Standards for Big Data in the Cloud

International Cloud Symposium
15/10/2013
Carola Carstens (Project Officer)
DG CONNECT, Unit G3 – Data Value Chain
European Commission
Outline

1) Data Value Chain Unit
2) Activities in the Area of Big Data
3) Discussion
**DVC Mission:** Foster commercial and social added value based on the intelligent use, management and re-use of data sources in Europe

**G3: Data Value Chain (DVC)**

G1: Converging Media and Content  
G2: Creativity  
G4: Inclusion, Skills and Youth  
G5: Administration and Finance
Data Value Chain Vision

• aims at extracting the maximum value from data
• by building on the intelligent use of multilingual data sources across Europe

This will lead to
• increased business intelligence and efficiency of private and public sectors
• world class applications and services
• new business opportunities involving SMEs
• new ways of tackling societal challenges
Questions around the Data Value Chain

- Who/what is producing data in the EU?
- How is data captured?
- How is data shared?
- How is data reused?
- What is the relevant maturity of data technologies and processes?
- How deep and diverse is the EU data supply chain?
- How can the EU offer support?
Data has Become a Resource

**Types**: Publicly funded data (like statistics, environmental and geographical data, meteorological data, business information, legal information), institutional and private data

**Use and reuse**: Combination of different types of data (e.g. geo, traffic and tourism; business and open data)

**Competitive advantage** is to offer the right data to the right people at the right time
Where data comes from

Humans (data volumes growing slowly)
- Professional
- Personal/social

Machines (data volumes growing fast)
- Earth observation (satellite/environment monitoring)
- Scientific equipment (DNA sequencers, telescopes, particle accelerators, ...)
- Sensors (Internet of Things, RFID, Smart Cities)
Big Data

• “big data” is when the size of the data itself becomes part of the problem

• “big data” is data that becomes large enough that it cannot be processed using conventional methods
Challenges

- Growing data quantity (big data)
- Data-friendly legal and policy environment
- Skills in the area of data
- Multilingualism
- Multimodal information
- Integration of business data with open data
- Unstructured data
- Heterogeneity of data and data sources
- Complexity
- Data quality
- Interoperability
Big Data and Cloud Environments

- Big data software frameworks available in cloud
- Big data analytics applications available in cloud
- (Big) data sets in the cloud are easily accessible for applications
- Having data in one place reduces bandwidth costs and time
- Having data in the cloud increases the potential of data re-usage and establishing a data market

→ Cloud as an important infrastructure for big data
Activities of the Data Value Chain Unit

Research and Innovation (R & I):

• Building a European data ecosystem
• Supporting R & I fostering the intelligent use, management and reuse of complex and large amount of data for better decision making, efficiency and knowledge management

Framework conditions:

• Promoting Open Data policies across the EU (Revised Directive on the Re-use of Public Sector Information, Open Data Portals)
• Supporting data skills
• Enhancing interoperability, standardisation
Big Data Projects I

SemaGrow

- [http://www.semagrow.eu/](http://www.semagrow.eu/)
- Develops novel algorithms for querying distributed triple stores
- Develops scalable and robust semantic indexing algorithms for extremely large datasets
- Based on W3C recommendations such as POWDER, SPARQL
Big Data Projects II

LDBC

• Linked Data Benchmarking Council
• [http://ldbc.eu/](http://ldbc.eu/)
• Develops scalability benchmarks for RDF and graph database technologies
• Involvement of industry (RDF and Graph database vendors)
Approach to Standards

- Promote the usage of standards in Research & Innovation projects
- Promote the usage of standards for Open Data
  - in European Open Data portals
  - via Thematic Network on Open Data standards (under negotiation)
- Support benchmarking activities (standardised measures)
Questions to the Audience

• At which level would you see the need for increased involvement of the EC in these processes?
  • assessing standards landscape
  • coordinating activities between different stakeholders
  • benchmarking activities
  • regulations
  • ...

• Which **types of standards** need to be developed (syntactic level, semantic level, processes for bringing big data into the cloud,...)?

• Which **industries** are likely to operate in the cloud and to see a big benefit in agreeing on standards (as a basis for using specialised analytics services)?
Further information

Data Value Chain Unit (DG CONNECT G3)
• http://cordis.europa.eu/info-management/
• http://cordis.europa.eu/fp7/ict/language-technologies/
• Email: cnect-g3@ec.europa.eu

ICT 2013
06.-08.11.2013 in Vilnius

European Data Forum 2014
19.-20.03.2014 in Athens
http://2014.data-forum.eu/
Follow-up

Join the LinkedIn group EU Data Ecosystem: http://www.linkedin.com/groups/EU-Data-Ecosystem-4925185

Follow us on Twitter: @EUDataEcosystem