Panel Session

Common issues on benefits and challenges of Big Data sources

Emanuele Baldacci

Italian National Institute of Statistics (Istat)
Head, Department for Integration, Quality, Research and Production Networks Development (DIQR)
Outline

- Common methodological issues and quality concerns
- Common privacy issues: legal frameworks, ethical guidelines and trusted technology solutions to safeguard privacy
- Common issues on partnerships: data acquisition, division of responsibilities
- Common IT issues: cloud computing
Common methodological issues and quality concerns

- Data quality and suitability of statistical methods (Accuracy/Timeliness)
- Framework for data quality: bias, coverage, provenance
- Data linkage and profiling methods for Big Data
- From inference to data discovery: new framework for sound statistics
- Small area estimators for geospatial data
- Machine learning (using surveys as training set)
- Big Data analytics tools (visualisation)
Common privacy issues: legal frameworks, ethical guidelines and trusted technology solutions to safeguard privacy

- Access and use of Big Data
- Policies and directives about data management and protection
- Personal data storage and privacy by design
- Privacy-preserving algorithms for access and dissemination
- Reuse statistical framework for data access, process and dissemination
- Innovative/Interactive way to inform people on the use of personal information from Big Data
- Managing public trust and acceptance of data reuse and its link to other sources
Common issues on partnerships: data acquisition, division of responsibilities

- Regulators (Privacy Authority, Digital Agenda)
- Private sector (data providers, IT industry)
- Academia and research Institutes (analysis, new skills)
- Users (researchers, data scientists, data journalists, common users)
- Official Statistics community (NSIs, ESS, UNECE HLG)

Emanuele Baldacci. Beijing, 30 October 2014
Common IT issues: cloud computing

- Accessing and storing unstructured data (control mechanisms)
- Distributed processing and computing
- Multi-media data processing
- Web scraping techniques (text and data mining algorithms in the estimation phase)
- Technological platforms for Map-Reduce tasks (Map-Reduce algorithms - Hadoop)
- Semantic web techniques
- Linked-Open Data (RDF with statistical ontologies)
Thank you for your attention

Contacts:

baldacci@istat.it

www.istat.it

Emanuele Baldacci. Beijing, 30 October 2014