

EDF2015 Submission:

BigDataEurope - Showcasing the Societal Value of Big Data

Contributor names and short CVs: Sören Auer and Simon Scerri

Prof. Dr. Sören Auer heads the Enterprise Information Systems group at the computer science department at University of Bonn. He is also member of the leadership council of Fraunhofer-Institute for Intelligent Analysis and Information Systems (IAIS) and scientific head of the Fraunhofer IAIS department Organized Knowledge. Sören is coordinator of the H2020 project BigDataEurope, member of the board of directors of Big Data Value Association.

Dr. Simon Scerri is a senior postdoctoral researcher in the “Organized Knowledge” department at Fraunhofer IAIS and in the “Enterprise Information Systems” group at the University of Bonn. His research interests range from Knowledge Engineering, Information Extraction and Linked Open Data, to Human-Computer Interaction, Collaborative Systems and Context-Awareness.

Type of the presentation proposed:

Impact contribution

Title of the presentation:

BigDataEurope - Showcasing the Societal Value of Big Data

Summary of the presentation

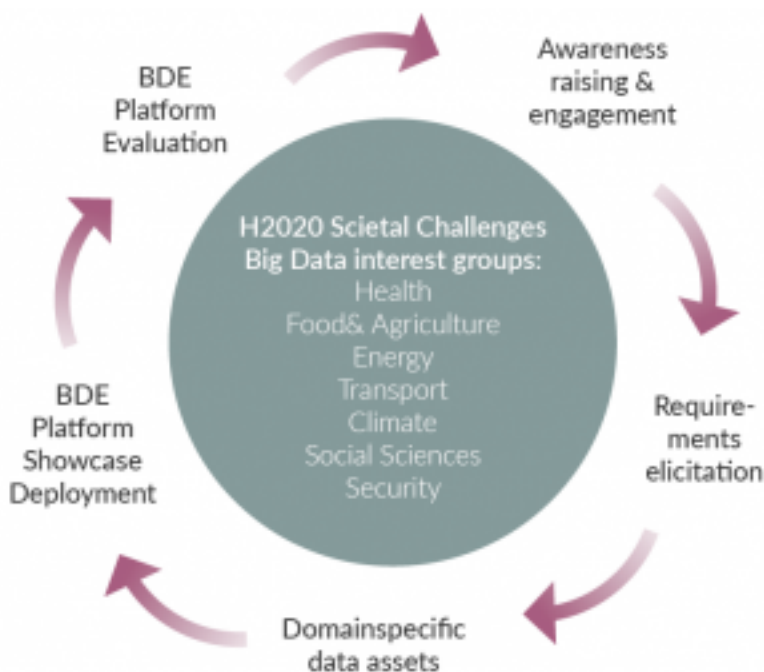
In this presentation, we will discuss strategies for showcasing the societal value of Big Data technology. In particular, we will present the architecture of a Big Data Platform currently developed by the H2020 SA BigDataEurope, which builds on current Big Data technology industry standards and facilitates the aggregation and semantic integration of large-scale, evolving heterogeneous Big Data. We will also present the BigDataEurope stakeholder working groups, which support the requirements and large scale deployment of the platform in the H2020 societal challenge domains Climate, Energy, Food, Health, Transport, Security, and Social Sciences.

Extended abstract of the presentation

The growing digitization and networking process within our society has a large influence on all aspects of everyday life. Large amounts of data are being produced permanently, and when these are analyzed and interlinked they have the potential to create new knowledge and intelligent solutions for economy and society. Big Data can make important contributions to the technical progress in our societal key sectors and help shape business. What is needed are innovative technologies, strategies and competencies for the beneficial use of Big Data to address societal needs.

Climate, Energy, Food, Health, Transport, Security, and Social Sciences – are the most important societal challenges tackled by the European Union within the new research and innovation framework program “Horizon 2020”. In every one of these fields, the processing, analysis and integration of large amounts of data plays a growing role – such as the analysis of medical data, the decentralized supply with renewable energies or the optimization of traffic flow in large cities.

In this presentation we will present strategies for enabling European companies to build innovative multilingual products and services based on semantically interoperable, large-scale, multi-lingual data assets and knowledge, available under a variety of licenses and business models. The following figure illustrates the overall approach followed by the H2020 BigDataEurope project in this regard.

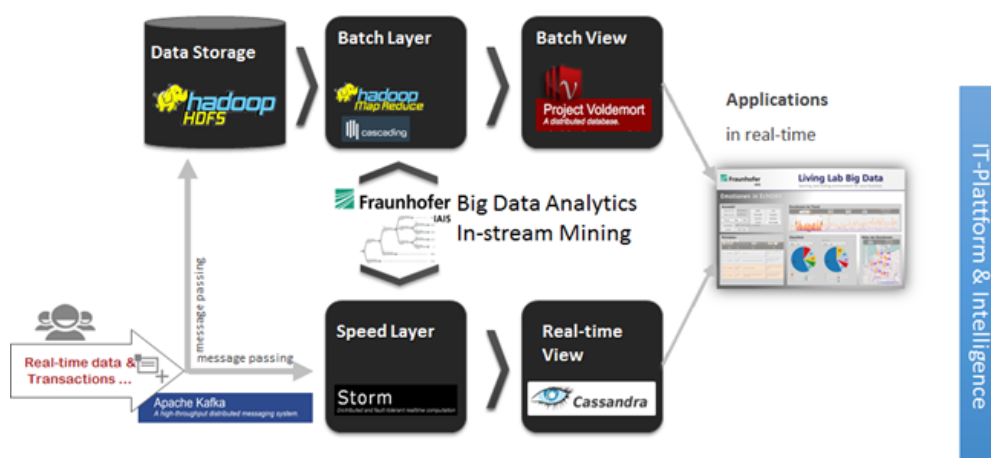


The two main strategies to increase the societal value pursued by BigDataEurope and discussed in this presentation are:

- Collect requirements (via societal challenge stakeholder groups) for the ICT infrastructure needed by data-intensive science practitioners tackling a wide range of societal challenges; covering all aspects of publishing and consuming semantically interoperable, large-scale, multi-lingual data assets and knowledge.
- Design and implement an architecture for an infrastructure (Big Data aggregator platform) that meets requirements, minimizes the disruption to current workflows, and maximizes the

opportunities to take advantage of the latest European RTD developments, including multilingual data harvesting, data analytics, and data visualization.

The Big Data aggregator platform architecture will build on existing Big Data industry best-practices in the spirit of the Lambda-Architecture and standard Big Data component distribution. However, as first discussions with stakeholders from the societal challenge domains have shown, the data variety is the most challenging aspect of Big Data management and applications in these domains and the current focus of Big Data platforms on volume and velocity does not yet satisfy this requirement. Consequently, we will demonstrate in this talk strategies, how semantic representation and integration can be added into Big Data technology components such as Hadoop, Storm and Flink. The strategies will comprise means for (1) ingesting and mapping data in semantic and non-semantic formalisms into the big data platform, (2) preserving the semantics, provenance and metadata during the Big Data processing and (3) exploiting the semantics for visualization, exploration and analytics. The following figure gives an overview on a Lambda architecture, into which the variety supporting and semantics preserving features will be integrated.



As a result, large-scale community-driven Big Data platforms aggregating a variety of different sources will become much easier to realize. With OpenPhacts (for the pharma and life-sciences domain) and LinkedGeoData (for the mobility and transport domain) we will present examples of prospective BigDataEurope platform deployments.