



United Nations Economic Commission for Europe  
Statistical Division

## Workshop on the Modernisation of Statistical Production and Services

November 19-20, 2014

### The Role of Big Data in the Modernisation of Statistical Production and Services



# Results of the Task Team Privacy

Peter Struijs

Statistics Netherlands



# Members of the Task Team

- Shane Weir, chair, and James Chipperfield, Australia 
- Pascal Jacques, Eurostat 
- Jörg Drechsler, Germany 
- Josep Domingo-Ferrer, Italy 
- Peter Struijs, Netherlands 
- Anna Nowicka, Poland 
- Vicenc Torra, Spain 
- Luis Gonzalez and Shaswat Sapkota, UNSD 
- Monika Jingchen Hu, USA 



# The tasks

- To give an overview of **existing tools** for risk management in view of privacy issues
- To describe how **risk of identification** relates to Big Data characteristics
- To draft **recommendations** for NSOs on the management of privacy risks related to Big Data



# Existing tools for privacy risk management

- Disclosure risk for **estimates** versus **micro-data** access
- **Risk of attempt** at disclosure versus **risk of success** of attempt
- **Micro-data access**: strategies for micro-data dissemination, onsite analysis and remote access
- **Databases**: distinguish between owner privacy, respondent privacy and user privacy
- **Managing privacy**: trade-off between disclosure risk and utility



# Big Data characteristics and privacy risk

- **Big Data characteristics:** *volume, velocity, variety, veracity*
- **Also relevant to privacy:** *availability, aggregation, society awareness, flexibility, provider infrastructure, geographical differences*
- **Task Team looked at:** *GPS location data, tool assessment in light of characteristics, onsite analysis versus remote access, feasibility of re-identification experiments*



# Recommendations on information integration and governance

- Monitor **database activity**
- Apply best practices for **security of IT systems** and business practices:
  - separation of duties
  - separation of concerns (modular approach to functionality)
  - principle of least privilege (no more access rights than needed)
  - defence in depth (multiple security mechanisms/layers)
- Apply best practices of **security of transportation**
- Apply **data encryption**



# Recommendations on statistical disclosure limitation/control

- **Preserve confidentiality** by restricting data access and data release
- **Ensure access** to useful data
- **Balance data utility and disclosure risk.** Use not only traditional approaches, but also modern techniques such as data records simulation using statistical models



# Recommendations on managing risk to reputation

- Enforce **ethical principles in the supply chain**, including accountability and informed consent
- Establish strong **compliance control**
- **Monitor threats** to reputation
- Be **transparent towards stakeholders**, and organise a dialogue with the public
- Create a **crisis communication plan**





# Conclusions

- Existing tools are well-developed
- Privacy risk can be linked to Big Data characteristics
- Recommendations have been formulated on:
  - information integration and governance
  - statistical disclosure limitation/control
  - managing risk to reputation
- **But: not much experience yet with Big Data privacy issues**

