WORKSHOP REPORT

BIG DATA EUROPE

# First Community Workshop for Societal Challenge 5 Climate Action, Environment, Resource Efficiency and Raw Materials

***Location*** *- Vertretung des Landes Nordrhein-Westfalen bei der Europäischen Union, Brussels*
***Date*** *– June 15, 2015*

## 1. AGENDA

12:00 - 13:30

| | |
|---|---|
| Welcome – | M. Vlachogianni |
| Round Table Introduction | |
| Introductory Talks | |
| Big Data Europe | S. Auer |
| Big Data Europe Technology Overview | A. Ikonomopoulos |
| Breakout Group Sessions | S. Andronopoulos |

*13:30-14:00 Lunch Break*

14:00-15:15 Keynote Presentations

Ioannis MALLAS (ECMWF) Big Data in ECMWF

Rasmus E. BENESTAD (Norwegian Meteorological Institute) Climate Analysis and Big data

Antonio S. COFIÑO (Group of Meteorology and Computing. University of Cantabria)
*End-user gateway for climate services and data initiatives*

*15:15-15:30 Coffee Break*

15:30-16:20 Keynote Presentations

Manolis KOUBARAKIS (Department of Informatics and Telecommunications, University of Athens) *Big, Linked and Open Earth Observation Data: the projects TELEIOS and LEO*

Alexis SANCHO REINOSO (Centre for Global Change and Sustainability, BOKU University)
*Improving climate research societal benefit: JPI Climate strategy on Open Knowledge*

16:20-18:00 Breakout GROUP SESSION

- Data-centric initiatives in Climate
- Technologies and data
- EU Policy Requirements

18:00-18:30 Summary, Outreach & Farewell

## 2. EXPECTATION AND BACKGROUND

Climate research is heavily based upon computer models that simulate the earth's climate for time periods spanning several decades. These three-dimensional global models discretize the entire earth surface and atmosphere to a resolution that has recently gone down to a few tenths of kilometres resulting in billions of grid cells. In this framework, millions of weather observations (including Earth Observation data) are collected and assimilated on a daily basis, while past observational data are re-analysed and climate simulations are performed producing massive amounts of data of the order of terabytes per day. Repeated climate simulations are carried out considering different scenarios of world-wide emissions of anthropogenic and natural pollutants to study their effects in the climate and the climate change societal impacts.

As the atmosphere and its interactions with the land and surface constitute a complex dynamical system, a lot remains to be understood about the earth's physical processes. Given the abundance of climate data from model simulations, Earth orbiting satellites and in situ observations, scientific efforts are heavily placed to narrow the knowledge gaps by directly handling these large data sets. Currently, the progress in climate science induced by Big Data is slow although this field has become one of the most data rich domains in terms of volume, velocity and variety.

Therefore management and manipulation of climate models simulations' results is a Big Data challenge for the organizations engaged in climate research and services and involves techniques and tools for storage, analysis and visualisation in order to extract useful conclusions. It also requires techniques and tools for combination of climate models results with data from other areas, e.g., agricultural production, population distribution, economic activities, etc. Big Data management and analytics of global climate models' results can be used to address real world impacts of climate change.

The aim of the first BigDataEurope SC5 workshop was the identification of current and future challenges for Big Data and data management in the Climate and integration of Earth Observation data domain. The Big Data of SC5 focus mainly on real-time monitoring, stream processing and data analytics. In the workshop, real examples of the challenges and complexities of using Big Data in the domain were presented and discussed.

The workshop focused on the elicitation of the user requirements to support the design and realization of the necessary ICT infrastructure on which the deployment and use of the BigDataEurope platform (aggregator) could be based. The platform targets the facilitation of Big Data usage in real world examples. The BigDataEurope platform will offer the opportunities of the latest RTD developments to the interested participating third parties, including real time streaming, multilingual data harvesting, data analytics and data visualisation.

The workshop was structured with three sections. The first involved a general introduction to the BDE background, objectives and targets, as well as an overview of the tools and technologies envisaged within the project. An extended summary of the BDE technology overview is provided in the next section of this report (Section 3). The workshop participants were further on introduced to three sets of questions selected to map the user requirements and contribute to the Requirement Elicitation (RE) and the Requirement Specification (RS) activities in work package 2.

In the second workshop section, five keynote presentations were given by invited speakers, illustrating Big Data and data management activities and experiences from the data services, academia and research sectors:

- The first presentation (by I. Mallas) introduced the audience to the current Big Data activities at the European Medium Weather Forecast (ECMWF) involving data acquisition, processing and provision to a great number of weather and climate services worldwide.
- The second presentation (by R. Benestad) discussed the climate and weather analysis carried out currently at the Norwegian Meteorological Institute and presented an in-house built open source Big Data tool for data analysis, statistical downscaling and visualization.
- The third presentation (by A. S. Coffino) provided information on the End User Gateway for climate services from the Meteorology Group of the University of Cantabria and other data Initiatives (e.g. ECOM, COST “VALUE”, WCRP-CORDEX) operating at European and World-wide level.
- The fourth presentation (by M. Koubarakis) presented highlights of the two EC projects TELEIOS and LEO for managing big, linked and open earth observation data, coordinated by the University of Athens .
- The fifth presentation (by A.S. Reinoso, BOKU University, Vienna), discussed the Joint Initiative Programme (JPI) of Climate Strategy on open knowledge and data transparency for improving the climate research societal benefit.

The third section constituted the interactive part of the workshop during which the participants were split into different breakout sessions according to three types of stakeholder groups to serve the user requirements elicitation process. The results of the discussions are summarized in the following sections of this report (Sections 4, 5 and 6).

## 3. BDE Architectural Overview, Background and Expectations

The BDE platform is built upon existing Big Data industry best-practices making use of the Lambda-Architecture[1] that constitutes a generic, scalable and fault-tolerant data processing architecture. The envisaged implementation integrates mature, existing, open-source components into a comprehensive software stack suitable for serving and consuming interoperable data. The platform will be available as an open source implementation maximizing software re-usability and community involvement, while paving the new comer path to data products and services.

The architecture of the BDE platform is tailored to consume high-volume streams of real-time data (e.g. sensor measurements, social network activity, mobility data) and process them in two parallel pipelines, namely the:

- **batch pipeline:** that handles data at preset time intervals (e.g. hourly/daily) using Map-Reduce algorithms to provide aggregated views and
- **real-time pipeline:** that interactively manipulates incoming data and provides data views up to a certain timeframe

Both pipelines provide an integrated Big Data view during short and long time periods. An early platform configuration is an archetypal blueprint that serves as the basis for revisions and improvements in accordance to stakeholder requirements from the seven societal challenges. The tight alignment of the development process with the societal challenge needs ensures the seamless and transparent integration of the proposed implementation in existing solutions.

The BDE project assumes **Hortonworks**[2] as its starting point: Hortonworks integrates components developed within the Apache Software Foundation[3] governance model to implement the Lambda architecture. This will be extended by the project into a powerful solution for managing and automating large-scale infrastructures that supports, among others, role-based access control and reporting. Tools will be integrated as part of the core aggregator and the ecosystem around it while additional components will be included for domain-specific analysis and visualization tasks. The outcome will accommodate machine deployment on any cloud platform to which it has access.

[1] Nathan Marz and James Warren: Big Data: Principles and best practices of scalable realtime data systems, Manning Publications Co., April 2015
[2] http://hortonworks.com/innovation
[3] http://www.apache.org

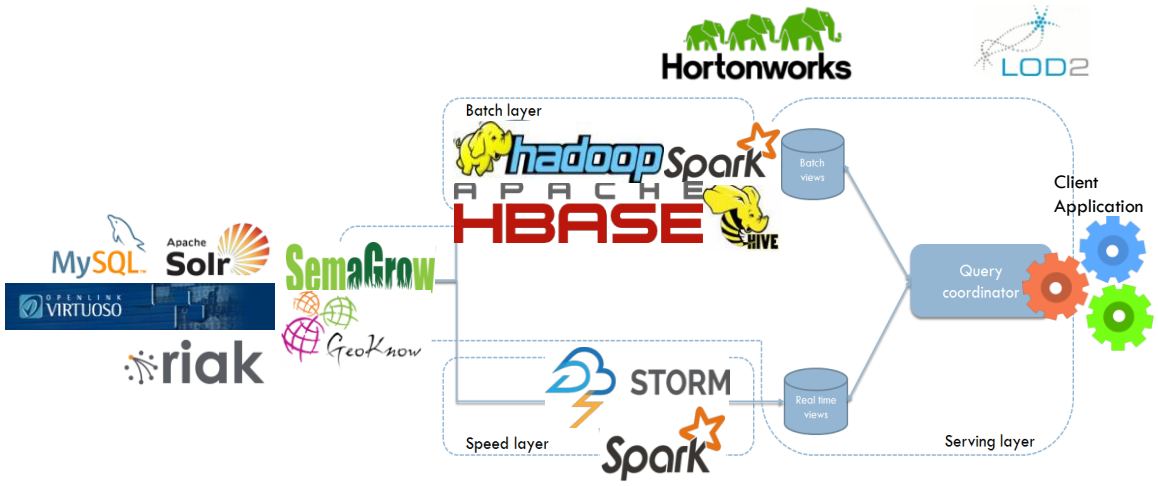


**Figure:** BDE Architecture

The figure outlines an initial set of core components and their position on the Lambda Architecture. Besides the components integrated by Hortonworks for batch and stream processing (Apache Hadoop[4], Apache Spark[5], Apache Storm[6]) as well as distributed storage (Apache HBase[7], Apache Hive[8]), the figure presents additional components that will be integrated at both ends of the core pipeline. These include well-established data management solutions such as the Apache Solr[9] text database, the Riak S2[10] object store, along with the OpenLink Virtuoso[11] and MySQL[12] databases.

The most notable extensions will however be solutions that facilitate heterogeneous data integration. Members of that group are the **SemaGrow**[13] federated querying infrastructure that transparently relegates and translates queries from a unifying end-point to a heterogeneous federation of *relational* and *numerical* data services as well as the **GeoKnow**[14] framework that integrates *relational* and *GIS* data into a scalable spatial knowledge store. The BDE platform functionality will be further augmented with the **LOD2**[15] component stack, a collection of tools that support Linked Data publication and analysis.

[4] http://hadoop.apache.org
[5] https://spark.apache.org
[6] https://storm.apache.org
[7] http://hbase.apache.org
[8] https://hive.apache.org
[9] http://lucene.apache.org/solr
[10] http://basho.com/products/riak-s2
[11] http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main
[12] http://www.mysql.com
[13] http://www.semagrow.eu
[14] http://geoknow.eu
[15] http://lod2.eu

## 4. SUMMARY OF BREAKOUT GROUP - 1

**Group 1 – Report of the Discussion Regarding Data Centric Initiatives in Climate**

## Preamble

The basic task of Group 1 was to elaborate on the current state of the art of Big Data in SC5, discussing current and past projects, the gaps that need to be addressed and use cases.

Group members' knowledge and experience of Big Data and data management in the climate sector was fairly high-level, so the discussion covered mainly broad topics and ideas rather than specific examples of work being done in this sector.

The group discussed the general attitudes of people involved in the climate sector towards big data usage and potential. The points raised were the following:

- Access of users to Big Data is an issue of concern, in terms of their ability to access in a form that they can integrate in their processes and ability to store and manipulate.
- There is a need to develop a common language based on what the users need.
- The users must have the ability to build their own case using predefined questions.
- The big data users can be categorized into different levels depending on their use cases.
- Building a small inventory of questions of what people need can help the user to access the data needed.
- There is a general sense of urgency about moving forward.
- From a knowledge worker's view what should be realized is the facilitation of the data collection and integrated management of large (Tb to Pb) datasets from diverse knowledge domains for interdisciplinary research, thus allowing researchers to concentrate on knowledge rather than on data management.
- Various tools and management practices are common:
    - GIS data management and analysis tools, statistical tools
    - Custom-developed data processing tools and procedures developed in R, Python, Matlab
    - General purpose libraries and suites, e.g., NetCDF Tools, NCO, CDO
- Tasks involved in handling large datasets from users were stated to be:
    - Data collection from more than 30 global and regional models and observational data
    - Data processing for impact modelling analysis
    - Remote sensing research, agro-environmental modelling,
    - GIS analysis
- What is lacking are reliable methods and (semi) automatic procedures to discover and integrate heterogeneous data from different domains and also,

as a basis for that, the integration of currently scattered semantics / ontologies from diverse domains with sufficient detail to be able to describe sources on the data level.

- There is a great benefit for improving and incorporating the Big Data approach as it is a technology challenge awareness and assessment.
- The Big data management solution can be used by:
    - Internal researchers and external project partners and collaborators
    - Researchers and modellers in data-intensive research areas (e.g. life-sciences)
    - The climate change impact assessment community.

  Such a solution would dynamically collect, re-format and pre-process diverse datasets at a large scale in order to prepare a single integrated dataset appropriately formatted for custom-made analysis tools. If such pre-processing can be reliably and efficiently replicated, this alleviates the need to persist large-scale derivative datasets for the purpose of reproducing experiments.
- In some organizations, there is no procedure or best practice for administrating Data Management procedures.

## The Data Situation

The group then discussed the data situation.

- The kinds of data of the users mentioned were the following:
    - Climate Data. i.e. CMIP5 /CMIP6, CORDEX, SPECS and many others from different providers.
    - Agro-environmental data, remote sensing data, model input/output
    - Observational data
- Data for users from partners, external service providers, open data.
- The data format is important. The tendency is towards GRIBB. NetCDF is preferable for analysis. Tables of parameters, if local (e.g. from integrated models), constitute a problem. The World Meteorological Organization (WMO) is responsible for such guidelines.
- Popular formats are: Structured e.g. Multidimensional arrays. Netcdf3, Netcdf4, HDF5, GRIB1, GRIB2, and ASCII formats, GIS-databases, shape, JSON. Unstructured like documents.
- Data encoding must draw special attention.
- The format conversion is a problem to the users and not as such the format itself. It was mentioned that ECMWF is working on encoding, as it has been requested by users, on Buffer data.
- Regarding the metadata, the data exchange between different geographical regions and communities must be standardized.
- Currently there are metadata and multilingual problems. In this context, libraries or services should facilitate data access.

- Pre-knowledge must be available for metadata. Routines must be built-up to test if the data makes sense. Visualization tools need metadata not only at the final stage of processing to the users.
- The tools used for presenting the output were mentioned to be:
    - client tools like R, Python or Matlab
    - maps, graphs, integrated viewers and dashboards

The group then discussed possible pilot use cases as a test bed for the BDE platform. A typical example for a climate researcher was proposed for the geographic area of the Iberia peninsula. Large datasets from modelling experiments like CORDEX, CIMP5 are available to set up a use case touching upon issues like infographics (analysis and visualization), metadata and formats.

Finally, there was the comment that the Big Data efforts can primarily provide the means to support researchers in data-intensive science fields, in shifting the majority of their work from the data and information level towards the level of domain knowledge.

## 5. SUMMARY OF BREAKOUT GROUP - 2

**Group 2 – Report of the Discussion Regarding the Technologies and Data**

### Preamble

The basic task of Group 2 was to discuss functional and non-functional requirements for the BDE platform. Group members' knowledge and experience of Big Data and data management in the climate sector was such so that the discussion covered the following:

### Data Acquisition

During the data acquisition phase, climate domain experts identified data heterogeneity as something that BDE could address. Experts typically work with data maintained and provided in remote repositories and in different formats. The experts then are typically required to manually discover and download datasets of interest, before they process them to match their procedures, locally. Federated querying tools, such as SemaGrow16, as well as their integration with tools commonly used for pre- and post-processing (e.g. Hadoop17 MapReduce) could aid experts in this regard.

### Analysis and Processing

Climate experts indicated that they typically make use of in-house analysis tools in order to carry out their work. As a follow-up activity to that of data acquisition, this requires that data are shipped over to a local computing facility before analysis can take place. A question that was raised was whether it would be possible for analysis to take place next to the data, and therefore either avoid shipping data locally, or the expert having to cater for differences in types of data and remote or local infrastructures. Whereas BDE is not suited to adapting analysis software to domain-specific requirements, it could potentially provide for more streamlined data analysis procedures by including and, wherever possible, integrating analysis software in the tools aggregator. Solutions such as Hadoop are used in order to process massive amounts of data in-situ. BDE can therefore investigate whether and which analysis tools can be integrated with a Hadoop ecosystem – integrating components such as DFS, MapReduce, HBASE18, Parquet19, etc – and therefore be made available within the BDE aggregator.

[16] http://semagrow.eu
[17] http://hadoop.apache.org
[18] http://hbase.apache.org
[19] https://parquet.apache.org

## Storage and Curation

Regarding storage and curation, a number of items were raised, such as the frequent need for standards and ontology alignment. Already from the data-acquisition phase of the discussion it was made clear that alignment of data takes places manually, after data from different sources have been transferred locally. Coupled with a variety of analysis tools, experts end up having to cope with heterogeneous datasets which cannot be readily processed and analysed further or compared, etc. BDE is therefore called to investigate ways to bridge such differences via storage and curation solutions, in order to facilitate further processing, analysis and archiving more effectively. Semantic maps, as these will result from further discussions within BDE, in conjunction with the use of distributed querying and annotation tools, is a field BDE should investigate.

Further to the semantic alignment of datasets, the climate community also discussed the issues of data versioning and data quality. Due to the nature of climate data, being streamed time-series, as well as to the diverse processing and variety of data products, data versioning was brought forward as something desirable by the community. Identifying and dealing with incomplete and temporarily erroneous data sets was also discussed. These seemingly unrelated issues are important to big data scientific processing as well as to other fields and can be addressed by a combination of technologies and approaches, such as the use of persistent identifiers, semantic annotation and metadata as well as by the general field of data provenance. As these are still large and research-active fields and the solutions that exist are typically domain-specific, in BDE we propose to address versioning and data annotation incrementally, depending on priority within as well as the overlap across BDE communities.

## 6. SUMMARY OF BREAKOUT GROUP - 3

**Group 3 – Report of the Discussion Regarding Big Data Legal and Policy issues**

## Preamble

The basic task of Group 3 was to discuss legal and policy issues related to Big Data. The discussion was oriented around the following questions:

*What is your domain strategic vision regarding data management?*

In H2020 an Open Research Data Pilot is applied as an option to all EU-funded projects (European Commission, Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, v. 1.0, 11 December 2013, pp. 14). In this respect the EU-funded projects are asked to apply an open-access policy to research data (data produced or collected by the project). This will be optional, so that projects may opt out (totally or partially). Participating projects are asked to develop a Data Management Plan, which should specify what data will be open (European Commission, Guidelines on Data Management in Horizon 2020, Version 16 December 2013). More specifically the DMP should describe: (a) what types of data are produced or collected, (b) what standards are used, (c) whether and how it will be exploited or made accessible for verification and re-use, and (d) how it will be curated and preserved. The data access policy will depend on the owners. A template of a DMP is described in the document A.CARTIER, M.MOYSAN, N.REYMONET, Réaliser un plan de gestion de données : guide de rédaction (V1, 09/01/2015).

The value of data was pointed out as a factor that affects the data management plan. Usually “academic” and “industrial” stakeholders follow different approaches. In general academics favour the open access policies whereas industrial people are in favour of restrictions in data availability in order to protect their commercial value. However this is not always true, as in many cases academic people impose restrictions to access to data, in order to gain some advantage due to earlier publications and competitiveness for funding (a usual data management strategy is the “embargo” on data for a short period of time before they become freely available).

Satellite images have been mentioned as an example of data with great commercial potential. The challenge would be to combine an open-data policy with commercial benefit. This could be seen in analogy to open source software that has given rise to an entire economy.

*Which political framework or domain specific strategic factors can influence your decisions on your data management strategy/systems in use?*

- The EC Digital Single Market
- The existing legal obstacles, IPR, copyright legislation

- The incentives to promote open data policies
- The capacity / potential given by legislation to profit

*Can data be private and secure yet useable for research?*

- Yes, it is possible but difficult. Issues about anonymity must be considered.

*Which legal concerns could be encountered in the usage of (Big) Data management solutions?*

- There are big issues with the copyright and the IPR. Legislation is currently in preparation. It was mentioned, as an example, that use of a model in cloud computing may be violating IPR.

*What needs to be done to influence policy requirements?*

- Lobbying, public consultation – in connection to Digital Single Market.

# Appendices

## Background Material and Preparation:

### Invitation Letter

**Dear Sir/Madam,**

In the framework of the BigDataEurope H2020-ICT-2014 CSA, we would like to invite you to a Workshop on **Big Data and Data Management in the H2020 Societal Challenge (SC5) Climate action, environment, resource efficiency and raw materials,** in **Brussels,** on **June 15, 2015**. As a recognised stakeholder in the sector, you will have the opportunity to influence the design, and ultimately benefit from, the Big Data platform that the BigDataEurope project will deliver.
The aim of this workshop is the identification of current and future challenges for Big data and data management in the Climate and integration of Earth Observation data domain. In the workshop, real examples of the potential, challenges and complexities of using big data in this domain will be discussed.
The workshop output will support the design and realization of the required ICT infrastructure on which the deployment and use of the BigDataEurope platform will be based. The platform targets the facilitation of Big Data usage and Data Management in real world examples and it will consist of an architecture, components, guidelines and best practices. BigDataEurope platform will offer to the interested participating third parties the opportunities of the latest European RTD developments, including real time streaming, multilingual data harvesting, data analytics and data visualisation.
**Event Registration**: More information about the Workshop registration, agenda, venue, etc. can be found on: http://www.big-data-europe.eu/event/sc5-brussels-2015/
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 (please contact mandy@ipta.demokritos.gr 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 and data management in your sector. Where possible, we would be grateful if you could focus on informatics professionals, who have current

experience on Big Data and Data Management on a daily basis. When forwarding the invitation, please include mandy@ipta.demokritos.gr in the carbon copy field.

We look forward to your participation!

Kind regards,

Diamando Vlachogiannis,
Senior Researcher, Environmental Research Laboratory, NCSR "Demokritos"
Societal Challenge 5 Representative on behalf of the BigDataEurope Consortium
mandy@ipta.demokritos.gr
www.big-data-europe.eu

## Draft Workshop Agenda

**SC5 WORKSHOP**

The aim of this workshop, the first of a scheduled series in the BigDataEurope project, is the identification of current and future challenges for Big data and data management in the Climate, integration of Earth Observation data and Smart Cities and Sustainability domain. In the workshop, real examples of the potential, challenges and complexities of using big data in this domain will be discussed.
The workshops output will support the design and realization of the necessary ICT infrastructure on which the deployment and use of the BigDataEurope platform will be based. The platform targets the facilitation of big data usage in real world examples and will consist of the architecture, components, guidelines and best practices.
BigDataEurope platform will offer to the interested participating third parties the opportunities of the latest European RTD developments, including real time streaming, multilingual data harvesting, data analytics and data visualisation.

**Preliminary Agenda**
**SC5 CLIMATE WORKSHOP June, 15, 2015, Brussels**

- **12:00 -13:30 Welcome and Introduction**
    - **Round Table Introduction**
        - Name & Affiliation
        - Role in organisation
        - Connection to big data & data management
        - Expectations for the workshop (what to take home)
    - **Introductory Talks**
        - Big Data Europe
        - Big Data technology for Climate applications
        - Big Data & Data Management in Climate domain

- **13:30-14:00 Lunch Break**

- **14:00-17:30 Interactive Sessions**

- **14:00-15:30 Session 1: Data-centric initiatives in Climate**
    - Identification of projects and initiatives addressing the issue of Big Data in relation to Climate Research & Applications
    - Big Data use-cases in Climate, Pilots**:** discussion on potential use cases and selection of pilot cases for BDE

- **15:30-15:45 Coffee Break**

- **15:45-16:45 Session 2: Technologies and data**

- **16:45-17:30 Session 3: EU Policy Requirements**
    - Legal issues around (Big) data, Governance, Data Portability
    - Other Requirements

- **17:30-18:00 Summary, Outreach & Farewell**
    - Summary, Q&A session
    - Closing note, outreach plans

## Follow-up Message

Follow-up message sent to stakeholders highlighting comms channels, survey and workshop summary

Dear Friends,
We would like to thank you for your participation in our workshop on Big Data in the H2020 Societal Challenge Climate action, environment, resource efficiency and raw materials calls held on **June 15, 2015**.
We are working on the full summary report of the workshop and will be in touch with a link to the report soon. For now we'd like to draw your attention to other ways you can engage with, and help the BigDataEurope project.

- To keep in touch with the latest BigDataEurope project developments, please visit our website and subscribe to our Newsletter;
- The slides presented at the workshop will be available on our SlideShare
- We have established W3C Community Groups for each Horizon 2020 societal challenge, please join us there
- Participate in the other planned key Stakeholder Workshops,
- Participate in one of our BigData Pilots - please contact us directly for details
- Follow us on the Social Media (Twitter, LinkedIn, SlideShare)

If you have 12 minutes to spare, we would love to have your input to our big data survey and help set the requirements for our big data aggregator platform. Alternatively if you

have a little more time to spare, please me directly for a more in depth discussion.

Thanks for your participation!
Kind regards,

Diamando Vlachogiannis,
Societal Challenge 5 Representative on behalf of the BigDataEurope Consortium
mandy@ipta.demokritos.gr
www.big-data-europe.eu

## Photos

Photos from the workshop ( BDE Flickr site)

## Slides & Presentations

Material for slides ( Slideshare)

## Group Questions

Questions per group based on the WP2 elicitation spreadsheet:

**Group 1 Data centric initiatives in Climate**

1. Which data management tools and practices are currently common in your domain/your organisation? Are they future proof?
2. What tasks using large (internal and/or external) datasets would you like to perform? What are the problems in getting, processing, analysing and storing the data?
3. What use cases would be of interest to you? Can we setup 2 pilot cases?
4. What are the most important gaps?
    a. What could be accomplished if those gaps were filled?

**Group 2 Technology and Data**

1. What are the requirements for the big data platform?
    a. Both functional and non-functional
2. What kind of data do you need to process?
3. What tools and approaches have been applied?
    a. Tools/systems/databases/resources -gaps and pitfalls
    b. Which of them are future proof?
4. What restrictions and limitations might impede access and processing of the data?
    a. Practical or technical restrictions

**Group 3 EU Policy Requirements (Legal issues around (Big) data, Governance, Data Portability, other requirements)**

1. What is your domain strategic vision regarding data management?
2. Which political framework or domain specific strategic factors can influence your decisions on your data management strategy/systems in use?
3. Can data be private and secure yet useable for research?
4. Which legal concerns could be encountered in the usage of (Big) Data management solutions?
5. What needs to be done to influence policy requirements?
    a. Identify key roadblocks, and how to highlight their impact

*Report by NSCRD BDE workshop organization team:*
*D. Vlachogiannis, Spyros Andronopoulos, A. Ikonomopoulos, I. Klampanos, V. Karkaletsis, S. Konstantopoulos.*