META-FORUM 2015

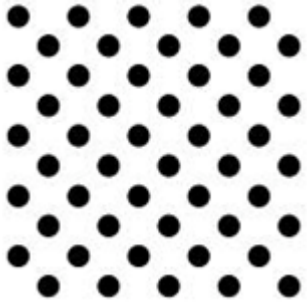
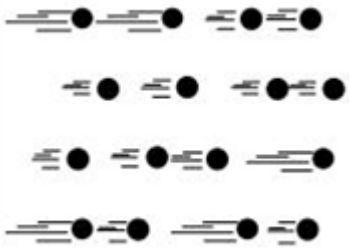
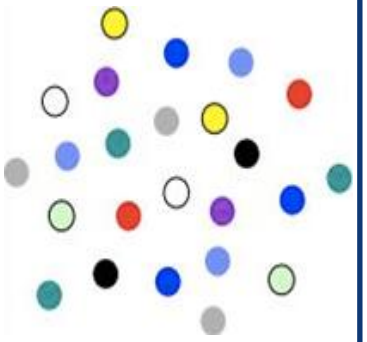
CHALLENGES, SOLUTIONS AND VISIONS FOR THE MULTILINGUAL EUROPEAN DATA ECONOMY

Laure Le Bars
Christian Lieske
SAP

BIG DATA
WHAT'S IT ALL ABOUT

When is Data 'Big'?

BDV BIG DATA VALUE

Volume	Velocity	Variety	Veracity	Value
				
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**	**Data into Money**
Terabytes to exabytes of existing data to process	Streaming data, requiring mseconds to respond	Structured, unstructured, text, multimedia,…	Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception	Business models can be associated to the data

Adapted by a post of Michael Walker on 28 November 2012

“*Value*”

- Creating transparency
- Discovering needs, expose variability, improve performance
- Segmenting customers
- Replacing/supporting human decision making with automated algorithms
- Innovating new business models, products services

McKinsey Global Institute

- + Combining data together – Corporate data, Social Data, Sensor Data
- + Privacy, Consumer Protection, Skills
- + Engaging with other to develop together

BDVA

Big Data – Not Just Data

Big Data is not only about 'Big Data'!

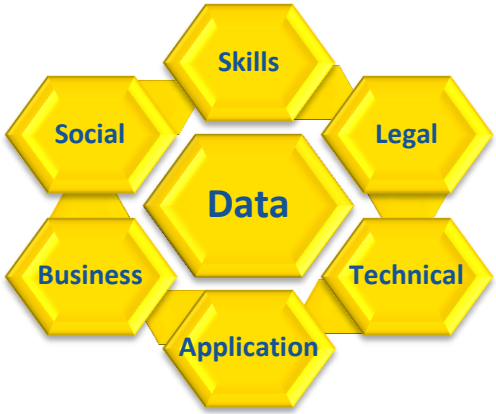
- **ie Volume**,

But also about

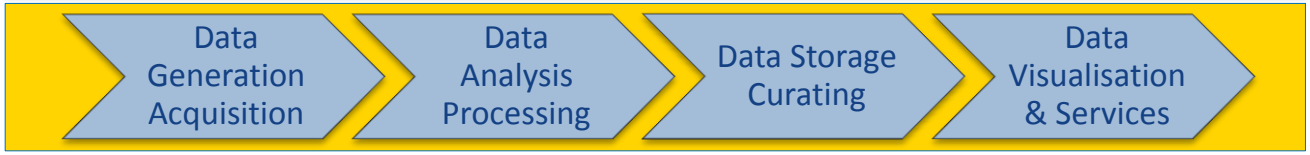
- **Velocity**
- **Variety**
- **Veracity**

And mainly the **VALUE** achieved from all this

Big Data Value has **Multiple Dimensions**



Big Data Value requires new **Business Models**, vibrant **Eco Systems** with strong European players along the entire **Value Chain**



BIG DATA VALUE
CPPP & BDVA

What is the BDV cPPP about

The Objective of the PPP is:

- The cPPP shall create results that have **IMPACT** on **members**, **participants**, **industry, economy** and **society**…

The Strategy needs to be:

- The main focus is the **transfer of technology and application** via the “instruments” designed for the PPP

The Operationalization :

- Establish a coherent set of **projects** that complement each other without being dependent.
- Define a **Framework** for projects that ensure that “results” is fit for the purpose of the next step in the innovation cycle and complies with the end user demand

BDVA – dimensions & activities

R&I Projects

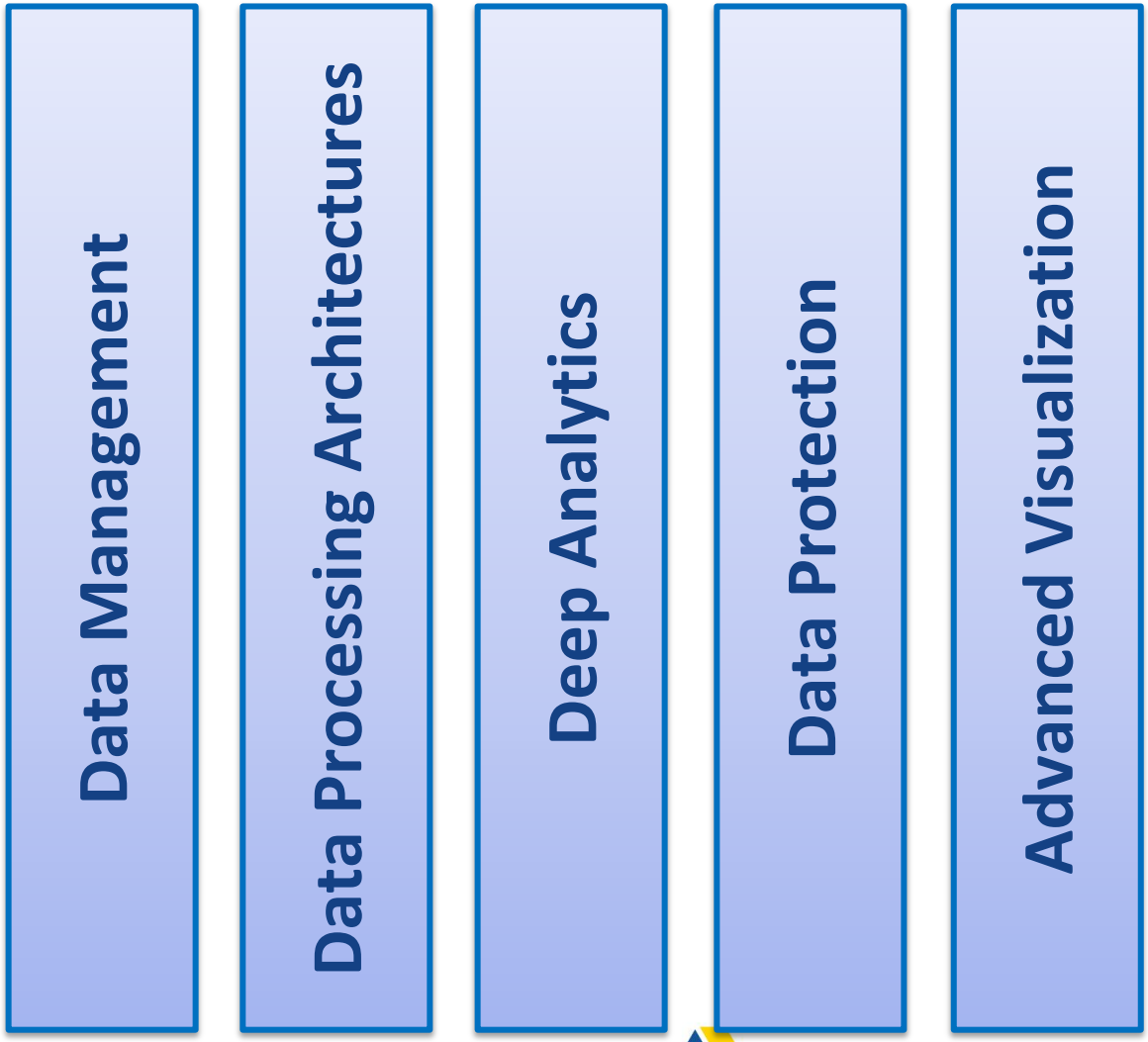
Large Targeted research and innovation projects, delivering foundational Big Data technology

Innovation Spaces

Hubs for bringing data, technology and application developments together; catering for development of skills, competence, and best practices.

Lighthouse Projects

Large scale demonstrations focusing on certain sectors and domains



European Innovation Spaces – a core component of the cPPP

European i-Spaces

- Serve as hubs for bringing the technology and application developments together and cater for the development of skills, competence, and best practices.
- Offer new and existing technologies and tools from industry and open source software initiatives as a basic service
- Facilitate the access to data assets.
- Allow an interdisciplinary approach along the various dimensions – technology, applications, legal, social, and business, data assets and the building up of skills.

Lighthouse projects – a mechanism for large-scale demos and awareness

Lighthouse Projects

- The major mechanism for Europe to demonstrate Big Data Value ecosystems and sustainable data marketplaces
- Running data-driven large scale demonstrations
- Propose replicable solutions by using existing technologies or very near to market technologies that could be integrated in an innovative way and show evidence of data value
- Create high level impact and broadcast visibility and awareness driving towards faster uptake of Big Data Value applications and solutions

BIG DATA
& LANGUAGE TECHNOLOGIES

The importance of the Language Technology community for Big Data

Language Technology and Big Data meet in areas such as :

- Search
- Processing of unstructured, multimodal, uncertain data
- Validation and cleansing
- Semantic analysis (including sentiment analysis)
- Enrichment and curation
- Inference (including statistical prediction)

Data Management
Technical Priority

Background & Challenges

1. How to **semantically annotate** unstructured and semi-structured data without imposing extra-effort to data producers?
2. How to unlock data silos by creating **interoperability standards** and technologies for storing and exchanging of data?
3. How to improve and assess the **data quality** from the various domains?
4. How do manage the **data lifecycle** as well as enforcing consistent quality as the data scales in volume, velocity and variability?
5. How to ensure consistent **data provenance** along the data value chain?
6. How to bundle and provision data, software and data analytics results to **ensure reuse of intermediate results**?

Outcome

1. Techniques and tools for handling **unstructured** and semi-structured **data**
2. Languages and techniques for **semantic interoperability** such as standardized data models and interoperable architectures for different sectors enriched through semantic terminologies.
3. Languages, techniques and tools for measuring and assuring **data quality**
4. Methods and tools for a complete **data management lifecycle** ranging from data curation and cleaning to long term storage and data access.
5. Languages and tools for **data provenance,** control and IPR
6. Principles for a clear **data-as-a-service model and paradigm** fostering the harmonization of tools and techniques

How to engage with LT community to address the data management challenge?

- How to handle the **multilingualism of data sources**? (i.e. existing tools can often not be used as they are generally provided only in English language)
- How can LT community help to achieve **semantic interoperability**?
- How can LT assure **data quality**? How can LT support cleaning and curation of data?
- How to align **data provenance** approaches of both communities?

How to engage with LT community to foster data-driven innovations?

- How to **make multilingual language resources available** to develop new products for all European languages?
- Are **best practices for publishing** linguistic resources available that can be reused to facilitate the access to data assets?
- How to **speed up implementation** time of LT-based products?

How to engage with LT community in running large scale demonstration ?

- How to support the BDV community in building sector-focused large scale demonstrations, for instance by progressing existing language technologies to the **understand the sector specific terminology**, such as the „medical language“

Where BDVA and the Language Technology community can meet (1/2)

Assist to define and prioritize real-world **functional and non-functional requirements**

- Any language
- Any modality (although focus on speech and text might be OK)
- Any quality
- Any domain/sector

Build-in **robustness, scalability, performance,** and **maintainability**

Based on **standards** (e.g. from International Standards Organization, or the World Wide Web Consortium)

Taking into account **statistical** and **knowledge-based approaches** (e.g. drawing on Linked Open Data)

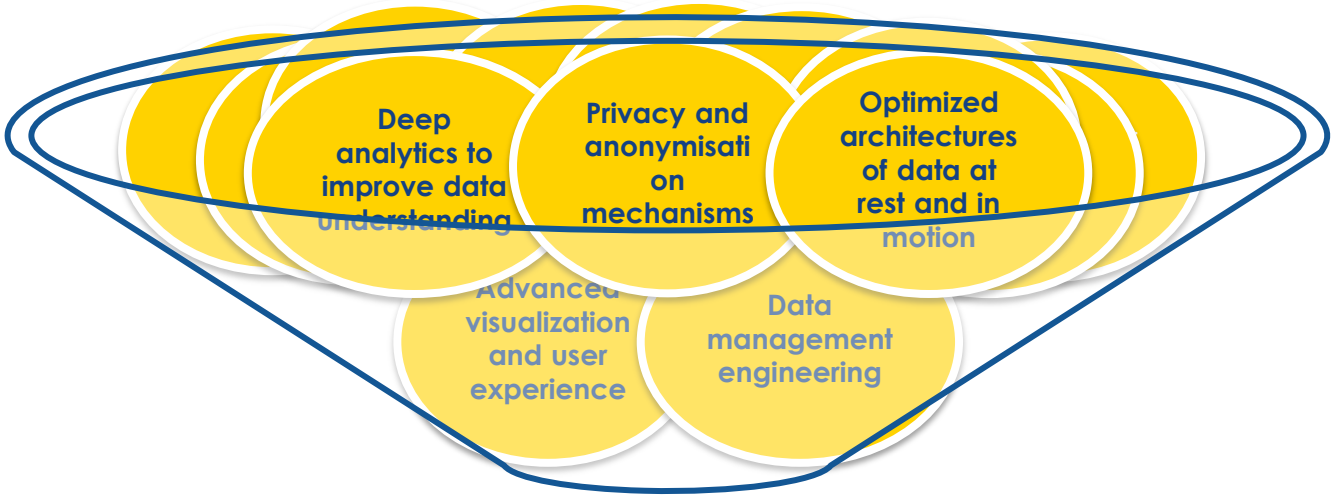
Where BDVA and the Language Technology community can meet (2/2)

- Work with **data** available to & relevant for the BDVA
- Discuss the role of **freely available**, highly **reusable** multilingual assets and technologies vs. company-specific, paid offerings
- Ensure that activities are **transparent**, and not fragmented
- Help to define and run **lighthouse projects**
- Co-create project sketches for the **demonstrations**
- Aim to turn the demonstrations into **foundations** that can be build upon
- Define or verify **business models**

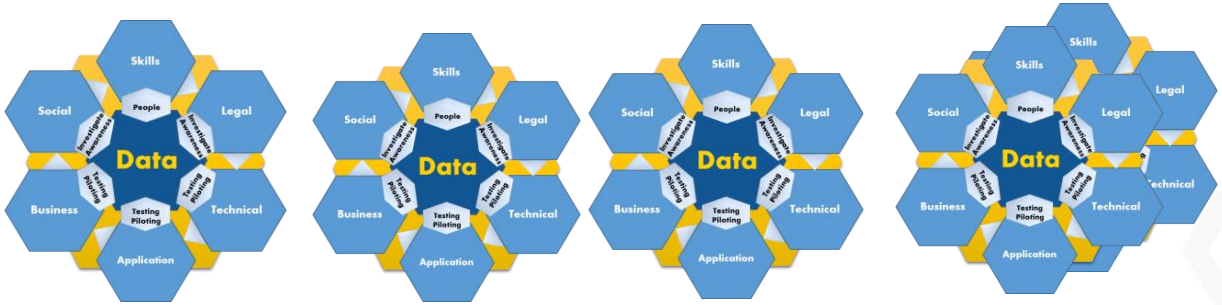
BDVA APPENDIX

The BDVA main elements

R&I Projects



I-Spaces Projects



Light-house Projects

Manufacturing | Personalised Medicine | Logistics | Energy

CSA - Projects

Stakeholder Platform Governance

Legal Environment

Business Models

Skills

Societal Acceptance

BDV BIG DATA VALUE

www.bdva.eu


BDV
BIG DATA VALUE ASSOCIATION
Home
Blog
News
Events
Downloads
Get Involved
FAQ
About us
The Big Data Value Association (BDVA) is organizing its first Summit in Madrid from 17th till 19th June 2015. The aim of the BDVA Summit is to inform about the BDVA activities, present the main expected funding opportunities in the field of Big Data and to set-up dialogues to discuss opportunities and challenges offered by Big Data among the European Big Data Community.

SAVE THE DATE!

- Get informed about the Big Data Value PPP
- Learn about funding opportunities in the field of Big Data
- Engage in activities of the Big Data Value Association (BDVA)
- Discuss opportunities and challenges offered by Big Data
- Network and cooperate on research, development and innovation
- Enjoy beautiful Madrid

June 17-19, Madrid, Spain
www.bdva.eu