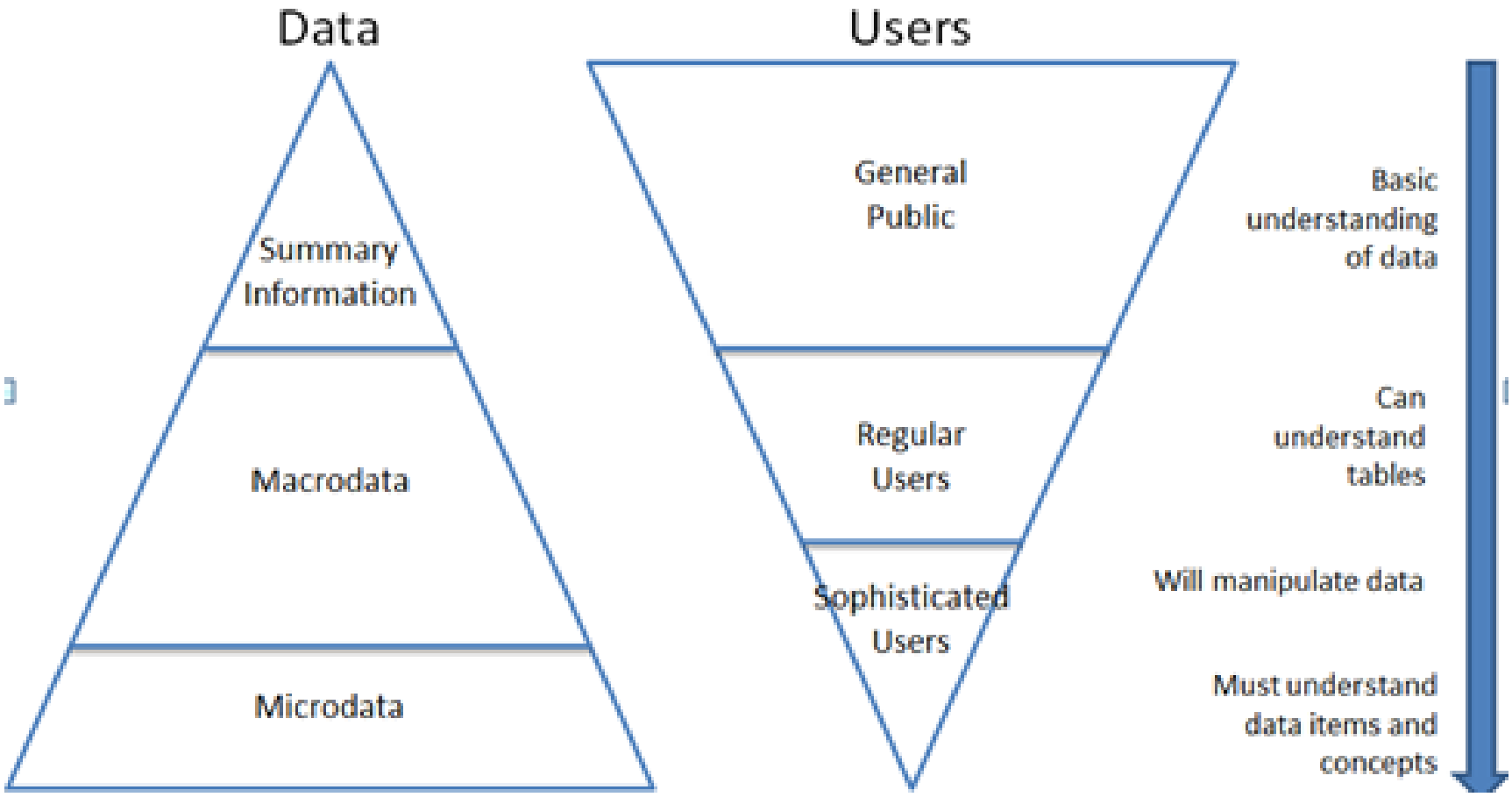
How do we communicate metadata

- Terminology
  - Basic DDI concepts
- Portals
- Dissemination products
  - Statbank
  - Publications
  - Communication
- Search Box

Inverted news pyramid



Different data needs



User types

- User behavior (eg Farmer, Miner and Tourist)
- Sector (eg ministries, authorities, media, academia)
- Knowledge e.g. of statistics (experts or new users)

”Words don’t come easy” *(F.R.David)*

- Even professionals may have difficulty with the understanding of our concepts and definitions - I wonder how our users understand?
- How do we ensure that the language understood as intended - of different target groups?

“Lost in Translation”

- Can we get something from search engines on the web?
- Can specialists from our Information Services help from the queries received?

"I only believe in statistics that I doctored myself" *(W. Churchill)*

- Explanations, definitions and special considerations are essential for the understanding of statistical data
- The quality of the metadata is often as important (or more) as the quality of the statistical data
- Communication, perception and understanding are essential quality dimensions
- Metadata may be an integral part of the statistical data regardless of how those are presented

One bite at a time

One bite at a time

Sometimes perhaps two..