



# **The Statistical Data Warehouse**

**Lars Thygesen**

Septembe 2007



# Outline

- Centralised vs. decentralised statistical system
- The OECD Statistical Information System
- Metadata management
- Lessons learnt - Do's and don'ts



# Official Statistics – raison d'être

- because all good decisions are based on facts
- the basic figures on society – a common reference point
- foundation for **democracy** and **economy**
- a lighthouse in the information jungle
- independence

# Centralised vs. decentralised

## - good or bad ?

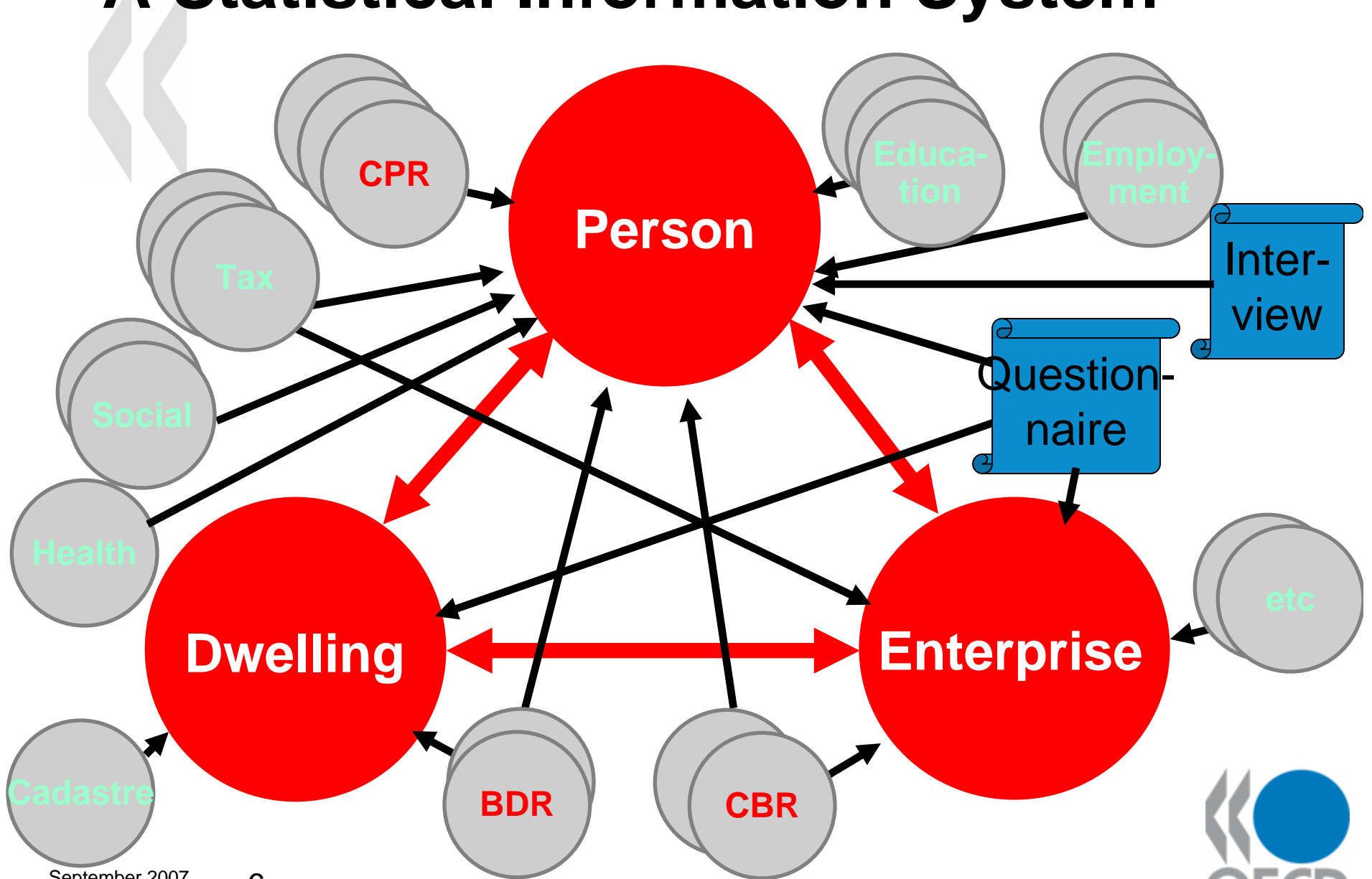
- knowledge and sensitivity to needs of users and decision makers
- independence
- credibility - statistical services integrated with policy making organisation
- statistical quality, mainly coherence, usability and findability



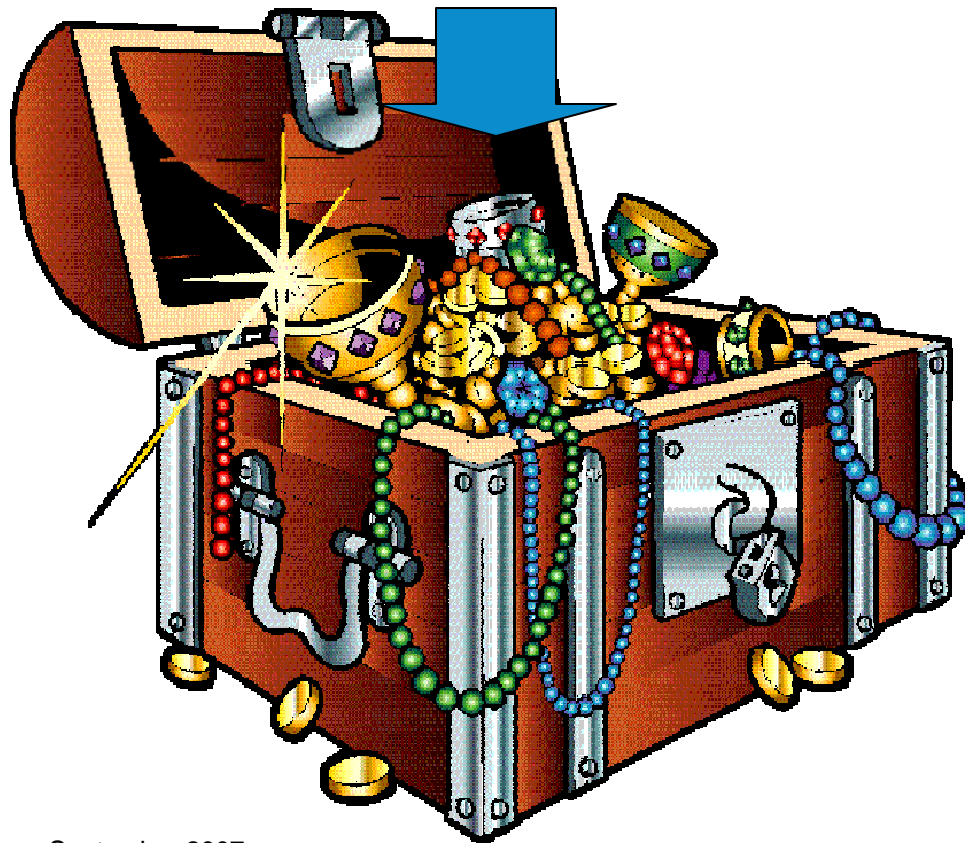
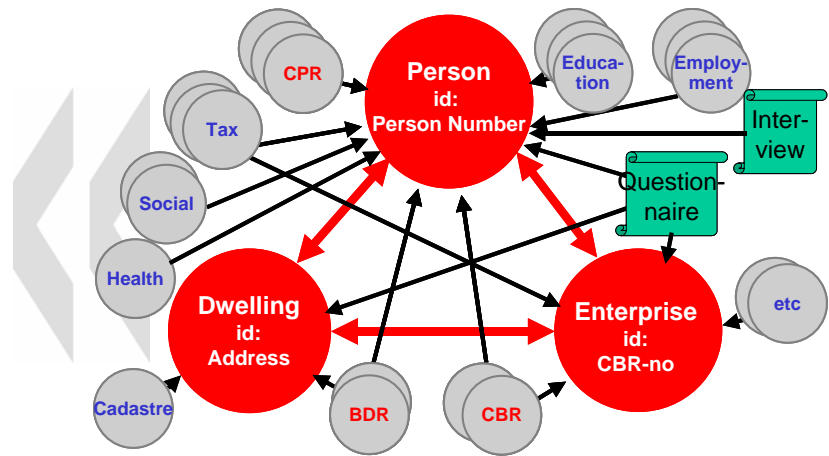
# Internal organisation

- stove-pipes or independent production lines
- individual and different dissemination systems or channels
- The Data Warehouse

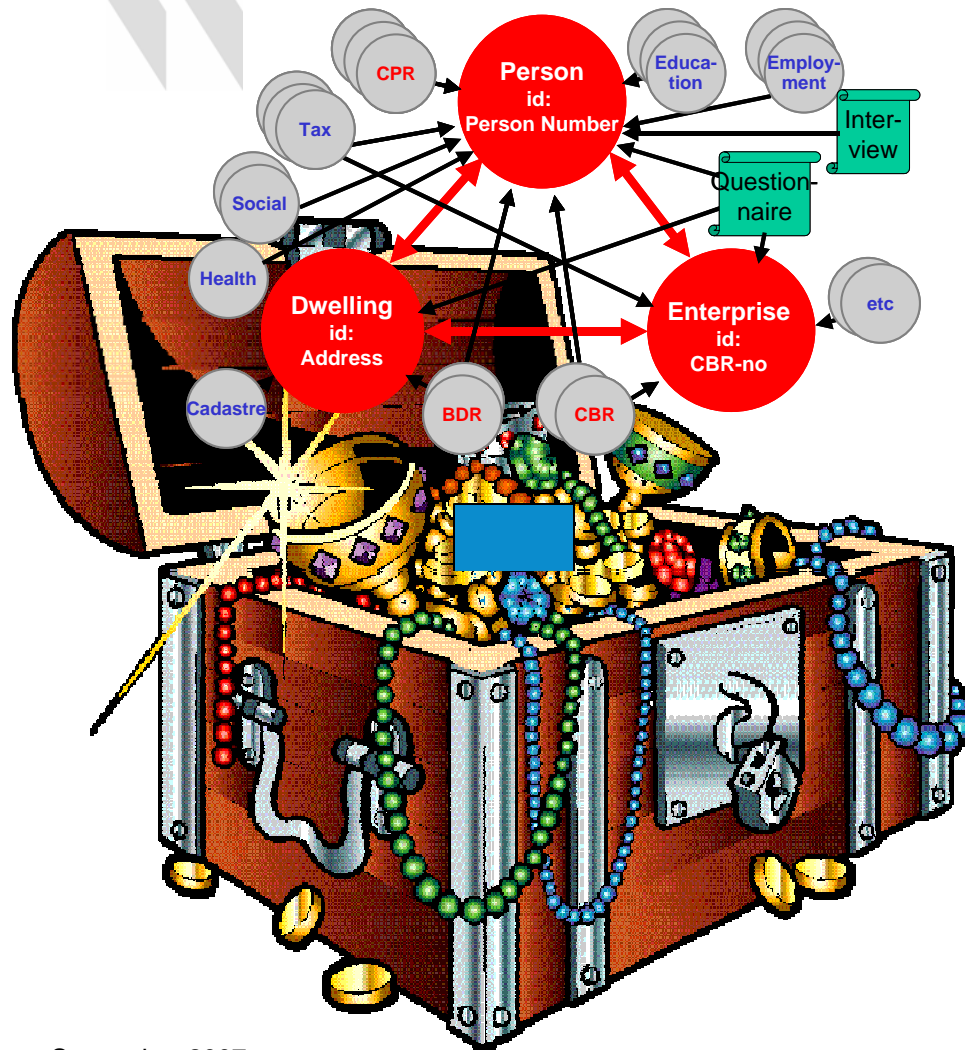
# A Statistical Information System



# A treasure!



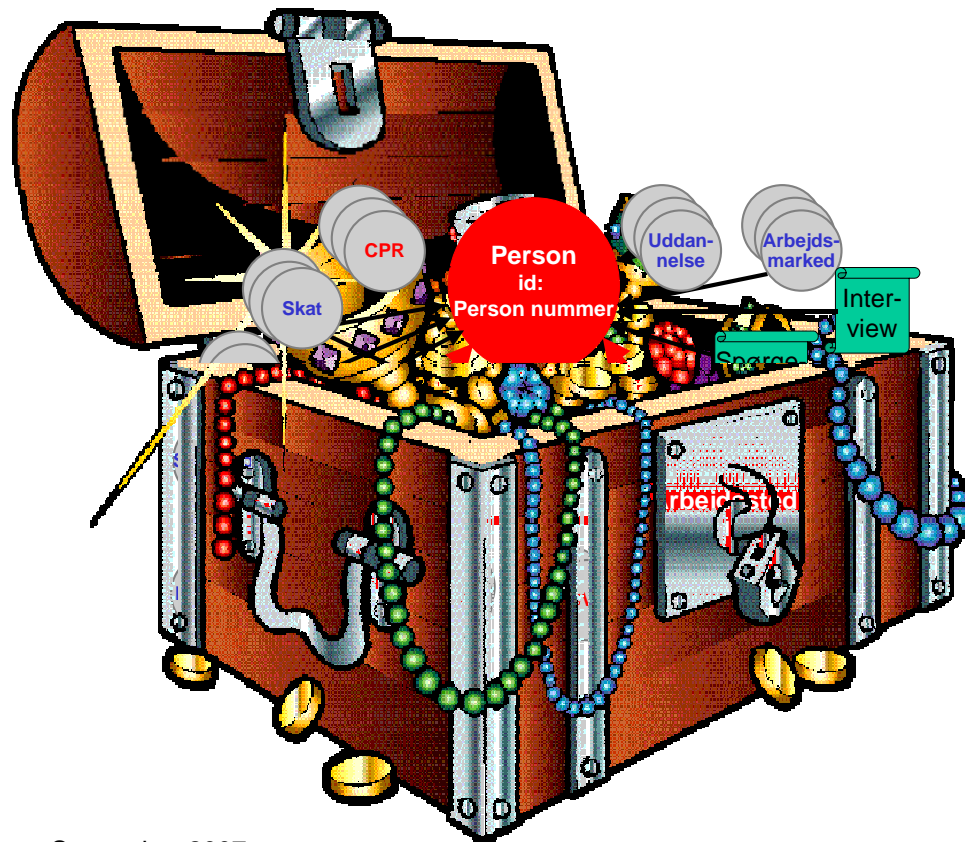
# A treasure!





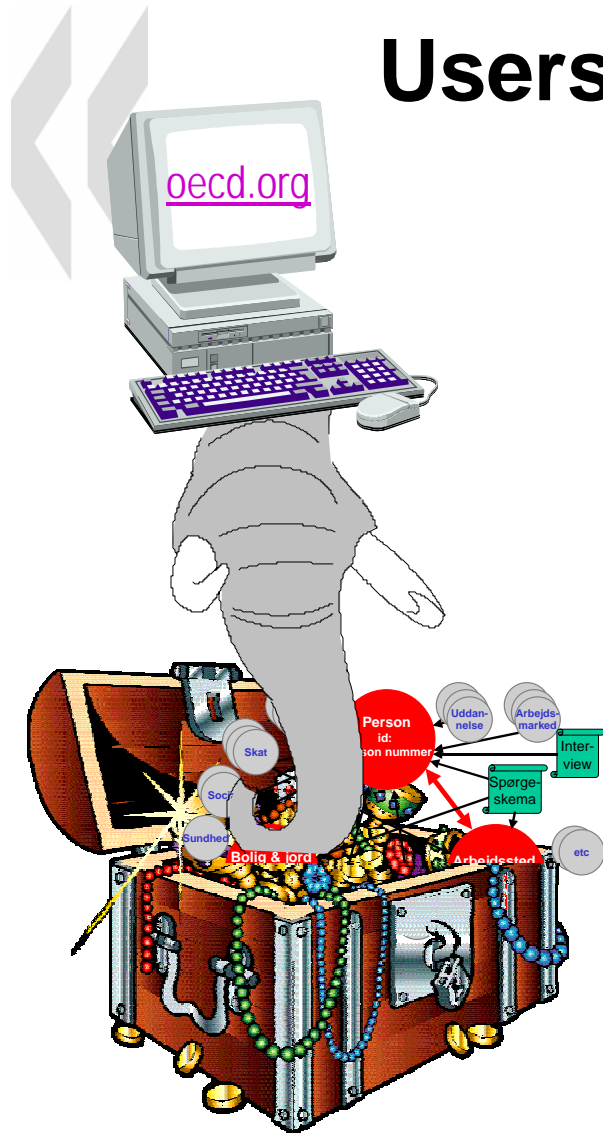
# A treasure!

- All domains
- Same method
- Good metadata
- Coherence
- Combining across

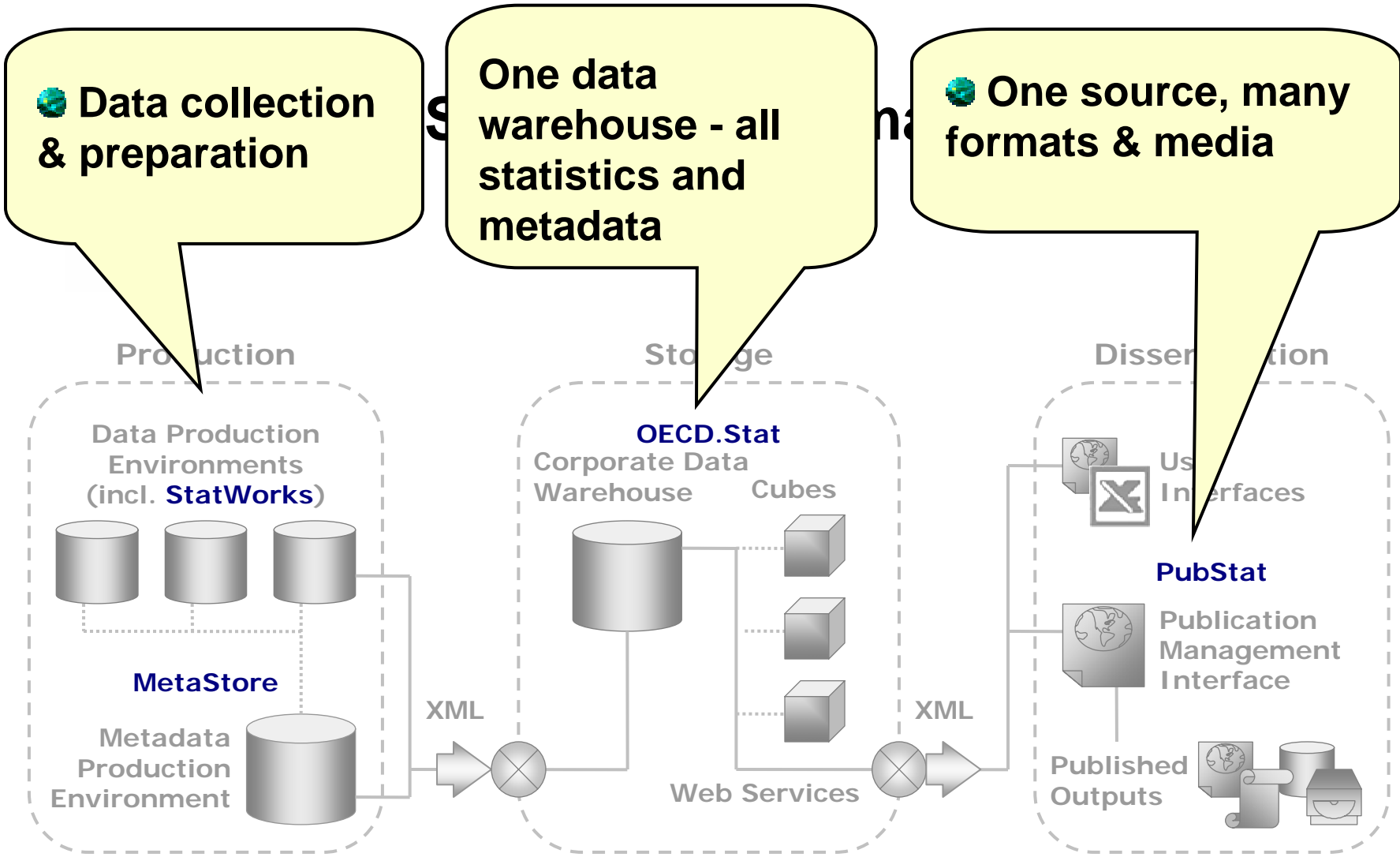


<http://stats.oecd.org/wbos>

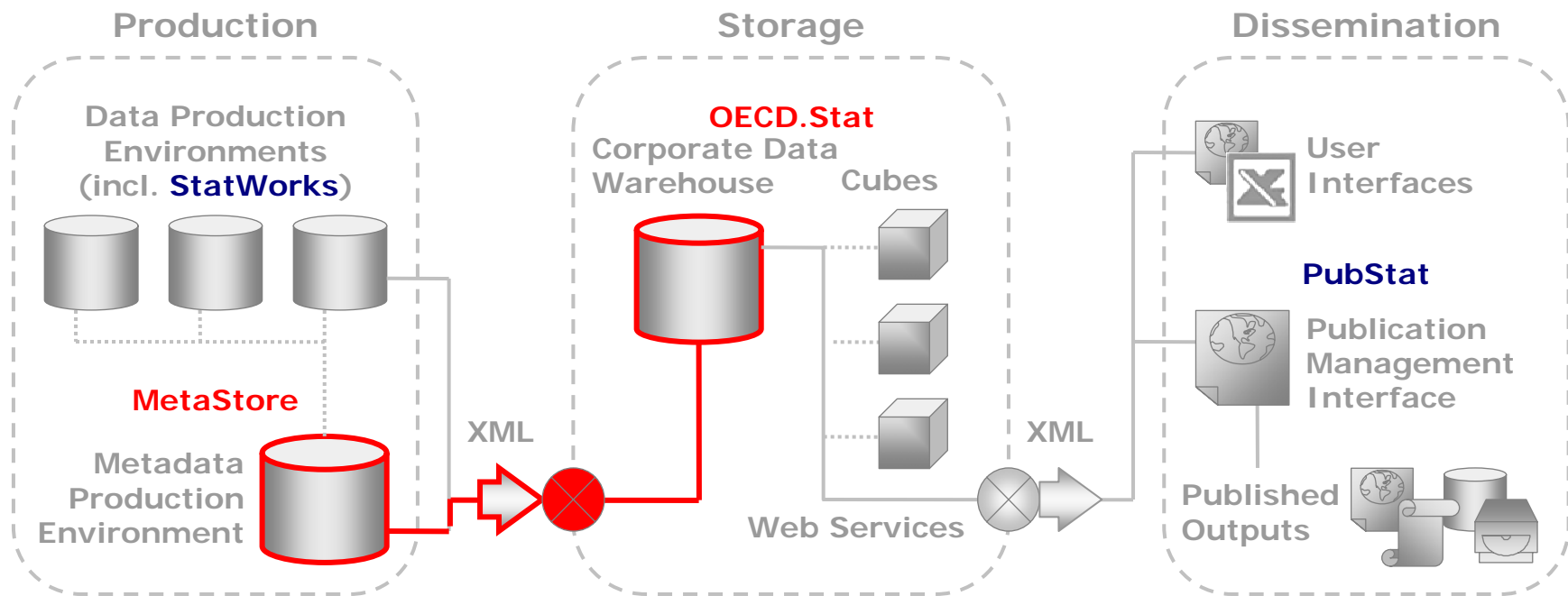
# Users want to get quality data



- Billions of figures
- Fully updated
- Flexible query
  - by theme
  - by full-text search
  - by stored query
  - links to queries
- Other access methods
  - from FAME & SAS
  - SDMX web service

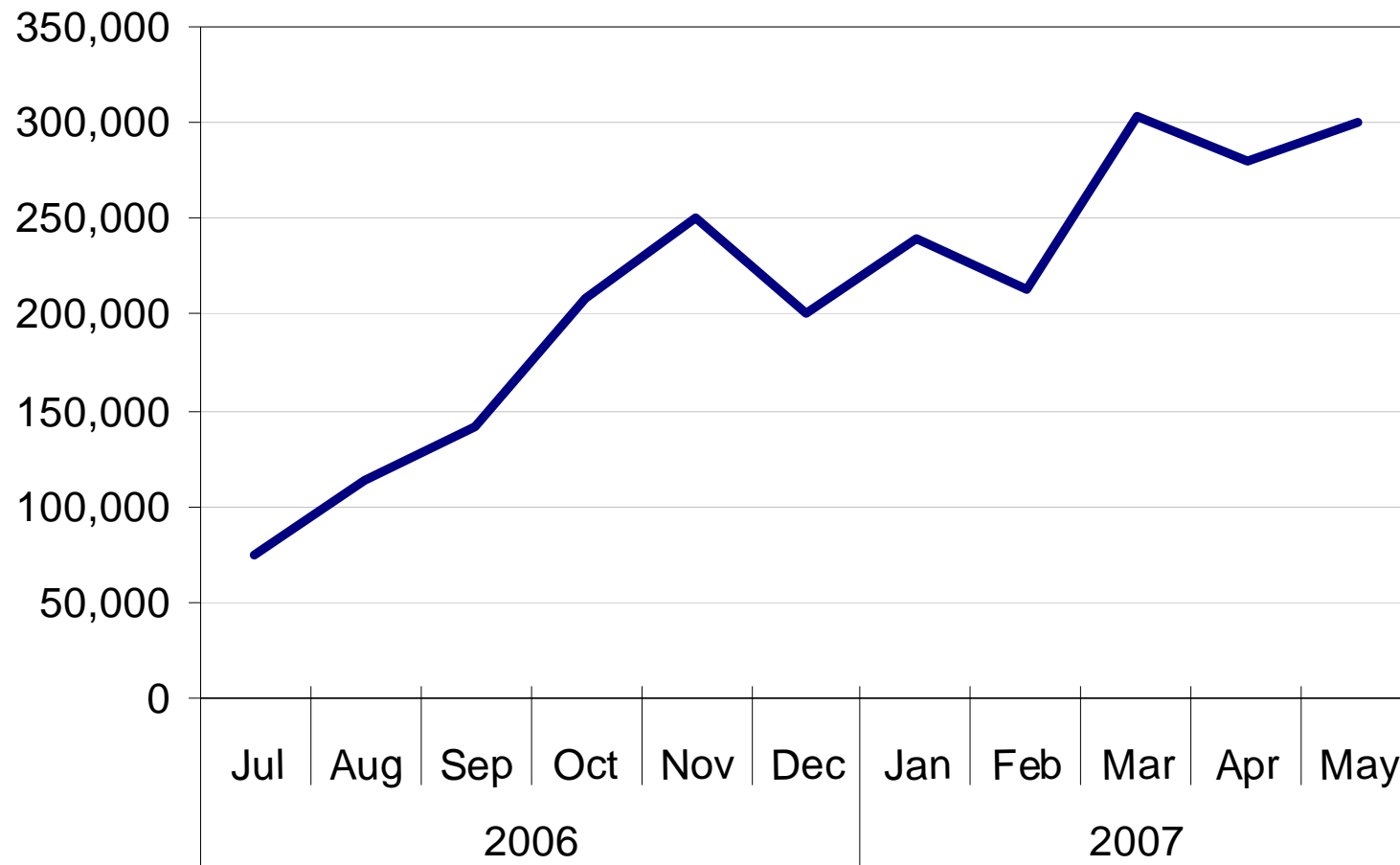


# The OECD Statistical Information System





# External access to OECD.Stat





# The data model

- Multidimensional tables (cubes)
  - each owned by one database administrator
- with metadata attached everywhere
- common dimensions
  - country
  - frequency
  - time
  - age
  - sex
  - ...
- Arranged in themes and subthemes



# Metadata principles

- 42 subcategories or metadata items
- how the metadata should be presented
- reuse of metadata : promote standardisation and coherence across domains
- supported by MetaStore




## Some lessons learnt

- Engage users of the information
- Study best practice and steal as much as possible.
- Top management engagement is essential
- Spread support and understanding
- Reallocation of resources – a problem
- Offer tempting tools
- Offer central assistance



## - and more do's and don'ts

- “Stalinist” measures necessary
- Set up governance rules defining responsibilities
- Set up data content standards,
- Set up metadata standards
- Start migration with the managers having a positive attitude
- Database managers keep ownership
- Hard and continued work



“there is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things. Because the innovator has for enemies all those who have done well under the old conditions, and lukewarm defenders in those who may do well under the new.

This coolness arises partly from fear of the opponents, who have the laws on their side, and partly from the incredulity of men, who do not readily believe in new things until they have had a long experience of them.

Thus it happens that whenever those who are hostile have the opportunity to attack they do it like partisans, whilst the others defend lukewarmly, in such wise that the prince is endangered along with them.”

***Niccolo Machiavelli, 1515***



Tour  
Europe  
OECD

The End