Abstract: This deliverable reports on the project’s progress, targeting the general public. It will focus on the impact of the conducted work across Europe’s seven Societal Challenges, exemplifying the contribution and importance of Big Data Europe towards facing these 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.
### History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Reason</th>
<th>Revised by</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>28/11/17</td>
<td>Draft</td>
<td>Alexandra Garatzogianni (FhG IAIS)</td>
</tr>
<tr>
<td>0.1</td>
<td>07/12/17</td>
<td>Updating SC4</td>
<td>Andrea Toth</td>
</tr>
<tr>
<td>0.2</td>
<td>09/12/17</td>
<td>Updating section 2</td>
<td>Ivan Ermilov</td>
</tr>
<tr>
<td>0.3</td>
<td>10/12/17</td>
<td>Updating SC4</td>
<td>Luigi Selmi</td>
</tr>
<tr>
<td>0.4</td>
<td>12/12/17</td>
<td>Updating SC7 (final)</td>
<td>Michele Lazzarini</td>
</tr>
<tr>
<td>0.5</td>
<td>13/12/17</td>
<td>Updated (final)</td>
<td>Fragiskos Mouzakis</td>
</tr>
<tr>
<td>0.6</td>
<td>14/12/17</td>
<td>Updating SC2 (final)</td>
<td>Valeria Pesce (FAO), Panagiotis Zervas (AgroKnow)</td>
</tr>
<tr>
<td>0.7</td>
<td>14/12/17</td>
<td>Updated section on SC5</td>
<td>Iraklis Klampanos (NCSRd), Andreas Ikonomopoulos (NCSRd), Spyros Andronopoulous (NCSRd)</td>
</tr>
<tr>
<td>0.8</td>
<td>14/12/2017</td>
<td>Updating SC6</td>
<td>Ivana Versic (CESSDA), Martin Kaltenboeck (SWC)</td>
</tr>
<tr>
<td>0.9</td>
<td>16/12/17</td>
<td>Finalized section 2</td>
<td>Ivan Ermilov, Aad Versteden</td>
</tr>
<tr>
<td>0.91</td>
<td>23/12/17</td>
<td>Added SC1 section</td>
<td>Kiera McNeice</td>
</tr>
<tr>
<td>0.10</td>
<td>05/01/2018</td>
<td>Updating SC4</td>
<td>Zane Mezdreija</td>
</tr>
<tr>
<td>0.11</td>
<td>20/01/2018</td>
<td>Internal Review</td>
<td>Partners</td>
</tr>
<tr>
<td>0.12</td>
<td>25/01/2018</td>
<td>Updating SC4</td>
<td>Luigi Selmi</td>
</tr>
<tr>
<td>1.0</td>
<td>20/01/2018</td>
<td>Final Review</td>
<td>Simon Scerri, Alexandra Garatzogianni</td>
</tr>
</tbody>
</table>

### Author List

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Name</th>
<th>Contact Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>FhG</td>
<td>Alexandra Garatzogianni</td>
<td>Alexandra. <a href="mailto:Garatzogianni@iais.fraunhofer.de">Garatzogianni@iais.fraunhofer.de</a></td>
</tr>
<tr>
<td>FhG</td>
<td>Simon Scerri</td>
<td><a href="mailto:scerri@iai.uni-bonn.de">scerri@iai.uni-bonn.de</a></td>
</tr>
<tr>
<td>Organization</td>
<td>Name</td>
<td>Email</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>FAO</td>
<td>Valeria Pesce</td>
<td><a href="mailto:valeria.pesce@fao.org">valeria.pesce@fao.org</a></td>
</tr>
<tr>
<td>CRES</td>
<td>Fragiskos Mouzakis</td>
<td><a href="mailto:mouzakis@cres.gr">mouzakis@cres.gr</a></td>
</tr>
<tr>
<td>Ertico</td>
<td>Zane Mezdreija</td>
<td><a href="mailto:zane.mezdreija@mail.ertico.com">zane.mezdreija@mail.ertico.com</a></td>
</tr>
<tr>
<td>InfAI</td>
<td>Ivan Ermilov</td>
<td><a href="mailto:iermilov@informatik.uni-leipzig.de">iermilov@informatik.uni-leipzig.de</a></td>
</tr>
<tr>
<td>SatCen</td>
<td>Michele Lazzarini</td>
<td><a href="mailto:Michele.Lazzarini@satcen.europa.eu">Michele.Lazzarini@satcen.europa.eu</a></td>
</tr>
<tr>
<td>TF</td>
<td>Aad Versteden</td>
<td><a href="mailto:aad.versteden@tenforce.com">aad.versteden@tenforce.com</a></td>
</tr>
<tr>
<td>FhG</td>
<td>Luigi Selmi</td>
<td><a href="mailto:luigi.selmi@iais.fraunhofer.de">luigi.selmi@iais.fraunhofer.de</a></td>
</tr>
<tr>
<td>NCSRD</td>
<td>Iraklis Klampanos</td>
<td><a href="mailto:iaklampanos@iit.demokritos.gr">iaklampanos@iit.demokritos.gr</a></td>
</tr>
<tr>
<td>NCSRD</td>
<td>Andreas Ikonomopoulos</td>
<td><a href="mailto:anikon@ipta.demokritos.gr">anikon@ipta.demokritos.gr</a></td>
</tr>
<tr>
<td>NCSRD</td>
<td>Spyros Andronopoulos</td>
<td><a href="mailto:sandron@ipta.demokritos.gr">sandron@ipta.demokritos.gr</a></td>
</tr>
<tr>
<td>Open PHACTS Foundation</td>
<td>Kiera McNeice</td>
<td><a href="mailto:kiera@openphactsfoundation.org">kiera@openphactsfoundation.org</a></td>
</tr>
<tr>
<td>W3C/ERCIM</td>
<td>Rigo Wenning</td>
<td><a href="mailto:rigo@w3.org">rigo@w3.org</a></td>
</tr>
<tr>
<td>VUA</td>
<td>Ronald Siebes</td>
<td><a href="mailto:rm.siebes@few.vu.nl">rm.siebes@few.vu.nl</a></td>
</tr>
<tr>
<td>CESSDA</td>
<td>Ivana Versic</td>
<td><a href="mailto:ivana.versic@cessda.eu">ivana.versic@cessda.eu</a></td>
</tr>
<tr>
<td>SWC</td>
<td>Martin Kaltenböck</td>
<td><a href="mailto:martin.kaltenboeck@semantic-web.com">martin.kaltenboeck@semantic-web.com</a></td>
</tr>
<tr>
<td>Agroknow</td>
<td>Panagiotis Zervas</td>
<td><a href="mailto:pzervas@agroknow.com">pzervas@agroknow.com</a></td>
</tr>
<tr>
<td>FhG</td>
<td>Mohamed Nadjib Mami</td>
<td><a href="mailto:mami@cs.uni-bonn.de">mami@cs.uni-bonn.de</a></td>
</tr>
</tbody>
</table>
Executive Summary

The third deliverable in this series provides an overview of the main project results accomplished in the third and final year of the project by the BigDataEurope (BDE) consortium. During this time, the dissemination efforts targeted at the seven communities were intensified; focusing on the pilot results and demonstrating their value with respect to the whole data value chain within each Societal Challenge; from data collection, processing, storage and visualisation to the development of specific data services. The last series of workshops, webinars and hangouts organised throughout 2017, including the strong presence of the BDE Consortium as Sponsor in the International Semantic Web Conference (ISWC\(^1\)) and the European Big Data Value Forum (EDVF\(^2\)), have brought together these communities by disseminating and showcasing the project’s results, while increasing the project’s impact and enabling the sustainability of its output via the creation of synergies.

The above ‘Coordination’ efforts were complemented by concrete ‘Support’ by way of technical results that have a broader scope than the project’s 7 showcased pilots. To that effect, in November 2017 the final release of the implemented BDE platform was released and announced in a very strongly promoted Webinar. This new version took into account the improvement requests received, including additional features and providing the opportunity to further intensify BDE’s dissemination and community building activities for the promotion of the integrator platform, its components and take-up by stakeholders, while showcasing the results accomplished in each one of the seven Societal Challenges.

The reported activities in this deliverable are structured along two sections. The first summarizes the various networking activities that were carried out within each of the seven societal challenge communities, explaining the efforts performed in 2017 and the next steps planned beyond the project’s duration. The second section describes advances made by the technical team during the second 12-month period and outline future plans. This report is public and shall be made available for unrestricted download on the BDE webpage [http://www.big-data-europe.eu/results/](http://www.big-data-europe.eu/results/).

\(^1\) [https://iswc2017.semanticweb.org/](https://iswc2017.semanticweb.org/)
Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDE</td>
<td>Big Data Europe</td>
</tr>
<tr>
<td>SC</td>
<td>Societal Challenge</td>
</tr>
<tr>
<td>H2020</td>
<td>Horizon 2020</td>
</tr>
</tbody>
</table>

Table of Contents

1. Domain Summaries for the Third Year (2017) ................................................................. 7
   1.1 Societal Challenge 1 - Health, Demographic Change and Wellbeing .............................. 7
       General Objectives .............................................................................................................. 7
       Pilot Use Case ................................................................................................................... 7
       Stakeholder Workshop ....................................................................................................... 7
       Engagement, Outreach, Dissemination .................................................................................. 8
       Next Steps beyond the Project’s Duration ........................................................................... 8
   1.2 Societal Challenge 2 - Food, Agriculture, Forestry, Water and Bioeconomy ............... 9
       General Objectives .............................................................................................................. 9
       Pilot Use Case ................................................................................................................... 9
       Stakeholder Workshop ....................................................................................................... 9
       Engagement, Outreach, Dissemination .................................................................................. 10
       Next steps beyond the Project’s Duration ........................................................................... 10
   1.3 Societal Challenge 3 - Secure, Clean and Efficient Energy ........................................... 11
       General Objectives .............................................................................................................. 11
       Pilot Use Case ................................................................................................................... 11
       Stakeholder Workshop ....................................................................................................... 11
       Engagement, Outreach, Dissemination .................................................................................. 12
       Next Steps beyond the Project’s Duration ........................................................................... 12
   1.4 Societal Challenge 4 - Smart, Green and Integrated Transport .................................... 13
       General Objectives .............................................................................................................. 13
       Pilot Use Case ................................................................................................................... 13
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement, Outreach, Dissemination</td>
<td>14</td>
</tr>
<tr>
<td>Next Steps beyond the Project’s Duration</td>
<td>15</td>
</tr>
<tr>
<td><strong>1.5 Societal Challenge 5 - Climate, Environment, Resource Efficiency and Raw Materials</strong></td>
<td>15</td>
</tr>
<tr>
<td>General Objectives</td>
<td>15</td>
</tr>
<tr>
<td>Pilot Use Case</td>
<td>15</td>
</tr>
<tr>
<td>Stakeholder Workshop</td>
<td>16</td>
</tr>
<tr>
<td>Engagement, Outreach and Dissemination</td>
<td>16</td>
</tr>
<tr>
<td>Next Steps beyond the Project’s duration</td>
<td>17</td>
</tr>
<tr>
<td><strong>1.6 Societal Challenge 6 - Inclusive, Innovative and Reflective Societies</strong></td>
<td>17</td>
</tr>
<tr>
<td>General Objectives</td>
<td>17</td>
</tr>
<tr>
<td>Pilot Use Case</td>
<td>17</td>
</tr>
<tr>
<td>Stakeholder Workshop</td>
<td>17</td>
</tr>
<tr>
<td>Engagement, Outreach, Dissemination</td>
<td>20</td>
</tr>
<tr>
<td>Next Steps beyond the Project’s duration</td>
<td>20</td>
</tr>
<tr>
<td><strong>1.7 Societal Challenge 7 - Secure Societies</strong></td>
<td>21</td>
</tr>
<tr>
<td>General Objectives</td>
<td>21</td>
</tr>
<tr>
<td>Pilot Use Case</td>
<td>21</td>
</tr>
<tr>
<td>Engagement, Outreach, Dissemination</td>
<td>22</td>
</tr>
<tr>
<td>Research Achievements</td>
<td>24</td>
</tr>
<tr>
<td>Semantic Data Lake</td>
<td>24</td>
</tr>
<tr>
<td>Sparkall</td>
<td>24</td>
</tr>
<tr>
<td>Sansa</td>
<td>25</td>
</tr>
<tr>
<td>Engagement, Outreach and Dissemination</td>
<td>26</td>
</tr>
</tbody>
</table>
1. Domain Summaries for the Third Year (2017)

1.1 Societal Challenge 1 - Health, Demographic Change and Wellbeing

General Objectives
In the third year of the project, SC1 focused on implementing and evaluating the second and third cycles of the Pilot, and continued engaging with stakeholders, in particular to consider how the pilot might be expanded.

Pilot Use Case
The second and third cycles of the Pilot Use Case has been completed, with the Open PHACTS Discovery Platform functionality fully replicated on the BDE infrastructure, replacing the commercial components of the Open PHACTS Platform with open source alternatives. It is now possible for any user to deploy an instance of the Open PHACTS Platform in-house, an important consideration for security-conscious pharmaceutical researchers.

Using the open source Virtuoso RDF Docker component, the Pilot can answer a total of 18 of the original 21 use cases for the Open PHACTS Discovery Platform. The remaining three cannot be answered without access to non-open datasets.

As the Pilot’s technology is fundamentally domain-agnostic, the possibility was raised of connecting to data from other domains. SC2 (Food Security and Sustainable Agriculture) was considered a good target, as the effects of chemistry on biology in this domain are an important consideration, and a connection could exploit the linked data and services currently within the Open PHACTS Platform. Pursuing this idea proved to be beyond the scope and timeframe of the BDE project, but the Open PHACTS Foundation will continue to look for ways to expand the platform to connect to new kinds of data.

Data refreshes also turned out to be a more complex challenge than anticipated, given that the Pilot integrates data from multiple sources, with multiple release schedules and systems. However significant progress has been made towards automating refreshes of Pilot data, work which will continue beyond the end of the Big Data Europe project.

Stakeholder Workshop
The third SC1 Stakeholder Workshop was held in December 2017 at KoWi in Brussels. As this was the last workshop to be held during the Big Data Europe project, the workshop served the dual purposes of giving a final live demonstration of the Big Data Integrator, as well as demonstrating the SC1 Pilot and discussing next steps for using the BDI in areas of health beyond the rest of the project. Attendees came from a variety of backgrounds related to big data
in health. Three external speakers were invited to present other projects on big data in health for discussion:

- MIDAS, a project aimed at connecting big data in health and presenting it in such a way as to be accessible and useful for public health policy makers;
- BigMediylitics, a project aiming to take a holistic view of healthcare, and find ways to improve productivity in the healthcare sector by reducing costs, improving quality, and delivering better access;
- IASIS, a project aiming to connect data to support personalised medicine in two pilot areas of lung cancer and Alzheimer’s disease.

Following this we held an open discussion on the future of big data in health, full details of which will be published in the workshop report. An early point to note was that questions of privacy, ethics, and consent where patients’ personal data are concerned came up again and again in the context of improving and personalising healthcare. Building up trust among stakeholders will be key to addressing these questions. The importance of linking data across different domains will also be key to improving health outcomes, as health depends on so many factors beyond the individual – infrastructure, food, environment, etc.

**Engagement, Outreach, Dissemination**

As well as hosting the third Stakeholder Workshop, SC1 has been engaged in several other outreach activities this year.

Big Data Europe was invited to present a partner session at eHealth Week in May 2017, where the conference theme was “Data for Health”. Around 35 delegates attended the presentation of the Big Data Europe platform, and the progress that had been made in the SC1 pilot. SC1 also hosted a webinar in April 2017 demonstrating the BDE infrastructure, and how to set up and use the SC1 pilot on a local machine or network. In December 2017, SC1 hosted a panel discussion on the challenges of updating public databases, where representatives of major health data infrastructures engaged with the community about developing best practice solutions. Finally, in January 2018, a one day tutorial on OpenPHACTS and the BigDataEurope pilots was held in Amsterdam, The Netherlands as part of the BioSB/ELIXIR-NL course: Managing and Integrating Life Science Information (5th edition).

**Next Steps beyond the Project’s Duration**

Refreshing the datasets used in the SC1 pilot turned out to be a much more complex challenge than initially anticipated. Work will continue beyond the Big Data Europe project on identifying ways to facilitate and automate refreshes of the SC1 pilot data, if at all possible in a way that can be Dockerized and compatible with the BDI infrastructure.
1.2 Societal Challenge 2 - Food, Agriculture, Forestry, Water and Bioeconomy

General Objectives
Activities in SC2 in the third and last year of the project revolved around two main areas: stakeholders’ engagement and implementation and presentation of the big data pilot for SC2.

Pilot Use Case
The SC2 Pilot Cycle 2 and 3 were completed. SC2 Pilot Cycle 2 demonstrated the ability of BDE proposed technologies to extend and scale up the processing workflows to handle a variety of data types (beyond bibliographic data) relevant to Viticulture. To this end, the Flume/Kafka pipeline was extended to handle other than bibliographic data (e.g. sensor/weather data) and inclusion of use case scenarios which combine/link more heterogeneous data sources. The BDE proposed technologies that were evaluated successfully during the SC2 Pilot Cycle 1, proved to be equally capable of handling other data types also in the case of SC2 Pilot Cycle 2. Therefore, the choice of the technologies used for the purposes of the SC2 Pilot proved to be successful. SC2 Pilot Cycle 3 provided an engaging, intuitive graphical web interface addressing key data-oriented questions relevant to the Viticulture Research Community, as well as intuitive interfaces for end-users for sharing and linking their on-the-field generated data. User-feedback was used to fine-tune cycle 3, however no new functionalities were implemented.

Stakeholder Workshop
The main event around stakeholders’ engagement in the third year was the third societal challenge workshop, held in Brussels on the 31st of March. The SC2 pilot partners (Agroknow and FAO) gathered together a selected group of EC representatives (from DG AGRI, DG CONNECT, DG RTD, the eROSA project but also other initiatives) that are working on projects related to open data and big data in general or more specifically to the use of data for food and agriculture, as well as a number of actors working in agricultural research and industry and already collaborating in international open data initiatives. The intention was to give the opportunity to EC representatives to share with the community representatives the most recent developments in terms of funding agendas or projects relevant to big and open data in agriculture and food, and also to help projects in Europe to better shape and align their agendas in order to better serve the policy and strategy objectives of the EC.

Additionally, a follow-up workshop was held in Brussels on the 11th of December in conjunction with e-ROSA project, supported by the Global Open Data in Agriculture & Nutrition (GODAN) initiative and its GODAN Data Ecosystem Working Group. This
workshop was a first-hand opportunity to present to the participants (mostly EC representatives and community stakeholders) the next steps regarding the SC2 pilot uptake.

**Engagement, Outreach, Dissemination**

Agroknow and FAO have continued with the BDE webinar series held on the Webinars@AIMS 3 platform. In April 2017, a webinar on “the BDI platform and the VITIS pilot introduced to food and agriculture experts” was conducted by Dr. Hajira Jabeen and Dr. Maritina Stavrakaki. In May 2017, a webinar on “How much data is enough in Precision Agriculture” was conducted by Dr. Mike Duncan of Niagara College (NC) Agriculture & Environmental Technologies Innovation Centre. A third webinar in November focused on the presentation of the “Big Data Europe Integrator platform” by Dr. Hajira Jabeen to the agricultural research environment. People from Alterra participated in this webinar and they were highly impressed by the work done.

News about BDE (webinar, workshop, reports) have been disseminated to the community through various channels (the FAO AIMS website, the GFAR website, the Agroknow’s blog and related newsletters). Eight thematic articles/blogs on data for food and agriculture have been posted to the BDE website.

The BDE SC2 Pilot was presented at a workshop which was held at the Agricultural University of Athens in February 2017 4, with the participation of delegates from the Chinese Academy of Agricultural Sciences (CAAS). The pilot was also presented at the 9th Plenary Meeting of the Research Data Alliance (RDA) 5 which took place in Barcelona, Spain, in April 2017, as well as at the second “Open Harvest” workshop held in Chania, Crete, in May 2017, where representatives from different types of institutions in the area of food and agriculture (research, private sector, international) met to discuss elements of a roadmap for a data infrastructure for food and agriculture. In order to better disseminate the features of the SC2 Pilot, a video 6 was also prepared which outlines the main functionalities of the pilot.

**Next steps beyond the Project’s Duration**

After the end of the project, the main priority will be the continuous engagement of the SC2 community and the sustainability of the SC2 pilot. Especially, with regards to the uptake of the pilot, it is of high importance that it will be further exploited and extended through a new H2020 research project, called BigDataGrapes, that kicks off in 2018. The upcoming project will continue and extend the software stack that has been produced in the context of SC2 pilot.

---

6 [https://www.youtube.com/watch?v=RQL0cuq81Ng&t=160s](https://www.youtube.com/watch?v=RQL0cuq81Ng&t=160s)
BigDataGrapes will evolve further and integrate this work in order to put in place an improved model for orchestrating and documenting different contributions to this hub, which come from different teams and other relevant projects and contributors.

1.3 Societal Challenge 3 - Secure, Clean and Efficient Energy

General Objectives
The main domain related objectives during the third year regarded the work on the third cycle of pilot case and the continuation of the dissemination actions.

Pilot Use Case
The efforts during the third year were focused on the identification of an appropriate use case in the energy domain with not only significance to the community but also with open data to support its realization. In the Energy domain it is evident that the future scene will be dominated by new energy production and distribution models in which the consumer will be actively part of it, and this is where the new IT tools are expected to be exploited. Having the above in mind and compiling the findings of the first two workshops and especially the presentations and discussions related to energy production and grid operation challenges the concept for a use case regarding the forecasting of renewable energy production in country level was followed. The concept incorporates two important aspects, namely the forecasting at unit level and the utilization of simulation data for the topography effects. The involved data comprise a variety of sources from terrain data, weather forecast data, energy production unit data and meteorological on-line data. The pilot delivers forecasts in unit level and regional aggregates continually. In the scene of the new energy bidding system the forecasting tools are of primary importance, and their evolution is continuous. Although forecasting itself is available as a service, the implementation of customized forecasting procedures, tailored to the assets of the producer or the resources of the retailer (for instance specific on-line data from owned assets) will make the difference in the market.

Stakeholder Workshop
The third SC3 workshop was held in Amsterdam on the 28th of November 2017 within the major annual event for wind energy that is WINDEUROPE Conference and Exhibition 2017. The workshop was thematic and mainly regarded the prospects of exploitation of BDE in the field of system monitoring. The workshop was divided in several sessions regarding BDE scope and opportunities, challenges in wind energy asset management and system monitoring, BDE Open Platform technical aspects, the pilot cases and round table discussion. The workshop was addressed by Mr. Mark van Stiphout (Deputy Head of Unit DG ENER C.2) by presenting the EU Research & Innovation priorities on energy and digitalization setting the framework in
which BDE can be exploited in the future. BDE project was presented by Prof. Maria Ester Vidal (Fraunhofer IAIS, University of Bonn), the technical part was presented by Mr. Ivan Ermilov (University of Leipzig) and the pilot case by CRES. Keynote presentations in the field of asset management (data management & analytics) were given by Mr. M. Asgarpour from a major utility (VATTENFALL), by Mr. A. Kyritsis from a service provider representing also a RES producer (ALTSOL/TERNA) and by Mr. P. Clive from WoodGroup (analytics). In system monitoring the keynote presentations were given by Prof. Jan Helsen (Drivetrain monitoring and Big Data coordinator, OWI Lab), Mr. G. Gainza (testing and monitoring service provider). The workshop concluded with discussion on future data management and analytics challenges in wind energy, the BDE prospects and future research opportunities. The workshop was attended by 35 persons.

Engagement, Outreach, Dissemination
The dissemination activities were focused on the presentation of BDE and pilot case in specialized events. These presentations allowed the access to a large audience with strong thematic interest and direct contact with the community and as such proved successful substitutes for the foreseen hangouts. In all events the presentations were lengthy with follow up discussions. The events were:

a) 2nd Annual Wind Power Industry Big Data and IoT Forum in Amsterdam on the 31st of March 2017
b) 3rd International Conference on Digital Data Integration & Management in Berlin on the 26~28th of September 2017 and
c) 4th Edition Wind Power Big Data and IoT Forum in Berlin on the 8~9th of November 2017
Furthermore, during the WindEurope event (in which the workshop was held) CRES had a poster presentation in the conference and hold a booth with BDE posters and project and technical leaflets available to the attendees of the exhibition. Dissemination was also supported by thematic blog posts in BDE web page.

Next Steps beyond the Project’s Duration
The development of the RES energy forecasting pilot is a major project and at its completion will be part of CRES supporting role to the regulatory agencies and operators in Greece. This effort is set as a CRES internal project to be built upon BigDataEurope platform. The work will focus on a) the implementation of the developed components on BDE platform (support from the technical core team will be pursued under new agreement), b) exploitation of available BDE tools (GIS), c) enrichment of the background simulation CFD data base internally and/or by
community support, d) enrichment of the connected on-line assets, e) inclusion of solar energy (PV) forecast component.

1.4 Societal Challenge 4 - Smart, Green and Integrated Transport

General Objectives
The activities in the transport domain (Societal Challenge SC4) focus, in particular, on using machine learning in large historical databases containing floating car data for building spatio-temporal data models able to predict with accuracy the traffic conditions in the city of Thessaloniki, with an average error of 4-5 km/h.

ERTICO - ITS Europe has the task to organise three workshops, as well as to provide support in dissemination and community building activities with regular blog posts and the organisation of multiple webinars.

Pilot Use Case
CERTH-HIT collects data in real time from 1,200 taxis in Thessaloniki with a frequency of 5-6 seconds, generating large amounts of data. These historical databases contain the knowledge about the spatial and temporal relations between the traffic status in the different road sections of Thessaloniki. Machine learning techniques are used for building a prediction tool able to consume years of time series in thousands of links and predict with an accuracy below 5 km/h the traffic status in the next time period (15 minutes).

The architecture has been designed in order to handle many data sources, mostly stream data, process it in time windows and store the result so that critical features such as geographic information and time can be indexed enabling low latency for search and queries. The pilot is based on the microservices architecture where different software components act as producers or consumers of data that can communicate through a messaging system. The messaging system used in the pilot is Apache Kafka. Kafka can be configured to store the messages as long as it is needed. The processing tasks are handled by Apache Flink that provides support for stream and batch processing, and libraries to work with graphs and machine learning algorithms. The producers read the Floating Car Data (FCD) records containing location, speed and orientation of a vehicle, from a file or web service, change the format from text to a binary one, based on a known schema, and finally write the data in a Kafka topic. The consumers read the data from the topic and match the location of the vehicles to the Thessaloniki road network through a service based on R scripts and PostGis. The map-
matching step adds a field to the original FCD record containing an identifier of the road segment on which each vehicle is being driven. The road segments’ identifiers are used to distribute the computation of the number of vehicles and of the average speed within a time window to different Flink workers. The information about the status of each road segment in different time windows is finally stored in Elasticsearch, a distributed document database, where the data can be analyzed and easily shown on a map. All the components have been made available in the BDE platform as docker images.

Stakeholder Workshop
The third SC4 stakeholder workshop took place on 14th September 2017 in Brussels. Invitations were sent out via the BigDataEurope SC4 contact list, which includes over 300 contacts, and the ERTICO Newsletter, which reaches over 11,000 people. Ultimately, 40 persons attended the workshop, including representatives from the industry, research institutes, the European Commission and different public administrations.

The objective of the workshop was to showcase the BigData Integrator Platform, in particular as concerns the transport pilot, as well as to collect feedback from stakeholders requiring transport solutions on how the platform can be extended to increase its value and facilitate further applications. The workshop also included a series of speeches discussing the transformation that big data is bringing to the transport industry and presenting some of the current EU funded projects working with transport and big data.

Alongside BDE consortium partners, speakers included Ms Maria Rautavirta (Finnish Ministry of Transport and Communications), Mr Rodrigo Castineira (INDRA), Mr Arnaud Burgess (Panteia), and Mr Victor Corral (ATOS). Ms Rautavirta spoke at length about the success of introducing the MaaS (Mobility-as-a-Service) concept in Finland, whereas the other three speakers gave introductions to the projects TransformingTransport, LeMO, and AutoMat, respectively. Thus the workshop presented an ideal opportunity for various current EU-funded projects to come together and learn from each other.

The diverse background of the participants helped facilitate lively debates in the afternoon session. After the workshop, participants remarked that the workshop had provided an excellent opportunity for networking by gathering people from different backgrounds, but all of which work with big data and transport.

Engagement, Outreach, Dissemination
In the third year of the project, SC4 also organised a webinar entitled “Empowering Mobility Management with Big Data”. The webinar took place on 14th June and gathered over 20 people. The purpose of the webinar was to give a status update of the BigData Integrator Platform and
SC4’s work on the pilot. A live demo highlighted the potential of streaming sensor networks and geospatial data integration using the BigData Integrator Platform.

BigDataEurope was represented at KnowMe: the 1st International Workshop on Knowledge Discovery from Mobility and Transportation Systems, which was organised as part of the European Conference on Machine Learning & Principles and Practice of Knowledge Discovery in Databases. A live demonstration of the Thessaloniki mobility centre was given at the event and emphasis was put on the traffic prediction components that have been developed in the framework of the BigDataEurope project.

Last but not least, SC4 has been active on the BigDataEurope website as well, where a total of 12 blog posts have been added throughout 2017. SC4 has established fruitful cooperation with the three projects that took part in the SC4 workshop in September 2017 – TransformingTransport, LeMO, and AutoMat – all of whom have kindly provided additions to the BigDataEurope transport blog.

**Next Steps beyond the Project’s Duration**

CERTH will adopt the Big Data tools developed within the project and integrate them into the mobility dashboard it hosts in Thessaloniki. The tools will be implemented in the virtual servers provided by GRNET in Greece and will support the existing processes of the mobility living lab in Thessaloniki.

Once the service will be fully operable at CERTH-GRNET facilities based on Floating Car Data, other sources will be added, such as speed measurements from cameras and loops or travel time estimation from Bluetooth detectors. The addition of these sources will enrich significantly the model and an improvement of the accuracy is expected.

### 1.5 Societal Challenge 5 - Climate, Environment, Resource Efficiency and Raw Materials

**General Objectives**

During the third year of the project SC5 implemented and evaluated the second and third Pilot cycles, while also planning and carrying out dissemination and workshop activities.

**Pilot Use Case**

During the second phase the Pilot demonstrated the following use case: A radioactive substance is released in the atmosphere that results to increased readings in one or more monitoring stations on the ground. The user accesses a user interface provided by the pilot to define the
locations of the monitoring stations (the readings locations) as well as the measured values (e.g. gamma dose rate). The platform utilises a weather matching algorithm in order to find similar pre-computed weather patterns to current weather and then a dispersion matching algorithm that finds similar pre-computed substance dispersion patterns compatible with the current monitor readings.

The third phase of the Pilot extends on the above use-case and takes place when the matching algorithms return their results. The pre-computed substance dispersions are enriched with semantic information by introducing linked and heterogeneous data in the platform. Using GeoNames data, the pilot is able to help quantify the effect of such scenarios by visualizing the population under the plume’s affected area. Using OpenStreetMaps, the pilot is able to visualize the nearest hospitals under the dispersion plume while providing additional information such as communication information, wheelchair accessibility, etc. These additions were implemented to aid decision makers in cases of such emergencies.

**Stakeholder Workshop**

The 3rd BDE SC5 Workshop on Big Data in Climate Action, Environment, Resource Efficiency and Raw Materials took place in Brussels on 6/11/2017. During this workshop, we’ve had the opportunity to present the 2nd SC5 pilot on rapid source estimation during nuclear or radiological events.

In addition to reporting on progress and receiving feedback, we’ve had the opportunity to attend talks about a number of big data-related applications, from emergency response during adverse events in cities to precision medicine. We would like to thank our invited speakers: Patrick Armand (CEA), Stella Moehrle (KIT) and Maria-Ester Vidal (Fraunhofer IAIS).

**Engagement, Outreach and Dissemination**

After two successful webinars the SC5 represented Big Data Europe in various conferences during the third year of the project.

In April 25 of 2017 the second SC5 pilot was presented in the "European Geosciences Union General Assembly 2017" after the submission of an extended abstract.

In May 18 of 2017 the second SC5 pilot was presented and demonstrated in the NERIS Workshop (NERIS is the European Platform On Preparedness For Nuclear and Radiological Emergency Response and Recovery, Proceedings can be found here).
In October 23 of 2017 the third pilot was demonstrated live in the 16th International Semantic Web Conference demo session after a demo submission.

The live demonstration of the third phase of the SC5 pilot resulted in winning the "People's Choice Best Demo Award" in ISWC 2017.

**Next Steps beyond the Project’s duration**

The SC5 pilot completed the objective of estimating an unknown source during a substance dispersion using a sufficient geospatial grid and configurations. The next steps revolve around the upscale of geospatial grid to more fine resolutions that may provide better dispersions. Pre-computed dispersions settings could also be upscaled in terms of emission hours for more realistic results considering radiological incidents. Last but not least, weather data could get upscaled to include more historical reanalysis data as well as streamlining real time weather data.

**1.6 Societal Challenge 6 - Inclusive, Innovative and Reflective Societies**

**General Objectives**

The activities in the SC6 domain (Societal Challenge SC6- Europe in a Changing World: innovative, inclusive and reflective societies) focused on coordination of the Societal Challenge 6 and potential users of big data in the fields of social sciences and (digital) humanities (SSH), and on building of that particular interest group, collecting requirements, and assisting in the development of the big data infrastructure access point for social sciences and humanities.

In the course of the third year of BDE project, the domain lead CESSDA along with SWC has organised the final workshop, and has provided support in dissemination and community building activities with regular blog posts, as well as the organisation of 3 webinars.

**Pilot Use Case**

The pilot is driven by CESSDA Consortium (Consortium of European Social Sciences Data Archives) and Semantic Web Company with (at the beginning) support by NCSR-Demokritos. The overall idea and objective of the Pilot was to create an online Dashboard on Economic Data - this means:

- **Harvest data** from several sources in diff. formats
- **Normalise the data** (RDF)
- **Link & map the data** (attributes, structure, languages)
- **Analyse the data** (comparisons, predictions etc.)
- **Visualise the analysis** on an online dashboard
including help & infos to understand data & analysis

- **Provide raw data** (for further use as open data)

**Overall Objective:** Can we make budgets more useful for citizens, researchers and decision makers?

The 4Vs of Big Data are represented as follows in the Pilot of SC6:

- **Variety**: requirement based on the harvesting of budget data and budget execution data from several sources, available in different structures and formats.
- **Volume**: requirement regarding the growing amount of open budget data available as well as of budget execution data
- **Velocity**: requirements regarding budget execution data that is provided on continuous basis by the publisher (daily, weekly, monthly).
- **Veracity**: Veracity refers to the biases, noise and abnormality in data. Even within the same country there are differences on the published data because often it: a) comes from different systems or b) public accounting standards are not enforced absolutely uniformly (across e.g. different municipal departments)

The architecture of the SC6 Pilot System is as follows:
Technical Components of the SC6 Pilot are:

- Apache Flume, [https://flume.apache.org/](https://flume.apache.org/) (data ingestion)
- Apache Kafka, [http://kafka.apache.org](http://kafka.apache.org) (messaging service)
- Apache Spark, [http://spark.apache.org](http://spark.apache.org) (distributed analysis, transformation)
- SWCs’ PoolParty Semantic Suite, [http://poolparty.biz](http://poolparty.biz) (data consolidation, curation, mapping)
- SWCs’ PoolParty Graph Search, [link](http://poolparty.biz) (search, browse & visualise data)
- OpenLink s’ Virtuoso, [http://virtuoso.openlinksw.com](http://virtuoso.openlinksw.com) (triple store: data storage, data endpoint / API)
- Apache HTTP, [http://httpd.apache.org](http://httpd.apache.org) (linked data serving)
- D3 JS Library, [https://d3js.org/](https://d3js.org/) (visualisation of RDF data using SPARQL queries)

In year three the main activities in SC6 Pilot work was

- Continuous improvement the analytics dashboard for financial ratios (e.g. debugging financial ratio calculations, improvement of visualisation etc)
- Improvements of mappings between budget items of the municipalities account system
- Extending the Pilot by a data API in form of a SPARQL endpoint
- Specification and implementation of a 2nd analytics dashboard on top of budget attributes of the budget (execution) data

The systems is available publicly online via

- Dashboard 1: [https://bde.poolparty.biz/GraphSearchSC6/](https://bde.poolparty.biz/GraphSearchSC6/)
- Dashboard 2: [https://bde.poolparty.biz/GraphSearchSC6-2/](https://bde.poolparty.biz/GraphSearchSC6-2/)

**Stakeholder Workshop**

The third and the final workshop in domain of the EU Societal Challenge 6 - Europe in a Changing World: Inclusive, Innovative and Reflective Societies (SC6) was held on 11 September 2017 in Amsterdam, The Netherlands. The title and topic of the workshop was: The importance of Metadata & Big Data in OpenScience – where are we and where are we going to. It was organised by CESSDA ERIC and the Semantic Web Company (SWC), both
beneficiaries in the Big Data Europe project, as an official workshop of the SEMANTiCS 2017 conference. Around 45 participants joined the workshop – 2/3 officially registered and 1/3 participating from the SEMANTiCS 2017 audience ad-hoc by registering on-site.

The workshop offered high-quality talks by several experts in the fields of social sciences, smart statistics and open science as Ron Dekker (Director of Cessda), Peter Doorn (Director of DANS) or Fernando Reis (Big Data team Eurostat) as well as an open panel discussion with Q&A from the audience and finally a presentation of the SC6 pilot implementation on ‘Citizen Budget on Municipal Level’ and was overall a great success.

**Engagement, Outreach, Dissemination**

In the third year of the project, SC6 has organised 3 webinars (out of 6 planned for the whole duration of the project for SC6 domain). The 4th SC6 webinar entitled "Insight into Virtual Currency Ecosystems" (By making use of Big Data Technology) took place on 16 February 2017. The key speaker was Dr. Bernhard Haslhofer works as a Data Scientist at the Austrian Institute of Technology. It introduced virtual currencies and their technical characteristics, presented a selection of analytics features and described their technical implementation, focused on the computational (Big Data) challenges with building such a platform, and finally, it pointed out some possible future research directions.

Fifth SC6 webinar was held on 30 May 2017. Theme of the webinar was “European Open Science Agenda: Quo Vadis?” presented by Ron Dekker, Director of CESSDA. Presentation covered Open Science, its objectives and its tools, followed by the European Commission agenda on Open Science Policy. Then on the re-use of data and importance of standardised Data Management Plans. The presentation ended with a description of CESSDA’ objectives and its work in the context of the European Open Science Cloud.

On 23 November 2017, the 6th and final webinar of the Big Data Europe SC6 team took place with a presentation by Fernando Reis, member of the Big Data Task Force at Eurostat. Fernando gave interesting insights into the field of big data and smart data and how these new principles and technologies can be used in / applied to the domain of statistics. The title of the webinar was: From Big Data to trusted Smart Statistics.

Finally, SC6 domain has been active on the Big Data Europe website as well, with a total of 12 blog posts have been added throughout 2017. SC6 has established fruitful cooperation throughout consortium and external partners of SC6 domain (i.e. Eurostat) and secured their support and further developments beyond the BDE project.

**Next Steps beyond the Project’s Duration**

The current approach for update and sustainability is to bring SC6 Pilot system to the Cessda Data Platform later in 2018 and/or to bring and integrate the Pilot system into a new research
project with the topic of EOSCloud / Research & Science data that Cessda and SWC are working on at the moment that is expected (if approved for funding) to be started by end of 2018.

1.7 Societal Challenge 7 - Secure Societies

General Objectives
In the framework of the BigDataEurope project, SatCen is the domain leader addressing the “Secure Societies” Challenge focusing on Big Data opportunities and requirements as well as exploring and evaluating relevant Big Data end-to-end management approaches and techniques. The “Secure Societies” H2020 Societal Challenge is related to the protection of freedom and security of Europe and its citizens. A major activity in supporting the primary aims of this Societal Challenge (in particular to enhance the resilience of our society against natural and man-made disasters, to develop novel solutions for the protection of critical infrastructure, to improve border security and to support the Union’s external security policies) is the provision of geospatial products and services, mainly resulting from satellite data.

Pilot Use Case
The third and the last phase of the pilot developed by SatCen (societal challenge leader), University of Athens (technical domain leader) and NCSR “Demokritos” (technical support), aimed to add new data (from Remote Sensing and Social Sensing sources) to the previously developed pilot based on the fusion and integration of satellite data and news. The aim was to increase the information available for the final users with respect a specific event and to extend the sources for the cross-validation.

Through a user-friendly interface (Sextant), the user can analysis satellite images to detect areas with changes (Change Detection workflow) on land cover or land use and integrate the output extracting information from social media and news items (Event Detection workflow). In line with the project recommendations, the Change Detection workflow is based on the use of open data, such as the satellite images provided by ESA Sentinel 1 and Sentinel 2 (added in the third phase). The Event Detection workflow uses publicly available content from the Social Media (e.g. Twitter, and in the third phase Flickr) and from news agencies (e.g. Reuters).

Stakeholder Workshop
The community was consolidated with the 3rd “Big Data in Secure Societies” Workshop, held in Brussels on 3rd of October: 40 participants (in particular, stakeholders from EU entities, International Organizations, Industry and Academia) attended this event, representing a successful case of stakeholders engagement. Moreover, two Hangout sessions (held in May and December respectively) allowed to present the BigDataEurope activities through dedicated webinars.

**Engagement, Outreach, Dissemination**
Different activities were mainly focused on raising awareness on “Big Data in Secure Societies” amongst the stakeholders of the Security domain. In the framework of these Community Building activities, the Secure Societies pilot was presented during key events of the security domain.

BigDataEurope project and the related SatCen activities were presented during international conferences and SatCen internal events. Presentations were given at the 3rd International Security Symposium (Frascati, 13-14 February), at the World Cover Conference 2017 (Frascati, 14-16 March) and at the 2017 Conference on Big Data from Space (BiDS’17) (Toulouse, 28-30 November). In the framework of internal events, the project was presented in the SatCen Expert User Forum (Madrid, 8th February, 31st May and 4th October). The SC7 pilot was also presented in the Big Earth Observation Data workshop that took place at the European Big Data Value Forum (Versailles, November 23rd).

**Steps beyond the Project’s Duration**
Several initiatives are in place for the future sustainability of BigDataEurope SC7 pilot activities. In the framework of three H2020 projects, EVER-EST (2015-2018), NextGEOSS (2016-2020) and BETTER (2017-2020), the Change Detection chain with satellite data is going to be refined and new data sources will be added. The knowledge acquired during BigDataEurope on Big Data and Cloud Computing technologies will also support a possible internal development of a SatCen platform for heterogeneous data exploitation. New tools for data analysis (e.g. Machine Learning) will also seek through the application of future H2020 calls.

In the last year of the BDE platform development, most of the efforts were concentrated on the support of the existing BDE components. Overall, the final version of BDE platform provides over 30 various docker images including Hadoop, Hive, HBase, Spark, Flink, Zeppelin and others. Additionally, we develop the user interfaces for the platform such as BDI IDE, Workflow Builder and Workflow Monitor. Also, we have migrated to the last version of Docker and Docker Swarm. The easy installation and setup of Docker has eliminated the need for chef recipes/ansible playbooks for cluster-wide deployment of frameworks but for convenience we still provide an ansible playbook\(^7\).

The graphical user interfaces developed within BDE for assisting several deployment and management tasks for the BDE platform can be categorised as follows.

**BDE IDE.** Allows to deploy BDE instances and execution of different pipelines on a cluster. It has different UIs that combine to form BDE Integrator Platform.

**Stack Builder.** This allows users to import an already built docker-compose.yml definition from the BDE git-repository describing the services to be deployed in the working environment. The components within the pipeline are then accessible for editing and specialised changes in the Stack Editor.

**Stack Editor.** This allows users to create a personalised docker-compose.yml by updating the imported file definition.

**Swarm UI.** After the docker-compose.yml has been created in the Stack Builder, it can be pushed into a git repository. From the Swarm UI, users can clone the repository and launch (start, stop, restart, scale, etc.) the containers using docker swarm from a graphical user interface.

**BDE Logger.** The logger service provides logging of all the HTTP traffic generated by the containers and pushes it into an Elasticsearch instance, where it can be visualised with Kibana. To enable HTTP log collection for particular containers the user simply needs to set up the logging=true container label. When visualizing data in Kibana, please make sure to specify the hars* pattern for the index so that the data can be discovered.

**Workflow Builder.** This helps to define a specific set of steps that have to be executed in sequence, as a "workflow". The workflow builder reports the startup flow to "init daemon" (microservie) that can validate whether a specific component can start based on the initialisation status reported by the other components. The workflow needs to be described per application

---

\(^7\) [https://github.com/big-data-europe/ansible](https://github.com/big-data-europe/ansible)
stack as it specifies the dependencies between services and indicates where human interaction is required.

**WorkFlow Monitor.** This allows a user to follow the initialisation process. It also offers the option to the user to manually abort a step in the pipeline if necessary.

**KOBE Open Benchmarking Engine.** This allows a user to specify the experimental setup (datasets, processors, parameters) of data science experiments and to publish such specifications as repeatable and reproducible data processing experiments.

The architecture of the platform as well as semantification layer were published on top scientific conferences ICWE’2017, ISWC’2017, KESW’2017 and thus received the acknowledgement of the scientific community. The architecture of the BDE platform is also publicly acknowledged as a reference Big Data architecture by Big Data Value Association (BDVA) in many webinars and presentations.

### Research Achievements

**Semantic Data Lake**

Members of the technical team drove this idea throughout the period of the project. Although very ambitious and technically-challenging, there has been some success, and more is on the agenda. The idea is that we have recently made a huge leap in terms of data formats, data modalities, and storage capabilities. Dozens of data storage techniques have been created as a result. Today, we are able to store cluster-wide data, and to choose a storage technique that suits our application needs, rather than the opposite. If different data stores are interlinked and integrated, this data can generate valuable knowledge and insights. The pull of data where data results in its original form and format is called Data Lake. Semantic Web technologies can be used to solve the heterogeneity problem, mainly by mapping data schemata to ontology terms, and use those terms to query the data. This process renders the Data Lake a *Semantic Data Lake.*

**Sparkall**

Sparkall is an implementation of the Semantic Data Lake concept, using Apache Spark as the underlying query engine. It leverages Spark’s connectors, which are wrappers capable of extracting data from various data sources, and loading them into data structures queriable using SQL. Those data structures are called DataFrames and are distributed across a cluster and operated on in parallel. Sparkall provides a SPARQL interface to the outside. It accepts and decomposes a SPARQL query into sub-queries, detects relevant data from the Data Lake, and loads this data into DataFrames. It then creates SQL queries against those DataFrames, and

---

8 https://www.youtube.com/watch?v=zaR2I7rucWk
joins them to create the final results. In order to be able to extract data based on query terms, data has to be mapped to ontology terms. Sparkall supports the user in this mapping task, as well as query construction task by offering GUIs. The aim is to support non-RDF experts in interacting and, thus, extracting knowledge from Data Lakes.

Sansa

Most of the use cases gathered from the Societal challenges showed that a major research challenge is to perform scalable analysis of large-scale knowledge graphs to facilitate applications like link prediction, knowledge base completion and reasoning.

Analytics methods which exploit expressive structures usually do not scale well to very large knowledge bases, and most analytics approaches which do scale horizontally (i.e., can be executed in a distributed environment) work on simple feature-vector-based input.

Semantic Analytics Stack SANSA supports expressive and scalable semantic analytics by providing functionality for distributed computing on RDF data. SANSA is an open-source\(^9\) structured data processing engine for performing distributed computation over large-scale RDF datasets. It provides data distribution, scalability, and fault tolerance for manipulating large RDF datasets, and facilitates analytics on the data at scale by making use of cluster-based big data processing engines. SANSA provides:

- Specialised serialisation mechanisms and partitioning schemata for RDF, using vertical partitioning strategies,
- A scalable query engine for large RDF datasets and different distributed representation formats for RDF, namely graphs, tables and tensors,
- An adaptive reasoning engine which derives an efficient execution and evaluation plan from a given set of inference rules,
- Several distributed structured machine learning algorithms that can be applied on large-scale RDF data, and
- A framework with a unified API that aims to combine distributed in-memory computation technology with semantic technologies.

To achieve the goal of storing and manipulating large RDF datasets, we leverage existing big data frameworks like Apache Spark\(^10\) and Apache Flink\(^11\), which have matured over the years and offer a proven and reliable method for general-purpose processing of large-scale data.

---

\(^9\) [https://github.com/SANSA-Stack](https://github.com/SANSA-Stack)
Engagement, Outreach and Dissemination

The efforts to promote the Big Data Europe project and to disseminate its results continued successfully. The BDI is rather popular today because of its quality, but also because of its unique selling point: The seamless integration of linked data. The growing amount of high quality content on the project’s website attracts ever more visitors. The social network strategy was successful with the number of Twitter followers grown to over 1,300. Over 220 presentations have been uploaded to SlideShare and 13 videos have been uploaded to YouTube. The release of the final version of the Big Data Integrator (BDI) platform raised significant interest and the project has been promoted directly at high profile events, including the ISWC 2017 conference. This was one of the two high profile events that the project sponsored in 2017; in all around 12,000 people have heard about the project at different events during 2017. This is in addition to the webinars and workshops organised by the societal challenge partners.

The project’s flyer has been updated and two iterations of a separate handout giving an overview of the technology used in the BDI have been created. The project’s newsletter has over 1000 subscribers and the partners have collected over 1500 e-mail addresses of stakeholders with interests in each of the societal challenges. It is noted that more effort needs to be made to make the newsletter more attractive to boost engagement.

In the final year of the project, dissemination activities were combined with the community building activities of work package 2. The contact to coordinators of new projects beginning in 2017 with a view to offering the BDI platform as a means to handle their data and the partners were successful. Several new projects are taking up the platform: Data Ocean, SPECIAL, IASIS, DARE, and the Greek national Precision Medicine project.