



Big Data Benefits and Challenges

Big Data and the
International
statistical
community,
Beijing, 28 Oct 2014

Albrecht Wirthmann
@ec.europa.eu

Methodology

- New approaches relate to data mining, text analysis, machine learning
- Analysis of network data
- Summarize + visualise large and complex data
- Replicability and reproducibility of results
- Correlation is not causation
- Big Data Big Problems?
 - **Increased noise in the data**
 - **Hypothesis testing**
 - **Model articulation, model testing and model selection**
 - **Meta-analysis across heterogeneous data**
 - **Identifying false positives and false negatives**
 - **Validity of generalisation (overfitting of models)**

Quality

- Start from users' needs
- Revisit conceptualisation of information needs
- Analyse according to different quality dimensions
- Take multitude of data sources to observe differences
- Reliability of data source
- De-mystify big data

Legal Frameworks / Ethical issues

- Data protection and privacy law
- New developments might require new approaches
 - **Enable/empower individuals to manage their data**
- Enablers for access to data (Agreements, legislation)
- Review of statistical domain legislation
- Contradicting regulations
- Balance between legislation and non-legislative agreements (code of practice)
- Secure data processing and assure anonymised outputs

Commonalities

- Data sources very similar
- Opportunities for standardisation and harmonisation
 - **Methodology**
 - **Algorithms and software (open)**
 - **Statistical products**
 - **Quality assessment**
 - **Skills**
 - **Ethical principles**
 - **Recommendations for legal approach**
 - **IT infrastructures**
 - Common use
 - Best practice and guidelines for planning and implementation
 - **Communication of statistics**
- Analysis of country specific situations necessary