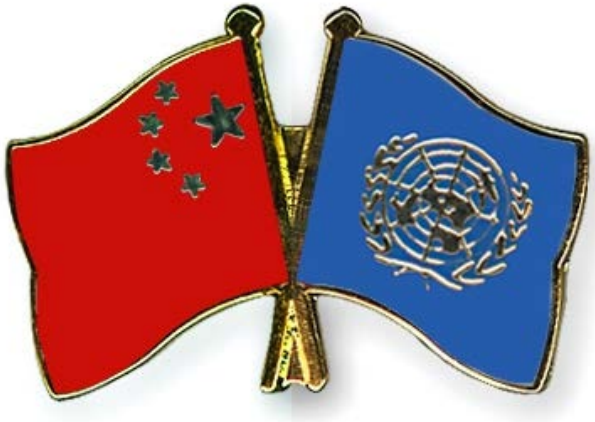
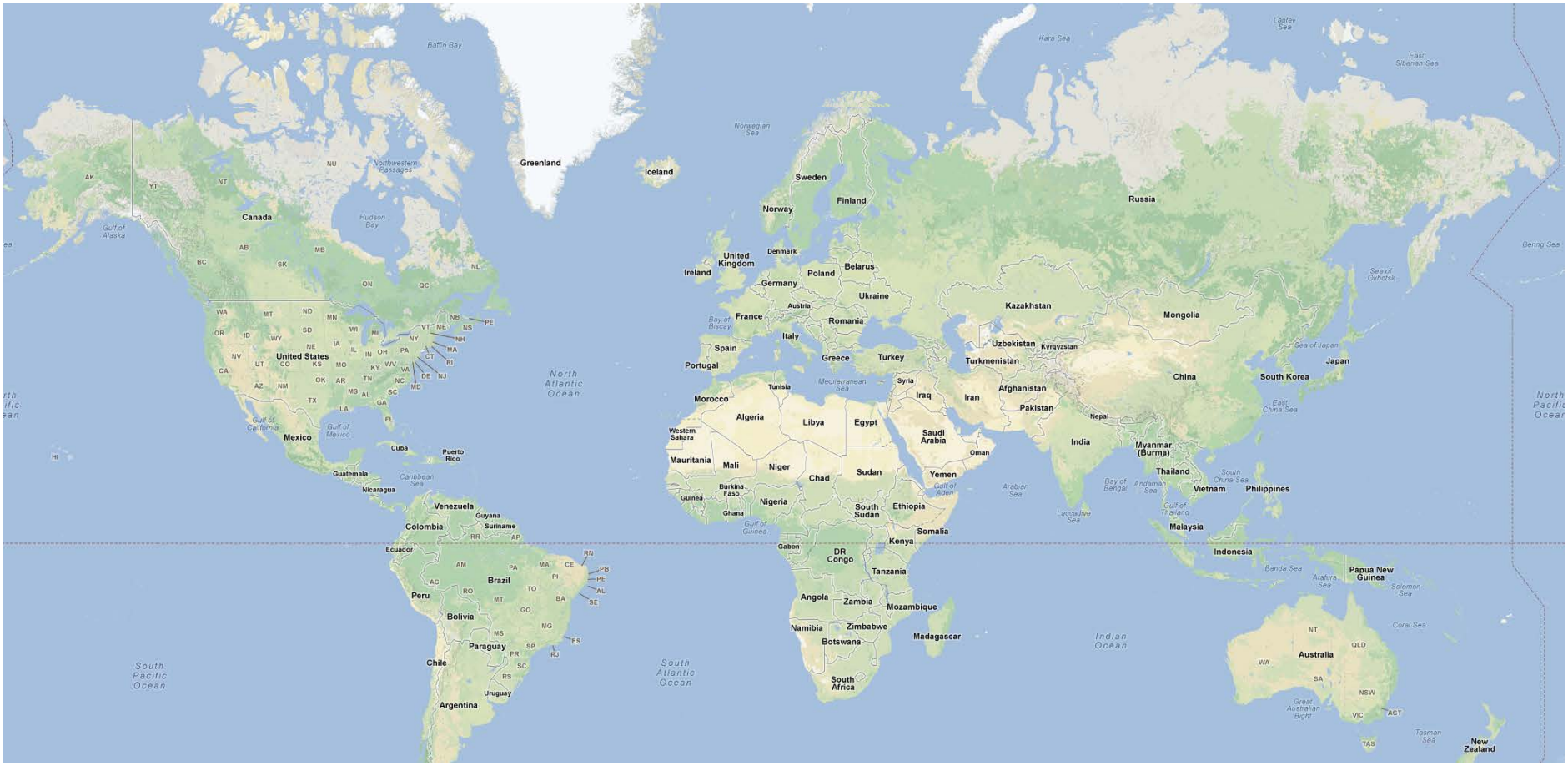
United Nations Economic Commission for Europe
Statistical Division

UNECE Big Data Activities



Steven Vale
UNECE
steven.vale@unece.org

Introducing UNECE Statistics



Introducing UNECE Statistics

Introducing the HLG

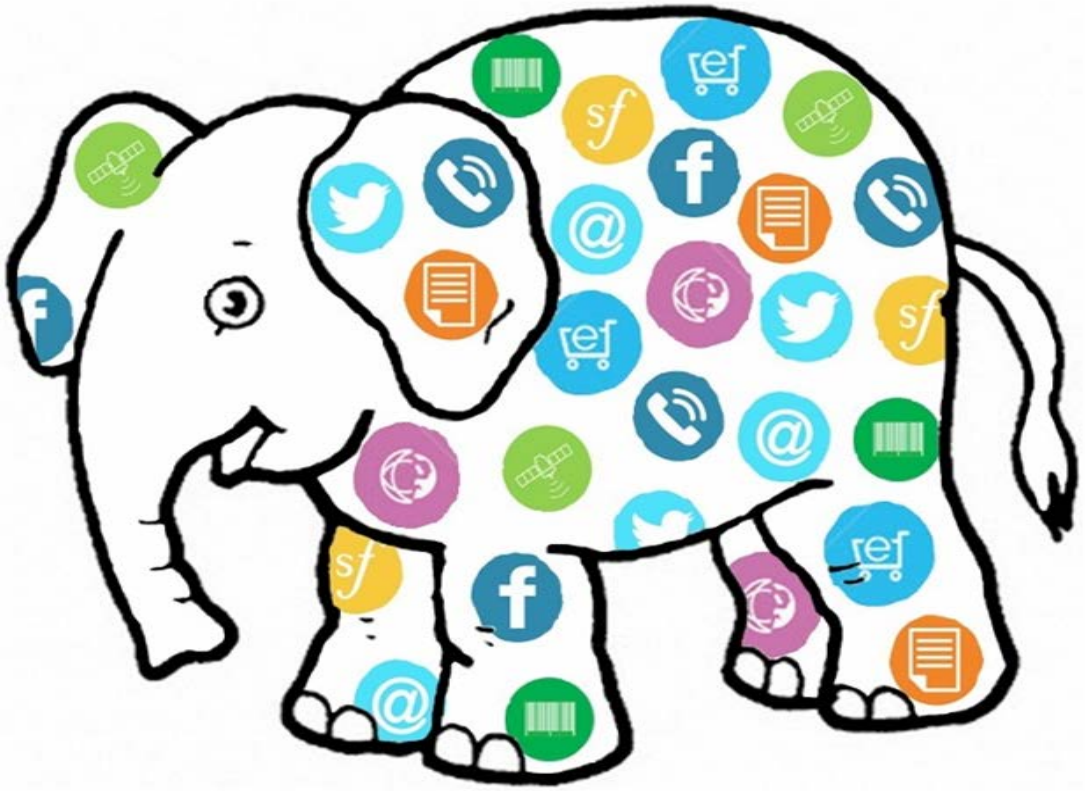
- High-level Group for the Modernisation of Statistical Production and Services
- Created by the Conference of European Statisticians in 2010
- Vision and strategy endorsed by CES in 2011/2012

HLG

Expert Groups

Projects

Big Data



In the last 2 years more information was created than in the whole of the rest of human history!

The story so far

- High-level Seminar on Modernization of Statistical Production and Services (St. Petersburg, 2012):
  - "Big data is an increasing challenge. The official statistical community needs to better understand the issues, and develop new methods, tools and ideas to make effective use of Big data sources"
- 2013 – Task Team of international experts
  - Paper: "What does Big data mean for official statistics?"
- 2014 – International collaboration project
  - "The Role of Big Data in the Modernisation of Statistical Production"

Virtual sprint

Consultation

Physical sprint

Task teams

Guidelines

UNECE

Virtual sprint

We are here

Consultation

Physical sprint

Task teams

Guidelines

Areas of work

- Key challenges identified:
  - Quality framework for Big Data
  - Privacy and data security
  - Partnerships with suppliers, processors and users
  - Methodology and IT for Big Data
  - Skills needed to use Big Data

Skills survey

- 107 responses
- 37% of respondents working with big data
  - 43% planning to start soon
- Important skills
  - IT: databases and Hadoop
  - Statistical: methodology and data mining
  - Other: problem solving, data governance, ethics
- **Training needed in all these areas!**

Methodological experiments

- Scanner data – price statistics
- Web scraping – job vacancies and price statistics
- Mobile phone data – mobility / tourism
- Smart meters – energy use, clustering
- Traffic loops – mobility
- Social media – sentiment, mobility

Next steps

- Discuss draft outputs at HLG Workshop
- Finalise and publish all outputs by end 2014

- Decide the future of the sandbox
- Possible extension in 2015 for additional experiments

- Coordinated activities with UNSD and Eurostat

More Information

- HLG Wiki:

www1.unece.org/stat/platform/display/hlgbas

- Big Data Wiki:

www1.unece.org/stat/platform/display/bigdata

- LinkedIn group:

“Business architecture in statistics”

You are welcome to join HLG activities!