



BIG DATA EUROPE

Support Action

## Big Data Europe – Empowering Communities with Data Technologies

Project Number: 644564

Start Date of Project: 01/01/2015

Duration: 36 months

### Deliverable 2.5 Report on Interest Groups Workshops II

Dissemination Level	Public
Due Date of Deliverable	M12, 31.12.2015
Actual Submission Date	M13, 18.01.2016
Work Package	WP2, Community Building & Requirements
Task	T2.1
Type	Report
Approval Status	Approved
Version	1.0
Number of Pages	43
Filename	D2.5_Report_on_Interest_Groups_Workshops_II.pdf

**Abstract:** This report summarises the organization and derived results from the four Interest Group workshops organized during the reporting period (Societal Challenges 2 - Food & Agriculture, 4 - Transport, 6 - Reflexive & Inclusive Societies, 7 - Security) and carried out by each group associated with each societal challenges.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability.



Project funded by the European Union's Horizon 2020 Research and Innovation Programme (2014 – 2020)



## History

Version	Date	Reason	Revised by
0.0	01.08.2015	Placeholders	Simon Scerri
0.1	30.10.2015	SC7 Report	Sergio Albani, Michele Lazzarini
0.2	30.11.2015	SC2 Report	Valeria Pesce, Nikos Manouselis
0.3	30.11.2015	SC4 Report	Maxime Flament, Maria De Rycke, Andrea Toth, Simon Scerri
0.4	15.12.2015	SC6 Report	Eleanor Smith, Ivana Ilijasic Versic, Nina Bakanova, Martin Kaltenböck, Timea Turdean
0.5	15.01.2015	Cross-reviewing with contributors	Valeria Pesce, Andrea Toth, Sergio Albani, Ivana Ilijasic Versic, Thomas Thurner

## Author List

Organisation	Name	Contact Information
Fraunhofer	Simon Scerri	simon.scerri@iais.fraunhofer.de
SWC	Thomas Thurner	t.thurner@semantic-web.at
SWC	Martin Kaltenböck	m.kaltenboeck@semantic-web.at
SWC	Timea Turdean	t.turdean@semantic-web.at
FAO	Valeria Pesce	valeria.pesce@fao.org
AgroKnow	Nikos Manouselis	nikosm@agroknow.com
ERTICO	Maxime Flament	m.flament@mail.ertico.com
ERTICO	Maria De Rycke	m.derycke@mail.ertico.com
ERTICO	Andrea Toth	a.toth@mail.ertico.com
EU SatCen	Sergio Albani	sergio.albani@satcen.europa.eu
EU SatCen	Michele Lazzarini	michele.lazzarini@satcen.europa.eu
CESSDA	Eleanor Smith	eleanor.smith@cessda.net
CESSDA	Ivana Ilijasic Versic	ivana.versic@cessda.net
CESSDA	Nina Bakanova	nina.bakanova@cessda.net



# Executive Summary

In this deliverable we provide an in-depth report and material associated with the first round of workshops that have taken place in the second 6 month-period of the BDE project. The reports include information about the participants, the sessions organised, the talks and discussions as well as the gathered results (input for requirement elicitation). In addition, material associated with the workshop, such as the agenda and the original invitation letter, is also included.



## Abbreviations and Acronyms

<b>SC</b>	Societal Challenge
<b>EC</b>	European Commission
<b>RE</b>	Requirement Elicitation
<b>RS</b>	Requirement Specification
<b>WP</b>	Work Package



# Table of Contents

1. Introduction .....	8
2. First Round of Societal Workshops (II) .....	8
2.1 SC2.1 - Food, Agriculture and Forestry .....	8
2.1.1 Agenda .....	9
2.1.2 Expectation and Background .....	9
2.1.3 Setting the Scene .....	10
2.1.4 Summary of Breakout Groups .....	11
SC2.1 - Group 1: Data .....	12
SC2.1 - Group 2: Technology .....	13
2.1.5 Conclusions .....	14
2.1.6 Appendices .....	16
2.1.6.A Workshop Invitation Letter .....	16
2.1.6.B Slides & Presentations .....	17
2.1.6.C Photos .....	17
2.1.6.D Follow-up Post .....	17
2.1.6.E Attendees .....	17
2.2 SC4.1 - Smart, Green and Integrated Transport .....	18
2.2.1 Agenda .....	19
2.2.2 Workshop Scope And Structure .....	19
2.2.2.1 Background .....	20
2.2.2.2 Objectives .....	20
2.2.2.3 General information .....	20
2.2.3 Data-Centric Initiatives in Transport .....	20
2.2.4 Big Data Use Cases in Transport .....	21
2.2.5 Technologies and Tools Used and Envisaged .....	22
2.2.6 Summary of Breakout Groups .....	22
SC4.1 - Group 1: Policy .....	22
SC4.1 - Group 2: Business .....	23
SC4.1 - Group 3: Data & Technology .....	23
2.2.7 Conclusions .....	24
2.2.8 Appendices .....	25
2.2.8.A Workshop Invitation Letter .....	25



---

2.2.8.B Slides & Presentations.....	26
2.2.8.C Photos .....	26
2.2.8.D Follow Up Post .....	26
2.2.8.E Attendees .....	26
2.3 SC6.1 - Societies in a Changing World .....	27
2.3.1 Agenda.....	28
2.3.2 Expectation and Background.....	28
2.3.3 Summary of Breakout Groups .....	29
SC6.1 - Group 1: Data in place in the Social Sciences and Humanities .....	29
SC6.1 - Group 2: Risks and challenges of successful data management in the Social Sciences and Humanities .....	30
SC6.1 - Group 3: Technological demands of data in the SSH.....	31
SC6.1 - Group 4: Legal and policy demands of Data in the Social Science Sciences and Humanities.....	31
2.3.4 Appendices.....	33
2.3.4.A Invitation Letter.....	33
2.3.4.B Advertised Workshop Description and Agenda .....	33
2.3.4.C Attendees .....	34
2.3.4.D Follow-up Post and Message .....	34
2.3.4.E Photos .....	35
2.3.4.F Slides & Presentations.....	35
2.3.4.G Group Questions .....	35
2.4 SC7.1 - Secure Societies .....	36
2.4.1 Agenda.....	37
2.4.2 Expectation and Background.....	38
2.4.2.1 Background.....	38
2.4.2.2 Workshop aims .....	38
2.4.2.3 Workshop general information .....	38
2.4.3 Introductory Talks.....	39
2.4.3.1 Invited talks.....	39
2.4.4 Summary of Breakout Groups .....	40
SC7.1 - Group 1 : Secure Societies: Pilot Implementation Challenges.....	40
SC7.1 - Group 2: Technical Requirements.....	41
2.4.5 Conclusions.....	41
2.4.6 Appendices.....	42

---



---

2.4.6.A Invitation Letter .....	42
2.4.6.B Attendees .....	42
2.4.6.C Follow-up Post.....	42
2.4.6.D Photos .....	42
2.4.6.E Slides & Presentations.....	43
3. Summary.....	43



# 1. Introduction

This deliverable contains four reports for the first round of BigDataEurope workshops held in the second six month period of the project:

1. [SC2: Big Data for Food, Agriculture and Forestry](#)
2. [SC4: BigDataEurope Workshop on Smart, Green and Integrated Transport](#)
3. [SC6: The challenges of big data for societies in a changing world](#)
4. [SC7: Big Data in Secure Societies](#)

A summary and a copy of a detailed workshop report (including elicited requirements for the Big Data aggregator platform) is provided for each in the next Section. The report has, or will be circulated to all participants and other identified stakeholders. The communication will take place via multiple channels, including directly by email, project website and the respective W3C interest groups which have been set up.

## 2. First Round of Societal Workshops (II)

The four below-described workshops, in addition to the first three workshops which were held in the fifth and sixth month of the project, complete the first round of BDE workshops. The workshops Invitations were sent to the identified stakeholders, in multiple rounds. The workshops were designed around the blueprint provided in Deliverable 2.1, with minor adjustments to reflect the needs and characteristics of each community. A summary of workshop details, plus the full workshop report, are included below.

### 2.1 SC2.1 - Food, Agriculture and Forestry

The following table includes a summary of the workshop:

Date	22.09.2015
Venue	Research Data Alliance - Interest Group on Agricultural Data (IGAD) collocated event, INRA headquarters, 147 rue de l'Université, Paris, France
Invitations Sent	75
Invitations Accepted (Registrants)	48
Attendees (Total)	36
Attendees (Project Consortium & Project Officer)	7
Attendees (Other)	29





Breakout Sessions	2 sessions, 3 groups
-------------------	----------------------

### 2.1.1 Agenda

<ul style="list-style-type: none"><li>◎ 14:00 - Welcome &amp; Introduction, 30 mins<ul style="list-style-type: none"><li>○ Tour de Table</li><li>○ Pascal Neveu (INRA), introductory presentation: “Introduction to INRA's big data perspective and implementation challenges”</li></ul></li><li>◎ 14:30 - Data Session<ul style="list-style-type: none"><li>○ Lightning Talks, 30 mins<ul style="list-style-type: none"><li>❖ Tim Verhaart (Wageningen UR, LEI) "Big data opportunities for marketing of horticultural products"</li><li>❖ Elisabeth Arnaud (CGIAR Bioversity International) "Big data analytics in the CGIAR research portfolio and the Biodiversity perspective"</li></ul></li><li>○ Interactive Session, 45 mins - split in 3 groups with 1 facilitator each</li><li>○ Report to Plenary, 15mins</li></ul></li><li>◎ 16:00 - Technologies Session<ul style="list-style-type: none"><li>○ Lightning Talks, 30 mins<ul style="list-style-type: none"><li>❖ Rob Lokers (Wageningen UR, Alterra) “Big data challenges and solutions in agricultural and environmental research”</li><li>❖ Valeria Pesce (UN FAO &amp; GFAR) "A global linked and open data infrastructure for agricultural development"</li></ul></li><li>○ Interactive Session, 45 mins</li><li>○ Report to Plenary, 15mins</li></ul></li><li>◎ 17:30 - Wrap up<ul style="list-style-type: none"><li>○ Sören Auer (FhG) on BDE big data technologies</li><li>○ Stefano Bertolo (European Commission), position statement on the big data research agenda and portfolio of the EC</li><li>○ Q&amp;A session</li><li>○ Closing</li></ul></li></ul>
---

Workshop summary<sup>1</sup> posted on the BDE website 7 October 2015; slides available in the BDE area in slideshare and also linked to this report.

### 2.1.2 Expectation and Background

FAO and Agro-Know, the BDE partners responsible for bringing in stakeholders and eliciting requirements from the food and agriculture community, have already convened stakeholders in the past to discuss challenges and opportunities around data and therefore had a rough idea

<sup>1</sup> <http://www.big-data-europe.eu/bigdataeurope-workshop-on-big-data-for-food-agriculture-and-forestry-opportunities-and-challenges-held-in-paris-on-22-september-2015-the-main-challenge-is-variety/>



of the level of awareness and the expectations around “big data” in the food and agriculture community.

However, the Big Data Europe project needed far more precise indications and requirements than the general scenario that past consultations could provide. Normally, Big Data in agriculture are associated with information collected by sensors, satellites or drones combined with genomic information or climate data, which can all help farmers to optimize their farms’ operations. In addition, challenges and opportunities have been identified by existing communities of data managers in this area also around the heterogeneity of the data that need to be combined and integrated for both fostering new research and innovation and providing meaningful information for decision making.

As an example of generic needs expressed by actors in the broad domain of “agriculture” (which includes food, forestry, fisheries and biotechnologies), a much envisaged big-data-empowered scenario would be the generic ability to deliver better added-value integrated services that can answer the needs of different types of actors. Examples of such integrated systems go from alert systems (pests, disasters) to image -recognition-based plant / pest identification systems to food tracking systems to global research information systems to any conceivable advanced decision making platform combining climate, soil, crop, pest, price, political and social data.

The objective of convening a workshop on big data in this area was that of checking if this scenario still reflected the perspectives of key actors and above all that of asking them more precise questions in order to elicit more specific requirements. This was the background for the BigDataEurope workshop on “Big data for food, agriculture and forestry: opportunities and challenges”, held in Paris on 22 September 2015. The workshop was organized by Agro-Know, FAO, GFAR and the Big Data Europe project and hosted by the Institut National de la Recherche Agronomique (INRA).

The workshop was intentionally co-located with another event that was going to bring together a good number of key actors in food and agriculture--related research: the pre--meeting of the Research Data Alliance Agricultural Data Interest Group. The Research Data Alliance (RDA, <https://rd-alliance.org/>) is an international initiative started in 2013 by a core of group of agencies (the European Commission, the US National Science Foundation and National Institute of Standards and Technology, and the Australian Government’s Department of Innovation). The Agricultural Data Interest Group (ADIG) is a domain oriented interest group in RDA, co--chaired by FAO, to work on all issues related to data that are relevant to agriculture and related domains. This ensured the participation of more than 40 experts from around the world, representing universities, research institutions, private companies, international organizations, international projects and the European Commission (see participants’ list in the Annexes).

### 2.1.3 Setting the Scene

The workshop was organized around a short introductory session and two discussion sessions, one on data and one on technologies. While the main results of the workshop are to



be found in the summary of the discussion groups, where interesting requirements and challenges for BDE came out, a few presentations in the three sessions had the objective of setting the scene, describing current experiences and plans on big data in institutions working in agricultural research. The workshop opened with Pascal Neuveu, senior research engineer at the French National Agronomic Institute (INRA) and director of the MISTEA Laboratory. He gave the big data perspective and implementation challenges that such a large and distributed agronomic research organisation has (INRA has 18 different centers spread across over 40 geographical locations, and more than 10,000 people from which about 4,000 are researchers). Pascal also presented a specific case study, the one of high throughput phenotyping at five large, open-air experimental fields, two greenhouses with controlled environments, and two fully equipped laboratories for carrying different types of omics analyses. Variety and Velocity were identified as the most challenging dimensions (velocity especially for streaming images). After this initial presentation Stefano Bertolo from the European Commission raised the type of question that he kept raising at different points during the workshop: precise numbers in order to understand the order of magnitude of these data and the actual need for big data technologies. E.g. what is the largest number of environments in which 1 geno/phenotype has been tested from start till today? Answers from the audience (200 locations in different countries, 50k-60k per location?) tended also to point out difficulties in answering this question (e.g. considering also years of experiments, depending also on data stream rates...). A similar question was asked about the number of queries that can be handled.

Tim Verhaart from the socio-economic research institute of LEI (Wageningen UR) talked about “Big data opportunities for marketing of horticultural products”, introducing a very interesting public-private partnership through which the fruits and flowers industries are calling on big data technology researchers to help them do business better. Rob Lokers from the inter-disciplinary environmental research institute of Alterra (Wageningen UR) brought forward the high complexity of using big data as input for complex agricultural and environmental modelling, which is then generating new (big) data, information, and knowledge that supports research and policy making – in his talk on “Big data challenges and solutions in agricultural and environmental research”. Elizabeth Arnaud from Bioversity International provided an excellent insight into the CGIAR Big Data Analytics Platform and the CGIAR big data plans. Valeria Pesce from the Global Forum on Agricultural Research (GFAR) presented where we stand today with a global linked and open data infrastructure for agricultural development. At the end of the workshop, Soren Auer described the Big Data Europe project and Stefano Bertolo of the European Commission gave an overview of the big data research agenda and portfolio of the EC, also raising again some issues regarding the more precise identification of quantitative indicators that will justify the use of big data technologies in food and agriculture.

#### 2.1.4 Summary of Breakout Groups

Two discussion sessions of 45 minutes each were organized in three breakout groups. The first session was more around data (identifying the most relevant types of data and the related main challenges) while the second one discussed technologies (currently used solutions, promising technologies, gaps).



## SC2.1 - Group 1: Data

The three groups were asked to answer specific questions on data and all answers were captured on flipcharts. Questions:

1. What are the most important data sources for agriculture and food?
2. What is most important for such data in terms of: Volume, Velocity, Variety, Veracity?
3. What are the challenges around these data sources (and the data) in terms of: availability, legal issues, policy issues, skills?

### 1. Overall Results

Question 1: *What are the most important data sources for agriculture and food?*

Participants were invited to think especially of data types and data sources that are perceived as “big”. The main types of data identified across the three groups were very similar:

- Sensor / drone data
- High-rate image streams
- Genomic data
- Phenotypic data
- Combined cross-disciplinary data (climate, economic, social...)

Question 2: *What is most important of the 4 Vs for such data?*

Across all three groups, the most important V was definitely Variety: even when speaking about an apparently homogeneous type of data (genomic, sensor), the shared opinion among participants was that a) those data were not particularly useful if not combined with other related data; b) even data of the same type don't come in a homogeneous form. Particular importance was also given to veracity (especially for genomic data). Volume was of course recognized as a big challenge for sensor data; velocity was mentioned mainly for high-rate image streams.

Question 3: *What are the challenges in terms of: availability, legal issues, policy issues, skills?*

Most participants complained about the lack of availability of data they would need due to policy or legal issues. In particular, it was noted that even when data are available what is often lacking is easy discoverability (vs. granting of special access). The lack of skills also came up across all three groups.

### 2. Other Findings and Discussions

Other relevant types of data identified:

- Food tracking (volume)
- Food prices (volume)



- Multimedia for different purposes
- Public administration data (tax, customs, traffic)
- Historical data
- Model outputs

Other challenges identified:

- Availability: little availability of historical data; field phenotyping data are closed; lack of nutrients database; issue of competitive advantage
- Skills: lack of skills in all sorts of “omics”
- No data type on its own is useful, need to combine
- Again the issue came up in discussions of how “big” these data are and what the threshold is beyond which a normal powerful server is not enough.

Regarding some challenges identified by participants in terms of availability of needed data, Stefano Bertolo from the EC highlighted both the existence of public European services providing different types of public utility data for free (e.g. regarding high speed imaging, Copernicus provides terabytes of data daily for open use) and the role of the EC in putting pressure on governments and public research if public data are not made available.

### **SC2.1 - Group 2: Technology**

The three groups were asked to answer questions with specific reference to technologies for data processing, representation, acquisition and visualization.

Questions:

1. What kind of data management solutions are in place at the moment & where are the main issues with those solutions at the moment?
2. Are there promising technologies you have identified for the future?
3. What are the challenges around these data technologies regarding: data acquisition, data processing, data infrastructure, data publication, security and/or privacy issues?

### **2. Overall Results**

An initial answer to questions 1 and 2 was provided by Rob Lokers’s presentation, in which several technologies commonly used in agricultural research were presented:

- RDBMS
- Geo-databases
- Various “old & proven” programming languages (esp. for modelling, data processing)
- Remote sensing: dedicated tools & environments for processing and analysis, ENVI, R, GDAL etc.
- Harmonized information / data models (but still per discipline)
- High Performance clusters / grids
- Still experimental (ICT research for agriculture):
  - RDF databases
  - Vocabularies and ontologies (no alignments)
  - NLP algorithms

Question 1: *Data management solutions in place now and main issues?*



Most of the technologies mentioned by Rob were also mentioned across all three groups. Many participants reported that technology solutions are often produced in house and that researchers are still using Excel for managing their data. All participants seemed to have a clear perception that the current technologies they are using are not enough for certain data processing needs.

Question 2: *Are there promising technologies you have identified for the future?*

Many participants seemed to agree on a number of promising technologies identified:

- NOSQL
- SPARQL
- Semantic tools
- NLP
- HADOOP
- Elastic search

No participant seemed to be already using a stack of big data technologies.

Question 3: *Challenges in terms of acquisition, processing, infrastructure, publication, security / privacy issues?*

Acquisition: legacy

Processing: performance; need for adequate processing for decision making

Publication: lack of incentives; no standard tools; preservation

## **2. Other Findings and Discussions**

Other technologies that were mentioned include:

- In use:
  - Dataset sharing platforms: Data Cite, Figshare
  - Global information systems: e.g. GRIN for germplasm
  - Pattern recognition software
- In use, with issues:
  - Spreadsheets
  - In house developed software
- Promising:
  - OpenStreet Map
  - REST services
  - Neural networks
  - Linked Data Fragments
  - Federated search
  - Annotation tools
  - ODK for surveys
  - GOBII (high density marker data)

### 2.1.5 Conclusions

Overall, what came out of the discussions was that the special thing about big data in agriculture is its extreme variety. This is what you get, if you contrast the four V's of big data to the data types and sources that are typically used in agricultural and food research. In most



cases we are not talking about an extremely large Volume (other domains have much more voluminous data); it is not that data come with an extremely high Velocity, especially compared to other domains. In many cases, their Veracity is quite high. But in food and agriculture, data Variety matters the most: you need to combine multiple, heterogeneous data types and formats from several sources, trying to solve the information problems and support decision making of the relevant stakeholders. However, real cases where volume and velocity were high were reported. The discussion indicated that it will be important for the BDE project to more precisely understand the order of magnitude of the data types discussed in this workshop (e.g. streaming data over decades, genomic data...) and the actual need for big data technologies.





## 2.1.6 Appendices

### 2.1.6.A Workshop Invitation Letter

Dear Sir/Madam,

as part of the next meeting of the Research Data Alliance (RDA) Interest Group on Agricultural Data (IGAD) that will take place between September 21st-22nd in Paris, the BigDataEurope (<http://www.big-data-europe.eu>) action is organising a session dedicated to big data in agriculture, forestry and food. The IGAD meeting will take place at the head office of the Institut National de la Recherche Agronomique (INRA), as a co-located event of the Research Data Alliance (RDA) 6th Plenary.

The big data session of the IGAD meeting is expected to investigate the characteristics of data sources & repositories that our community has, in order to identify cases that offer a high degree of variety, velocity and volume - and therefore good candidates for the application of big data technologies. In addition, an introduction to such technologies is going to be given and a discussion on their relevance and applicability for the wider range of topics related to food security, sustainable agriculture and forestry.

As a recognised stakeholder in the food and agriculture sector, we believe that you would be interested to join this session, share experiences and ideas on how big data applies for our sector, report on your relevant projects and activities, and ultimately find out ways in which the BigDataEurope action can support you.

Event Registration: send an e-mail expressing your interest to attend to the IGAD meeting organiser Imma Subirats (FAO of the UN, [Imma.Subirats@fao.org](mailto:Imma.Subirats@fao.org)), cc-ing Nikos ([nikosm@agroknow.gr](mailto:nikosm@agroknow.gr)).

In addition, as a representative of one of our target communities, we would like to engage with you in the long term, in order to identify your big data technology needs, challenges and requirements. We therefore invite you to give due consideration to our participation offer by considering one or more of the following different levels of stakeholder engagement:

- Subscribe to the BigDataEurope Newsletter;
- Join your respective W3C Community Group;
- Participate in the planned key Stakeholder Workshops, including the one announced above;
- Participate in one of our BigData Pilots (subscribe for more information);
- Follow us on the Social Media (Twitter, LinkedIn, SlideShare).

This invitation is extended to others in your network with an interest in the challenges of Big Data in your sector. Where possible we would be grateful if you could focus on information and data managers who are working with big data on a daily basis.

We look forward to your participation!

Kind regards,

Valeria Pesce (FAO/GFAR)

Nikos Manouselis (Agro-Know)

Societal Challenge Representatives on behalf of the BigDataEurope Consortium





### 2.1.6.B Slides & Presentations

1. [Big Data Europe \(BDE\) - Project Overview & Food Workshop](#)
2. [INRA's Big Data perspectives and implementation challenges](#)
3. [A global linked and open data infrastructure for agricultural development](#)
4. [Big Data challenges and solutions in agricultural and environmental research](#)
5. [Big Data in CGIAR](#)
6. [Big Data opportunities for marketing of horticultural products](#)

### 2.1.6.C Photos

Photos are available in the respective workshop folder [here](#).

### 2.1.6.D Follow-up Post

A follow-up blogpost/message was shared on the BDE [website](#) and W3C [group](#).

### 2.1.6.E Attendees

The following table is the list of registered attendees for the workshop:

<b>Name</b>	<b>Institution/Company</b>	<b>Country</b>
Soren Auer	Fraunhofer	Germany
Nikos Manouselis	Agro-Know	Greece
Valeria Pesce	FAO / GFAR	Italy
Martin Kaltenböck	Semantic Web Company	Austria
Timea Turdean	Semantic Web Company	Austria
Stefano Bertolo	European Commission	Brussels
Rob Lokers Alterra	Wageningen UR	Netherlands
Marie Angélique	Laporte Bioversity International	Italy
Rosemary Shreshta	CIMMYT	Mexico
Nicolas Saby	INRA	France
Pascal Neveu	INRA	France
Nicolas Tremblay	Agriculture and Agri-Food Canada	Canada
Giovanni L'Abate	Consiglio per la Ricerca e la sperimentazione in Agricoltura (CRA)	Italy
Caterina Caracciolo	FAO	Italy
Robert Davey	The Genom Analysis Centre	United Kingdom
Ruth Bastow	Global Plant Council	
Nicolas Gengler	University of Liège	Belgium
Elizabeth Arnaud	Bioversity International	Italy



Tim Verweert	LEI, Wageningen UR	Netherlands
Daniel Martini	Association for Technology and Structures in Agriculture (KTBL)	Germany
Dana Tomic	IEEE	
Anahita Nafissi	Forschungszentrum Juelich	Germany
Hugo Besemer	Wageningen UR	Netherlands
Elf Pavlik	Freelancer	
Cyril Pommier	INRA	France
Simon Scerri	Fraunhofer	Germany
Dimitrios Zisis	Institute of Plant Genetics, Polish Academy of Sciences	Poland
Pawel Krajewsky	Institute of Plant Genetics, Polish Academy of Sciences	Poland
Harris Moysiadis	Future Intelligence and QUHOMA project	Greece
Kalin Muldzhawski	Linked Farm	
Cristian Vasquez	Pierre Larmande IRD	France
Aravind Venkatesan,	IBC	France
Erick Antezana	Bayer CropScience NV	Belgium
Karna Wegner	FAO	Italy
Ruthie Musker	UC Davis/ ETH Zurich	US/Switzerland

## 2.2 SC4.1 - Smart, Green and Integrated Transport

The following table includes a summary of the workshop:

Date	07.10.2015
Venue	ITS WC 2015 collocated event. Palais des Congrès de Bordeaux, Bordeaux, France
Invitations Sent	14,000
Invitations Accepted (Registrants)	147
Attendees (Total)	35
Attendees (Project Consortium & Project Officer)	4
Attendees (Other)	31



Breakout Sessions	3
-------------------	---

## 2.2.1 Agenda

<ul style="list-style-type: none"><li>⊙ 9:00 Official Welcome (Maxime Flament, ERTICO – ITS Europe)</li><li>⊙ 9:15 Opening Session<ul style="list-style-type: none"><li>○ 9:15-9:30: Big Data for Transport – Why Does it Matter? (Paul Timmers, European Commission)</li><li>○ 9:30-9:40: What can big data do for transport? (Simon Scerri, Fraunhofer IAIS)</li></ul></li><li>⊙ 9:40 Session 1: Data-centric Initiatives in Transport<ul style="list-style-type: none"><li>○ 9:40-9:50: Big Data Sources For/From Intelligent Road Transport: An overview (Evangelos Mitsakis, CERTH)</li><li>○ 9:50 – 10:00: The use of big data for public transport performance measurement (Roberto Baldessari, NEC)</li><li>○ 10:00 – 10:15: The role of social media in transport (Dave Marples, Technolution)</li></ul></li><li>⊙ 10:15-10:25: Coffee break</li><li>⊙ 10:25: Session 2: Big Data Use-cases in Transport<ul style="list-style-type: none"><li>○ 10:25-10:35 : Open Logistics (German Herrero, ATOS)</li><li>○ 10:35-10:45: Real-time Traffic information and traffic management (Nick Cohn, TomTom)</li><li>○ 10:45-10:55 How to deal with all the data that is generated by the vehicle sensors? (Sean Gaines, Vicomtech)</li></ul></li><li>⊙ 10:55: Session 3: Technologies and Tools used and envisaged<ul style="list-style-type: none"><li>○ 10:55 – 11:05: Existing tools and technologies used across other sectors (Simon Scerri, Fraunhofer IAIS)</li><li>○ 11:05-11:15: Data Analyses in Transport (Helena Gellerman, Chalmers University)</li></ul></li><li>⊙ 11:20-12:20 Breakout sessions (Business, Policy, Data &amp; Technology)<ul style="list-style-type: none"><li>○ Leader of the Business group: Dave Marples, Technolution</li><li>○ Leader of the Policy group: Giselle Roesems, DG CONNECT</li><li>○ Leader of the Data &amp; Technology group: Sean Gaines, Vicomtech</li></ul></li><li>⊙ 12:20 Coffee break</li><li>⊙ 12:30 Breakout session reports</li><li>⊙ 12:55 Closing note</li></ul>
--

## 2.2.2 Workshop Scope And Structure

In the framework of the BigDataEurope (BDE) project, ERTICO-ITS Europe organised a workshop on “Big Data for Smart, Green and Integrated Transport” on the 7th October 2015, collocated with the ITS Congress in Bordeaux.



### **2.2.2.1 Background**

BDE will undertake the foundational work for enabling European companies to build innovative multilingual products and services based on semantically interoperable, large-scale, multilingual data assets and knowledge, available under a variety of licenses and business models. The Transport Societal Challenge set its eyes to contributing to the following objectives: resource-efficient transport that respects the environment; better and more informed door-to-door mobility; less congestion and unforeseen delays; safer and more reliable multi-modal mobility; secured exchange of personal information; global leadership for the European transport industry; and socio-economic and behavioural research and forward looking activities for policy making.

### **2.2.2.2 Objectives**

Participants to this workshop had the opportunity to influence the design, and ultimate benefit from the Big Data platform that BDE will deliver. This platform aims to facilitate Big Data usage in real world examples, and will consist of an architecture, components, guidelines and best practices to make the best of Big Data in this case in the setting of transport. The workshop focused on the elicitation of requirements for Big Data management within the intelligent transport domain, by encouraging European stakeholders to participate in the definition of Big Data needs and requirements for the intelligent transport domain in Europe.

### **2.2.2.3 General information**

Over 30 participants attended the workshop, which was more oriented to road transport rather the entire domain. The workshop hosted speakers from various backgrounds be it policy, industry, research institutions or universities; who presented on a wide range of topics ranging from the role of social media in transport, over open logistics and traffic management to data analyses techniques. The workshop consisted of three sessions dedicated respectively to data-centric initiatives in transport, big data use-cases in transport, and technologies and tools used and envisaged. Talks in these sessions were followed by three breakout sessions; dedicated to policy, business, and data and technology respectively. The workshop was concluded following a summary of discussions and results in the three parallel sessions.

## **2.2.3 Data-Centric Initiatives in Transport**

The first of three sessions dealt with data-centric initiatives in transport. The use of big and open data in the transport sector is relevant for governments (traffic control, planning and modelling, route planning, congestion management, etc.), for the private sector (travel industry, route planning and logistics, competitive advantages, etc.) and for individuals (route and travel planning).

Big data plays an important role in how smart cities obtain their transportation targets, meaning how smart cities use and deploy ICT to enhance their transportation networks. Big data in



transport will lead to improved multi-source traffic and travel data availability and processing, and to tools to enhance multi-source traffic and travel data fusion for, for instance, improved traffic and mobility management. Combining big, open and linked data will leverage innovation and economic benefits.

Big data is important for traffic management: it provides new insights into traffic patterns, real-time traffic data to information service providers. In public transport, big data helps to understand travellers' journey patterns, informing transport agencies and operators on how different social groups use the public transport system. Big data for private travellers enables personalised information (such as delays). Finally, big data can also be used for asset maintenance offering new opportunities to identify problems more quickly and reduce costs.

Big data is highly valuable to monitor of public transport performance. Currently data analytics use public transport Key Performance Indicators such as Excess Waiting Time (EWT) in order to measure the contractual performance of the public transportation companies. They reveal the worst performing routes, key bottlenecks on the routes, causes for dwell time at bottlenecks and time table improvement margins. Based on this, bus operators and municipalities/authorities can have an EWT profile based on their public transit scores. Simple analytics derive hot spots to attack in order to reduce EWT.

Social media play an important role in transport with continuous, inbound information (stimulus to traffic control) and outbound information (stimulus to road user). Open questions relate to how to decide which social media inflows are reliable and to what degree, what degree does the fact that social media sources are electively published give users open rights to use them, how to prevent spoof outflows from confusing or malicious users, and how to prevent beneficial services encouraging inappropriate use of social media apps.

## 2.2.4 Big Data Use Cases in Transport

The first use-case discussed in this session was Open Logistics. The use of big data can lead to efficiency gains in the logistics chain. Manufacturers are looking at big data as a catalyst for greater collaboration, enabling more complex supplier networks that focus on knowledge sharing and collaboration as added value. Big data and advanced analytics are being integrated into optimisation tools, demand forecasting, integrated business planning and supplier collaboration & risk analytics at a quickening pace. Big data also contributes to the delivery process.

The second use-case considered was Real-time Traffic Management. Big data allows warnings ahead of jams, HOV lane handling, traffic flow predictions, etc. Going from big data to active traffic management requires merging big data with data from fixed sources. The use of archived data allows to improve individual route planning, to measure bottlenecks and delays, to measure system reliability, to determine priorities for infrastructure improvement, and to analyse the impact of the investments made.



The third use-case focused on how to deal with the data generated by vehicle sensors. Data can come from external and internal cameras each with 360° viewing. Data fusion beyond 720° viewing allows Holistic Driving Models, Personalised Driving Models, HMI response, Continuous Monitoring, etc.

## 2.2.5 Technologies and Tools Used and Envisaged

In the context of data analysis in transport, the tools and process in the Field Operational Test (FOT) analysis platform were presented. FOTs generate complex data, subjective and contextual data. As a next step, data is processed using HPC and is hosted in databases and files for video. A future big data platform must allow real-time data analysis (instead of research), include visualisation tools that allow data mining and analysis results, automatic video coding data sent over GPRS/Wi-Fi to cloud based storage, efficient data structures to allow efficient data extraction, and will need to support open repositories with high quality context information.

The ultimate goal of the BDE is to design and build a platform that can benefit not only the transport domain but also a variety of other societal challenges. Therefore the presentations of the workshop were wrapped up by Fraunhofer, as project coordinator. Big data involves the analysis of historical data, analysis of actual data with low latency in “real-time”, and interactive analysis by online queries. The question is how this can all be put together in one big data management system. Fraunhofer presented a blueprint of the data aggregator platform that follows the typical Lambda architecture, and utilizes existing solutions such as BigTop components.

## 2.2.6 Summary of Breakout Groups

The three sessions were followed by three parallel breakout sessions focusing on policy, business and data and technology. Each breakout session saw the participation of between 5 and 12 participants.

### **SC4.1 - Group 1: Policy**

The Policy session discussed privacy issues, the role of the public authorities to make data available to the service providers, issues regarding the use of data (how to allow to link and dig into these data), and the public/private role regarding standardisation. If a common standard will be developed, it should be a non-discriminatory standard with open APIs. Policy makers should also provide clarity on the re-use rights of data. The session discussed the issue of contract obligations and whether guidance from policy makers is required to bind a user to one single service. Participants agreed that there should be a “free flow of data initiative” in the EU. The EU should also promote the use of open data. The session discussed the public-private role with regard to data centre regulations: should the market decide or should there be a service level agreement. The policy sessions discussed the tools needed and better guidelines, on how to anonymise and aggregate data; better guidance,



understanding and control of big data; principles of privacy design, and a checklist for testing if privacy for design is respected. The EU should also provide guidance on whether the transport sector should have a transport-specific privacy framework or keep the general framework.

**SC4.1 - Group 2: Business**

The Business breakout session discussed how current businesses are challenged by big data. An issue is to have the right mindset to make data available in the first place. Making data available also involves a risk factor. An education/outreach process to the public at large is required. Through use-cases, people must be made aware that good things can happen. The process will require a political leap of faith and brave policy makers. It is also important to publicise the fact that specific data is available. Business models will arise from those who can turn data into usable information. The market must become better at turning data into information. The quality of the output is important: the higher quality the higher the market value of the data. Current businesses will be challenged by big data. Ultimately the user will choose which services will survive.

**SC4.1 - Group 3: Data & Technology**

The Technology break-out session discussed the cross-domain nature of transport data use-cases, which involve psychologists, road engineers, vehicle engineers etc. Discussions revealed that the BDE project’s role is to quiz the societal challenge domains, including the transport sector, to propose technical solutions without their clear understanding of the underlying problem. The results will be converted to architectural and design requirements. The session discussed the three transport data dimensions and data types including Infrastructure (dynamic maps/attributes, maintenance, operational, foresight), Vehicle (location, driver monitoring, performance and telemetry) and Users (events, state, behaviour and personal activity). Each of these dimensions and data types were evaluated against the three core Big Data V’s: Volume, Velocity and Variety. The results are summarised in the table below. Open issues relate to unforeseen events, diverse data formats, common understanding (lack of), and scalability / volume issues.

Dynamic Maps/Attributes	+	-	+
Maintenance	-	-	+
Operational	+	+	+
Foresight	?	?	?
Location	+	-	+
Driver Monitoring	0	0	+





Performance	+	+	+
Telemetry	+	+	-
Events	?	+	+
State	-	-	+
Behaviours	-	+	+
Personal Activity	-	+	+

## 2.2.7 Conclusions

This workshop was organised in order to elicit and better understand the requirements that the Transport Societal Challenge poses on a future Big Data infrastructure in Europe. This is not always straightforward since frequently, people involved in the sector do not even have a complete understanding of the big data challenges, and therefore struggle to identify ideal solutions. The outcome of the workshop will therefore feed the requirements elicitation phase of the BDE project, ahead of developing the first prototypes and the selection and initiation of pilots that are to be implemented. The workshop results indicate that there is a clear need for Big Data solutions in Transport, but that there are also very diverse areas for application. Moreover, this first workshop mainly focussed solely on road transportation. At the same time, it was evident that stakeholders are asking some questions which remain unanswered. This confirms that there is a need for more Big Data expertise to contribute to the transport fields with guidance and best practices. There was also an emphasis on understanding how to better control and measure the quality of the information.

In view of the foreseen Big Data architecture and implementation, it is important to eventually propose generic tools and expertise which will answer a broad scope of transportation topics from logistics, public transport to traffic management. A basis is to convince the transport stakeholders to contribute to the creation of large pools of well-documented and accessible road data, i.e., open and with known velocity, volume and variety. In some sectors within Transport and some regions, this is well understood and we see that opening the data often brings added-value services. By providing expertise, more flexible tools and demonstrations, the BigDataEurope project will help to convince authorities and other players to open their data and focus on building new services for the end users.

One of the major cross-sectoral concerns is the protection of private data, especially any geolocated data. Here, the project could help by outlining best practices of how to deal with privacy protection while preserving the data's value.





## 2.2.8 Appendices

### 2.2.8.A Workshop Invitation Letter

**Dear Sir/Madam,**

We would like to invite you to a workshop on **Big Data in the H2020 Societal Challenge: Smart, Green and Integrated Transport**, to be held in **Bordeaux** on **7 October 2015 8h30-13h30**. As a recognised stakeholder in the European Transport sector, you will have the opportunity to influence the design, and ultimately benefit from, the Big Data platform that the [BigDataEurope](#) project will deliver. This platform aims to facilitate Big Data usage in real world examples, and will consist of an architecture, components, guidelines and best practices to make the best of Big Data in the setting of Transport.

The [Transport Societal Challenge](#) set its eyes to contributing to the following objectives: resource-efficient transport that respects the environment; better and more informed door-to-door mobility; less congestion and unforeseen delays; safer and more reliable multi-modal mobility; secured exchange of personal information; global leadership for the European transport industry; and socio-economic and behavioural research and forward looking activities for policy making.

The workshop targets European stakeholders to participate in the definition of the **Big Data needs and requirements for the European intelligent transport domain**.

The results will be used to design and realise the required ICT infrastructure and support the use and deployment of the platform – maximising the opportunities of the latest European RTD developments, including multilingual data harvesting, data analytics, and data visualisation.

For further information on the workshop and the preliminary agenda, please [click through](#).

To register for the workshop please follow the registration process on the [ITS World Congress website](#).

We would like to further invite you to:

- Join the discussion on the BDE-Smart, Green and Integrated Transport Community Group <https://www.w3.org/community/bde-transport/>
- Subscribe to the BigDataEurope [Newsletter](#)
- Follow us on Social Media ([Twitter](#), [LinkedIn](#), [SlideShare](#), [Flickr](#))

This invitation is **extended to others in your network** with an interest in the challenges of Big Data in your sector. Where possible we would be grateful if you could focus on informatics professionals who have current experience of working with Big Data on a daily



basis. When forwarding the invitation, please include [m.flament@mail.ertico.com](mailto:m.flament@mail.ertico.com) in the carbon copy field.

We look forward to your participation!

Kind regards,  
Maxime Flament  
ERTICO

**2.2.8.B Slides & Presentations**

1. [Big Data Europe & Transport](#) (Simon Scerri)
2. [Logistics and big data](#) (Germán Herrero)
3. [Data analyses in transport](#) (Helena Gellerman)
4. [Existing tools and technologies](#) (Simon Scerri)
5. [Vehicle sensors](#) (Seán Gaines)
6. [Traffic management](#) (Nick Cohn)
7. [Role of social media in transport](#) (Dave Marples)
8. [The use of big data for public transport performance measurement](#) (Roberto Baldessari)
9. [Big data Sources for/from Intelligent Road Transport](#) (Evangelos Mitsakis)

**2.2.8.C Photos**

Link to the [Photos](#) slideset on the public BDE Flickr account.

**2.2.8.D Follow Up Post**

A follow-up blogpost/message was shared on the BDE [website](#) and W3C [group](#).

**2.2.8.E Attendees**

#	Name		Organisation
1	Agerholm	Niels	Aalborg University
2	Baldessari	Roberto	NEC Europe
3	Bradley	Richard	Atkins
4	Christoforou	Zoi	Ecole des Ponts ParisTECH
5	Cik	Michael	Graz University Of Technology
6	Cohn	Nick	TomTom
7	Defreyne	Peter	VIM



8	Dinh	Tu-Uyen	Ministry of Transport, France
9	Gaines	Seán	Vicomtech-Ik4
10	Gellerman	Helena	Chalmers
11	Heong	Tan Kian	SMRT-Singapore
12	Herrero	German	ATOS
13	Hoefs	Wolfgang	European Commission
14	Hyen Chee	Kwek	SMRT-Singapore
15	Ito	Hiroshi	Japan Automobile Research Institute
16	Langheim	Jochen	STMicroelectronics
17	Laoide-Kemp	David	National Roads Authority
18	Lefevre	Pauline	Microsoft Norway
19	Marples	Dave	Technolution B.V.
20	Mitsakis	Evangelos	CERTH
21	Namaki Araghi	Bahar	City Of Copenhagen
22	Otaegui	Oihana	Vicomtech-Ik4
23	Pattara-aikom	Wasan	National Electronics and Computer Technology Center - Thailand
24	Potters	Paul	Cachelot
25	Raper	Jonathan	TransportAPI
26	Riederer	Markus	Swiss Federal Roads Office
27	Roesems	Gisele	European Commission
28	Sannomiya	Chihiro	Toyota InfoTechnology Center
29	Scerri	Simon	Fraunhofer IAIS
30	Tierolf	Jan-Willem	Rijkwaterstaat
31	van Gaever	Alain	European Commission
32	Vlassenroot	Sven	Flanders Institute Mobility
33	Xiu Juan	Angel Chan	SMRT-Singapore
34	Kalasek	Daniel	H-Comp Consulting
35	Canel	Annie	ASFA

## 2.3 SC6.1 - Societies in a Changing World

The following table includes a summary of the workshop:

Date	18.11.2015
------	------------



Venue	EDF2015 co-located event, Eurostat, BECH building, Luxembourg City, Luxembourg
Invitations Sent	60
Invitations Accepted (Registrants)	68
Attendees (Total)	42
Attendees (Project Consortium & Project Officer)	9
Attendees (Other)	33
Breakout Sessions	4

### 2.3.1 Agenda

<ul style="list-style-type: none"><li>⊙ 08:30 - 09:00 Registration and Coffee</li><li>⊙ 09:00 Introductory Session<ul style="list-style-type: none"><li>○ 09:00 - 09:10 Welcome by Big Data Europe &amp; Eurostat (Michail Skaliotis, Head of the Eurostat Big Data Task Force)</li><li>○ 09:10 - 09:30 The European Commission's Perspective on Big Data Challenges, Kimmo Rossi, European Commission, DG CONNECT</li><li>○ 09:30 - 09:45 The rationale behind Big Data Europe and this workshop, Sören Auer, Fraunhofer IAIS, Coordinator of Big Data Europe</li><li>○ 09:45 - 10:15 Big Data Europe platform requirements &amp; draft architecture: Results of the online survey, Martin Kaltenboeck, Semantic Web Company</li><li>○ 10:15 - 10:45 The challenges and requirements for efficient data management in the domain of official statistics, Fernando Reis, European Commission, Eurostat</li></ul></li><li>⊙ 10:45 - 11:15 Coffee break</li><li>⊙ 11:15 - 12:15 2 Parallel Sessions (2 rooms / 2 groups)<ul style="list-style-type: none"><li>○ Parallel Session 1: Data in place in the Social Sciences and Humanities</li><li>○ Parallel Session 2: Risks and challenges of successful data management in the Social Sciences and Humanities</li></ul></li><li>⊙ 12:15 - 13:15 Lunch break</li><li>⊙ 13:15 - 14:15 2 Parallel Sessions (2 rooms / 2 groups)<ul style="list-style-type: none"><li>○ Parallel Session 3: Technological demands of data in the SSH</li><li>○ Parallel Session 4: Legal and policy demands of data in the SSH</li></ul></li><li>⊙ 14:15 - 15:00 Q&amp;A, Wrapping up</li></ul>
--

### 2.3.2 Expectation and Background



Role of CESSDA as networking partner in the BDE consortium is to coordinate the Societal Challenge 6 Interest Group: “Europe in a changing world - Inclusive, innovative and reflective societies”, and potential users of big data in the fields of social sciences and humanities (SSH). Apart from building the interest group and collecting its requirements, CESSDA should assist the building of the ICT big data infrastructure access point for social sciences and humanities, explore and evaluate the input data, and discover the implications for the future of big data in social sciences and humanities. It is done in collaboration with the SC6 technical partner, SWC from Austria.

Being a consortium itself with a large network of members, observers and engaged non-members, CESSDA aims to ensure that all relevant stakeholders in this domain, in Europe and beyond, are reached out to and engaged with the BDE project using internal communication channels, as well as the W3C Community infrastructure. As a recognised stakeholder in this field, participants in the project will have the opportunity to influence the design, and ultimately benefit from the platform that the project aims to deliver.

The first of three annual workshops envisaged within the Work Package 2 “Community Building & Requirements” took place on 18 November 2015 and introduced the background, covered the main challenges, and sought real examples of the potential, challenges and complexities of using big data in our societies. Workshop should provide additional information for the social sciences and humanities’ requirements elicitation process, their definition of and prioritization regarding the specific nature of the SC6.

### 2.3.3 Summary of Breakout Groups

#### **SC6.1 - Group 1: Data in place in the Social Sciences and Humanities**

In this session, participants were asked to list the most important ‘big’ data sources they work with, in an attempt to understand the characteristics and complexity of data sources in the SC6 community. Of all the societal challenges SC6 is perhaps the most open-ended in terms of a core domain. This was reflected by the variety of the participants’ background, whose interest lied in making sense of data (in particular figures and statistics) coming from various fields and sectors, e.g. Health, Transport and Governance, to provide information and services to society as a whole. For this purpose, participants were also asked to describe their main use-case for collecting and analysing data from the identified sources. The former ranged from cultural heritage, job market monitoring and policy making to consumer behaviour, trend detection and crime prevention. Sources included national data sources and government data, as well as social media and data coming from other societal domains such as health, transport and mobility.

Following the outlining of the data sources and their principal use-case(s), participants discussed which in their view is/are the most challenging dimension(s) of big data, based on the four V’s: Volume, Variety, Velocity, Veracity. In contrast to other SCs, here again there was no clear winner in terms of the most pressing dimension, although velocity emerges as the least perceived challenge. The former observation could again be attributed to the fact



that SC6 has no one primary sector but is instead longitudinal against all sectors. Following the technology session, the last observation was put into question, given that the community representatives (who in their majority came from a statistical background) were not very familiar or involved in technologies dealing with high-velocity data streams, such as social media, in their use-cases.

### **SC6.1 - Group 2: Risks and challenges of successful data management in the Social Sciences and Humanities**

This session was divided in two parts: the first one was devoted to risks and challenges of successful data management in the Social Sciences and Humanities, and the second one in determining the potentials and opportunities from data management. Participants were split into three groups of five to six people and asked to brainstorm and then present the results to the rest of the group.

#### **1. Risks and Challenges**

Among the various risks identified during the session, access to data was emphasized (as well as access to the same data for several usage and that also private companies could provide more data), which can vary depending on data provider, legal restrictions of the respective country, data type, level of access granted, etc. The always-present issue of open vs. privately owned data was also discussed. Another linked issue brought up during the session was the ethical concern when accessing data, which also depends on the level of data anonymisation already carried out on the data. Also the issue of costs of anonymisation was raised. The issue of whether an analysis was replicable was brought up as anonymisation of data in association with stable data sources (along with metadata) make it possible to repeat an analysis and confirm or disprove previous findings, or even add value by adding additional variables in the set. However, if only one of those pre-requisites were to be missing, this would render replicability impossible.

Work in interdisciplinary and multilingual teams has exploded over the past 5-10 years leading to more profound and overarching results, but also confronting data management with several issues: providing multilingual data sets is time and effort consuming, leading to the lack of adequate skills to address it, and at the same time trying to cope with the speed of technological advancement, as well as encouraging people to work together from different areas of expertise (research and academic, data & IT experts, the private sector, etc.). Even with all other requirements fulfilled, data sets are not always completely usable and some degree of “cleaning” is necessary before conducting any analysis. Participants also questioned whether Europe faced a skill gap.

Further remarks concerned data silos (isolated repositories of fixed data, not openly accessible) as barriers to effective operations and a barrier to collaboration, accessibility and efficiency as well as the fact that technologies are changing so rapidly that it is hard to follow the latest trends.

#### **2. Potentials**



Most of the data in social sciences and humanities is publically available at virtually no cost. The huge amounts of data potentially available and the speed at which it can be accessed are assets. The management of such data requires considerable effort as well as new skills, which in turn leads to new job opportunities, generating new ideas and new applications for consumers. Such a propulsive environment enables monitoring and maintaining of data to be done in a completely different way. Hypothesis analysis is easier to conduct in data accessible surroundings, enabling increased reproducibility of outcomes, and finally leading to better forecasting and predictive analytics. With such mechanism in place, data produced can better inform policy & decision making, thus creating easier access to different markets (via e.g. adaptive pricing).

#### **SC6.1 - Group 3: Technological demands of data in the SSH**

Participants in this session were asked to list big data technologies that they are either already using or familiar with, and those that they are aware of and would like to learn more about, or plan to include in the architectures in the near future. The results of this exercise were discussed in an attempt to classify them according to existing big data technology groups or clusters, while keeping them generic enough for the audience.

A distinction between the listed technological components immediately became apparent: a significant amount of participants listed conventional tools that are not necessarily able to handle big data in its true sense. In fact, only the Storage components indicated are designed for or support big data (e.g. Hadoop, Virtuoso). The analysis and processing tools and solutions mentioned to an extent precede the advent of big data, and reflect the amount of technological components available (especially statistical tools like SPSS, Matlab) in this domain that have been around for decades. The question arose as to whether these tools can successfully be integrated in big data architectures. The participants struggled with naming technology that can support the import, export (and representation) of multi-source and multi-format data, as well as its fusion, re-use and sharing (and in particular, provenance).

Following the ensuing discussions, a number of big data bottlenecks and critical issues have been identified. Apart from an identified lack of awareness of velocity-related components by the community, challenges included difficulties of dealing with the increasing volumes of textual data for analysis and geo-locational tagging.

#### **SC6.1 - Group 4: Legal and policy demands of Data in the Social Science Sciences and Humanities**

In this parallel session, participants were split into three groups of five to six people and asked to brainstorm and then present the results to the rest of the group.

The reform of the General Data Protection Regulation is currently in “trilogue” phase between the three European institutions based on the document from June 2015, with a final agreement on the data protection package expected in December 2015. Practical implications are still missing as proposals regarding data transfer and storage are by no





means clear and ready to be used. Moreover, the legal framework differs in different countries and largely depends on the type of data collected. For certain types of data (e.g. medical data) it takes too long to get necessary permissions for collecting data. Another issue, especially with official statistics, is the data sources stability and trust (are the resources reliable, will repeated measurements be valid, variations in tested samples, etc.). When it comes to the data ownership, the issues of open data versus enterprise owned data emerges: some countries (e.g. Belgium) started the process of voluntary sharing of data owned by private companies (e.g. mention of the Harmonized Index of Consumer Prices - HICP regulation).

Another acute issue in scientific and official databases is data provenance; it is crucial for the validation of data. The possibility of copying data and transforming it has made it increasingly difficult to trace the origins of a data set. With this mind, participants underlined that the provenance of the information was a clear requirement for big data management. The stability of data sources and data ownership were also raised as important legal issues. Data profiling is becoming a more and more pressing issue for data protection law to address, especially as it often takes place without the individual's knowledge or consent. Although it is proposed in 2013 that the GDPR should include a definition of 'profiling' together with additional provisions to protect data subjects, the outcome is still uncertain.

Data anonymisation was also addressed. It was suggested that it should be taken into account in the research planning stage. The management of both direct and indirect identifiers before undertaking data collection produces better informed consent and requires a less resource intensive process when doing data anonymisation (e.g. administrative data, although without great research value, represents personal and sensitive information and need for its collection and later anonymisation has to be determined prior to data collection).

Final remarks were regarding the need for the constant update of the regulations since the data research community environment is very dynamic, as well as the need for constant upgrade of tools and skills in anonymised data management.





## 2.3.4 Appendices

### 2.3.4.A Invitation Letter

#### **Dear Sir/Madam,**

Big Data Europe Workshop: The challenges of big data for societies in a changing world  
November 18, 2015 @ 09:00 - 15:00

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.

The European Commission's Horizon 2020 societal challenge "Europe in a changing world – Inclusive, innovative and reflective societies" seeks to address a number of crucial challenges for the future which include inequality, social exclusion, poverty, unemployment and the economic and financial crisis. At the same time, it recognises that there is great potential for Europe through new forms of innovation and by the engagement of citizens.

We would therefore like to invite you to the following (free of charge) workshop:

WHAT: The challenges of big data for societies in a changing world WHEN: 18 November 2015, 9:00 – 15:00 WHERE: Eurostat, BECH building, Luxembourg  
AGENDA: [Agenda-BigDataEurope-Workshop-18Nov2015](http://agenda-bigdataeurope-workshop-18nov2015)

### 2.3.4.B Advertised Workshop Description and Agenda

Dear Sir/ Madam,

Thank you for registering to our Big Data Europe workshop: The challenges of big data for societies in a changing world

18 November 2015; 9:00 - 15:00, Eurostat, BECH building, Luxembourg

We are pleased to inform you that the full agenda is now online.

It is important that you let us know if you are no longer able to participate, as the event is sold out and we have a long waiting list. Many thanks in advance.

If you are intending on attending our workshop, we would like to kindly ask all participants to fill in the following form by 30 October at the latest: <http://goo.gl/forms/S7IRYfgxWB>

The purpose of the form is to help us better prepare the workshop and it also gives you a chance to contact us if you have any questions.



We would also like to remind you that you will need to bring with you a valid form of identification (passport, ID card, driver's licence) on the day to be granted access to the venue of the workshop, the Eurostat building.

We are very much looking forward to meeting you in Luxembourg!

Ivana Versic (CESSDA) & Martin Kaltenböck (SWC)

The BigDataEurope Consortium  
<http://www.big-data-europe.eu>

Big Data Europe Project Coordinator  
Fraunhofer IAIS  
Schloss Birlinghoven  
53757 Sankt Augustin  
Germany

#### 2.3.4.C Attendees

42 participants attended the workshop: total of 63 had registered with 28 on a waiting list. Participants were relatively evenly split up between domain experts (academics), strategic decision makers (EU/EC bodies representatives) and technical experts, with few participants from the industry sector. The workshop hosted speakers with a policy and data management/providing background (European Commission, Eurostat) and a technical background from within the BDE project (Fraunhofer IAIS, SWC). The workshop consisted of an opening joint session, followed by two parallel sessions and a final joint wrapping up session.

#### 2.3.4.D Follow-up Post and Message

A follow-up post was shared through the BDE [Website](#) and on the [W3C Group](#).

A follow-up message sent to stakeholders highlighting comms channels, survey and workshop summary.

Dear Sir/Madam,

Thank you for your participation and proactive contributions to our Big Data Europe workshop:

**The challenges of big data for societies in a changing world**, which took place on 18 November 2015, from 9:00 - 15:00 at Eurostat, BECH building, in Luxembourg. We have received a lot of positive feedback on the event and have also learned a great deal



concerning the requirements, challenges and risks as well as opportunities and potentials of (big) data management in the

social sciences and humanities for the Big Data Europe project.

As a follow-up to our workshop, we have put together **a brief summary blogpost** providing the following information:

- Workshop Report
- Workshop Slides
- Workshop Flickr photo collection

In case you have any questions on this material and/or the workshop or the BDE project itself, please do not hesitate to get in touch with us.

To stay in touch with the BDE project and its results you can use the following resources:

- Visit the [BDE Social Science Website](#)
- Subscribe to the [BigDataEurope Newsletter](#)
- Join our social sciences [W3C Community Group](#);
- Social Media ([Twitter](#), [LinkedIn](#), [SlideShare](#), [Flickr](#)).

We are already planning our activities for 2016 and will publish information regarding the future workshop and webinar on the website and via social media when it becomes available, so do follow us.

Thank you again for supporting and participating in our Big Data Europe workshop.

We are very much looking forward to seeing you again soon.

Yours sincerely,

Ivana Versic (CESSDA) & Martin Kaltenböck (SWC)

The BigDataEurope Consortium  
<http://www.big-data-europe.eu>

#### **2.3.4.E Photos**

Photos from the workshop have been shared on [Flickr](#).

#### **2.3.4.F Slides & Presentations**

1. [From your data to data stories](#) (Anna Triantafillou and Vangelis Karkaletsis)
2. [Big data \(phenomenon\) challenges and requirements in official statistics](#) (Fernando Reis)
3. [Big Data Europe platform requirements and draft architecture: The results of the online survey](#) (Martin Kaltenböck)
4. [What can big data do for you?](#) (Sören Auer)
5. [Official Statistics in the Age of Big Data](#) (Michail Skaliotis)

#### **2.3.4.G Group Questions**

Questions per group based on the WP2 elicitation spreadsheet:



## Key Questions of the interactive workshop sessions

### Session 1 - Data in Social Sciences

- YDS intro - Pilots/invitation (if possible in time) - Are there data stories you can think of that could be Pilots for BDE in the form of a co-operation?
- What are the most important data sources in social sciences available / you are using (open / closed)?
- How are the characteristics along the 4 Vs of Big Data regarding such sources (Volume - Variety - Velocity - Veracity)

### Session 2 - Challenges & Risks

- What are the most important challenges in data management in social sciences
- What are the most dangerous risks you can think of regarding data management in social sciences
- If possible: start a SWOT - Data Management in Social Sciences
  - Strength
  - Weaknesses
  - Opportunities
  - Threats

### Session 3 - Technology

- What technologies are in place in your organisations?
- What technologies are on your roadmap - or are you evaluating at the moment?
- What are the most critical technological issues?

### Session 4 - Legal & Policies

- Open Versus Closed data in social sciences?
- What are the most important legal issues in place?
- What needs to change regarding Policies to enable more efficient data management in social sciences?
- Please tell us cases about sensitive data in social sciences & humanities

## 2.4 SC7.1 - Secure Societies

The following table includes a summary of the workshop:

Date	30.09.2015
Venue	Spanish Office of Science and Technology (SOST), Rue du Trône 62, Brussels, Belgium



Invitations Sent	150
Invitations Accepted (Registrants)	68
Attendees (Total)	45
Attendees (Project Consortium & Project Officer)	8
Attendees (Other)	37
Breakout Sessions	2 interactive sessions

## 2.4.1 Agenda

- 10.30 - 10.40 Welcome and Workshop presentation (Gisele Roesems-Kerremans, EC CNECT.G3)
- 10.40 – 11.40 Introduction Talks
  - 10.40 – 11.00 Big Data and Societal Challenges (Gisele Roesems-Kerremans, EC CNECT.G3)
  - 11.00 – 11.20 The BigDataEurope project (Sören Auer, Fraunhofer IAIS)
  - 11.20 – 11.40 Big Data and Secure Societies (Sergio Albani, EU SatCen)
- 11.40 - 12.40 Invited Talks
  - 11.40 – 12.00: Space Data for Secure Societies (Sakellaris Hourdas, EC GROW.I3)
  - 12.00 – 12.20: Social and Open Source Data for Secure Societies (Jean-Dominique Nollet, EUROPOL EC3)
  - 12.20 – 12.40: Big Data challenges in Cybersecurity and Trust (Pierre Chastanet, EC CNECT.H4)
- 12.40 – 14.10 Lunch
- 14.10 – 15.10 Interactive Session 1: Big Data Technologies in Secure Societies (Chair: Vangelis Karkaletsis, NCSR Demokritos)
  - Topic 1: Technical challenges and issues
  - Topic 2: Data science and advanced analytics
  - Topic 3: Architectures and infrastructural layers
- 15.10 – 16.10 Interactive Session 2: Big Data Users in Secure Societies (Chair Manolis Koubarakis, University of Athens)
  - Topic 1: User requirements and needs
  - Topic 2: Pilots, use cases and scenarios
  - Topic 3: On-going projects and initiatives
- 16.10 – 16.30 Summary, Outreach and Feedback (Sergio Albani, EU SatCen)
  - Closing Note and Farewell (Gisele Roesems-Kerremans, EC CNECT.G3)



## 2.4.2 Expectation and Background

The EU SatCen in the framework of the BigDataEurope (BDE) project organised a workshop on “Big Data in Secure Societies” for the “Secure Societies” Horizon 2020 Societal Challenge on 30 September 2015 in Brussels; the workshop was the first of a scheduled series in the BDE project for the Secure Societies domain.

### 2.4.2.1 Background

The “Secure Societies” Horizon 2020 Societal Challenge is related to the protection of freedom and security of Europe and its citizens. An example of major activity in supporting the primary aims of this Societal Challenge is the provision of geospatial products and services, mainly resulting from satellite data; the datasets used in the Space and Security domain comply with the definition of Big Data in terms of volume, variety, velocity, veracity and value.

### 2.4.2.2 Workshop aims

With dedicated sessions on specific topics led by specialists from the BDE project and the Security domain, the workshop aimed to:

- Identify current and future challenges for Big data and data management in the “Secure Societies” societal challenge;
- Build a Security community involving relevant entities through the collection of user requirements and Big Data implementation strategies;
- Show real world examples and use cases;
- Highlight the current state of the BDE project and the pilot case for the Security domain;
- Support the design and realization of the necessary ICT infrastructure on which the deployment and use of the BigDataEurope platform will be based.

The aim was to provide stakeholders in the Security domain with the opportunity to influence the BDE project and the development of the Big Data platform for security as well as to address a wide audience comprising data users from a variety of fields in the Security domain.

### 2.4.2.3 Workshop general information

The workshop was organised by the EU Satellite Centre and hosted by the Spanish Office for Science and Technology (Rue de Trone, 62) in Brussels on 30 September 2015 between 10.30 and 16.30. The workshop was divided in two main sessions: the morning session consisted of an introductory part followed by invited talks while the afternoon session consisted of two interactive discussions. The workshop had a 66% turnout, with 68 registered participants and an actual number of 45 attendees from EC (DG CNECT, DG GROW, DG HOME, DG RTD), JRC, EDA, EUROPOL, EDA, FRONTEX, ESA, CDTI, DLR, ASD-EUROSPACE, EARSC and a number of other entities and private companies covering the domains of Space and Security, Cybersecurity, Fight against Crime and Data/Infrastructures. Working sectors were represented as follows: 33 % Space and Security; 29 % Data/Infrastructure; 21 % Cybersecurity; 17 % Fight against Crime.



### 2.4.3 Introductory Talks

The Introduction Talks session aimed to offer an overview of Big Data challenges and opportunities in Europe, focusing on the societal challenge “Secure Societies”. The workshop welcome and first introductory talk was given by Gisele Roesems-Kerremans (EC CNECT.G3); her presentation highlighted the importance of the BDE project for the EC, in line with the Big Data communication action defined in 2014 to boost jobs in data economy.

The second presentation was given by Soren Auer (Fraunhofer IAIS), BDE Project Coordinator. He showed how BDE aims to leverage the societal value of the Big Data for the seven Horizon 2020 Societal Challenges through three main activities: gathering the user requirements, developing a technical platform and proposing possible applications/pilots.

The most suitable technical solutions to build the platform have to be adopted considering the large variety of tools available on-line. The main challenges of Big Data are the so-called 3Vs: Volume, Velocity and Variety. The state-of-the-art approach, called lambda architecture, is structured to address the volume (historical data that are processed by the batch layer) and the velocity (real time data that are processed by the speed layer); however, variety is missing. BDE will cover this gap by introducing a semantic layer for variety, which will make the data integration more flexible. BDE will also develop a platform that will require less technical skills for dealing with Big Data.

The Introduction Talks session was concluded by the presentation of Sergio Albani (EU SatCen), Secure Societies domain leader; his presentation illustrated the EU SatCen role in the BDE project. SatCen is addressing the “Secure Societies” Challenge focusing on the building of a Secure Societies Community (eliciting needs and requirements of relevant stakeholders related to Big Data management and exploitation) as well as on planning, developing and evaluating the instantiations of the BDE stack through appropriate pilot trials in real-world scenarios for Secure Societies. Preliminary user requirements were collected from interviews with SatCen staff, EU entities, international organizations and industry representatives. The requirements were categorized in accordance with the data value chain (management and exploitation) and the platform characteristics. These requirements will lead to the development of the first pilot in collaboration with the BDE Secure Societies technical partners (University of Athens and NCSR-Demokritos). The pilot will consider two different workflows of data: the first one is related to the detection of changes in areas of interest using satellite images and successive verification using (social) media information; the second workflow is related to the event detection using (social) media and the verification using satellite images.

#### 2.4.3.1 Invited talks

The Invited Talks session intended to cover the main areas of the Secure Societies challenge: Space and Security, Fight against Crime and Cybersecurity.





The presentation on Space and Security was delivered by Sakellaris Hourdas (EC GROW.I3), who presented the Copernicus programme with a focus on Secure Societies and Big Data challenges.

A presentation on Fight against Crime was delivered by Jean-Dominique Nollet (EUROPOLEC3).

EUROPOL is the EU's law enforcement agency with the main goal to help achieve a safer Europe by fighting against terrorism, cybercrime and organized crime; EC3 is responsible for addressing cybercrime in the European Union and for protecting European citizens.

The talk on Cybersecurity issues was held by Pierre Chastanet (EC CNECT.H4), who described how cybersecurity is a fundamental sector in security scenarios, ranging from IT infrastructures development to data security.

## 2.4.4 Summary of Breakout Groups

### **SC7.1 - Group 1 : Secure Societies: Pilot Implementation Challenges**

The first interactive session was chaired by Vangelis Karkaletsis (NCSR Demokritos), technical partner in the development of the Secure Societies pilot. The aim of the session was to show an overview of the technical challenges and issues for the platform development and pilot implementation, in order to discuss with the audience the architecture to adopt and the advance analytics tools to use.

The lambda architecture was described along with the available existing technologies (not only the well-known Hadoop technologies, but also tools already implemented in other EU projects). A first selection amongst the Big Data technologies in the frame of the BDE project has already been done, considering the requirements of the pilot.

In particular, the Security pilot will cover all the challenges of Big Data: volume (wide-satellite images), variety (heterogeneous sources of data, as images and text from news and social media), velocity (fast-paced social data - news stream), veracity (cross-verification of the sources) and value (adding useful information). Apart from the established Big Data tools in the frame of the lambda architecture, specific tools for image processing (e.g. change detection), text mining (e.g. clustering, crawling), geo-data storage & access (Semagrow) as well as a GUI (Sextant) will be considered.

The first workflow of the pilot is defined as change detection workflow and it is based on remote sensing data: tools for satellite images selection, download and processing (such as co-registration and change detection) are already available on the Sentinel Scientific Data Hub and Sentinel 1 Toolbox/SNAP. The first phase of the pilot will integrate these tools into the BDE platform, scaling their operations to large volumes of data by transforming them into a parallel procedure. If a change is detected using satellite images, the social sensing part of the workflow will be activated in order to add information to the satellite information and to verify the event. The second workflow of the pilot is called event detection workflow and it is based on social sensing data: it extracts information from news and social media, whose outcomes activate the change detection tool, focusing its processing on a specific area. The





weight that will be assigned to the two sources of information (remote sensing and social sensing) in the final decision making is still under discussion. During the session a presentation was given by Mark Last (Ben-Gurion University of the Negev), expert on text mining for security and cybersecurity applications.

### **SC7.1 - Group 2: Technical Requirements**

The second interactive session was chaired by Manolis Koubarakis (University of Athens), Technical Leader of the Secure Societies domain. This session aimed at presenting some ongoing projects and initiatives in order to collect more ideas for the development of the Security pilot.

The session was also dedicated to collect requirements from the audience in addition to those highlighted during the day (e.g. integration of ethical issues through ethical guidelines in the project, use of data encryption tools, use of additional data sources to increase the variety of the data such as tracking systems and communications logs).

During the session a presentation was given by Mihai Datcu (DLR), who described a tool for data mining / information extraction from satellite images (mainly high resolution SAR images).

## **2.4.5 Conclusions**

The conclusions were presented by Sergio Albani (EU SatCen). The event was successful with several participants covering different sectors of the Secure Societies domain and speeches given by key representatives from EC, EUROPOL-EC3 and the BDE project. The interactive sessions, led by the BDE Secure Societies technical partners, gave the audience the possibility to provide useful recommendations and comments which will be considered in the platform setting, in the pilot development and for future initiatives. It was also announced that most of the material will be available online and the participants were encouraged to check the BDE website in order to follow up the project and to be actively involved in it. The workshop was then officially closed by Gisele Roesems-Kerremans.



## 2.4.6 Appendices

### 2.4.6.A Invitation Letter

**Dear Sir/Madam,**

In the framework of the BigDataEurope project, we would like to invite you to the 1st Workshop on “Big Data in Secure Societies”; the event will be held the 30th of September 2015 in Brussels at the Spanish Office of Science and Technology (SOST), Rue du Trône 62.

Participating to this interactive workshop you will have the possibility to:

- Identify current and future Big Data needs and challenges in the H2020 Secure Societies Societal Challenge;
- Contribute to the building of a Security Community aiming at collecting Big Data user requirements and sharing Big Data strategies;
- Show and discuss real user scenarios and use cases, e.g. related to the provision of products and services in the Space and Security domain;
- Influence the BigDataEurope project and define the pilots for the Security domain;
- Support the design and realization of the necessary ICT infrastructure on which the deployment and use of the BigDataEurope platform will be based.

*Event Registration:* More information about workshop registration, agenda and venue can be found here: <http://www.big-data-europe.eu/event/sc7-brussels-2015/>

This invitation is extended to others in your network with an interest in the challenges of Big Data in the Security domain.

We look forward to your participation!  
Best Regards,

Sergio Albani (EU SatCen)  
Secure Societies Domain Leader

### 2.4.6.B Attendees

The workshop had 45 attendees from the European Commission (DG CNECT, DG GROW, DG HOME, DG RTD), JRC, EDA, EUROPOL, EDA, FRONTEX, ESA, CDTI, DLR, ASD-EUROSPACE, EARSC and a number of other entities and private companies covering the domains of Space and Security, Cyber-security, Fight against Crime and Data/Infrastructures.

### 2.4.6.C Follow-up Post

A follow up message was share on the BDE [Website](#) and the respective W3C [group](#).

### 2.4.6.D Photos

[Photos](#) from the workshop are available online.



#### 2.4.6.E Slides & Presentations

1. [Towards a data-driven economy in Europe](#) (Gisele Roesems - Kerremans)
2. [The Big Data Europe project](#) (Soeren Auer)
3. [Big Data and Secure Societies](#) (Sergio Albani)
4. [Space Data for Secure Societies](#) (Sakellaris Hourdas)

## 3. Summary

The workshop reports provided in this deliverable cover the related BDE WP2 activities in the second six month period of the project. These reports supplement the first three covered in the first deliverable in this series (D2.2 Report on Interest Groups Workshop I).