BIG DATA VALUE cPPP

ANNEX 2 to the Monitoring Report 2018

Detailed explanations of KPIs

Version 1. July 2019

BDV BIG DATA VALUE ASSOCIATION

ABOUT THIS DOCUMENT

This document is part of the Big Data Value contractual Public-Private Partnership (BDV cPPP) Monitoring report 2018 and complements the information provided in the main report and Annex 1 to the main report.

In particular Annex 2 Includes detailed information per KPI: methodology, response rate, detailed input and assessment. It also provides information about the specific questionnaires used to gather information and the list of stakeholders who provided input.

Version control:

Version 1.0 (July 2019) Annex 2 to the first official version of the Big Data Value cPPP Annual Monitoring Report 2018.

***Disclaimer**: This document has been prepared by the Big Data Value Association (BDVA) and it reflects the views only of its authors.*

Table of Contents

1 INTRODUCTION

When referring to key performance indicators (KPIs) we use the following scheme: A Roman number refers to the annex in which the KPI is listed, i.e. I for common, II for specific KPIs and III for contribution to programme level KPIs. This is followed by an Arabic number indicating the number of the indicator in the respective table, e.g. KPI II.1 refers to the first KPI specified in the list of specific KPIs.

The list of KPIs for this cPPP, description, and target values are defined by the following documents:

1. KPI_note_for_annual_reporting-cPPPs.pdf released by DG RTD in February 2018, defining, describing and providing a methodology for the 4 common KPIs to all the cPPPs.
2. BDV cPPP contractual agreement[1].
3. BDV cPPP SRIA[2].

BDVA leads the production of the Monitoring Report 2018 of the Big Data Value cPPP as part of its contractual obligations in the cPPP. The work is developed by the BDVA TF2 (impact). To produce this report the Association has gathered input from:

1. **All the running and selected Big Data cPPP projects** during year 2018 (**16 projects** from the Call 1, 2016, **17 projects** from Call 2, 2017, and **10 additional projects** from the calls evaluated in 2018). The **32**[3] projects selected and funded during 2016 and 2017 are the main contributors to most of the KPIs. The projects selected in 2018 started mostly in January 2019 and therefore only contributed to provide some input in forecasted contributions to the BDV cPPP roadmap.
2. **All for-profit project partners** from the projects active during 2018: **236 for-profit companies** from which **131 are SMEs** (including the **18** start-ups funded by the DataPitch project and **29** start-ups funded by EDI during 2018, using the modality of cascade funding)**.**
3. **Members of BDVA**: **207 organisations**[4] from which **98 are private for-profit** companies.
4. BDVA Task Forces and BDVA Office.
5. EC DG CNECT G1.
6. The European Data Market Monitoring Tool[5] from which this report has gathered information for the Macro-Economic KPIs.

The information was collected in the following way:

For projects active in 2018:

---

1 http://www.bdva.eu/sites/default/files/BDVPPP_Contractual_Arrangement_.pdf

2 BDV cPPP SRIA was updated to version v4 in November 2017 and therefore it does not reflect the new re-arrangement of KPIs defined by DG RTD. The KPI catalog has been updated and it will be added to the new version of the SRIA planned for Sep-Oct 2018.

3 One of the projects from the 2016 Calls (KPLEX, ICT-35-2016 Enabling responsible ICT-related research and innovation, Sister-project of the cPPP) is counted as input for the financial contribution from the EC, This project ended in 2018 and it particular contribution to the KPIs listed in this document is very minor. For this reason this project is not counted as provider of input to the KPIS, and the totals of projects running in 2018 counted as 32)

4 BDVA members during 2018

5 SMART 2016/0063 – Study "Update of the European data marketMOnitoring Tool", IDC and Lisbon Councils

- Projects from the Call 2016 (AEGIS, BDVe, BigDataOcean, Data Pitch, DataBio, e-Sides, euBusinessGraph, EW-Shopp, FashionBrain, SLIPO, SODA, SPECIAL, TT, QROWD, MH-MD),

And,

- Project from Calls 2017: BigDataGrapes, BigDataStack, BigMedilytics, BODYPASS, BOOST 4.0, CLASS, Cross-PPP, DataBench, E2Data, EDI, FANDANGO, I-BiDaaS, ICARUS, Lynx, TheyBuyForYou, Track and Know, TYPHON.

The following tools were used to collect information:

- **Online questionnaire** with 80 questions to gather information for each of the KPIs and specific sections of the document (*Questionnaire for PPP projects running in 2018*).
- **Excel questionnaire** to report on delivered, ongoing and planned **contributions to the different technical priorities** to assess coverage of the SRIA implementation. (Technical priorities table for projects - Questionnaire*)*
- **Excel questionnaire** to collect information in relation to **innovations to market** ***(****Significant Innovations to Market – Questionnaire***)**

For projects from the call 2018 (Selected in 2018 but not active until 2019): CloudButton, Data Market Services, ELASTIC, EXA MODE, ExtremeEarth, INFORME, MOSAICO, MUSKETEER, Safe-DEED, SmartDataLake.

Information was collected through an **Excel questionnaire** to report on planned **contributions to the different technical priorities** to assess coverage of the SRIA implementation. (Technical priorities table for projects - Questionnaire).

For all Industrial partners (for-profit organisations) from call 2016 and call 2017 projects (including cascade funding) and BDVA members, information was collected through an **Excel questionnaire** to gather information for the private investments and SME mapping. (Investments *questionnaire)*

For BDVA members (all members): Information was gathered through an **Online questionnaire** with 40 questions mainly covering aspects of job creation, job profiles and skills development, Innovations and technical results, contribution to Standards, contribution to Big Data Ecosystem, and success stories. (*Questionnaire for BDVA members).*

Task Force and Subgroup leads from BDVA, BDVA office and the European Commission (DG CNECT G1) provided direct input to specific parts of the report and the annexes.

All questionnaires are available in section 13 of this annex.

2 MOBILISED PRIVATE INVESTMENTS (I.1)

Through this KPI we attempt to understand and capture/show the level of industrial engagement within the BDV cPPP.

All **for-profit organisations** participating in **cPPP projects** active during 2018[6] **and** all for-profit organisations **members of BDVA** were called to provide input to this KPI through an investments survey or questionnaire.

In total **296 organisations** were requested to provide input:

6 Project partners, linked third party of cascade funding

- **236 for-profit companies** from the projects running in 2018 from which **131 are SMEs** (69%) (including the **18** start-ups funded by the DataPitch project and **29** start-ups funded by EDI during 2018, using the modality of cascade funding)**.**
- **98 for-profit** companies from members of BDVA (from which 64 are SMEs, 65% of the total). 38 of these 98 organisations participated in projects 2018 (therefore counted under the 236 companies represented in projects).

2.1 Methodology

For the Annual Monitoring Report 2018 the European Commission suggested a common structure and methodology for all the contractual PPPs. This methodology identifies different categories of investments as follows:

- **EC**: EU contribution received for projects under the cPPP
- **Direct leverage** (A = A1 + A2):
  - **A1**: Direct contributions from the industry to the cPPP projects, as predetermined in the projects (co-funding)
  - **A2**: Additional investment or resources spent by your company for the execution of cPPP projects, beyond the direct contributions under A1 (these are typically due to shortfalls in fixed overheads for a project, infrastructure costs not covered by direct funding, consumables and maintenance costs, equipment costs, additional personnel costs, other costs).
- **Indirect leverage** (B = B1 + B2)
  - **B1**: Additional investment or resources spent as a follow-up to the activities and results related to cPPP projects (e.g. to exploit or scale-up results) and that are not supporting directly the execution of these projects. This can include additional funding mobilised thanks to the cPPP projects such as venture capital, other public funding, etc.
  - **B2**: Additional investments or resources spent that support the objectives of the cPPP and that have been triggered/encouraged by the existence of the cPPP, but that are not following up activities or results of the cPPP projects as such (hence, not counted in A1, A2 and B1).

Figure **1** visually shows the structure of this KPI as proposed by the EC.

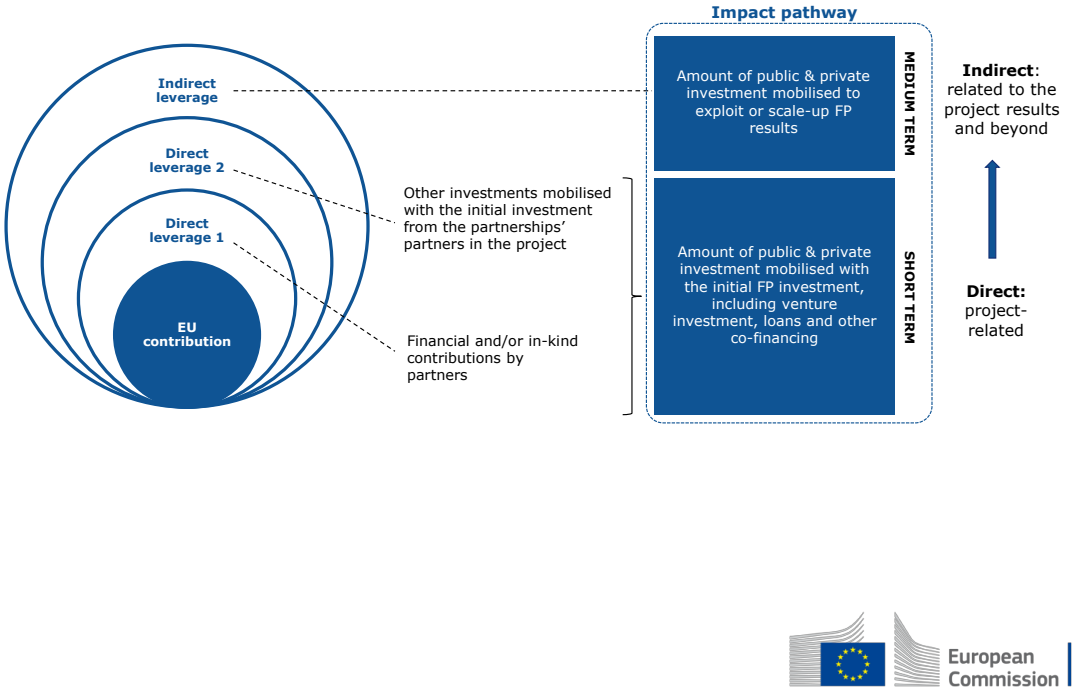


Figure 1: Methodology and KPI structure proposed by EC for MR2018 (cPPPs)

This new structure to present leverage investments is in line with the structure and methodology used by BDVA in previous annual Monitoring reports, allowing consistency in presenting the evolution of numbers over the years. The mapping in between the sub-KPIs (A1, A2, B1 and B2) from this year report and previous reports can be seen in Table 1.

New Leverage investments Categories (MR 2018)	Leverage investment categories used in MR2017 (and before)
B2: Additional investments or resources spent that support the objectives of the cPPP and that have been triggered/encouraged by the existence of the cPPP, but that are not following up activities or results of the cPPP projects as such (hence, not counted in A1, A2 and B1).	Amount of R&D expenses that are related to the Big Data PPP but are not related to EC funded projects. (This **excludes** any expenses that are funded by the EC by definition)
	Amount of R&D expenses resulting from follow-up investments of projects funded by the EC that are topic-wise related to the Big Data PPP however initiated outside the Big Data PPP (in FP7 or in H2020). **(This excludes any expenses that are funded by the EC by definition)**

B1: Additional investment or resources spent as a follow-up to the activities and results related to cPPP projects (e.g. to exploit or scale-up results) and that are not supporting directly the execution of these projects. This can include additional funding mobilised thanks to the cPPP projects such as venture capital, other public funding, etc.	Amount of R&D expenses resulting from follow-up investments of Big Data PPP projects (This excludes any expenses that are funded by the EC by definition)
A2: Additional investment or resources spent by your company for the execution of cPPP projects, beyond the direct contributions under A1 (these are typically due to shortfalls in fixed overheads for a project, infrastructure costs not covered by direct funding, consumables and maintenance costs, equipment costs, additional personnel costs, other costs)	Additional investments in execution of PPP projects (in BDV CPPP projects). (This excludes any expenses that are funded by the EC by definition)

Table 1: Mapping KPI structure proposed by EC for MR2018 and the one used in previous years

This mapping allows to present evolution of figures over the years and also to calculate cumulative figures and overall leverage and leverage factor.

A questionnaire was sent out the 296 companies[7]. The questionnaire requested the following relevant information[8]:

- Name of the company and cPPP projects where the company is involved (if any)
- Turnover, employment and overall R&D Expenses (regional coverage Europe)
- Estimation of the total amount of R&D expenses that are related to Big Data and Data-drive AI (BDV SRIA and roadmap).
- Estimation of the amount of R&D expenses that are related to the Big Data and Data-drive AI (BDV SRIA and roadmap) but are not related to EU-funded projects. **Answer from companies to this question is aggregated to calculate B2.**
- Estimation of the amount of R&D expenses resulting from follow-up investments of projects funded by the EC that are topic-wise related to the Big Data PPP however initiated outside the Big Data PPP (in FP7 or in H2020) (excluding any expenses that have been funded by the EC). **Answer from companies to this question is aggregated to calculate B2.**
- Estimation the amount of R&D expenses resulting from follow-up investments of Big Data PPP projects (Excluding expenses that are funded by the EC) (only for participants in cPPP projects). **Answer from companies to this question is aggregated to calculate B1**.
- Calculation of additional investment or resources spent by the company for the execution of cPPP projects, beyond direct contributions or co-funded reported to EC (in BDV CPPP projects). (only for BDV CPPP project partners**): Answer from companies to this question is aggregated to calculate A2**.

7 Questionnaire can be found in section 12 or this annex 2.

8 This same questionnaire was used to gather input about SME evolution.

The data was mainly requested for year 2018 and companies were also asked to estimate values 2019-2022 when possible (mainly for the follow up investments in cPPP projects).

The companies were further asked to provide the following information for their figures:

- the regional coverage of the figures;
- whether or not their organisation is an SME according to the Horizon 2020 participation rules.

SMEs were asked to provide turnover and employment data from 2014, and age of the company, but this historical data was used only to input I.3 (impact in SMEs).

The questionnaire sent to companies can be found in section 12 of this annex 2.

No extrapolations are done, nor external sources used (e.g. investments in the sector). Investments in Big Data Value does not relate only to the companies developing technologies, but those ones integrating them and bringing them to the different verticals, industrial user communities adopting these new technologies and data providers (industrial users or public sector). The BDV cPPP incorporates all this kind of companies throughout the different projects and BDVA members, however it is difficult to find external sources that combine investments of all involved stakeholders in many different sectors. In future reports, and once the number of companies increases and so do the maturity of the projects it might be relevant to assess investment figures per sector or per role in the Big Data Value chain.

The figures for EU funding and A1 values (in relation to the specific public investments and private co-funding for the year 2018) is meant provided by the EC[9]. Committed accumulative based on projects funded in Calls 2016, 2017 and 2018 (and therefore committed) is known and included.

2.2 Responses rates

In total **121 for-profit organisations provided input** to this KPIs. 99,2% of these organisations are partners or are linked to projects running in 2018.

- Considering only the cPPP projects as scope, the **response rate** is **51,3%.**
- Considering the overall ecosystem (cPPP projects and BDVA members) the **response rate** is **40,9%**.

Not all companies provided input to all the questions needed to calculate aggregated figures for the different categories and therefore the responses rates for each of the sub-KPIs is different.

New Leverage investments Categories (MR 2019)	Response rate (% total projects)
B2: Additional investments or resources spent that support the objectives of the cPPP and that have been triggered/encouraged by the existence of the cPPP, but that are not following up activities or results of the cPPP projects as such (hence, not counted in A1, A2 and B1).	31% (24% considering also companies not involved in projects)
B1: Additional investment or resources spent as a follow-up to the activities and results related to cPPP projects (e.g. to exploit or scale-up results) and that are not supporting directly the execution of these projects. This can include additional funding mobilised thanks to the cPPP projects such as venture capital, other public funding, etc.	7%

9 Current version of the document works with estimates as those numbers have not been provided to the Association by the EC.

A2: Additional investment or resources spent by your company for the execution of cPPP projects, beyond the direct contributions under A1 (these are typically due to shortfalls in fixed overheads for a project, infrastructure costs not covered by direct funding, consumables and maintenance costs, equipment costs, additional personnel costs, other costs)	19%

Table 2: Response rate per investment sub-KPI

The low response rate in input to B1 (7%) is logical, considering no cPPP projects have finished yet so only specific investments triggered by project activities (but not part of the projects) are included and logically the companies leave that question empty.

2.3 Overall results

Table 3 shows the EU contribution granted by the EC in calls for all the years since the launch of the cPPP and the committed private contribution as per Grant agreements. Projects active in 2018 are result of the calls 2016 and 2017.

In total the EC has allocated 201,3 MEuros in projects of the Big Data Value cPPP.

To calculate the amount corresponding to the execution of projects during the year 2018, we estimate a 2,5-year project duration average, and a flat distribution of funding along that time. Considering the contribution granted in 2016 and 2017 (projects running in 2018) we estimate the EU contribution received from projects of the cPPP during 2018 **is 63,52 MEuros.**

Description	KPI	Unit	2015	2016	2017	2018	Total
Total EU contribution for projects granted during reporting period (starting the year after)	EC	MEUR	N/A	69,90	88,90	42,50	**201,30**
Total Private contribution for projects granted during reporting period (starting the year after)	A1	MEUR	N/A	13,70	10,40	0,00	**24,10**

Table 3: Total EU and private contribution for projects granted since the launch of the PPP

Table 4 shows the overall results of private investments for 2018.

Description	2018 (Meuros)	Response rate (%)
EC: EU contribution received for projects under the cPPP	63,52	Calculated over 100%
Direct leverage (A = A1 + A2)	**19,16**	

A1: Direct contributions from the industry to the cPPP projects, as predetermined in the projects (co-funding)	9,64	Calculated over 100%
A2: Additional investment or resources spent by your company for the execution of cPPP projects, beyond the direct contributions under A1 (these are typically due to shortfalls in fixed overheads for a project, infrastructure costs not covered by direct funding, consumables and maintenance costs, equipment costs, additional personnel costs, other costs)	9,52	19%
Indirect leverage (B = B1 + B2)	**448,31**	
B1: Additional investment or resources spent as a follow-up to the activities and results related to cPPP projects (e.g. to exploit or scale-up results) and that are not supporting directly the execution of these projects. This can include additional funding mobilised thanks to the cPPP projects such as venture capital, other public funding, etc.	7,57	7%
B2: Additional investments or resources spent that support the objectives of the cPPP and that have been triggered/encouraged by the existence of the cPPP, but that are not following up activities or results of the cPPP projects as such (hence, not counted in A1, A2 and B1).	440,74	31%

Table 4: Investment results 2018

Direct leverage (A = A1 + A2)

Direct leverage includes the direct contributions from the industry to the cPPP as predetermined by the projects (co-funding) and the additional investments or resources spent by companies for the execution of the projects beyond the known direct contribution.

A1 - **Direct contributions from the industry to the cPPP projects, as predetermined in the projects (co-funding).** This number is known by the EC, but not to the Association that has calculated a number based on the following assumptions.

Total amount of private funded committed for projects granted during 2016 and 2017 (so active in 2018) is 24,1 MEuros (13,7Meuros in projects granted 2016 and 10,40 MEuros in projects granted in 2017). Considering an average duration of 2,5 year per project, the amount of private contribution during reported period 2018 is **9,64 MEuros.**

A2 - **Additional investment or resources spent by companies for the execution of cPPP projects**, beyond the direct contributions under A1 (these are typically due to shortfalls in fixed overheads for a project, infrastructure costs not covered by direct funding, consumables and maintenance costs, equipment costs, additional personnel costs, other costs).

This input is provided though the questionnaires. Companies have reported **9,52 MEuros** of additional investment in project execution (with a response rate of 19%). Extrapolation to the 100% of the companies would take us to an additional private investment of **49,9 Meuros**. However, extrapolations are not considered in this exercise but only aggregated numbers.

In total companies have reported investments in project execution of 19,16 Meuros (30% of the EC contribution). With extrapolated values total private investments would reach 59,55 Meuros (93,5% of the EC for the year 2018)

Indirect leverage (B = B1 + B2)

B1 - **Additional investment or resources spent as a follow-up to the activities and results related to cPPP projects** (e.g. to exploit or scale-up results) and that are not supporting directly the execution of these projects. This can include additional funding mobilised thanks to the cPPP projects such as venture capital, other public funding, etc.

7,57 MEuros have been reported for this KPI, with a 7% of response rate. The BDV cPPP projects started in 2017 and no project had ended in 2018 it is normal to have a small figure close to 0, as it is mainly the Start-ups in some of the incubators, and some specific concrete areas already delivered in projects that are subject to follow up investments.

B2 - **Additional investments or resources spent that support the objectives of the cPPP and that have been triggered/encouraged by the existence of the cPPP**, but that are not following up activities or results of the cPPP projects as such (hence, not counted in A1, A2 and B1).

73 companies provided input to the specific questions linked to this KPI so a **31% response rate** considering all the for-profit organisation in project, and a 24,6% if we consider all the for-profit organisations in projects + Association.

The total based on aggregated numbers from those 73 companies **is 448,31 M€**:

- **70,40 M€** relate to as follow-up investments for EC-funded projects topic-wise related to Big Data
- **370,33 M€** as investments related to the Big Data PPP, but not related to EC-funded projects

Considering the response rate, we can assume this number is much higher, but as in previous years we only report on aggregated numbers based on real individual inputs with no extrapolation. As the methodology used this year to calculate this KPIs is the same that the one used in previous Monitoring Reports exercises we can show the evolution.

Table 5 shows the evolution of reported numbers in private investments from 2015 to 2018.

KPI	Description	Unit	2015	2016	2017	2018	Total
B2	Estimate the amount of R&D expenses that are related to the Big Data PPP but are not related to EC funded projects. (This excludes any expenses that are funded by the EC by definition!)	MEUR	245,80	289,10	388,67	370,33	1.293,90
B2	Estimate the amount of R&D expenses resulting from follow-up investments of projects funded by the EC that are topic-wise related to the Big Data PPP however initiated outside the Big Data PPP (in FP7 or in H2020). (This excludes any expenses that are funded by the EC by definition!)	MEUR	35,10	49,40	83,55	70,40	238,45
B1	Estimate the amount of R&D expenses resulting from follow-up investments of Big Data PPP projects (This excludes any expenses that are funded by the EC by definition!)	MEUR	N/A	N/A	1,09	7,57	8,67

A2	Additional investments in execution of PPP projects (in BDV CPPP projects). (This excludes any expenses that are funded by the EC by definition!)	MEUR	N/A	N/A	3,46	9,52	12,97
A1 (estimated 2018)	Annual private contribution (estimated for reporting period 2018)	MEUR			5,48	9,64	15,12

Table 5: Evolution private investments on BDV cPPP over time

Aggregated to the numbers reported in 2015 (280,9 M€), 2016 (338,5 M€) and 2017 (482,25 M€) the amount of mobilised private investments since the launch of the cPPP is **1.569,1 M€ (1,57 B€).** Considering the amount of EU funded allocated to the cPPP so far (201,30 M€) the BDV cPPP ends 2018 with a **leverage factor of 7,8** much higher than the leverage factor of 4 committed contractually.

3 NEW SKILLS AND/OR JOB PROFILES (I.2, II.8)

3.1 KPI I.2 (Number of jobs created)

This KPI considers new skills and job profiles created or forecasted within a reasonable timeframe in the future. This KPI is split in 2 different sub-KPIs:

- Job creation (quantitative value and qualitative description of causality)
- Job profiles established and skills created by 2023 (quantitative value and qualitative description).

BDV cPPP has 3 additional specific KPIs contributing to the job creation, profiles and skills:

- II.2 (Number of Data Companies) (Macro-economic)
- II.5 (Increased number of Data workers) (Macro-economic)
- II.8 (High establishment availability of big data value creation skills development)

A combination of the input provided in the 4 KPIs (I.2, II.2, II.5 and II.8) is needed to have a full overview about how job profiles and skills are being created within the activities and sectors covered by the BDV cPPP, the number of workforce affected in the European Union and forecasts of its evolution on the employment markets[10].

Target and response rates

All Big Data cPPP projects active in 2018 (32 projects) and all the BDVA members were asked to contribute with input to these KPIs. Responses rate:

- 30 over 32 projects provided input to these KPIs (**94%**)

10 KPI_note_for_annual_reporting-cPPPs.pdf

- 35 over 207[11] BDVA members provided input to these KPIs (**17%**)

3.1.1 Job creation

77% of the BDV cPPP projects indicated their project will contribute to job creation by 2023, with an estimation in accumulated numbers of thousands.

Projects indicated in a qualitative way the causality in between the project activities and the job creation, and some provided an estimation in numbers (although several indicated that a serious estimation is difficult to do at this stage). Job creation in discrete numbers is not possible to calculate precisely at this stage but **estimated numbers go over 7.500 new jobs created by 2023 linked to project activities** and many more considering indirect effect.

BDV cPPP projects contribute to job creation in Europe 1) **increasing of market share of Big Data Technology providers** in Europe; 2) developing **new job profiles** that generate new jobs; the creation; 3) **developing new opportunities for entrepreneurs and startups** in the new Data Economy; 4) generating job opportunities by **increasing data sharing**; 5) creating new **jobs** already **during the life time of the project**; and, 6) forecasting jobs created as a **follow up of project results**.

1) **Increase market share of Big Data Technology providers in Europe,** as for example:
   - **DataBio** is expected to increase the market share of Big Data technology providers in bioeconomy and, consequently, these providers will create jobs for data workers, other ICT and Earth Observation professionals, sales and field professionals, to work on this new business A quick survey of the DataBio partners revealed that 42 new jobs are expected to be directly created by them. 80% of them will be data workers in Europe, to work on follow-up commercial and research projects. The remaining 20% of them will be sales and field professionals (e.g. agronomists), to support commercial sales and provide expert advice.
   - **FashionBrain**'s solutions will increase the revenue of the European stakeholders involved in the project, which in turn will create more jobs. Moreover, our solution requires the use of European crowdsourcing.

2) Increase because of the **new job profiles developed**, as for example:
   - **MHMD** already contributed to the creation of new jobs: blockchain developers with specific expertise in the health domain have been hired/trained by more than one company in MHMD. Such figures are expected to grow in the future.
   - **BigMedilytics** will open the door to new services and projects in the context of Health Sector using Big Data and Artificial Intelligence solutions. Currently, there is not a relevant number of job offers for data scientists in this sector, as it is complex to access and process data. Bigmedylitics addressed specifically the latter challenge. In addition, GDPR experts for data provided from wearable devices (medical and well-being context)
   - **BodyPass** helps defining in the garment industry the requirements for a new type of product designers: "design engineering analyst"; The success of this project, could improve the commercial approach to clients, so we will need more qualified person to follow up the market needs.

3) **Creation, support and new opportunities for entrepreneurs and startups** in the new Data Economy (particularly stimulated by the Data incubators as DataPitch, but also mentioned by other projects such as BigDataGrapes and BigMedilytics). For example:

[11] Total number of members accounted in 2018

- **BigDataGrapes** project expects to support the creation of new agri-tech startups via the project Data Marketplace.
- Part of the **BigMedilytics** project's results are currently aiding the start-up of a venture within the company. Due to the data-driven proposition of the start-up, the venture will employ more people with a data science background. There will also be other data science workers employed outside the venture to support the activities of the venture.
- It is estimated that companies embarking in **EDI incubator** will create new jobs.
- **DataPitch** has seen an impact on job creation from the startups already accelerated. These companies have provided over 60 new jobs to the EU economy. In 2019 DataPitch will be taking on 29 new startups. Estimation is based on similar survival projections to those we have seen in past projects e.g. ODINE.

4) **Job opportunities generated by increasing data sharing**, as for example:

- Some of the expected outcomes of the **BOOST 4.0** project are increasing the data analysis and data sharing on the industrial sector, which will need new workers specialised in these technologies, as well as data scientists to allow these activities. That means new job profiles for many SME's that will have to hire new workers to cover, especially in tech companies that will have to develop and deploy new solutions, but also in the industrial sector, where it will be needed to analyse the data generated.
- The **E2Data** framework will enable new services across the fields of Health, Finance, Security. These services were not feasible to be implemented with currently available technologies.
- Key motivation of Cross-CPP project is to give cross-sectorial industries access to the great spectrum of sensor data coming from high volume products from various industrial sectors (vehicles, smart home devices, etc.). With the increasing number of connected sensors and actuators within such mass products, this number will rise in short-term.

5) **New jobs created already during the life time of the project**, as for example:

- AEGIS: The project already created more than 20 new jobs. The main technological result of the AEGIS project is an online platform providing data analytics as a service. It reduces the barrier for organisations to start using big data analytics but requires from them to have data analysts – the target users of the AEGIS platform.
- BigDataOcean has set up a new organisation exploiting the project's results with new employees.

6) New jobs created as a follow up of project results, as for example:

- **EW-Shopp** expects that by 2022 new jobs will be created directly as a result of this project.
- **SODA** foresees involving internal use and business development by project partners as well as uptake by third parties of the technology.
- **BigMedilytics** expects to create more positions for either PhD students and/or FTE to expand further the developments of the project in the area of healthcare by leveraging the innovations developed that might be applied to other scenarios.
- **TT** assumes that beyond the project, especially those industrial partners focused on the development of software products and tools for big data processing would benefit from the knowledge acquired during TT and they would invest resources on continuous improving the product hence promoting the creation of jobs thanks to the research initiative.
- **Big data Stack**: the LeanXcale company may leap forward following the research and development that this company is doing; The IBM research has already developed a product at beta level that will most probably lead to job creations; The various universities participating (UPRC, UPM, Glasgow) advance their research and contribute to bigdata know-how; The various use case companies involved in this project may get the knowledge and experience that will help them adapt to bigdata.

- **TheBuyForYou** stimulates jobs creation for data workers in public administrations in the field of procurement intelligence, data journalists, using the procurement intelligence tools.
- **Typhon** technologies will lead to the creation of a new skills profile of hybrid data management expert having the skills in hybrid data stores, data management, and development of new innovative applications that combine and exploit the diverse big data already collected by enterprise and government systems.

Some projects that have reported no contribution to Job creation focus normally on developing more efficient technologies e.g. SLIPO project that expects to lowers the human resources/costs for data integration or Lynx project that does not expect to create jobs but changes in current practices, improvement in the efficiency of certain work methodologies related to compliance.

On the other hand, **40% of the BDVA members stated that their participation in the BDVA/BDV CPPP has already contributed** directly or indirectly **to job creation**, either directly because of the hiring of new experts to develop H2020 projects (Big Data Engineers, Computer scientists, data scientists, big data architects, big data analysts and programmers), new PHDs (e.g Big Data in bioinformatics for diabetes profiling), startups created (e.g Startup Logical clocks by RISE), new profiles hired to develop operations (Big data researchers, data center experts, Data Scientists, Analytics Experts, Big Data Architects. Big Data Tools administrators, Big Data Tools Experts, Big Data Application developers).

3.1.2 Job Profiles established / Skills created

30% of the projects running in 2018 reported contribution to the creation of new Job profiles. Projects reported **48 job profiles** were created or identified in 2018, and **106 new job profiles** are reported as expected to be created from 2019 onwards and by the end of the project[12] linked to the project activities.

Projects directly contributing to new job profiles are DataPitch, SODA, TT, MH-MD, BigMedilytics, Boost4.0, E2Data, EDI, and TYPHON.

The profiles have now been identified but not in all cases defined in full during 2018.

Specific Job profiles identified/created (by 2018):

In the Healthcare domain (MH-MD and BigMedilytics projects):

- Blockchain developers
- Data curation, anonymization and treatment
- To manage the cluster installed as well as the secure channel to access the cluster
- Data scientist profiles
- Big data analytics, cloud computing and distributed real-time streaming systems, etc
- Project management, ethics and regulatory submissions, patient contact, consenting patients, patient follow up etc.
- Knowledge Engineers: experts in knowledge representation, ontologies, controlled vocabularies in the biomedical domain, rule mapping languages
- Knowledge Extraction and Mining: experts in techniques for extracting and annotating knowledge from textual and non-textual material

12 End of 2019 considering the length of the 16 active projects in 2017

- Knowledge Management and Representation: experts in techniques for integrating knowledge extracted from heterogeneous data sources, and representing and exploring all these knowledge in knowledge graphs
- Data scientists with special expertise in health sector. Skills in data analytics, big data, data processing, databases, programming, administration of databases, business, economic aspects of healthcare and healthcare processes in a hospital.
- Machine Learning: experts in machine learning, natural language processing, statistical analysis
- Data scientist, big data engineer, and data ethics specialist
- Researcher track and trace, track and trace analyses

In the manufacturing domain (BOOST4.0 project):

- Software developers specialised in graph data
- IoT Engineer – engineering of the IoT platform and integration with machines
- Data scientist – profound knowledge in data analysis and machine learning
- Research associate – strong technical background and skills on data analytics, deep learning, semantic
- IT manager
- Maintenance engineering
- Data analyst
- Project manager-Data science
- Software engineer
- Senior AI researcher

DataPitch incubator estimates at least 18 new job profiles from the 18 start-ups accelerated in 2018 - 1 job profile per funded start-up (specifics not entirely known yet by the programme)

EDI incubator has identified 3 new job profiles:

- Big Data Infrastructure Engineer
- Data Scientist
- User Experience Engineer

Typhon project:

- Hybrid Data Manager - having new skills of hybrid data management for hybrid data stores, hybrid data management, and development of new innovative applications that combine and exploit the diverse big data already collected by enterprise and government systems.

Additionally, although not reported as new job profiles but new skills developed, the DataBio project has developed skills to offer Big Data services to the bio-economy sectors, that need workers that effectively combine a number of skills. These workers need to combine expertise in Big Data technologies (e.g. data analytics and visualization), Earth Observation technology (satellite, airplane and drone based) and services, domain knowledge (in agriculture, forestry or fishery, but also in the particular regions, e.g. crops cultivation

in Northern Greece). In addition, they should be able to support an integrated service to the end-users. The pilot trials of the project involved these skill sets.

67% of the projects running in 2018 reported **contribution to the generation of new skills** by end of the project.

In addition to the skills linked to the new job profiles reported above new skills are expected to be developed in cross sectorial domains (e.g new skills in the form of 'privacy aware data processing' and 'privacy aware big data innovation' as reported by the SPECIAL project) and in specific sectors (e.g analysis techniques using weather data, reported by the EW-SHOPP project).

Some projects will contribute to the combination of skills and expertise needed to support users in an end-to-end scenario (e.g. DataBio project, where workers need to combine expertise in Big Data technologies, Earth Observation technology and services, domain knowledge (in agriculture, forestry or fishery). The BDV cPPP incubators (DataPitch and EDI) help developing skills of startups in both technical and non-technical skills needed to develop business in the Data Economy.

Not enough quantitative data has been provided to provide discrete numbers in relation to the skills creation but details of activities developed by the projects, and specific skill sets develops show the **richness of the programme in skills development**. In particular:

- The project **AEGIS** organised several training sessions to master students to learn how to use the AEGIS platform and integrated it in Big Data analytics tools.
- **BigDataOcean** is helping to develop new data and business analyst skills in traditional sectors (e.g. energy, maritime).
- **Data Pitch** incubator provides startups and SMEs with training and support e.g in the development of skills related to data policy and privacy. The programme has created a toolkit to up-skill startups within this area: https://datapitch.eu/privacytoolkit/. These are areas in which the startups have requested this kind of knowledge transfer. To impact on the wider startup ecosystem, the programme shares its materials via our Youtube channel: https://www.youtube.com/channel/UCHA-yGRbHgAdB8KqfTs6gyw
- **EDI** incubator also provides skills development for the startups of its incubation programme and in particular developing skills for big data engineers, data scientists, sales experts, marketing staff, user experience engineers.
- In the **DataBio** project the developed skills help to offer Big Data services to the bioeconomy sectors, that need workers that effectively combine a number of skills. These workers need to combine expertise in Big Data technologies (e.g. data analytics and visualization), Earth Observation technology (satellite, airplane and drone based) and services, domain knowledge (in agriculture, forestry or fishery, but also in the particular regions, e.g. crops cultivation in Northern Greece). In addition, they should be able to support an integrated service to the end-users. The pilot trials involved these skill sets.
- **e-Sides**: Big Data Analytics Ethical and Legal review specialists.
- **euBusinessGraph:** Working with company data in new ways.
- The analysis techniques using weather data are not widespread and **EW-Shopp** will contribute to refine these techniques and create new skills; also, the development of analytics platform dealing with large-scale events is essentially a cutting-edge topic and we aim to help the creation of tools and capabilities to tackle this successfully.
- **FashionBrain**: Web developing using database capabilities.
- **SODA**: Multi Party Computation (MPC) application development. Academic partners include MPC in the curriculum. MPyC framework specifically targets education.

- Considering that strong anonymisation reduces data utility, **SPECIAL** aims to demonstrate to companies that it is possible to support innovative product and services that use personal data by obtaining consent for the processing from data subject and providing transparency with respect to the processing and sharing of personal data. In this context SPECIAL aims to support new skills in the form of 'privacy aware data processing' and 'privacy aware big data innovation'.
- **BigMedilytics**:
  - Machine Learning/Deep Learning, Hadoop, Spark
  - Application of machine learning techniques for processing Electronic Health Records
  - Infrastructure deployment for trustful EHR analysis and data sharing
  - EHR anonymization in distributed environments
  - Expertise in writing SAS and R code
  - Expertise in applying machine learning techniques
  - Expertise in setting up a project structure
  - The project contributes to a expansion of radiologists expertise in becoming "machine learning literate" and able to deploy, exploit, and even co-develop machine learning based systems, frameworks or their implementation in the clinical context. This is an essential skill, as diagnostic workflows increasingly involve the complex interaction between human experts, and machine based information processing, analysis, prediction, and identification of corresponding markers.
  - On the technology side, skills developed triggered by this project, are computer scientists, data scientists, and engineers skilled in communicating with medical expert stuff, and jointly developing solutions relevant for daily routine.
  - New skills related to Big Data in bioinformatics for diabetes profiling.
  - New skills related to the Personal data analytics
  - New skills related to Data protection from wearables
  - data stewards and data scientist
  - Integrating structured and unstructured data into a knowledge graph
  - Defining ontologies for modelling knowledge in the knowledge graph, as well as mapping rules for transforming extracted knowledge into facts in the knowledge graph
  - Definition of methodologies for evaluating and curating the knowledge graph
  - Implementation of knowledge discovery methods, e.g., to discovery interactions between drugs and drugs and side-effects
  - Definition of techniques for distributing the knowledge graph and for efficiently accessing a federation of the distributed knowledge graphs
  - Processing with natural processing techniques open bibliography data, but also phone transcripts.
  - Predictive and descriptive analytics
  - Graph representation & mining techniques
  - Integrating structured and unstructured data into a knowledge graph
  - Defining ontologies for modelling knowledge in the knowledge graph, as well as mapping rules for transforming extracted knowledge into facts in the knowledge graph.
  - Definition of methodologies for evaluating and curating the knowledge graph.

    - Implementation of knowledge discovery methods, e.g., to discovery interactions between drugs and drugs and side-effects.
    - Definition of techniques for distributing the knowledge graph and for efficiently accessing a federation of the distributed knowledge graphs
    - The project creates skills in the combination of platform development with medical imaging analytics.
    - New approaches to understand and utilize the knowledge inside MRI images in breast cancer studies are needed. New technical applications are developed and utilized
    - Data science, big data engineering, and data ethics and GDPR
    - creation of business skills for data scientist and health care providers. Development of new research methods to examine the social aspects of the use of big data technologies. Training of research in creating impact by visualisation of scientific results.
- **BodyPass**: Technician, physicians and researchers to develop new skills concerning the methodologies and the potential application of 3D body scanner for medical/health uses in order to carry out basic and applied research. Creation of a research framework to implement new hard skills related to virtual reality, computational science, psychology, and the related intersection among these. Incubation of new professional roles for fashion and design companies, which are in between of 3D Designer and Data Analyst: design engineering analysis.
- By boosting the value of data in the industry, **BOOST4.0** is contributing in a very significant way to de development of new skills, mostly related with data analysis and data treatment. Some examples identified are:
    - Skills in semantics and graph data
    - ETL processes dedicated for time-series data
    - Machine learning for root-cause analysis
    - Interactive dashboard design
    - IoT technologies
    - Engineering management
    - Big data analysis skills
    - Knowledge in sensor technology
- **DataBench**: Big Data Analytics Benchmarking Specialists
- **Icarus**: Enhancement of the technical / technological skills of the consortium R&D partners in terms of Data Management, Data Analytics and Security
- Each of the following skills sets will be made possible by technologies developed in the **TYPHON** project: Polyglot Data Management, Polyglot Data Deployment, Polyglot Data Evolution.

Among the **BDVA members**, **51%** of organisations have reported contribution to the **creation of new Job Profiles** and **almost 60% contribute to the creation of new skills linked to the Big Data Value**.

Some examples of job profiles and skills developed by BDVA members:

- **RISE**: Big data with focus on streaming data
- **ATC**: data scientist, combining outstanding mathematical know-how with industry practice experience

- **UPV**: Programmers highly skilled in C++ and Python programming for developing AI based solutions. Big Data architects for maintaining big data architectures.
- **Software AG**: Low code data analyst. Matter expert from nin.IT department is enabled to do data analytics including prediction without deeper know-how in special data analysis tools. "Democratize data analytics".
- **ITA**: Natural Language Processing Engineering, expertise in development of NLP models for NER, PoS, Opinion Analysis and Machine Learning and Datamining Engineering, expertise in development of deep learning models from Big Data Projects.
- **TRAGSA**: Data analysts, R (statistical programming language) expert.
- **Tree Technology**: Two Big Data Engineer positions. This position is in charge of developing, maintaining, testing and evaluating big data solutions within organisations (our own, and especially our clients). These profiles are specialised either on Data Science and AI techniques (i.e. Machine and Deep Learning, Computer Vision, Image and Video processing and analysis, etc) or Distributed Architectures technology (i.e. Spark/Flink/Hadoop ecosystems; NoSQL; parallel and distributed computing; etc.). (Tree technology)

Some interesting initiatives by members of the BDVA include (non-exhaustive list):

- **Institute of Advanced Studies Kőszeg** is developing two projects using big data capabilities, one in environment management and one in health care. New skills developed are data integration specialist and data analytics.
- **RISE** focuses on the skills to handle streaming data.
- **Paluno, University of Duisburg-Essen** develop relevant skills via lectures, seminars. Bachelor and master thesis, students' skills on predictive data analytics (using traditional and advanced machine learning techniques) have been developed.
- **Engineering** had developed Internal training courses on skills related to the entire context of big data, from system level to data modelling and analytics.
- **UPV** has a BSc in Data Science, a MSc in Big Data Analytics, a MSc in Artificial Intelligence and a MSc in Grid and Parallel computing. Additionally, UPV offers short summer courses in Machine Learning for Big Data Analytics, in Cloud computing, in Deep Learning, in Business Intelligence, in Internet of Things and in Data Protection.
- **Huawei** continued efforts to enhance the big data stream processing engines with new Online ML prediction algorithms and advanced Geo-spatial stream analytics. In addition, they added new features mainly Stream-SQL CEP, real-time inference for Deep Learning models and dynamic cost estimator that dynamically assigns resources based on the job metrics and user optimization goals. Furthermore, they have also deployed big data and AI technologies to analyse large amounts of data generated by Data Centres and Cloud Platforms. In addition, through the project BigMedilytics, the health sector became more prominent in the entire company and that on a global scale more jobs in the area of health solutions and underlying technology fields have been created.
- As public research institute, **Inria** welcomes Master, PhD, and Post-Doc students. Moreover, Inria runs the Inria learning Lab (see https://learninglab.inria.fr/), which produces MOOCs in the domain of computer science and applied Mathematics. These include topics like security and privacy, semantic web and linked data, relational databases, python, and bioinformatics tools and algorithms.
- **ITA** develops skills activities on AI, especially NLP projects based on Big Data architectures. Skill, massive text processing, deep learning NLP models.
- The **TU Wien** began offering a new Data Science Masters course in October 2018. This masters course has modules on: - Fundamentals of Data Science - Machine Learning and Statistics - Big Data and High Performance Computing - Visual Analytics and Semantic Technologies - Domain-Specific

Aspects of Data Science. The TU Wien is also running a Continuing Education programme for companies on Data Science and Deep Learning (http://www.ifs.tuwien.ac.at/idsdl/). 24 companies participate in this programme. The module topics are: - Ethics and Equal Opportunity - Fundamentals of Data Science - Scalable Data Technologies - Deep Learning - Text Analysis and Word Embedding - From Data to Markets This Continuing Education Programme was nominated as one of the finalists in the category "Digital skills for ICT professionals" of the European Digital Skills Awards 2018.

- **University of Southern Denmark** has developed and delivers courses within Big Data and AI.
- **ITI - Instituto Tecnológico de Informática** develops skills on reinforced learning and by self-learning.
- The **University of Granada** is very active in Data Science and Computer Engineering. In concrete, they have an Official Master's Degree focused on Computer and Data Science. The Official Master's Degree in Data Science and Computer Engineering offers training in a research and professional profile that is in high demand today (and foreseeably in the medium and long term) in the field of Information Technology and Communications. The need to analyse large volumes of data makes Data Science one of the disciplines with the greatest demand for employment in the immediate future throughout the world. This huge volume of data to be processed requires high-performance computing platforms and integrated systems for specific applications in the field of Computer Engineering. This master's degree provides the student with the opportunity to train as a researcher in these two lines of work, thanks to a teaching staff integrated in different research groups of recognized international prestige. The Master's in Data Science and Computer Engineering includes two specialties: Specialization in Computer Engineering and Networks and specialty in Data Science and Intelligent Technologies. http://masteres.ugr.es/datcom/
- **Maxfone** delivers skills development activities on big data management (data storage, elaboration and analysis - new software application), machine learning for image recognition, and machine learning for sentiment analysis;
- **FEUGA** assess professional profiles for the future for the galician universities (ordered by the regional government) and skills related with big data are being considered in this study.
- **Prizztech Robocoast** has started Robocoast derived education program AI Academy in the Satakunta University of Applied Sciences (SAMK).
- Staff members of **Tree Technology S.A.** collaborate with MSc and post-doc programmes at Universities, in new careers related to Big Data.
- **Siemens AG** is conducting internal digitalisation training for their internal decision maker and experts. Siemens will conduct several AI trainings for their internal decision maker and experts Siemens employees are giving Big Data related lectures at Technical University in Munich and Berlin and at CERN. Siemens is supporting the Data Science Lab (see http://dsl.ifi.lmu.de/cms/) which fosters the Big Data skill development along concrete challenges and research projects. They see vocational education and training as very central and effective as it allows them to establish more direct link between education and employment. The pace of change is so fast that learning requires close relation to practice.
- **GMV** collaborates with higher education organizations such as the following from Madrid, Sevilla, Valencia and Valladolid respectively. In particular:
    - URJC's Data Science Master, where GMV offers internships in diverse application fields such as precision agriculture, prevention of banking fraud, cybersecurity, digital surveillance, detection of datacenter anomalies, monitoring and analysis of internet publicity campaigns, management of clinical and epidemiological data and evidence-based clinical rehabilitation.
    - Uni. Sevilla´s Data Science and Big Data Master, where GMV offers applied Data Science training as well.

- Utrecht Network summer school at University of Valencia, where GMV deals about machine learning's impact on society, and the growing concerns about its transparency and fairness.
- Uni. Valladolid´s Master in Big Data Science, where GMV is involved in Fundamentals and Applications of Data Science and Big Data.

Finally, **60% of the projects** and **51% of the BDVA members** have reported **contributions to the to the Skills Agenda for Europe**[13]. Some examples of specific contributions:

- **AEGIS** develops training of students in data analytics
- **BigDataOcean** develops training of students in data analytics
- **DataPitch's** Data Innovation Academy's curriculum has been developed to help data-driven innovation to startups and SME's
- **Databio** has indirectly contributing by executing a large number of pilots (26 throughout Europe) and by engaging a large number of stakeholders (100+), farmers, forestry workers and fishermen are trained in using new big data technologies and acquire new skills. Hackathons and stakeholder events have been important tools for that in 2018 in addition to the first phase of the pilot trials.
- **EW-Shopp** has promoted cooperation among education and industrial sector with particular reference to e-commerce, retails, and marketing SMEs and startups, engaging in specific activities of know-how sharing and skill development (seminars, educational talks, hands-on sessions, etc.)
- **BigMedilytics** has developed digital and non-digital skills: Extension of open source frameworks for real-time data processing, development of machine learning algorithms for prediction of risks, CEP applied to healthcare scenarios, training and exchange of know-how/best practices among the PhD community in the area of Bioinformatics, etc; Education of digital health experts; Key competences promoting entrepreneurial and innovation-oriented mind-sets.
- **BOOST 4.0** contributes through its partners, some of our which are highly involved in the Skills Agenda for Europe. For example:
  - INNO is working closely with GESTAMP in the 'Digital Skills and Jobs Coalition' action and also in the "Upskilling Pathways: New Opportunities for Adults" actions to support cooperation in the form of workforce re- and up-skilling in joint education, employment and industry programmes with the goal of improving the digital skills of the wider population, not just IT professionals. This is implemented as part of the European Network of Digital Innovation Hubs (DIH) initiatives and with a focus on the contributions to the Basque Industry 4.0 strategy recently recognised by the United Nations Industrial Development Organization (UNIDO) as a worldwide reference for digital transformation of Industry 4.0.
  - ATLANTIS Engineering, who is a member of the European Digital Learning Network (DLEARN), a non-profit association promoting the use and benefits of ICT tools in all levels of education and training. The DLEARN is a Digital Skills and Jobs Coalition Pledgers, and it interacts with certain DGs and professional association to monitor the current status and propose steps for the future. In fact, Dr. Ifigenia Metaxa is currently the Chairman of the VET working group.
  - EPCC contributes on the Upskilling adults, the Digital skills and de VET work groups (until UK leaves the EU)
  - RISC contributes to the Key Competence workgroup
- **Fandango** contributed to develop skills in data journalism

13 http://ec.europa.eu/social/main.jsp?catId=1223

- **ICARUS** mainly contributes and will contribute to the 7th key Action of the Skills Agenda for Europe, namely to the "Key Competences", by helping IT professionals, as well as graduate students and other industrial partners acquire or enhance their skills in Data Management and Data Analytics, and through the business model to be adopted, facilitating the development and delivery of added value services over the data that will be made available, to promote entrepreneurial and innovation-oriented mind-sets and skills

From the BDVA members some examples of specific contributions:

- **Institute of Advanced Studies Kőszeg** is supporting improving skills intelligence in automotive.
- **RISE** develops conferences and meetups. Development of tools for new technologies such as streaming data and distributed GPU clusters for training of AI models.
- **Paluno, University of Duisburg-Essen** develops and maintains the Big Data Value Education Hub (http://bigdataprofessional.eu/), which lists master and phd programmes related to big data value.
- **RSA FG** is analysing and sharing of best practice on brain flows
- **Decsis** has have cooperated with education institutions and employment agencies, improving digital skills of non-technical workers.
- **Engineering** IT & Management School is providing training courses on different ICT related topics, including Big Data
- **Inria** is partner in the classcode project, which develops pedagogical tools and support for teachers of 8-15 years old to help them understand digital technology and master the implication this has on their (digital) lives. Inria promotes curiosity driven research to attract and grow is base of the most talented researchers. At the same time, it closely cooperates with industrial and academic partners through mobility programmes, and/or shared projects to share/transfer knowledge and technology.
- **Fraunhofer-Gesellschaft** is contributing to discussions at a high-level towards a European Qualifications Framework (update of skills, upskilling based on national and European needs).
- **Maxfone** enables not IT professional to social and web big data insights identification
- During 2018 **Robocoast** has (and will) keep "automation coffee break" information and showcase lectures in order to promote automation, robotization and AI cases for the companies in Finland.
- **Siemens** is developing and conducting internal trainings as well as external trainings at Universities and in the context of PPPs (e.g. BDVA). Siemens has contributed and shaped the ERT brochure on digital skills brochure in order to emphasize the high relevance of digital skill building to prepare employers to the digital transformation.
- **GMV** is supporting Digital Skills and Jobs Coalition by collaborating with Higher Education Organizations to increase digital skills of students entering into the market with increased level of employability.

3.2 KPI II.8 Higher establishment availability of big data value creation skills development

This KPI is linked and complements the common KPI I.2 related to the creation of new job profiles and skills.

This KPI is composed by different sub-KPIs:

- Number of training programs established arising from cPPP.
- Number of European training programs involving 3 different disciplines arising from cPPP.

- Number of Master and Phd. students involved in PPP projects.
- Number of dissemination events, seminars, conferences organised in cPPP projects (including number of participants).
- Number of other actions that contribute to this topic and not reported in as part of I.2 or II.8.

3.2.1 Number of training programs established arising from cPPP

43% of the cPPP projects (13 projects) have developed training programs arising from the cPPP in 2018 and have reported **85 training activities during 2018** involving over **9700 participants**.

Some examples:

- ***DataPitch:*** B2B sales - Youtube. How to develop sales leads for startups, Pitch training - Webinar. How to create an investor ready pitch deck for startups seeking funding. Privacy and Policy Toolkit - Workshop. Anonymisation techniques and data policies to be aware of for startups. Getting Press - Webinar. How to attain word of mouth and online media exposure for startups. Setting Organisational values and culture - Webinar. How startups can create company values and an organisational culture has developed training sessions around a tutorial on crowdsourcing and a social computing module.
- ***DataBio:*** Webinars were delivered online, hackathons were delivered both online and face-to-face, while conventional training was delivered face-to-face. Below is the list of the events realized within DataBio framework in the year 2018: Webinar: Prague INSPIRE Hack 2018; oral presentation online; audience industry Webinar: Prague INSPIRE Hack 2018 – Available Tools and Data; oral presentation online; audience industry Webinar: Orleans/OGC INPSIRE Hackathon 2018; oral presentation online; audience industry Webinars on Spatial Data on the Web and INSPIRE; oral presentation online; audience industry and policy makers Webinar on Essential Biodiversity Variables and their metadata; oral presentation online; audience industry Webinar on Unifying data and metadata, linked data in agriculture, and metadata in Open Land Use; oral presentation online; audience industry Training in Genomic prediction and selection; face-to-face; audience scientific community (higher education - research) INSPIRE Hackathon 2018: Open Data & Agricultural Resources; oral presentation online; audience industry, policy makers INSPIRE Hackathon 2018: APIs & Security; oral presentation online; audience industry, policy makers Webinar I: Prague INSPIRE Hackathon; oral presentation online; audience industry, policy makers Webinar II:P rague INSPIRE Hackathon; oral presentation online; audience industry, policy makers Webinar III: Prague INSPIRE Hackathon; oral presentation online; audience industry, policy makers Webinar IV: Prague INSPIRE Hackathon; oral presentation online; audience industry, policy makers Webinar V: Prague INSPIRE Hackathon; oral presentation online; audience industry, policy makers Workshop on GI Data and Software in Prague; oral presentation online; audience scientific community (higher education - research), civil society Prague INSPIRE Hackathon 2019; FtoF oral presentation; audience industry, policy makers Workshop on remote sensing and sensor data; FtoF oral presentation audience industry, policy makers, customers Workshop on remote sensing and sensor data; FtoF oral presentation Grupo de Trabajo de la SECA - Working and commercial meeting with Environment Ministry; FtoF oral presentation.
- ***e-Sides:*** have organised informative workshops in several conferences.
- ***eBusinessGraph:*** OpenCorporates organised three training events with a total number of 130 participants. These are events, which were all directed at Media. They gave an introduction into using OpenCorporates to investigate companies, including 5 minutes on Corporate Events, and highlighted why access to this data matters. All three events face to face workshops: • The Economist Workshop • NICAR - Using OpenCorporates to investigate companies • The Financial Times Workshop

- ***Slipo:*** Applied Artificial Intelligence; Customers, industry TomTom Hackathon; Employees WIGeoGIS Knowledge Day; Customers.
- ***SODA:*** 1. Berry Schoenmakers, Crypto Circus: Zeroknowledge and MPC Explained, NTNU, Trondheim, Norway; 2. Berry Schoenmakers, MPyC assignment, Cryptographic Protocols course, TUE, Eindhoven, Netherlands; 3. Niek Bouwman, Web analytics – confidentiality, guest lecture; 4. Claudio Orlandi, Efficient Protocols for MPC, Crypto Innovation School, Schenzen, China; 5. MPC course, Aarhus University, Aarhus, Denmark; 6. Identity and Privacy, summer course, Alexandra Institute, Aarhus, Denmark
- ***MH-***MD: 1 workshop with hospitals' managers and doctors, 15 people.
- ***BigMedilytics:*** Intro to ML, Company internal; Seminars describing possibilities and training in RTLS; Seminars and training in data anonymization; Seminars and training of GDPR; Seminars and training in analytics; Medication quality circle, attended by physicians; Individual training on software and background of involved radiologists; AI/ML intro for consortium members; Workshop with partners and clinicians to define the clinical journey of the breast cancer patient; Rules and regulations for BIG data Technologies (Channel: workshops, Content: Rules and regulations for BIG data Technologies (ethics, legal, policy), Audience: participants of BIG medilytics); Defining KPIs to measure productivity gains (Channel: workshops, Content: balance score card, KPI's, data management, Audience participants of BIG medilytics); Business models for BIG data technologies (Channel: workshops, Content: Canvas, financial models, Audience participants of the project).
- ***EDI:*** Webinar 0: "Big Data Ecosystems"; Webinar 1: "Big Data Stack and Application Infrastructure integration: building Big Data apps through Lambda Architecture"; Webinar 2: Design; Webinar 3: Tools for Big Data analytics; Webinar 4: Data legislation; Webinar 5: Business Model Canvas; Webinar 6: Scaling-up; Webinar 7: Public Funding focused on SMEs; Webinar 8: Marketing and Communications: CRM Systems.

BDVA members have reported additional **96 training activities** with over **8500 participants** involved (out of the context of the cPPP projects but linked to the BDV SRIA topics). Some examples include:

- **Institute of Advanced Studies Kőszeg**: workshops on case stories of big data analysis.
- **RISE**: Outreach to Swedish large companies, SMEs and public sector through meetups and conferences and workshops.
- **Engineering**: training courses on Big Data available from the Engineering training school, in classroom or remotely via web. (e.g. "Artificial Intelligence and Big Data: develop application models", "Introduction to Big Data", "Big Data Analysis with Hadoop and Spark". "Big Data application development with Scala and Spark", "Big Data with noSQL and Hadoop". "Designing and Building Big Data Applications").
- **UPV**: BSc degree in Data Science, 75 students. MSc degree in Big Data Analytics, 32 students. MSc degree in Artificial Intelligence, 40 students. MSc degree in Grid and Parallel Computing, 20 students. Short summer course in Machine Learning for Big Data Analytics, 25 students. Short summer course in Cloud computing (on-line), +100 students. Short summer course in Deep Learning, 22 students.
- **Cineca**: Introduction to R for data analytics; Tutorial on high-throughput computations: general methods and applications using AiiDA; Introduction to Python programming; Advanced School on Computer Graphics for Cultural Heritage; Introduction to the DAVIDE OpenPower GPU cluster; High Performance Bioinformatics; Programming paradigms for GPU devices; School On Scientific Data Analysis and Visualization; Introduction to Deep Learning and Tensorflow.
- **ITI**: Machine Learning & Big Data short courses (32 hours) and seminars (2 hours).
- **GMV**: Current GMV collaboration with training programs has not arisen from cPPP activities.

- **Huawei Technologies Düsseldorf GmbH**: Lectures, training in digital fields such as BD, ML and AI at Universities, for example: Lectures on distributed trace analytics using AI and Big Data and 3 research projects with universities on the same field. A master course of Big Data Tools Technologies and trends. Support scientific and research IT programmes for doctoral candidates and Master's students, such as the 'Software Campus' initiative which is funded by the BMBF and the 'Huawei Innovation Research Program' (HIRP) as well as numerous collaborations with universities, such as the TUM (Munich) and TUB (Berlin). Huawei has just launched the 'DigitAll Explorer Program' to develop 3,000 local excellent talents in Europe. Organize College Career fairs in Europe.

3.2.2 Number of European training programs involving 3 different disciplines arising from cPPP

These interdisciplinary programmes are meant to contribute to developing the knowledge and skills needed to deal with the complexity of Big Data. 5 projects (17%) have contributed to this KPI and in total **16 developed interdisciplinary programs arising from the cPPP during 2018** with **around 250 participants** engaged in such programs. In particular:

- **DataPitch**: Courses and materials are available in 6 key areas; data science and skills; open data; skills for digital businesses, including marketing and cyber-security; business innovation, the legal aspects of sharing data. Between 15-20 participants per programme.
- **DataBio**: One-week training in genomic analytics in the framework of Erasmus and staff training mobility (refer to European University Association for details): participants were from DataBio, Turkish Eastern Mediterranean Agricultural Research Institute, and Abant Izzet Baysal University, Bolu-Turkey.
- **e-Sides**: Ethics legal constraints and technical architectures for Privacy Preserving technologies.
- **EDI**: EDI training is on technical and business development, ethical aspects have also been dealt within our programme.
- **BigMedilytics**: has also reported interdisciplinary training programs.

Additionally, EW-Shopp has reported that UNIMIB offered a Master Program in Data Science, which includes courses of different disciplines, such as computer science, statistics and economics. The course is not funded by the project, but at least 4 professors of the program are involved in the project, including the program coordinator (Prof. Carlo Batini) and the project coordinator (Matteo Palmonari). The project is referenced several times during different course and students have been given access to the project data.

3.2.3 Number of Master and Phd students involved in PPP projects

With this KPI we intend to measure the way the cPPP projects contribute to skills transfer between industry and academia by involving more Master and Phd students with industrial players in developing industry-driven solutions and deploying experimentation testing scenarios. During 2018 87% of the running projects reported involvement of Master students and 97% the involvement of Phd students involved in project development. Overall **396 equivalent FTEs** of the mentioned profiles (260 Master and 136 Phd).

3.2.4 Number of dissemination events, seminars, conferences organised in cPPP projects (including number of participants)

To mobilise stakeholders the projects have developed intensive communication and dissemination campaigns, with a result of **323 events in total organised by the projects during 2018** (including own events, seminars, conferences**,** etc.) outreaching over 630.000 participants with the objective of raising awareness

about their different activities, to engage new stakeholders, and to keep expectations about future upcoming results.

In relation to the number of people outreached and engaged in dissemination activities developed by the projects during 2018, projects have reported **7,8 Million people** outreached (over 630.000 have been outreached through events, conferences, etc; over 10.000 have been engaged in online activities such as webinars, online training, etc; over 5,8 Million people have been outreached via mass media, online or physical format; and over 1,3 Million people have been outreached via other ways of communication and dissemination).

A full report with all the details is provided in Annex 4.

4 IMPACT OF THE BDV CPPP ON SMES (I.3, II.18)

4.1 KPI I.3: Impact of a cPPP on SMEs (in Euros/Qualitative Analysis)

Through this KPI we attempt to understand the impact of the cPPP activities on SMEs.

The European Commission offered the cPPPs the possibility to address this KPI in 2 different ways:

- Calculating the actual turnover of cPPP participants, and in particular the evolution on the development of the turnover of SMEs should span throughout Horizon 2020, with the year 2014 as the baseline (The link between the activities of a cPPP and this evolution should be explained, in particular if it differs from developments at a macroeconomic level).
- Alternative option of mapping the SME community represented within the cPPP including data on the evolution of the number of SMEs, size distribution, and company age.

The BDV cPPP has worked with both approaches, as we think it is combination of both what can give a more accurate indication of the impact of the BDV cPPP on SMEs.

4.1.1 Methodology

All SMEs participating in cPPP projects active during 2018[14] and all SMEs members of BDVA were called to fill a questionnaire including the following relevant information[15]:

- Name of the company
- Age of the company
- Country of headquarters
- Annual Turnover from 2014 to 2018 (in €)
- Number of employees (annual figure from 2014 to 2018)
- cPPP projects they were involved in (if any)

4.1.2 Map of responses and responses rates

In total 69 SMEs that provided input to the Monitoring report 2018 for this KPI (87% non BDVA members and 13% are BDVA members):

14 Project partners, linked third party of cascade funding

15 Same questionnaire was used to gather information related to private investments (KPI I.1) so those fields are explained under I.1

- 12% are medium-size companies (8),
- 41% are small companies (28) and,
- 48% are micro companies (33)[16].

Average age of the SMEs is **12,1 years**, and in particular:

- 20% of the SMEs (0 to 4 years old)
- 36% of the SMEs (5 to 10 years old)
- 42% of the SMEs (> 10 years old)

Geographical distribution of the SMEs providing input to this KPI is large covering the following European Countries: Austria, Belgium, Check Republic, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Luxembourg, The Netherlands, Norway, Portugal, Romania, Spain, Switzerland and UK.

68 of the 69 respondents were SMEs linked to cPPP projects running in 2018[17]:

- 22 SMEs are part of the Data Pitch Innovation (5) and European Data Incubator (18) Programmes, the Europe-wide accelerator offering start-ups and SMEs funding and support for data-centric business ideas through cascading grants. In total the DataPitch accelerator funded 18 companies during 2018 and EDI funded 29. In total Therefore **47%** of the start-ups in incubators provided input to this KPI (27,8% for DataPitch, and 62% in EDI).
- 47 SMEs are full partners of cPPP projects running in 2018 (projects partners). The response rate is **53,4%** considering 84 SMEs were funded by projects running in 2018.

Summary of Responses rates:

- **Overall response rate** (respondents vs all SMEs in projects and Association): **57%**
- **Response rate by SMEs partners in projects**: **53,4%**
- Response rate in DataPitch start-ups: **27,8%**
- Response rate in EDI start-ups: **47%**
- Response rate in Association members: **23%** (**77%** considering members receiving project funding).

4.1.3 Overall results:

Total turnover reported for SMEs in 2018 is **260,4 M €**[18].

In terms or **turnover evolution** there is an **increase in aggregated data of 60,13% with respect to 2014** (beginning of H2020) and a **17,72 %** increase in the last year[19] (2018 compared with 2017).

Results (by age of the company):

- SMEs (0 to 4 years old):
    - **128,4%** increase in turnover with respect to 2015 (many of the companies reporting information did not have any turnover in 2014)

16 Criteria for classification following EC rules: http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en
17 Project partners, linked third party of cascade fundinhg
18 Aggregated total of the companies
19 In alignment with the macro-economic figures provided in KPI II.3

- **54,84%** increase turnover in the last year (2018 related to 2017)

- SMEs (5 to 10 years old):
  - **284,2%** increase in turnover with respect 2014
  - **24,2%** increase turnover in the last year (2017 related to 2016)
- SMEs (> 10 years old):
  - **45,4%** increase in turnover with respect 2014
  - **16,32%** increase in the last year (2018 related to 2017)

Results (by size of the company):

- Medium-size companies (8):
  - **55,8%** increase in turnover with respect 2014
  - **10%** increase turnover in the last year (2018 related to 2017)
- Small companies (28):
  - **63,62%** increase in turnover with respect 2014.
  - **27,7%** increase turnover in the last year (2018 related to 2017)
- Micro companies (33):
  - **88,2%** increase in turnover with respect 2014.
  - **31,5%** increase turnover in the last year (2018 related to 2017)

Employment

Increase in employment for the SMEs part of the PPP is of 75% with respect to 2014 and a growth of 11,83% in the last year (2018 compared with 2017).

Data incubators

Average age of the companies receiving cascade funding from the Data incubators (DataPitch and EDI) is 4,9 years, 41% of those SMEs younger than 5 years, 50% are in between 5 and 10 years, and only 9% are older than 10 years old. Companies reported an increase in turnover of the 315% with respect to 2014 and 48,8% in relation to 2017, and 118,5% increase in employment in relation to 2014 with a 22,4% increase in the last year.

Assessment

Turnover:

We can see a general increase in turnover in the companies' part of the cPPP with reported numbers of 60% increase in turnover with respect to 2014 and 17,7% in the last year. Young SMEs (5 and 10-year-old) show in average the biggest growth in turnover in relation to 2014 (up to 284%). The youngest companies (<5 years) show in average the largest growth in the last year (54,8%).

It is also important to mention that not all the SMEs involved in BDV cPPP projects are technology companies but Data users or providers, and the overall results and trend clearly indicate an **ongoing growth of turnover along the whole value chain**.

Employment:

The trend is positive in all companies' part of the PPP with an average in Increase in employment for the SMEs part of the PPP is of 75% with respect to 2014 and a growth of 11,83% in the last year (2018 compared with 2017).

Results show the BDV cPPP is benefiting a broad range of SMEs considering geographical distribution, age and size of the company. There is also variety roles in the data value chain[20]. SMEs participating in cPPP projects clearly show a trend of increase in turnover and increase in number of employees.

4.2 KPI II.18: Participation and benefits for SMEs

This KPI is included in the contractual agreement of this PPP. It is partially replaced by the new common KPI I.3 but not completely, so we consider this KPI complementary for that common KPI.

Number of SMEs participating in cPPP projects and Share of participation of SMEs in cPPP projects

Annex 3 provides details of the results of the call for proposals from **2018** with a result of **22 SMEs** participating in projects selected for funding in 2018 **25,3% of the partners are SMEs** that are taking **23,5% of the total budget** of the call).

During 2018 **84** SME companies were directly involved in the projects as partners and additional **47** benefited from cascade funding and data incubation services, adding to a total of **131** SMEs involved in projects during 2018, **69%** of the total number of enterprises involved in the programme.

Estimation of the increase in turnover in SMEs participating in the cPPP projects

Explain and reported under I.3:

- Increase of turnover: **60%** (baseline 2014)
- Last year increase of turnover (baseline 2017): **17,7** %

Estimation of the increase in number of employees for SMEs participating in the cPPP projects

Explain and reported under I.3:

- % annual increase (2018 with respect to 2017): 11,8%
- % increase on baseline (2014): 74,9%

Total reported FTEs:

- 2014: 1 124 FTEs
- 2015: 1 409 FTEs
- 2016: 1 563 FTEs
- 2017: 1 757, 6 FTEs
- 2018: 1 965, 6 FTEs

20 This information has not been specifically gathered through a questionnaire but it is known because of the role of the companies in the projects. In future exercise it could be useful to request this information also.

5 INNOVATIONS EMERGING FROM PROJECTS (I.4, II.7, II.17)

5.1 KPI I.4 Number of significant innovations to market

This KPI concerns all developed items that have a marketable value, including products, processes, instruments, methods, technologies, datasets, etc. It should involve all items directly linked to the cPPP projects as developed foreground, as well as any items beyond the scope of the project that is linked to the foreground.

The BDV cPPP has 2 additional specific KPIs (part of article 7 of the contractual agreement and described in the SRIA) contributing to innovation coming from projects (more details for each KPI can be found below):

- II.7: New economic viable services of high societal value
- II.17: New Systems and technologies

A combination of the input provided in the 3 KPIs (I.4, II.7 and II.17) is needed to have a full overview about all innovation created in the BDV cPPP projects that has an exploitable and marketable value, is a new system/technology beyond state-of-the-art and, as some of them, address also a high societal value.

Target and response rates:

All Big Data cPPP projects active in 2018 (32 projects) were asked to contribute with input to these KPIs. Responses rate: 31 over 32 projects replied (**97%**).

It has to be pointed out that Innovation might not come from the CSA Projects like e-SIDES (involvement of stakeholders to increase citizen's confidence in Big Data Technologies) and BDVe (the Support Action for the whole BDV CPPP Ecosystem); means that 30 projects in 2018 were the target group to create exploitable and marketable results. Therefore the response rate is **100%**.

Questionnaires to collect information for this KPI can be found in Annex 2.

Results:

In total, **106 innovations of exploitable value** are reported as **delivered in 2018** (or delivered earlier and further developed or enhanced during 2018). Considering the innovations reported as delivered in 2017 the cPPP has delivered so far **132 innovations of exploitable value by end of 2018**. Considering plans towards the end of the project (and considering half of these 30 projects will end in 2019) total number of innovations of exploitable value expected in 2019-2020 by this 30 projects is **249**.

From the 106 innovations of exploitable value during 2018, **63%** have **medium impact** and **37%** are considered innovations of **significant impact**.

Considering 4 types of innovation (incremental, architectural, disruptive and Radical), **50%** of the innovations delivered in 2018 are **incremental innovations**, **6% are architectural**, **36% are disruptive innovations**, and 1**% is radical innovation**). 8% are not included in any of these categories.

Innovations to Market	Total	Medium Impact	Significant impact	Incremental innovation	Architectural innovation	Disruptve Innovation	Radical Innovation	Not assessed
Delivered in 2017 and no progress	26	19	7	4	0	0	0	22
Existing as backgrond but enhanced during 2018	1	1	0	1	0	0	0	0
Delivered in 2017 and progress in 2018	55	34	21	20	4	29	1	1
Delivered in 2018	50	32	18	32	2	9	0	7
Expected 2019+	117	49	68	64	20	31	2	0
Total delivered during 2018	106	67	39	53	6	38	1	8
Total delivered by 2018 (2017-2018)	132	86	46	57	6	38	1	30
Total expected 2019+	249	135	114	121	26	69	3	30

Table 6: Innovations for market 2018 classified by Type of innovation

93% of the innovations delivered in 2018 have **economic impact** and **48% have societal impact**[21]. 6% of the innovations reported are not labelled under any of these categories.

Innovations to Market	Total	Economic impact	Societal impact	Other or not assessed
Delivered in 2017 and no progress	26	23	4	3
Existing as backgrond but enhanced during 2018	1	1	0	0
Delivered in 2017 and progress in 2018	55	53	33	2
Delivered in 2018	50	45	15	4
Expected 2019+	117	106	46	6
Total delivered during 2018	106	99	48	6
Total delivered by 2018 (2017-2018)	132	122	52	9
Total expected 2019+	249	228	98	15

Table 7: Economic and Societal input of delivered innovations

41% are technologies (including platforms), 32% are services, 7% products, 8% are methods, 8% are systems, 1% is software, 4% are components or/and modules, and 11% others, including frameworks/architectures, processes, tools and toolkits, spin-offs, datasets, ontologies, patents and knowledge.

[21] Note than many innovations have both economic and societal impact.

Innovations to Market	Total	Technology (icluding platforms)	Service	Product	method	system	Software	Components and modules	Others
Delivered in 2017 and no progress	26	0	7	2	1	3	8	2	5
Existing as backgrond but enhanced during 2018	1	1	0	0	0	0	0	0	0
Delivered in 2017 and progress in 2018	55	14	21	3	4	6	0	0	8
Delivered in 2018	50	28	13	4	5	3	1	4	4
Expected 2019+	117	36	35	32	10	9	1	0	9
Total delivered during 2018	106	43	34	7	9	9	1	4	12
Total delivered by 2018 (2017-2018)	132	43	41	9	10	12	9	6	17
Total expected 2019+	249	79	76	41	20	21	10	6	26

Table 8: Innovations to market

Sectors covered:

Innovations delivered in 2018 (and planned ahead) are providing solutions to many different sectors and application domains, demonstrating the cross-sectorial nature of the Big Data Value cPPP.

- **16% of the innovation delivered in 2018 are fully cross-sectorial** (same value expected for +2019)
- **75%** provide solutions to the **Transport, mobility and logistics sector** (the one with the best coverage in the PPP)
- **20%** of the innovations relate to **Public services and Smart Cities**
- **19%** to **industry and manufacturing**
- **14%** to the **bio-economy**
- **13%** are innovations linked to the **Telco sector**
- **12%** are linked to **Marketing activities**
- **8%** relate to **health and healthcare**
- **8%** to the **ICT market**
- **7% geospatial market**
- **5%** to **commerce** and **3% to others** (including fashion, retail, business services, energy, media, compliance, etc)

Innovations to Market	Total	Cross-sectorial	Transport, moblity, logisitics	Industry, Manufacturing	Bio-economy	Health and healthcare	ICT	Marketing	geospatial	Pubic services and Smart Cities	Telco	Commerce	Others
Delivered in 2017 and no progress	26	4	22	0	0	0	0	0	0	0	0	6	0
Existing as backgrond but enhanced during 2018	1	0	0	0	0	1	0	0	0	0	0	0	0
Delivered in 2017 and progress in 2018	55	9	76	4	0	7	6	11	6	19	1	4	1
Delivered in 2018	50	8	3	16	15	0	3	2	1	2	13	1	2
Expected 2019+	117	16	8	29	2	20	9	1	0	12	3	1	25
Total delivered during 2018	106	17	79	20	15	8	9	13	7	21	14	5	3
Total delivered by 2018 (2017-2018)	132	21	101	20	15	8	9	13	7	21	14	11	3
Total expected 2019+	249	37	109	49	17	28	18	14	7	33	17	12	28

Table 9: Innovations to market – sectors

This distribution will change over time as some projects end and new projects start (e.g the high % of innovations in the Transport, mobility and logistics sector is due to the lighthouse project TT and several projects from Call 2016 such as BigDataOcean, QROWD, and others delivering innovations in that sector).

Assessing the results coming from the cPPP projects it can be determined that in fact the majority of the new systems and technologies can be utilized in different sectors/markets and by that stimulating the use of Big Data technologies in many areas.

Innovations to Market	Total	TRL3	TRL4	TRL5/TRL6	TRL7	TRL8	TRL9
Delivered in 2017 and no progress	26	0	3	7	6	2	1
Existing as backgrond but enhanced during 2018	1	0	1	0	0	0	0
Delivered in 2017 and progress in 2018	55	1	5	12	28	4	1
Delivered in 2018	50	6	5	26	6	4	0
Expected 2019+	117	28	13	34	16	17	0
Total delivered during 2018	106	7	11	38	34	8	1
Total delivered by 2018 (2017-2018)	132	7	14	45	40	10	2
Total expected 2019+	249	35	27	79	56	27	2

Table 10: Innovations to market – Maturity levels or TRLs

In relation to the Maturity levels and TRLs, 7% of the innovations delivered are TRL 3 (experimental proof of concept), 10% are TRL 4 (technology validated in lab), 36% is TRL5 or TRL 6 (technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies), or technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies), 32% is TRL 7 (system prototype demonstration in operational environment), 8% is TRL 8 (system complete and qualified) and 1% TRL 9 (actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies; or in space).

Additionally, the Transforming Transport project (lighthouse project from the call 2016) reported Time-to-market information, with more than 60% of the innovations delivered expected to be in the market by the end of the project or just after it during 2019.

Overall picture of the BDV cPPP Significant innovations to Market

Figure 2 provides the full overview of the innovations delivered by the cPPP during the year 2018 combining level of significance, type of innovation (incremental, disruptive, architectural and radical) and the TRLs. **Figure 3** provides the evolution towards 2019 (including of accumulative results expected by 2019+ by the active projects in 2019).

Although a large amount of innovations are classified as incremental innovation of medium impact, it is remarkable the high % of significant innovations (and expected growth in the upcoming years), the large amount of disruptive innovations (similar magnitude order than incremental), and the high TRLs in some cases close to deployment (and therefore to real impact in the market and society).

Although at a lower level the cPPP is also delivering some architectural and radical innovations.

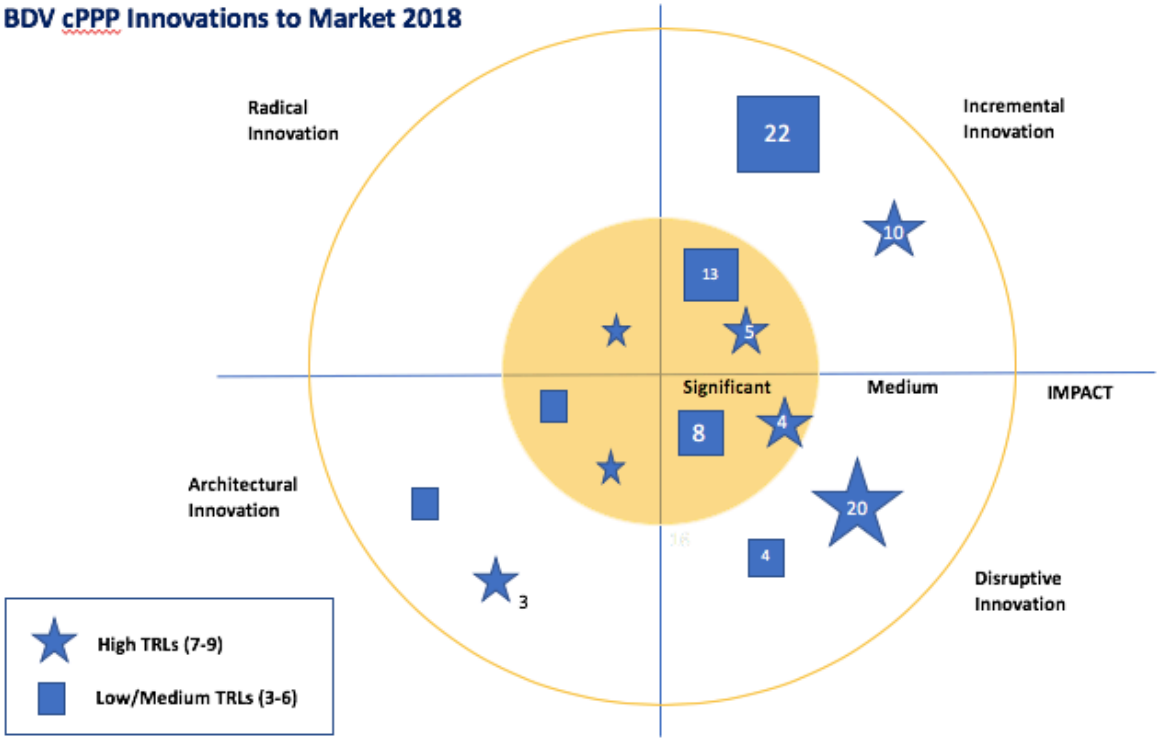


Figure 2: BDV cPPP innovations to market 2018

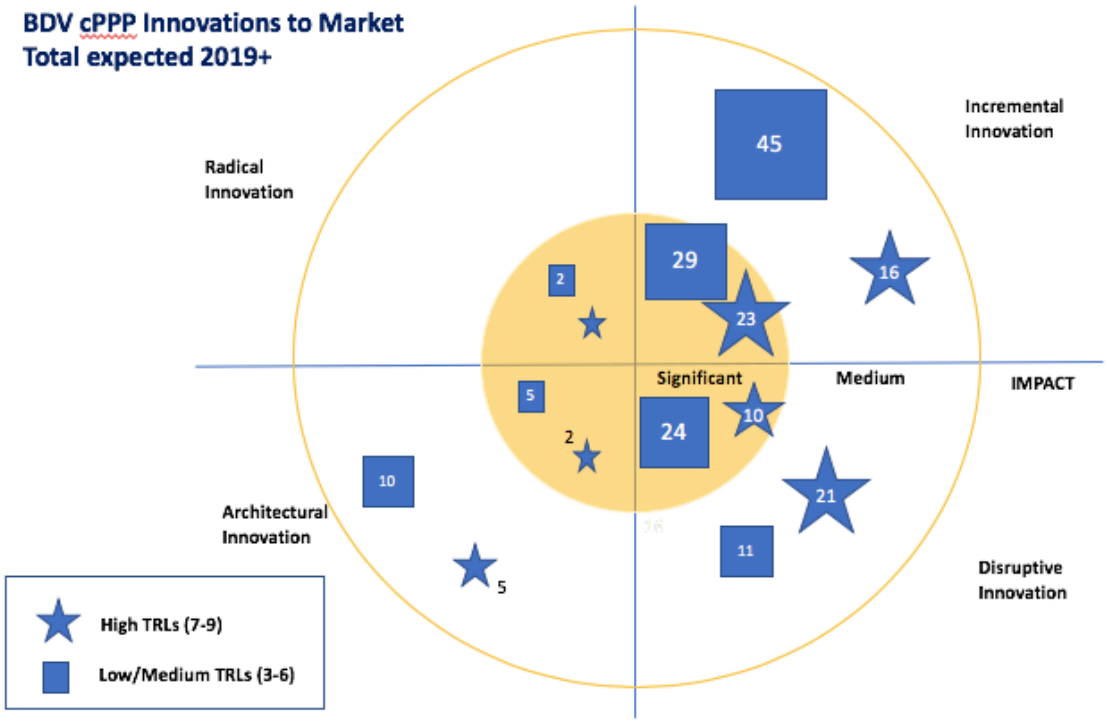


Figure 3: BDV cPPP innovations to be delivered in +2019 by active projects in 2018.

Project	Innovation name (short name if exists)	Type of innovation	Brief description (Including the marketable/exploitable value)	Impact (Significant/Medium)	Impact description	Impact type	TRL (if applies)	Maturity level	Sectors	Type of innovation (details in readme)	Timing
AEGIS	AEGIS Platform	technology, service	The AEGIS platform provides a multi-tenant data management and processing services for big data. The multi-tenancy behaviour allows different users and services to securely and privately access and process their data. The AEGIS platform enables users to share their data with other users on the platform and allow access for specific services. Also, users can use different data processing services that are supported by the platform to process and visualise their data. AEGIS is built on top of Hopsworks and Hops. It provides an integrated support for different data parallel processing services such as Spark, Flink, and MapReduce, as well as a scalable messaging bus with Kafka, and interactive notebooks with Zeppelin and Jupyter. Under the hood, the data is mainly stored in the AEGIS data store; however, the AEGIS data store APIs are kept hidden from users. Instead, the AEGIS platform provides a Project/Dataset service to allow users to upload/download, explore, and do analysis on their data in a secure way without interacting with the AEGIS data store directly. The AEGIS platform includes the Metadata Service responsible for managing the rich metadata associated with a particular dataset within the AEGIS platform and posing the foundation for it processing. It is based on the AEGIS ontology and vocabulary. Another important service in AEGIS is the AEGIS Harvester. It enables the data import to the AEGIS platform offering the transformation, harmonisation and annotation functionalities required within the context of the platform as well as the rich metadata generation for the imported data. Furthermore, the AEGIS platform includes several tools for users with limited technical background, but potentially useful for all, as they simplify and accelerate the work with data. Among them are: • Query Builder providing the capability to interactively define and execute queries on data available in the AEGIS system, • Visualiser enabling the advanced visualisation capabilities of the AEGIS platform and • Algorithm Execution Container accelerating analysis execution by simplifying the steps that data analysts perform, through eliminating the need to author code directly into the notebook. Additionally, the platform includes the Brokerage Engine acting as a trusted way to record and keep a log of transactions over the AEGIS platforms, which mostly have to do with the sharing of the different data assets. Taking the above issues into account, the AEGIS platform will be exploited as a service and as a software: • A live instance of the AEGIS platform will be hosted by KTH. It is going to be the AEGIS public demonstration instance and a public repository for collected data from the PSPS domain. This instance will provide only some restricted computational resources for demonstration purposes. • The productive instances of the platform software will be deployed in public or in private clouds by the organisations willing to use AEGIS. Each	Medium	The AEGIS platform is a holistic solution as a public/private cloud-based platform, the only European platform with small & big data integration (most of the competitors analysed above are either from the United States or outside Europe), allowing modular design, facilitating customization and therefore infrastructure/resources scarcity. Furthermore, AEGIS provide its users i) with a distribution of a powerful big data stack that can be deployed in a cluster hosted on-premise or over cloud-based infrastructure and ii) ability to put on public display datasets by operating in an environment of distributed AEGIS clusters and allowing partners of the ecosystems to monetize on them. Considering the opportunities, the current market offering is characterised by solutions that are mainly general purposes or science/research oriented, not specifically considering Public Safety and Personal Security (PSPS) and social value; consequently, AEGIS can still become the key venue and the standard for open, social innovation in big data for the PSPS domain, also facilitating a new market on PSPS and nurturing new ventures based on PSPS data. Consequently, AEGIS has the opportunity to enable a balancing between social value creation and economic value capture through value streams from industries which interested to PSPS datasets.	Economic and Societal	TRL7	Prototype	Public Safety and Personal security, Finance/insurance, eHealth, Smart Home, Automotive	Incremental	Expected in +2019

			of these private instances will be a private AEGIS cluster able to share data with the central demonstration node or copying data from it.								
BigDataOcean	BigDataOcean Platform	PRODUCT	BigDataOcean Platform - A state-of-the-art big data analytic engine for managing cross-sectoral maritime data of differing velocity, variety and volume in a trusted way	Significant	The BigDataOcean platform enables users to perfom analytics on blue data thus significantly minizing the cost and effort for the acquisition and analysis of diverse sources	Economic	7	PROTOTYPE	MARITIME	Incremental	Delivered mainly 2017 but progress in 2018
DataPitch - Predrepair	Spare part and mechanic activity recommendation engine	Product	The automotive after sales market is 44bn € in Germany alone. Our spare part recommendation engine allows car repair shops to deliver value with less human labor (increasing efficiency), allowing for higher value activities.	Significant	Is radically changing the way car repair shops work and will position themselves in the competition. Eventually, their entire business model might be disrupted	Economic		Supported by a prototype and MVP	Automotive	Architectural	Delivered in 2018

DataPitch - LexaTexer	Railway availability service desk	product	Machine learning / AI driven service desk to optimize Railway service operations	Significant	currently employees read large numbers of event messages daily, they need to match events manually (thousands of events) to identify root causes and to dispatch the matching repair team. Information gets lost or is not accounted for leading to a mismatch in dispatched teams and task, downtime, delays. We solve that problem by learning from past event-ticket matches, thus increasing availability, ream-placement.	Economic and Societal		prototype	Railway, Logistics		Already reported in 2017 and no new progress in 2018
DataPitch - Zevit	Data capture	Incremental		Significant		Economic		spin-off created	OPEX-heavy industrial service	Incremental	Delivered mainly 2017 but progress in 2018
DataPitch - Zevit	Data quality	Disruptive		Significant		Economic		spin-off created	OPEX-heavy industrial service	Disruptive	Delivered mainly 2017 but progress in 2018
DataPitch - Zevit	Data analysis	Disruptive		Significant		Economic		spin-off created	OPEX-heavy industrial service	Disruptive	Delivered mainly 2017 but progress in 2018
DataPitch - Zevit	Data share	Architectural		Significant		Societal		spin-off created	OPEX-heavy industrial service	Architectural	Expected in +2019
DataPitch - Ubiwhere	Multimodal data importation procedure	Technology	Microservice responsible for data collection, processing, harmonising and storage (Extract-Transform-Load operations)	Significant	Scalable new means of importing data from multiple sources allows Ubiwhere to quickly set up new instances of its solutions for mobility, smart cities and industry 4.0	Economic	8	Demonstrator		Incremental	Delivered in 2018
DataPitch - Ubiwhere	Customisable and responsive web dashboard	Technology	This component provides centralised dashboards and enables custom queries to the data available in the platform (concretely the geographic database)	Significant	This module allows for a quick setup of dashboards which can be customised per role, user or instance, easing the process of data validation and analysis for multiple sources of information	Economic and Societal	8	Demonstrator		Incremental	Delivered in 2018

DataPitch - Ubiwhere	Data analysis and harmonisation	Technology	Machine Learning component, where analytics and machine learning jobs run to extract the necessary knowledge from the incoming harmonised data;	Significant	The data analysis framework allows Ubiwhere to prototype, test and validate KPIs and allowing its solution to compare off-the-shelf with other sets of information as soon as they are integrated into the platform (such as weather conditions, nearby points of interest, cultural events, accidents and other urban interventions).	Economic	8	Demonstrator		Incremental	Delivered in 2018
DataBio	ZETOR Tractors, a.s.	Technology, service	Precise farming functions introduced to tractor	Medium	This innovation has medium impact as similar innovations are already on the market	Economic and Societal	TRL6	Supported by prototype and demonstrator	Agriculture	Incremental	Delivered in 2018
DataBio	Wuudis	Wuudis - A Forestry one-stop platform	Wuudis is a technically innovative, open, standardized and significantly cost effective solution for the entire forest ecosystem than current alternatives. Wuudis integrates data from several sources (forest management plans, satellites, drones and open data) by standardizing it, to digitalize the whole process from planting to harvesting and thereafter end-use.	Significant	Changes value creation in forestry	Economic and Societal	TRL6	Supported by prototype and demonstrator	Forestry	Disruptive	Delivered in 2018
DataBio	Metsak .fi eService	Technology, service	Metsään.fi is an eService provided by the governmental body, the Finnish Forest Centre (METSAK), to make forest resource information available for citizens free of charge. Metsään.fi as an eService serves forest owners and forestry service providers. Metsään.fi is a portal through which people who own forest property in Finland can conduct business related to their forests from their own desktops. The portal connects owners with related third parties, including providers of forestry services. This makes it easy to manage forestry work and to be in touch with forestry professionals.	Significant	Changes value creation in the forestry sector	Economic and Societal	TRL6	Supported by prototype and demonstrator	Forestry	Disruptive	Delivered in 2018

DataBio	Agri-insurance tools and services for the agriculture insurance market	Technology, service	Insurance in the agri-food sector deals with the increasing demand for agricultural insurance products and is expected to play a vital role in the forthcoming years as a tool for risk management. However, due to its multi-parametric nature, agricultural insurance is considered a special category in the insurance product portfolio. This innovation provides a solution for collection of enough valuable data for damage assessment and assessment of the complex biological processes that are incorporated in the crops growth stages. This set of tools and services aims to eliminate the need for on-the-spot checks for damage assessment and promote more rapid pay-outs. Through fusing heterogeneous data (EO, field) the assessment of climate-related damages can be achieved even at field level. The convergence of the aforementioned technologies in a single dedicated framework is expected to deal effectively with insurance market demands which require a smooth transition from traditional insurance policies (expensive, require human experts for damage assessment) to more flexible index based insurance approaches. For the insurance company, and from a business perspective, this methodology could highlight the benefits that index-based approaches could offer that might lead to new insurance products. New smart farming products, combined with agri-insurance aspects, are another business perspective that arises for NEUROPUBLIC, that leads the innovation activities. CSEM is involved in the experimentation for further enhancing the tools with computer vision methods.	Significant	Provides the bases for new insurance products for the agricultural industry	Economic and Societal	TRL3	Supported by prototype and demonstrator	Agriculture	Disruptive	Delivered in 2018
DataBio	AIS (Alien Invasive Species) - Preventions system for smart farm management	Technology, service	Real time Risk maps for preventing AIS introduction using of climatc and anthropogenic data to evaluate the AIS host capacity.	Medium	Real time Risk maps for preventing AIS introduction using of climatc and anthropogenic data to evaluate the AIS host capacity.	Societal		Supported by prototype and demonstrator	Forestry	Incremental	Delivered in 2018
DataBio	Arable Farming data integrator for Smart FArming	Technology, service	Using FOODIE data model and analysis of satellite data we will offer unique services. Using FOODIE standardized protocol service will be easy transferable into new countries	Significant	The use of data standards allows for introductzion to international market	Economic and Societal		Supported by prototype and demonstrator	Agriculture	Incremental	Delivered in 2018

DataBio	Data-driven genomic prediction and and selection of superior cultivars	Technology, service	Big data-driven genomic algorithms are developed and deployed to usefully and meaningfully substitute for conventional phenotypic and molecular breeding, in the process of crop improvement and development superior cultivars. A variety of information is fed to the genomic models including but not limited to whole-genome molecular data, whole-plant phenotypic and biochemical data, and big environmental data (IoT, etc..). The implementation of this genomic predictive and selection analytics proved promising in prior peer-reviewed works under international specialized settings wherein CREA is active. Advantages include but are not limited to: (1) significantly expediting the process of cultivar development, reducing the standard cycle down to half as long, (2) increased breeding velocity, accuracy, and simplicity, which translates into superior cultivars hitting market earlier, increased genetic gain (and hence, productivity) by unit time and cost, and higher revenues and satisfaction, respectively, to producers and consumers.	Significant	New process that inproves eficenency conidserably	Economic and Societal		Supported by prototype and demonstrator	Agriculture	Disruptive	Delivered in 2018
DataBio	Early detecction of irrigation Inhomogeneities for the calculation of water needs in agriculture	Technology, service	Develop a computer tool for the calculation of water needs that uses Big Data techniques to analyze in real time data from different sources	Medium	Obiously innovative and easily appreciated advantages to customer	Economic and Societal		Supported by prototype and demonstrator	Agriculture	Incremental	Delivered in 2018
DataBio	EO-based data analysis and maschine learning methodologies for CAP support	Technology, service	A set of EO-based services have been designed appropriately to support specific needs of the CAP value chain stakeholders including the farmer decision making actions during the submission of aid application and and more specifically leading to an improved “greening” compliance. The innovative part in this approach is the methodology as it deals with CAP demands for agricultural crop type identification, systematic observation, tracking and assessment of eligibility conditions over a period of time. Thus, it is fully aligned with the main concepts of the new EC's agricultural monitoring system that aims to the simplification and the reduction of the burden of controls through the adoption of a technology-driven methodology. Machine learning methods and geospatial data analytics with data coming from EO platforms are examined towards this direction by the involved partners. NEUROPUBLIC is leading the innovation activities, conducting industrial research and providing services. CSEM is involved in the experimentation activities for further enhancing the tools with computer vision methods. GAIA serves as the intermediate with the farmers. This is achieved through its associated network of Farmer Service Centers (FSCs) that provides collection and advisory services to the Greek farmers concerning the submission of the aid application for direct payments, including eligibility pre-check mechanisms for error reduction and proof provision. GAIA subsidy services are mainly oriented to aging small-sized farmers, which own 80% of the holdings in Greece. Over the last two annual periods, GAIA provided collection services and cross compliance checks to 76% of the holdings.	Significant	Significantly improves distribution of subsidaries in agriculture	Economic and Societal		Supported by prototype and demonstrator	Agriculture	Disruptive	Delivered in 2018

DataBio	Forest Health Management System for the early detection and monitoring of plagues/diseases affecting forests	Technology, service	Methodology and service using big data sources (satellite + aerial + UAV + field data) for the early detection and monitoring of plagues/diseases affecting forests	Significant	Significantly improved service for forest health management	Economic and Societal		Supported by prototype and demonstrator	Agriculture	Disruptive	Delivered in 2018
DataBio	GeoRocket Vizualisation & Aggregation for large amounts of spatial vector data	Technology	Efficient, web-based visualization and aggregation methods for large amounts of spatial vector data	Significant	Significant improved product	Economic		Supported by prototype and demonstrator	Generic applicability in many industries, where large amounts of geospatial data is required (agriculture, forestry, ...)	Incremental	Delivered in 2018
DataBio	Geospatial big data management tools for agriculture use cases	Technology	The innovation action exploits recent trends in big data management for optimally addressing several use case needs. It is tailored to handle large datasets consisting of vector and multi-temporal data assigned to each agricultural parcel (object-based methodology). Moreover, it allows the development of applications on top of it, supporting real business case needs for agricultural monitoring, fast querying, spatial aggregations and analytics. It fully exploits the capabilities offered by Fraunhofer's GeoRocket elasticsearch and integrates them with other (NEUROPUBLIC's GAIA cloud) big data solutions (and as such to explore the opportunity to improve them) that consist the backbone of specific agri-food business scenarios (smart farming, insurance pilots in Greece).	Significant	Significant improved processes and efficiency	Economic		Supported by prototype and demonstrator	Agriculture		Delivered in 2018
DataBio	Integration of machinery data with Telemetry	Technology	The focus is integrate telemetry data and data from machinery operations	Medium		Economic		Supported by prototype and demonstrator	Agriculture		Delivered in 2018
DataBio	Radiometric correction for improvement of the	Technology	Improvement to the ortophotographies provided by the National Service of Aerial Ortophotography - National Geografic Institute	Medium		Economic		Supported by prototype and demonstrator	Agriculture		Delivered in 2018

	ortophotographies										
DataBio	Smart farming pest/disease crop management services	Technology and service	Smart farming involves the continuous monitoring and prediction of diseases and pests breakouts and alerting in case these are found. Farmers can be facilitated in the effective application of fertilizers by having an overview of the risk indicator of infection or infestation of their crop by a specific disease or pest respectively, based on the environmental conditions that apply in the specific area. Hence, they can schedule the application of sprayings at the most appropriate time, maximising their effectiveness and sometimes reducing the number of required applications. This approach introduces strong innovation potential as it addresses the monitoring and prediction of specific disease and pest infection in olives, peaches, and grapes in Greece and the early/real-time triggering of alerts, based on crop-specific scientific models, once these are detected. Advice extrapolation in parcels within the same microclimatic zones can be seen as another, more business-oriented, innovation as the cost of infrastructure installation decreases and economies of spectrum arise. NEUROPUBLIC is leading this activity, as it serves both as infrastructure and advisory service provider, with all technological aspects getting more matured/expanded within the project and adapted to region-specific needs. IBM introduces a temporal reasoning aspect and allows to enhance the advisory services by triggering even earlier alerts. Finally, GAIA serves as the business partner of the collaboration, and can be seen as the single face to the customer. GAIA handles all communications with the end-users and its unique composition (representing 71 agricultural cooperatives and 152k farmers) effectively engages all the necessary stakeholders, securing direct access to the customers, who are also shareholders.	Significant	Has potential to significantly improve the efficiency in agriculture	Economic and Societal		Supported by prototype and demonstrator	Agriculture	Incremental	Delivered in 2018
DataBio	Yield timing for agricultural productivity	Technology and service	Defining the optimum timing for yielding sugar beet seeds based upon satellite imagery	Medium		Economic		Supported by prototype and demonstrator	Agriculture	Incremental	Delivered in 2018
euBusiness Graph	CED	New product	Corporate Events Data Access	Significant	New information not previously available in the way reported	Economic and Societal		prototype/ demonstrator	Business information	Incremental	Expected in +2019
euBusiness Graph	TDS	New product	Tenders Discovery Service	Significant	New service not previously available	Economic and Societal		prototype/ demonstrator	Business information	Incremental	Expected in +2019

euBusiness Graph	Atoka+	Extension of existing product	Atoka SaaS B2B Lead Generation Service	Significant	New business data and new services that use business data	Economic and Societal		prototype/ demonstrator	Business information	Incremental	Expected in +2019
euBusiness Graph	CRM-S	New product	CRM Service	Significant	New service not previously available	Economic and Societal		prototype/ demonstrator	Business information	Incremental	Expected in +2019
euBusiness Graph	DJP	New product	Data Journalism Product	Significant	New tool for data journalists	Economic and Societal		prototype/ demonstrator	Media	Incremental	Expected in +2019
euBusiness Graph	BR-S	New service	Registries API Service	Significant	APIs access to public data	Economic and Societal		prototype/ demonstrator	Business information	Incremental	Expected in +2019
euBusiness Graph	ER	New service	Event Registry	Significant	Events system	Economic and Societal		prototype/ demonstrator	General purpose	Incremental	Delivered mainly 2017 but progress in 2018
euBusiness Graph	WF	New service / technology	Service for semantic multilingual annotation	Significant	Annotation system	Economic and Societal		prototype/ demonstrator	General purpose	Incremental	Delivered mainly 2017 but progress in 2018
euBusiness Graph	GBA	New service / technology	Graph Based Analytics	Significant	Analytics services on top of the business graph data	Economic and Societal		prototype/ demonstrator	General purpose	Incremental	Expected in +2019
euBusiness Graph	Relation Tracker	New service / technology	Relation Tracker	Significant	Analytics services on top of the business graph data from news and events	Economic and Societal		prototype/ demonstrator	General purpose	Incremental	Delivered in 2018
euBusiness Graph	DG	New service / technology	Grafterizer 2.0 / DataGraft	Significant	Data-as-a-Service tool for knowledge graph generation	Economic and Societal		prototype/ demonstrator	General purpose	Incremental	Delivered mainly 2017 but progress in 2018
euBusiness Graph	ABSTAT	New service / technology	knowledge graph summaries with ABstraction and STATistics	Significant	knowledge graph summaries	Economic and Societal		prototype/ demonstrator	General purpose	Incremental	Delivered mainly 2017 but progress in 2018

euBusiness Graph	ASIA	New service / technology	Assisted Semantic Interpretation and Annotation of tables	Significant	table annotation	Economic and Societal		prototype/ demonstrator	General purpose	Incremental	Delivered mainly 2017 but progress in 2018
euBusiness Graph	EBG	Ontology	euBusinessGraph ontology	Significant	ontology for company data	Economic and Societal		published ontology	General purpose	Architectural	Delivered mainly 2017 but progress in 2018
euBusiness Graph	EBG-marketplace	New service	euBusinessGraph data marketplace	Significant	data maketplace	Economic and Societal		prototype/ demonstrator	Business information	Radical	Expected in +2019
EW_Shopp	ABSTAT: knowledge graph summaries with ABstraction and STATistics	technology	ABSTAT is a tool to create semantic profiles of knowledge graphs. Semantic profiles describe structural properties of knowledge graphs and the vocabulary used therein. Semantic profiles are applied to support different tasks: data understanding and exploration, data quality, vocabulary recommendation and analytics. ABSTAT is currently developed to profile RDF-based knowledge graphs but could be extended to profile knowledge graphs represented other data models (e.g., graph databases). ABSTAT is a free and open source software. Services on top of ABSTAT (deploy, customization, etc.) may be exploited.	Significant	ABSTAT is a first open source tool that allows users to generate knowledge graph profiles from a user interface. It is also the only tool that computes statistics about cardinality of schema patterns.	Economic and Societal	6	The software and its arechitecture are relatively stable. Immprovements concern distribution, scalability and services on top of it	ICT	Incremental	Delivered mainly 2017 but progress in 2018
EW_Shopp	ASIA with Data Reconciliation and Extension Services	technology	ASIA is a semantic table annotation tool to generate knowledge graphs and enrich tabular data with third-party data sources. It interoperates with semantic reconciliation services, to reconcile values against shared vocabularies and systems of identifiers, e.g., Wikipedia, GeoNames, GFK Products, and data extension services, to fetch data from, e.g., GeoNames, ECMWF, GFK Products. On top of Grafterizer, ASIA supports end-to-end data cleansing, transformation and semantic enrichment at scale. ASIA is a free and open source software. Customization, e.g., deploy on big data architectures, require expertise that is marketed as additional services.	Significant	ASIA is the first tool that embed semantic table annotation functionalities in an end-to-end tool for data manipulation, semantic enrichment and generation of RDF data using a selected vocabulary/ontology. It is also the first tool to natively support the above mentioned functionalities for large volumes of tabular data.	Economic and Societal	4	Main object of work for UNIMIB, aims at TRL 6 by December 2019	ICT	Disruptive	Delivered mainly 2017 but progress in 2018

EW_Shopp	Big Data workflow tool based on container orchestration with DataGraft and Grafterizer	technology	The Big Data workflow tool provides clustered infrastructure for deployment, management and scaling of data workflows. The solution is based on Docker containers, a container orchestration system and a distributed file system. It will provide a high-level interface to handle and monitor data workflows, a processing engine for carrying out and scaling the workflow steps and an extensible library of composable steps, including an interface to ArangoDB, to be used as part of the data workflows.	Medium	The solution allows for easy deployment of big data processing workflows on existing hardware infrastructures. It has the potential to make big data processing more accessible to companies that lack technical knowledge about the topic.	Economic	3	A prototype of the solution is available for internal use within the consortium. Parts of the templates for container orchestration are published on the project GitHub page.	ICT	Architectural	Delivered in 2018
EW_Shopp	Weather data API	technology	The developed weather data API is a wrapper around the official European Centre for Medium-Range Weather Forecasts (ECMWF) data archive access API. It simplifies access and retrieval of weather data for non-meteorological purposes by simplifying the code for ease of use and limiting its scope to just the weather values likely to be useful in business applications (e.g. temperature or relative humidity) as opposed to those of greater interest to meteorology (e.g. albedo or temperature of lake surface). The API is only usable with a valid ECMWF API key which can only be obtained at ECMWF.	Medium	The solution eases the use of weather data in data analytics applications	Economic and Societal	8	a freely-accessible, open-source API with documentation used in business case pilots	ICT, Marketing, Commerce, CRM, IoT, BI	Architectural	Delivered mainly 2017 but progress in 2018
EW_Shopp	Product Data Management Tools	technology	We developed two tools that use machine learning technology to automate different aspects of product data management. The first tool is the categorizer, which classifies products into existing (possibly hierarchical) taxonomies or categories based on their (non-structured) textual description. The second tool is the de-duplicator which identifies different textual descriptions of the same product and supports their reconciliation. Both of these are currently highly labor-intensive tasks in online stores or comparison shopping engines.	Medium	The solution provides classification and feature extraction for product catalogs	Economic	7	prototypes being tested in a business environment	ICT, eCommerce	Disruptive	Delivered mainly 2017 but progress in 2018
EW_Shopp	Event and Weather Analytics Toolset	technology	A set of tools and services built on top of the QMiner data analytics platform for building analytic models using events and weather data. These data features are joined with other data as external factors and their influence is modelled. An example of its application is modelling the effects of weather and major local events on sales in online stores. The tools support data ingestion, transformation, as well as model building and deployment (as a REST service).	Significant	The tool functionality can offer a significant competitive advantage to businesses in the e-commerce sector. By being specialized for its task it enables them to generate insights from their data faster than with general-purpose analytics platforms and the built-in prediction server allows rapid deployment to testing and production environments.	Economic and Societal	7	a freely-accessible, open-source API with documentation used in business case pilots	ICT, Marketing, Commerce, CRM, IoT, BI	Disruptive	Delivered mainly 2017 but progress in 2018

EW_Shopp	Keyword Clustering Tool	technology	A tool for clustering of keywords and keyword phrases based on their semantics. The tool uses pre-computed word embeddings built using deep neural networks. For multi-word phrases the embeddings are aggregated using an appoach based on Smooth Inverse Frequency (SIF) weighting. By unsing a metric in the embedding space the tool is able to find semantically related keywords for a given seed keyword. For example by using a category name (e.g. "Football") as the seed, the tool is able to find a set of keywords most likely belonging to this category (e.g. "football game", "referee", "FC Barcelona").	Medium	The solution helps dimensionality reduction in large-scale analysis of keyword-based marketing campaigns	Economic	5	prototype tested in a lab environment	ICT, Marketing	Incremental	Delivered in 2018
EW_Shopp	Event Data Management for Data Analytics	methods	An ontology and a methodology to support event data exhange for event-based analytics.	Medium	The solution eases exchange of event data to support event-based analytics using and extending the widely adopted Schema.org reference ontology and JSON-LD	Economic and Societal	4	Initial testing in the EW-Shopp ecosystem. Maturity will be reached at end of 2019	ICT, Marketing, Commerce, CRM, IoT, BI	Incremental	Delivered in 2018
EW_Shopp	Campaign BOOSTER	service	Digital marketing campaign impact strongly depends on the user behavior, which is directly affected by the weather conditions. In this scenario, JOT is developing a new service integrating weather forecast and marketing performance data enabling the implementation of 4 main functionalities: SCHEDULER, ALARM and IMPACT PREDICTOR (weather and event versions). They are based on keyword behaviour modelling (exploiting the campaign performance historical data base). This enables account managers to boost the impact of the campaign based on new insights generated by the correlation with the weather forecast and events and adapt the campaign bidding strategy accordingly, with the final goal of maximizing the number of impressions (as key impact factor).	Significant	The innovation created in EWSHOPP project will impact directly in the performance of the marketng campaigns, as it will allow the implementation of data driven campaign management including external factors not considered by the account managers so far. Due to the company business model, the better the performance indicators are, the more revenue is generated and more gross margin.	Economic and Societal	4	Initial pilot service implemented showing the adequance of the EWSHOPP approach and methodology	Marketing	Incremental	Delivered mainly 2017 but progress in 2018

EW_Shopp	EW-Shopp DataFlow Methodology	method / dataflow	This is a cross-domain, multi-actor methodology which enables to have a consistent workflow for data management and in particular multi-step dataflows. It concentrates on the logical steps (process) and 'actors' of a whole data management and sharing'food chain' and provides a general process for: 1) Data Integration; 2) Data Enrichment; 3) Data Analytics; 4) Data Visualisation. Additionally documentation can be enforced throughout all steps. The methodology is designed to allow/combine automatic, semi-automatic and human sub-flows and sub-processes as well as to support all phases of a dataflow starting from design up to implementation and running. Potential markets and exploitation scenarios are those where either an existing dataflow needs to be 'rationalised' (e.g. legacy dataflows) or ones where a new dataflow needs to be set-up. As it is the methodology is mostly logical and funcional, i.e. not tied to any particular tool or technology. To this end this approach was chosen so as to keep it open to differnt uses. It could of course be formalised / adapted to- several existing dataflow tools.	Medium	Impact can vary depending on the data-sets at stake and actors (human or technical) involved. In some cases the impact could be significant in terms of rationalising existing complex flows.	None	n.a.	Applied as a methodology for the evaluation of all EW-Shopp business cases. Also implemented in BC3 as a supporting methodology	All possible sectors dealing with complex dataflows	Incremental	Delivered mainly 2017 but progress in 2018
EW_Shopp	PERFORMANCE INSIGTHS: E-commerce Insights tool for shoppers, sellers and vendors	service	Performance Insights tool helps shoppers, sellers and vendors to use short term product search insights to support their smart decisions. The tool is build on top of years of user's uniquely integrated behavior data of searching products, weather data and different category events data. Shoppers will be able to use relevant purchase information and real time predictions to support their purchase decisions. Sellers will manage better pricing, category and campaign management and vendors get brand and SKU level purchase information in the shortest possible time to react and improve pricing, sales and campaign efficiency.	Medium	This solution will provide insights on the performance of products, brands and retailers based on the analysis of searches in a comparison shopping platform	Economic	3	Initial pilot Performance Insights tool implemented and shown to selected partners to gather first impressions and feedbacks. Weather aware web site widget for shoppers implemented and tested in real live environment	Commerce and E-Commerce, Retailers, Vendors, Shoppers	Disruptive	Expected in +2019

EW_Shopp	Contact Center Workforce Optimization Manager	service	Digital marketing campaign impact strongly depends on the user behavior, which is directly affected by the weather conditions. In this scenario, Browsetel is developing a new service integrating weather data with existing Contact Center services KPI data, to get an innovative traffic prediction model. This will enable account managers to predict the adequate time to launch new campaigns and to better predict the needed Contact Center workforce.	Medium	The solution will influence on decrising the probability of the Contact Center resources aggravation or saturation and incrising the probability of the call to be succesfully served to the customer. Consequently, we expect improvements at "Call Response / Number of Agents / Customer Satisfaction" rate.	Economic	3	Weather related prototype upgraded with Custom Event related-one, which has been build within a newely designed Custom Event Handling architecture and Custom Event API.	Contact Centers, CRM and CIM systems offering services for different industries e.g.: Telecommunications, Finance (Banking & Insurance), Commerce, Services & Tourism, Publishing, Energetics, Marketing & Research, ...	Disruptive	Delivered mainly 2017 but progress in 2018
EW_Shopp	Weather and Event-aware BI Sales Strategy Advisor	service	Due to lack of systemized evaluation on how weather and events influence daily business, the service will enable to determine which type(s) of marketing activity(es) give(s) the most optimal results in given conditions; which type of event has more influence on demand/sales variables; how to manage activites on different customer touchpoints/channels and possibly how a price discount affects sales (on category level). Therefore thorugh service we aim for optimization of Salesforce and resources within Marketing/Campaign and possibly also Category/Supply Chain Management.	Medium	The solution will advise decision makers on how to exploit weather and events to boost product sales	Economic	3	Initial pilot prototype developed. Full-fledged solution expected at the end of 2019	Marketing, Salesforce management (possibly also CM and SCM); Brand manufacturers/distributors /retailers in the segment of CE&Home Appliances.	Disruptive	Expected in +2019

EW_Shopp	Event and Weather Scout Dashboard	service	Our dashboard helps businesses allocate their advertising budget based on quantitative historical analysis of customer flows vs seasonality/weather trends and internal events. We analyzed the number of people who visited a particular store in past periods of time and defined the impact of a particular event on people's behavior measuring the impact of marketing campaign/events that happened in the store. The results are visually presented on the dashboard in the form of easy understandable figures that give a clue to our clients about the effectiveness of past campaigns and weather temperature vs number of visitor relations.	Medium	The solution will provide Measurence Scout customers with the capability of understanding and evaluating the impact of events and weather on the visits in their stores	Economic	4	Initial pilot prototype developed. Full-fledged solution expected at the end of 2019	Marketing, IoT, BI	Incremental	Delivered mainly 2017 but progress in 2018
FashionBrain	Flair	AI Natural Language Procesing Framework	A very simple framework for state-of-the-art Natural Language Processing (NLP)	Medium				beta, open source famework	all		Already reported in 2017 and no new progress in 2018
FashionBrain	FashionMNIST	dataset	dataset of fashion images annotated with class labels of 10 classes	Medium				open dataset	all		Already reported in 2017 and no new progress in 2018
FashionBrain	FEIDEGGER	dataset	dataset of images of dresses together with textual desciptions in German	Medium				open dataset	all		Already reported in 2017 and no new progress in 2018
FashionBrain	MonetDB	technology	The open-source MonetDB software suite wih FashionBrain extensions	Medium		Economic	TRL 6 - 9	Partly mature software product, partly prototype	IT	Incremental	Delivered mainly 2017 but progress in 2018
FashionBrain	FB Taxonomy: New comprehensive knowledge base for FashionBrain fashion items	product	The new taxonomy aggregates and extends the existing open source taxonomies such as Google, ebay and Amazon taxonomies. A key advantage of the FashionBrain taxonomy, a key advantage compared to existing taxonomies, is the redundancy reduction which is often caused by gender dependency	Significant	This taxonomy can be used as a backbone knowledge base for any fashion-related project to represent all fashion entities. Moreover, our taxomomy is easy to maintain, enrich and update.	Economic and Societal		Supported by a prototype: https://fashionbrain-project.eu/fashion-taxonomy/	Fashion		Delivered mainly 2017 but progress in 2018
FashionBrain	GenderBias DataSet	dataset	Dataset used in "Investigating User Bias in Image Search: A Cross-Regional Study". It contains 2,811 query-description comparisons for 281 different users.	Medium				open dataset	Fashion		Delivered in 2018

FashionBrain	Crowdsourcing logging interface	technology	Append-only, ephemeral in-memory logging REST interface.	Medium				supported by a prototype: https://fast-logging.herokuapp.com/			Delivered in 2018
FashionBrain	System TASTY	technology	High performing Named Entity Linkage System linking text to text data based on deep learning methods.	Medium				supported by a demonstrator: https://demo.datexis.com/tasty/	all		Delivered in 2018
FashionBrain	System IDEL	technology	In-database Entity Linkage system bases on MonetDB (and is currently further extended towards Exasol RDBMS). The system executes machine learning in a database with Tensor Flow	Medium				supported by a demonstrator: https://fashionbrain-project.eu/beuth-tasty-feat-idel-demonstration/	all		Delivered in 2018
SLIPO	SLIPO Workbench	System	Data integration software for Big geospatial data assets; supports and automates the entire workflow of data integration using Linked Data technologies	Medium	SLIPO Workbench enables world-scale geospatial integration at a fraction of the effort and cost, empowering geospatial value chain stakeholders with low-cost and high value data	Economic	7	Pre-commercial operation	Geospatial value chain (e.g., logistics, geomarketing, mobility, transport)	Architectural	Delivered mainly 2017 but progress in 2018
SLIPO	TripleGeo	System	Transformation software for geospatial data assets to RDF	Medium	TripleGeo enables the transformation of all types of geospatial assets to RDF, hence enabling the management, querying, and analysis via Linked Data technologies	Economic	8	Production version available since 2017	Geospatial value chain (e.g., logistics, geomarketing, mobility, transport)	Incremental	Delivered mainly 2017 but progress in 2018

SLIPO	LIMES	System	Interlinking software for linked data	Medium	LIMES is the leading interlinking software for Linked Data, supporting world-scale linking of data assets	Economic	8	Production version available since 2017	Geospatial value chain (e.g., logistics, geomarketing, mobility, transport)	Incremental	Delivered mainly 2017 but progress in 2018
SLIPO	FAGI	System	Data fusion software for linked data	Medium	FAGI is a generic-purpose framework for geospatial data fusion, enabling stakeholders to increase the richness and quality of their geospatial assets	Economic	8	Production version available since 2017	Geospatial value chain (e.g., logistics, geomarketing, mobility, transport)	Incremental	Delivered mainly 2017 but progress in 2018
SLIPO	DEER	System	Enrichment software for linked data	Medium	DEER enables the enrichment of Linked Data assets with other proprietary and open knowledge bases	Economic	7	Demonstrator	Geospatial value chain (e.g., logistics, geomarketing, mobility, transport)	Incremental	Delivered mainly 2017 but progress in 2018
SLIPO	SANSA	System	Scalable ML-based analytics software	Medium	SANSA is the first scalable analytics software combining ML and Linked Data analytics for the provision of value-added services	Economic	5	Demonstrator	Geospatial value chain (e.g., logistics, geomarketing, mobility, transport)	Disruptive	Delivered mainly 2017 but progress in 2018
SODA	Private data analytics for digital health platform	service	This innovation targets inclusion of MPC technology, i.e. secure computation and privacy preserving data analytics, in the healthtech cloud and data platforms which drive many of the service innovations.	Medium	enable evidence-based and analytics-based healthcare solutions based on combining data that previously could not be shared due to its privacy or business sensitivity	Economic and Societal	5	Concept demonstrator	healthcare	Disruptive	Expected in +2019
SODA	FRESCO – framework for efficient secure computation	service	FRESCO is a framework which enables developers to write applications that utilizes the benefits of secure computation without the developer having to be an expert in cryptography. FRESCO is the basis of many of its founder's and its spin-off companies' research prototypes and commercial solutions	Medium	enable data-driven business which data sensitivity previously prevented	Economic and Societal	7	Software platform succesfully used for number of business and spin-offs	any, finance, healthcare, other	Disruptive	Delivered mainly 2017 but progress in 2018

SPECIAL	Self-learning and identity-aware personal data catalog for enterprise privacy management and GDPR compliance	Technology	Tracking how personal data is being used across an enterprise is only possible if the controller knows what data it possesses, what truly constitutes personal data, and who such data belongs to. Unfortunately, the GDPR has rendered traditional rule-based and identity agnostic PII discovery solutions legally insufficient. TenForce intends to deliver an AI-driven Privacy Management platform which would fill the market gap and provide a data subject-centric view of the company data landscape.	Significant	The platform will fill a market gap	Economic and Societal		Demonstrator	Multiple	Disruptive	Expected in +2019
SPECIAL	On-Chain Enforcement of GDPR Compliance	Technology	By implementing the GDPR compliance algorithm in smart contracts we can take the first steps towards guaranteeing the compliant management of personally sensitive data while laying the technical foundations for a market in such.	Significant	Providing GDPR compliant management of sensitive data for industry	Economic and Societal		Demonstrator	Banking	Disruptive	Expected in +2019
SPECIAL	Consent based management for data driven products and services	Technology	We propose an innovative approach to consent management. By giving more control and transparency, we aim to reduce the cognitive overload on the data subject, while at the same time creating trust and enabling controllers to develop data-driven products respecting data subjects' decisions.	Significant	Enables the development of data-driven products that respect the data subjects' decisions	Economic and Societal		Demonstrator	Multiple	Disruptive	Expected in +2019
TT	HORUS DASHBOARD-AUSOL/NORTE LITORAL	Software	Customized dashboard for AUSOL and NORTE LITORAL based on HORUS software. It has been developed for helping and assisting road operators into operations and maintenance works in a road infrastructure.	Medium		Economic	TRL7	Prototype	Transport (Highways)	Incremental	Delivered in 2018
TT	MAID - Animal intrusion detection system	Technology	Passive Infrared detector has turned out as a reliable technology to detect persons and animals within short ranges. Scare-out systems based on ultrasound generators can create effective, invisible and non-physical fences only audile by medium to big animals like cows, wild pigs and dogs. Both systems work together thanks to the deep learning models that decide the better reaction to each species of animal	Medium		Economic	TRL7		Transport (Highways)	Incremental/Disruptive	Delivered in 2018
TT	Big Data - Descriptive Mobility Patterns	Knowledge	The results of the project can be exploited for the benefit of Cintra and Indra. The conditions of exploitation have not yet been determined but it is sure that the conclusions and the most important, the knowledge learnt along the project could be exploited for our benefit	Medium		Economic	TRL7		Transport (Highways)		Already reported in 2017 and no new progress in 2018

TT	S2N - O-D Matrixes and user loyalty (Bluetooth / Wi-Fi)	Knowledge	Traffic real time data is being gathered and loaded into the corporate IoT Platform to predict traffic flows and segment users Capture public information that vehicles and drivers make available themselves: Bluetooth and WIFI signals Compliance with the GDPR The information we get by scanning the RF space at the BT and WIFI frequencies, gives us an anonymous heat map of user's preferences, which is crucial when talking about user experience.	Medium		Economic	TRL9				Already reported in 2017 and no new progress in 2018
TT	Connected Cars Platform: Value added services on car data in the cloud.	Service	Descriptive and predictive analytics based on data coming from connected cars sensors.	Medium		Economic	TRL5	Prototype	Transport (Connected Cars)	Incremental	Already reported in 2017 and no new progress in 2018
TT	Improved Routing Service for Trucks	Component	Enhanced component to provide better and more detailed planning possibilities for truck routing and processes of Logistic Service Providers	Medium		Economic	TRL8		Transport (Connected Trucks)	Incremental	Already reported in 2017 and no new progress in 2018
TT	Predictive & automated diagnostic capabilities in the mainline rail industry	Software	Predicting and ultimately enabling the prevention of failure of assets within the mainline rail industry by consuming large amounts of data from many different sources	Significant	This innovation provides strong contribution to the operational efficiency as this helps plan the maintenance schedules better and safety as it decreases the time spent on tracks by the maintenance workers.	Economic	TRL4		Transport, (Railways)		Already reported in 2017 and no new progress in 2018
TT	Descriptive and predictive models for the infrastructure maintenance of railway high speed lines	Software models	These models will try to identify the relation between all the aspects of the railway exploitation in order to be able to improve the maintenance activities of the railway infrastructure focus on High Speed Lines. There are three main objectives at this pilot, which try to optimize the maintenance of these infrastructure elements, tracks, switch and crossing and slopes.	Significant	The algorithms used to obtain predictions related to tracks condition have achieved an accuracy of at least 85%, which means that most of failures related to tracks condi-tion can be predicted and, consequently, anticipate to them.	Economic	TRL4		Transport, (Railways)		Already reported in 2017 and no new progress in 2018

TT	Port and Terminal Productivity Cockpit	Process	The Port & Terminal productivity cockpit is a web-base tool for supporting a common set of shared and specific metrics to evaluate the logistic process of a por terminal. This logistic process could be depicted as a hub process where the clients are inputs/outputs from different transportation models: vessel, train, and truck. One of the main performance inidicators for a terminal is to provide the shortest possible time to leave or pick up containers at the lowest possible cost. This time-cost ratio is an overall indicator to measure performance that is transversal for the domain. The most widely used Key Performance Indicator (KPI) is the Truck/Train/Vessel Turnaround Time (T/VTT): the time a specific means of transport spends in the terminal to fulfil an order. At this point turnaround time is known in isolation, i.e. in a specific terminal, and calculated after the order is processed. Our solution integrates Big Data from several stakeholders involved into the containers management in orde rto provide better insights and metrics about the overall efficiency at both port and terminal levels. The defined KPIs provide benchmarking capabilities (e.g. related to costs and performance) that may indicate different levels of competitiveness. A key feature of theTerminal productivity cockpit goal is to achieve an improvement in KPIs as the Truck Turnaround Time by predicting its trend using analytical models. These models have been deveoloped using machine learning techniques and the whole set of data available. Therefore these models provide insights about current operations, thus helping stakeholders to understand how to improve the related KPIs.	Medium		Economic	TRL7	Prototype	Maritime Ports and Terminals	Incremental	Delivered in 2018
TT	Crane and Spreader Sensorization Infrastructure	Product	Sensorization of the Crane's Spreader to support the collection and transmission of signals in order to provide the necessary inputs for task such as operative monitoring or supporting maintenance tasks	Medium		Economic			Maritime Ports and Terminals	Incremental	Already reported in 2017 and no new progress in 2018
TT	Terminal Productivity Cockpit	System	The terminal productivity cockpit will exploit advanced data processing and predictive analytics to facilitate proactive decision-making and process adaptation. The productivity cockpit will leverage cutting-edge predictive business process monitoring solutions, i.e., real-time predictive big data analytics for terminal processes. Predictions are made for train delay based on real-time machine data.	Medium		Economic	TRL6	Prototype	Ports and Terminals	Incremental	Already reported in 2017 and no new progress in 2018

TT	Operation Management Predictive Optimization Module	Software module	This module will exploit predictive analytics with Passenger Flow data obtained in real time from airport and airline systems in order to facilitate proactive decision making in real time whenever there is any disruption over the initial plan	Significant	This innovation has facilitated proactive decision making in real time in the following areas: • To provide a more refined prediction of delays in departure flights ac-cording to late passengers. • Reduce the number of passenger missing connections. • To improve the efficiency of passenger processing stations. • To predict possible timesaving in the turnaround process.	Economic	TRL8		Transport, Airport management		Already reported in 2017 and no new progress in 2018
TT	Descriptive passenger behavior system	Software system	These models will identify the different type of passengers and their behaviors, to predict when they arrive to the airport, how they move across the airport terminal in order to dimension airport resources accordingly to the expected demand and optimize airport and airline procedures.	Significant	The following objectives have been significatly achieved: Track and predict boarding process Taxi in and out accuracy improvement Feed enhanced ETA (- 20 ') into Airport A-CDM platform based on historical operation data analysis	Economic	N/A		Transport, Airport management		Already reported in 2017 and no new progress in 2018
TT	Web page with Traffic related tweets for specific region and analysis of traffic cameras	Service	The asset is a web page, composed of twitter feeds related to traffic in a city or area.	Medium		Economic and Societal	TRL6		Transport, Urban Mobility		Already reported in 2017 and no new progress in 2018
TT	Parking Reservation APP	Software	The application which is built on top of the parking management system, contains information on parking places. Drivers and logistic operators can reserve parking spaces through the app. Parking wardens can verify through the parking management system the validity of the reservation.	Medium		Economic and Societal	TRL6		Transport, Urban Mobility		Already reported in 2017 and no new progress in 2018
TT	Urban Traffic Fluency Model	software model	The existing fluency model is remarkably improved by the use of floating car data (FCD) from the city buses.	Medium		Economic	N/A		Transport, Urban Mobility		Already reported in 2017 and no new progress in 2018

TT	Customized traffic incident alert	Service	Traffic incidents are automatically alerted to end users using Twitter as delivery channel	Medium		Economic and Societal	N/A		Transport, Urban Mobility		Already reported in 2017 and no new progress in 2018
TT	Traffic Model for freight transport in Valladolid City	Software model	A traffic model for particular areas in the city where freight transport has more impact to understand and gain insights about the patterns and attitudes of delivery areas in city centres	Significant	The following results have been achieved: - A reduction of 30% in time used by freight vehicles in the city centre, as a con-sequence of 30% reduction in average parking time. - 50% reduction in NOx emissions, which it is related to distance travelled.	Economic	TRL7		Transport, Urban Mobility		Already reported in 2017 and no new progress in 2018
TT	Dashboard for micro-simulation insights in Valladolid City Centre	Software model	Dashboard with micro-simulation traffic model results to show valuable insights from Data Analytics tasks done within TT project	Medium		Economic	TRL7		Transport, Urban Mobility		Already reported in 2017 and no new progress in 2018
TT	Integrated Platform	System	The integrated platform of WP10 in which all the algorithms and datasets are successfully integrated and a ready to use system is designed	Medium		Economic	TRL7		Dynamic Supply Networks, e-Commerce		Already reported in 2017 and no new progress in 2018
TT	Descriptive Analytics Service	Service	A descriptive analytics tool that will offer to the potential users: a) interaction with data through the selection of various criteria/dimensions, and b) Detailed analysis of data behavior in 3PL partners through a set of appropriate visuals, and deduction of patterns and trends.	Medium		Economic	TRL7		Dynamic Supply Networks, e-Commerce		Already reported in 2017 and no new progress in 2018
TT	Forecasting Analytics Service	Service	A forecasting tool that will allow the potential user to periodically predict the behavior of different groups of data depending on the selection of various criteria/dimensions, while presenting its forecasts through simple and convenient graphs	Significant	The following KPIs were greatly improved within the pilot scenarios: Time savings per delivery Reduction of fuel consumption and GHG emissions	Economic	TRL7		Dynamic Supply Networks, e-Commerce		Already reported in 2017 and no new progress in 2018
TT	Data Mining Approach	Method	An artifact, which is a clustering-based approach that accomplishes behavioral segmentation and characterization of shoppers' orders, by examining the product categories included in each customer delivery	Medium		Economic	TRL5		Dynamic Supply Networks, e-Commerce		Already reported in 2017 and no

											new progress in 2018
TT	Route Optimization Service	Service	Inventory routing in an omnichannel environment	Medium		Economic	TRL5		Dynamic Supply Networks, e-Commerce		Already reported in 2017 and no new progress in 2018
TT	Customer Reviews Analysis Service	Service	Extract online consumer insights with respect to logistics and delivery processes	Medium		Economic	TRL5		Dynamic Supply Networks, e-Commerce		Already reported in 2017 and no new progress in 2018
QROWD	QROWD Platform	Platform		Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Touristic Network services	Service	Provide travel time and delays to reach several touristic destinations (e.g. ski resorts), so tourists can choose the best option(s).Ensures better tourist customer experience; Allows City tourism authorities to suggest touristic hotspots to visit based on current estimated travel time;Reduces Traffic congestion.	Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	On-Street parking	Service		Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Historical Analysis Reporter	Service	Provides traffic reports to Mobility managers, to better understand specificities of traffic conditions in a concerned place, e.g. due to bad weather conditions, or man-made events. Improves road safety; Reduces traffic congestion; Improves drivers´ mobility experience.	Significant	Reduces traffic congestion	Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Road Event Reporter	Service		Significant	Reduces traffic congestion	Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	TomTom City	Service		Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Municipality Dashboard	Service	Dashboard for City staff showing the city´s mobility data and the results of analytics. Facilitates Decision making processes	Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018

QROWD	Citizen Dashboard	Service	Dashboard for citizens. It will be a subset of the Municipality Dashboard plus some extra services for citizens around mobility. Better-informed citizens	Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	iLog	Service (Data collection as a service)		Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Modal Split	Service	Allows City Management to run travel surveys using citizen's mobile phones instead of cumbersome Telephone interviewing or paper surveys. . Allows shifting from expensive, limited, years old static data to up-to -date cost-efficient data driven decision making, simulations, what-if analysis and policy formulation. Its systematic and timely usage generates disruptive positive changes in traffic management	Significant	disruptive improvement in traffic management, congestion reduction and ultimately reducing traffic-related pollution levels	Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Qrowd-DB	Analytics back-end	Back-end component that stores sensor streaming data (big data) and linked data. The i-Log (mobile app) requires it to collect data e.g. for the Modal Split calculation	Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Virtual City Explorer	Tool	Provides City managament a better way to generate maps of points or items of interest in a city that can be found in their Google Street View virtual representation.It solves the problem of completing mobility infrastructure records at a fraction of the cost of sending Municipality workers to do surveying/inventories/counting.	Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	QROWDSmith	Crowdsourcing platform	Allows city staff to define and run crowdsourcing tasks and citizen engagement challenges with gamification. Dramatic improvement in engaging citizens and data contributors to solve city issues	Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Data Acquisition Framework	Framework	Supports city staff in the acquisition of data and associated methodology. Improves data management	Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Crowd Feedback-aware link prediction	Tool		Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Analytics with Crowd Feedback	Tool		Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Analytics on Data Streams	Tool	Allows city staff to perform analytics on non-static and static data that comes in as a continuous data stream.Improve the performance of the analytics required to deliver the QROWD mobility services	Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018

QROWD	Spatio-temporal Analytics	Tool	Allows city staff to do analytics based on spatial and temporal relations like 'inside', 'near', 'after', 'before' and so on.Improve the performance of the analytics required to deliver the QROWD mobility services	Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
QROWD	Integrated Processing of Static and Stream Data	Method		Medium		Economic and Societal	7	Protoype/Demonstrator	mobility, transport, smart cities, logistics	Disruptive	Delivered mainly 2017 but progress in 2018
MHMD	Blockchain-based data exchange architecture and relevant smart contracts	Technology	Blockchain based infrastructure for health data acces and exchange. MHMD is leveraging on Hyperledger Fabric, which ensures high scalability. Additionally. all data remain in the original repositories. MHMD blockchain deals only with data acees and transactions recording. The system is equipped with diverse access and consent management interfaces (Web- and App-based). Blockchain usage in healthcare has seen many new initiatives starting between 2017 and 2018. MHMD is well positioned to validate the key features of this technology in the partecipating clinical centre. The exploitable value of this tool resides in its capability of enabling secure data exchange and data lifecycle tracking in a secure and efficient manner.	Medium	If successful, proper services for secure data exchange among clinical institutions might help untapping the potential of health data currently stored in isolated data siloes. This would enable the creation of new health data markeplaces.	Economic	6	Prototype available	Healthcare	Disruptive	Delivered mainly 2017 but progress in 2018
MHMD	AI-based system and method for secure data processing	Patent	Leveraging on Siemens Healthcare's DeepReasoner, MHMD developed innovative tools for performing advanced analytics and Big Data technologies over secure/encrypted data.	Medium	This innovation can create new opportunities for providing valuable clinical insight without the need of exposing the underlying data.	Economic	6	Prototype available	Healthcare	Incremental	Expected in +2019
MHMD	Secure Multi Party Computation protocol	Service	The secure multi-party computation protocol - developed on top of an open source software - can be used for enable cooperation among different data controllers for analysing wealth of data collectively generated, without revealing those data to all the members of the network (using secret sharing techniques). Such a solution can be exploitable in particualr in research contexts, creating a consortium/cooperative of data controllers, allowing queries on data across institutions, creating new value for data users.	Medium	This innovation can greatly facilitate the possibility of exploiting large set of data coming from different institutions, provinding the technical tools for gathering and analysing data while keeping totally secure the data. This would represent an incentive for data controllers to cooperate, for creating additional value (more than their individual datasets can provide), selling relevant data analytics to interested third parties.	Economic	6	Prototype available	Healthcare	Disruptive	Delivered mainly 2017 but progress in 2018
MHMD	Anonymisation/pseudonymisation tool	Product	The AMNESIA tools is a simple system that allows data controller to apply a variety of anonymisation/pseudonymisation techniques on data, prior sharing with third parties.	Medium	Facilitate data flow in a secure manner would improve the European data market, making easier for data controllers to contribute data in compliance with existing regulations.	Economic	7	Prototype available	Healthcare	Incremental	Delivered mainly 2017 but progress in 2018

MHMD	Synthetic data creation methodology	Method	This methodology allows to create perfectly secure and anonymous data starting from existing real databases, mantaining the granularity of "personal" data while not leaking any sensitive information, as the syntethic data is not linkable to any existing individual. If deployed as as service, this tool has enormous market potential, as it would allow to create very high quality datasets, perfectly cleared in terms of GDPR (because formally anonymous), and useful for research and testing.	Significant	Data owners could activate interesting revenue streams by selling high-quality synthetic data to research and industry. The impact is significant because - if successful in demonstrating the usability of synthetic data in a variety of research and development context - this system would open up to a new era of data abundance, by removing technical and regulatory challenges to data sharing.	Economic	4	Prototype available	Healthcare	Incremental	Delivered mainly 2017 but progress in 2018
MHMD	Semantic Data Mapping and Metadata catalogue	Service	MHMD "Data Harmonisation" layer integrates and normalizes all heterogeneus sources. For doing so, Data Modelling activities are performed for capturing and integrating all biomedical data following a dynamic Subjective-Objective-Assessment-Plan (SOAP) model of an Electronic Medical Record supporting vertical integration and temporal evolution. Well-established biomedical taxonomies and ontologies such as LOINC, SNOMED CT, ICD-10-CM, CPT, MESH, etc. areincorporated either directly or as semantic annotations. The resulting Comprehensive, dynamic and intelligent metadata catalogue allows external stakeholders to view and explore the data available in the network, thus serving as a direct window and user dashboard for businesses, other hospitals and researchers interested in appraising the value of the MHMD system	Medium	The most immediate impact of this innovation is to be found in its capability of allowing data controllers to "showcase" their available data without revealing any sensitive information, but only sharing metadata to describe what is available in a given dataset. This would facilitate third parties in browsing available data in a given network of data controllers, query the catalogue for specific categories of data, and submit data request. Potentially, this innovation would enhance the value of data currently "hidden" within isolated siloes.	Economic	6	Prototype available	Healthcare	Incremental	Delivered mainly 2017 but progress in 2018
MHMD	Dynamic consent app	Service	The application allows the connection/synchronisation with different data souces (e.g. other apps collecting data, wearable devices) and allows individuals to define his/her consent preferences (who can access data, for how long, for what purpose, and so forth) also allowing to withdraw consent at any time.	Significant	The app can address the already flourishing market of app for individual data control (as hu-manity or Solid) but can also be offered to hospitals as an innovative tool for administering and managing consent, while also keeping an always active communication route with patients. The same applies to clinical trial sponsors, that can use the app to facilitate and streamline consent and re-consent process during the trial. The impact is deemed to be significant because it can actually enable a completely new way of controlling and managing personal data and authorise their access and sharing among stakeholders, putting individual data owners at the centre.	Economic and Societal	7	Prototype available	Healthcare	Radical	Delivered mainly 2017 but progress in 2018

MHMD	Homomorphic Encyption scheme	Method	This innovative method uses a new approach to homomorphic encryption to make it more efficient and lightweight, at the same time allowing a very large number of operations over encrypted data. Some security issues need still to be addressed to make the solution usable in a real world environment, but if successful this tool may find application in a variety of use cases.	Significant	The impact is considered significant because the tool would immediately enhance the value of data by allowing third parties to make complex analytical operations without revealing anything of data itself, thus greatly facilitating data exploitation by removing relevant regulatory barriers.	Economic	4	Prototype available	Healthcare	Incremental	Delivered mainly 2017 but progress in 2018
BigDataGrapes	BigDataGrapes Platform as a Service	Service	BigDataGrapes will produce a wide range of tools and components for Big Data management and processing, making them available as a complete holistic solution readily deployable in different infrastructure providers (AWS, Google Cloud Platform, Microsoft Azure, Heroku, etc.) or in self-hosted environments. The advancements in resource usage and processing requirements foreseen in the project will extend application mobility, reduce maintenance and processing costs and allow the relatively effortless extension and upgrade of individual components	Medium		Economic	5	Concept demonstrator	Agriculture and food	Incremental	Expected in +2019
BigDataGrapes	BigDataGrapes Data Marketplace	Service	BigDataGrapes will open new directions for the management and operation of a marketplace via its results on the ingestion and integration research fields. It is expected that the novel approaches for handling and enriching heterogeneous content will have an immediate, direct impact in the respective processes. The Linked Data principles are generally well suited for managing diverse content and cultivating meaningful in-context knowledge from it. However, it tends to require high level of expertise in Semantic Web technologies and advanced programming skills. BigDataGrapes will cover the urgent need of toolkits and frameworks for describing and integrating legacy data as interoperable, harmonized and semantically rich Linked datasets, constituting Linked Data publishing accessible to non-experts and making the process transparent and intuitive to the data owner.	Significant	A Data marketplace for sharing and accessing large and heterogeneous grapevine-related data assets from both corporate and public organisations will have a significant impact for the competitive advantage of companies that serve with IT solutions these sectors	Economic	5	Concept demonstrator	Agriculture and food	Disruptive	Expected in +2019
BigDataStack	DATA_SKIP	technology	Accelerates SQL queries over Cloud Object Store	Significant	Enables IBM to beat competition	None		As of Feb 2019, incorporated at beta level in the IBM Cloud SQL service	Cloud Analytics	Disruptive	Expected in +2019
BigDataStack	SEAMLESS_ANALYTICAL_FRAMEWORK	technology	Executes SQL queries on datasets fragmented across multiple datastores	Significant	Federates query execution across different datastores, exploiting the benefits of their unique characteristics without compromising the performance of one for the benefit of the other				BigData Analytics	Architectural	Expected in +2019

BigDataStack	INFORMATION_DRIVEN_NETWORKING	technology	Enforces network policies based on big data requirements	Medium	Enables SiLo to beat competition	None		Not yet	Network Softwarization Technologies	Architectural	Expected in +2019
BigDataMedilytics	Hospital workflow optimization	spin-off	Spin-off providing solution using real-time location information of tagged entities in a hospital (e.g. patients, assets, staff). Real-time location information is used to optimize workflows within a hospital	Significant	Can help improve the workflow efficiency of hospitals by up to 20%. That has a significant impact on operational costs, patient and staff satisfaction	Economic and Societal	7	prototype	Healthcare	Disruptive	Expected in +2019
BigDataMedilytics	Risk stratification for population health	methods	Methods to identify high risk patients in primary care	Medium	Will help reduce the no. of patients who end up unnecessarily in secondary care which is much more expensive	Economic and Societal	6	prototype	Healthcare	Incremental	Expected in +2019
BigDataMedilytics	Analytics for Prostate cancer	product	Solution to integrate diverse medical datasets for decision making for care of prostate cancer patients	Significant	Will help in better decision making when performing surgeries on prostate cancer patients. Can have an a significant impact on long term care of patients, e.g. avoid the need to spend EUR50k on adult diapers over a 10-year period due to better surgery	Economic and Societal	7	prototype	Healthcare	Disruptive	Expected in +2019
BigDataMedilytics	Multiparty Computation (MPC)	technology		Significant		Economic		prototype		Radical	Expected in +2019
BigDataMedilytics	Predictor for HF comorbidities	method		Significant		Economic		prototype		Architectural	Expected in +2019
BigDataMedilytics	PRO Heart Failure Revalidation	process		Significant		Societal				Incremental	Expected in +2019
BigDataMedilytics	Adhere Heart Failure Revalidation	process		Significant		Societal				Incremental	Expected in +2019
BigDataMedilytics	Predicting inpatient treatments of patient with known heart	Method	Using all the available data about for insured members with known heart failure with complex analytical procedures (machine learning) to predict the probability of inpatient treatment for each individual. When high risk patients can be identified, actions might be possible to prevent inpatient treatments	Medium		Economic and Societal		Medium	Healthcare	Incremental	Expected in +2019

	failure with health insurance data										
BigDataMedilytics	Forta Score	Algorithm	The Forta score is based on an algorithm that can be computed based on routine clinical data and informs health care providers of potential over- and underuse of medications in persons aged 65 or older	medium	Signals improvement actions to health care providers, requires clinical judgement to assess specific actions in individual patients	economic	7	Prototype in deployment	Health care	incremental	Expected in +2019
BigDataMedilytics	SEARCH@MUW	service	on site deployed search system in a radiological routine environment. This enables the validation, testing of the search software as part of clinical routine	Medium	Initial deployment of the existing search engine, will be extended throughout project	Economic and Societal		demonstrator installed	healthcare	Architectural	Expected in +2019
BigDataMedilytics	GDM system (integrated solution)	System	Overall architecture design used for the project. It can be exported for any other medical solution, and to any other hospital	Significant	The hospital was previously using manual measurements. The time savings for the medical team, along with the reduce in human errors result in a significant impact	Economic and Societal		Prototype	Medicine	Disruptive	Expected in +2019
BigDataMedilytics	GDM mobile App	Product	Mobile App used by the patients to collect information. Can be exported to other hospitals, slightly modified for different use cases	Medium	Makes easier for the users to collect the data	Societal	5/6	Prototype	Medicine	Architectural	Expected in +2019
BigDataMedilytics	GDM medical portal	Product	Online dashboard where the medical team can monitor the patients. The design can be ported with minor modifications	Medium	Eases the medical team the monitoring of the patients	Economic	3	Prototype	Medicine	Architectural	Expected in +2019
BigDataMedilytics	GDM analytics	Service	Analytics for the process. It's a model based on real data that can be exported to other hospitals	Medium	Gives more personalized diagnosis, and helps medical team to adapt	Economic	3	Prototype	Medicine	Architectural	Expected in +2019
BigDataMedilytics	Biomedical Framework	technology		Medium		Economic		Supported by a prototype	Healthcare and Lifesciences	Incremental	This is background technology that was enhanced during 2018 and in 2019, we hope to further develop it for Radiomincs

BodyPass	Hyperledger 3D data exchange network	System	The Hyperledger 3D data exchange network is a blockchain-enabled framework to access to data from several providers from different sectors (tested in the health and consumer goods sectors). The framweork allows to connect with scattered data sources such as databases (accessed using PRESTO) or APIs. Hyperledger allows to make off-chain calls to retrieve data objetcs, while managing the data access and auditing without storing any data from the providers.	Significant	This approach allows to break data silos among different organizations from the same or different sectos, allowing data exchange in a secure and traceable manner	Economic	7	Currently prototype	Health / Consumer goods	Incremental	Expected in +2019
BodyPass	Anonymization of face information in full body 3D scans	methods	This methods will enable to replace the actual face of a 3D full body scan by an anonymous one. Deep learing technology will be used to detect the face traits of the person in the raw 3D pointcloud or mesh. Differnt algorithms such as smoothing or data-driven alternatives will be used to anonymise or crate the anonymous face content.	Medium	All the individual 3d data sets will follow this process to make possible the anonymous extraction of individual or aggregated data from BodyPass data pools.	Societal	6	prototype	Consumer goods /Health	Incremental	Expected in +2019
BodyPass	3D human body templates enabling 3D data aggregation from different sources	methods	We will explore and prototype diffrerent strategies using digital human body templates, non-rigid pose alignment and rigid alignment to enable the surface-to-surface compatibility between body scans coming from diferent data silos while preserving the know-how and specific data parametreisations used by each silo.	Significant	This approach allows to break data silos among different organizations from the same or different sectos, allowing data exchange in a secure and traceable manner	Economic and Societal	6	prototype	Health / Consumer goods	Incremental	Expected in +2019
Boost4.0	Blockchain for Volvo Supply Chain	Technology	We developed a blockchain solution based on Hyperledger Fabric and Hyperledger Composer for the track and trace of cab parts of Volvo trucks. Our solution includes the model, the blockchain backend, and a mobile user interface that allows each certified and authorized party in the network to initiate actions, finish their transactions, and track progress. It also includes the storing and tracking of the parts locations during the delivery process. The data stored on the blockchain is leveraged for the real-time analysis of this data to forecast potential delays and avoid unplanned stops on the assembly line.	Medium		Economic	5	first prototype	manufacturing, automotive	incremental	Expected in +2019

Boost4.0	6P-BD Method: Big Data Transformation 6Ps Method for Manufacturing Industry	Method	A methodology and toolset to support Big Data driven Digital Transformation DT of Manufacturing Industry. This is an instantiation for the Big Data domain of the MIDIH 6Ps method for Industry 4.0 in general. The 6 dimensions of the DT are related to Product, Process, Platforms (Technical Transformations) as well as People, Partnership, Performance (Socio-Business Transformations). This is an outcome of BOOST 4.0 WP9 and wsill be validated in a selected subset of Large Scale Pilots in WP4-WP8	Significant	Manufacturing Industry needs methods and tools to implement its DT pathways in the domain of Big Data Technologies. The Method is flexible and adaptable to any size, any sector and any location and encompasses Tools for BD Maturity Assessment, BD Strategy Definition and 6Ps pathways implementation. The final aim is to perform a large number of assessments and to implement several DT experiments so that to exctract precious benchmarks na d knowledge for EU Decision Makers. The Method and Tools should be associated to pan-EU surveys and benchmarks (e.g. Data Landscape) to better understand positioning and feasibility of DT pathways	Economic	6	Prototypes and validation experimets ns already exist (100+= but NOT tailored yet to the BD Technologies domain)	Manufacturing, all sectores	Disruptive	Delivered in 2018
Boost4.0	Digital Twin AGV	product	AGV virtualized simultaneously connected to the real AGV and to a simulation tool. Revenues generated by the sales of this new AGV type.	Significant	The connection with simulation tools and the use of Data Technologies will allow improvements of the models that are being used to calculate the number of AGVs for instalation, as well as optimize the distance travelled by the AGVs, the energy, the material resources. The virtualization of the AGV will allow offer the AGVs as a service.	Economic	5	demonstrator	Industrial	Incremental	Expected in +2019
Boost4.0	IDSA connected AGV	product	AGVs connected to the IDSA architecture. Revenues generated by data as a service pricing with data	Medium		Economic	4	demonstrator	Industrial	Incremental	Expected in +2019
Boost4.0	RISC framework Data Analytics	technology, methods, process	Newly developed methods of big data processing as well as methods from the fields of data engineering and data visualization or special methods of machine learning are tested, extended and integrated.	Significant	The Boost 4.0 project enables RISC to deepen its knowledge of big data technologies and to further develop its existing expertise in data science and artificial intelligence. The business processes focus primarily on research and development so that the existing RISC framework Data Analytics can be further developed.	Economic		prototype	Industry	Architectural	Expected in +2019

Boost4.0	Delay Forecasting Algorithms for Assembly Line Optimization	technology, methods	A predictive model to detect possible delays in VOLVO's transportation system of cabs. Furthermore, an assembly line optimization mechanism taking into account specific system constraints and the aforementioned cabs' predictions in order to minimize idle time of workstations and maximize productivity of the system.	Medium	Improving of transportation monitoring and reliability will lead to improve assembly line planning process by exploiting information from delay forecasting mechanism.	Economic	5	prototype	Industry and Supply Chain	Incremental	Expected in +2019
Boost4.0	Blockchain applied in the supply chain of a manufacturing industry	Technology, methods, processes	Blockchain supports the exhange of information in a network of several participants with a high security. In this pilot, blockhain is applied in a network with several actors participating in a supply chain transporting goods for manufacturing. The technical implementation will enable a higher transparency of the exchange of goods and information between the involved partnes and will ensure a real-time tracking in the supply chain. This will provide value in terms of reliability in receiving goods in time, more effective material handling to mention some.	Medium	It is an application of already existing technology to a new area. The impact lies in the manufacturing sector to receive a more reliable governance of their supply chain of goods and the exchange of information.	Economic		Developing	Manufacturing industry, supply chains, logistics	Disruptive	Expected in +2019
Boost4.0	FDT	product	Fault Detection Tool (FDT) is a combination of unsupervised fault detection approaches applied on streaming data.	Medium		Economic	7	supported by a prototype or demostrator	Industry	Disruptive	Delivered in 2018
Boost4.0	FPT	service	Fault Prediction Tool (FPT) is a combination of supervised Machine Learning fault prediction approaches.	Significant	FPT assists the adoption of the Predictive Maintenance to the Industry. The benefits are multiple: 1. Equipment downtime is decreased and the number of major repairs are reduced 2. Better conservation of assets and increased life expectancy of assets, thereby eliminating premature replacement of machinery and equipment 3. Reduced overtime costs and more economical use of maintenance workers due to working on a scheduled basis instead of a crash basis to repair breakdowns 4. Timely, routine repairs circumvent fewer large-scale repairs 5. Improved safety and quality conditions for everyone	Economic and Societal	6	supported by a prototype or demostrator	Industry	Architectural	Expected in +2019
Boost4.0	DSS-fusion	product	Decision Support System (DSS) enhanced with: 1. Machine Learning technique for the assisted/automated optimization of the defined rules. 2. Fusion of multiple fault detection and prediction outputs into a single outcome.	Medium		Economic	7	supported by a prototype or demostrator	Industry	Disruptive	Expected in +2019

Boost4.0	VW Assistant System in Production	Technology	Use data analytics results for assistant systems in production to support operators in their daily work.	Significant	Data analytics is an evolving field in the industrial real world enabled through ongoing connectivity between systems. Production assistance systems exploiting the results of data analysis approaches will have a significant impact in existing manufacturing systems.	Economic	6	Initial concept	Production	Incremental	Expected in +2019
Boost4.0	Dyamand	technology	DYAMAND (DYnamic, Adaptive MAnagement of Networks and Devices) is a software component that enables developers to integrate connected devices (e.g., IoT sensors and domotics) into their application	Medium			7	prototype, battle tested in production environments	N/A	architectural	Delivered mainly 2017 but progress in 2018
Boost4.0	Obelisk	technology	Higly scalable and optimized data storage platform for distributed and hyper heteregeneous sensor-based data sources	Medium			4	prototype, frequently used in research projects	smart city, Industry 4.0	architectural	Expected in +2019
Boost4.0	UNINOVA Big Data Architecture	Architecture/Framework	The UNINOVA BDA enables fast and easy configuration and deployment of known open-source Big Data technologies via a containerized architecture in a multi-node environment	Medium		Economic	NA	Supported by a prototype/ demo	All	Architectural	Expected in +2019
Boost4.0	UNINOVA Data Harmonization / Transformation IDS Big Data App	service	IDS-based Big Data Application enabling Data Harmonization / Transformation using defined schemas	Medium		Economic	NA	Supported by a prototype/ demo	All	Incremental	Expected in +2019
Boost4.0	Hybrid Twin approach for casting process	methods		Medium			4			Disruptive	Expected in +2019
Boost4.0	Ineractive visualization of large series	technology	Ability to visualize, explore, manipulate large series of data	Medium			4		Any	Incremental	Expected in +2019

Boost4.0	redBorder Active Industrial Network Monitoring Platform	product/ service	Technical Data Analytics, and Cybersecurity Big Data solution that is scalable up to the needs of Industry 4.0 cyber physical systems and industrial networks. redborder is the only platform that blends data from wired&wireless devices activity and can scale to millions of flows, logs, events, and users per second from thousands of devices without practical limits, from single-server to multi-site deployments. Crowd-sourced threat intelligence allows the platform to analyse, exploit, secure, and manage data from IIoT (Industrial Internet of Things) vendor protocols for making decisions in real time.	Medium	Deployment of protected communication chain from third party interference, where a series of "Industry 4.0 business rules" will be applied to trigger new events, alarms, or actions to prevent risks associated with the cooperation value creation network established between companies related to a smart factory. Visualisation of rel-time status of industry 4.0 assets and risk level based on big data stream monitoring	Economi c	7	Supported by a prototype/ demo	Manufacturin g	incrementa l	Expected in +2019
Boost4.0	M3 Big Data 3D Point Cloud Analytic Algorithms	product	Multisensor massive measurement platforms and algorithms for control of extremely large data sets comprised of 3D point clouds beyond 10 million points. Capability for fast processing and geometrical feature extraction and worker friendly visual analytics of multiple regions of interest with heterogeneous point cloud density. Particularly suitable for large parts in chassis and bofy in white applications in automotive	Medium	Easily store and share your dimensional information with great efficiency. Capability to share multi million 3D point cloud information anywhere and anytime, allowing easy access from different departments.Optimized to gain space without losing information and smart management	Economi c	6	first prototype	Manufacturin g	Disruptive	Expected in +2019
Boost4.0	IDSA TestLab	service	Facility and validation and verification framework for assessment and certification of IDSA component implementation	Medium	Ensure interoperbility of machines and big data plataforms used for implemnetation of AI algorithms based on industrial Data Space specifications	Economi c	5	Developing	Manufacturin g	incrementa l	Expected in +2019
Boost4.0	IDSA connected CMM	product	Coordinate Measuremen Machines connected to the IDSA architecture following QIF data model specifications. Revenues generated by data as a service pricing with data	Medium	Interoperbility of machining and quality control equipment in the implementation of zero defect manufcturing workflows	Economi c	4	demonstrat or	Manufacturin g	Incrementa l	Expected in +2019
Boost4.0	AITube Platform	Product	Zero defect mnufacturing IIoT and AI-based analytics and decision support support system	Medium	Multi dashboard visualization system for decision support based on process-centric information analysis. Rule based and event management support	Economi c	5	demonstrat or	Manufacturin g	Disruptive	Expected in +2019
CLASS	COMPSs	Technol ogy	An already existing computing framework used in HPC and Cloud computing to distribute big-data analytics workloads across the compute continuum (from edge to cloud), whilst providing timing guarantees on the overall execution of the workload. Moreover, the COMPSs framework is compatible with the Spark programming model and incorporates a proprietary tasking model.	Significa nt	It will allow incorporating distributed big-data analytics in systems with timing requirements (e.g. Automotive, smart cities, railway, smart manufacturing)	Economi c and Societal	TRL 6	Supported by a prototype	Smart cities, automotive, public transportatio n, Cloud computing, Edge computing	Architectur al	Expected in +2019

CLASS	EXPRESS	Technology	EXPRESS extends Lean OpenWhisk, an IBM event-driven development framework supporting the serverless concept, incorporating real-time directives to the programming language	Significant	It will support real-time execution in serveless computing	Economic and Societal	TRL7	Supported by a prototype	Smart cities, automotive, public transportation, Cloud computing, Edge computing	Architectural	Expected in +2019
CLASS	Rotterdam	Technology	Rotterdam is built on top of Kubernetes to support Container as a Service (CaaS). Rotterdam will incorporate Quality of Service (QoS) to support the real-time requirements guaranteed by COMPSs and EXPRESS when executing workloads at cloud level	Medium	-	Economic	TRL 4	Supported by a prototype	Cloud computing	Architectural	Expected in +2019
DataBench	Business KPIs motivating the Use of BDA techologies	Method	We have devceloped a classification methodology and process based on Business KPIs to benchmark the use of BDA techologies in business	Significant	Idem	Economic	N/A	Not yet	All		Delivered in 2018
E2Data	Heterogeneous JVM execution engine (Tornado VM)	Technology	The heterogeneous JVM powered by the Tornado framework, is capable of executing arbitrary Java code segments on heterogeneous devices such as GPUs and FPGAs. It aims to tackle the programmability challenge by enabling the hardware acceleration of high-level programming languages typically commonly found in Big Data frameworks. Additionally, by enabling dynamic application reconfiguration it also aims to partially address the partitioning and scheduling challenge at the node level. This is achieved by changing dynamically and at runtime the hardware device where a piece of code is being executed if a node has more than one heterogeneous device.	Medium		Economic	TRL 4	Prototype	ICT	Incremental	Expected in +2019
E2Data	Hardware-aware intelligent elastic scheduling (HAIER scheduler)	Technology	The Hardware-Aware Intelligent Elastic scheduler resides between the Big Data framework and the worker nodes. Its role is to decide dynamically which is the best possible, in terms of performance and/or power efficiency, hardware configuration for execution. The hardware-aware intelligent scheduler aims to tackle both the scheduling and the data processing timeliness chal- lenges at the cluster level.	Medium		Economic	TRL 3	Under development	ICT	Incremental	Expected in +2019

E2Data	Heterogeneous-aware Big Data framework (E2Data Framework)	Technology	Integration of the two aforementioned technologies into the Apache Flink framework: The Big Data framework of choice, Apache Flink, orchestrates the execution of Big Data applications on the heterogeneous hardware. E2Data will augment Apache Flink to perform dynamic data partitioning based on the capabilities of the underlying hardware configuration, aiming to tackle the data partitioning challenge at the distributed level.	Significant	The E2Data framework will provide compilation and runtime support for arbitrarily compiling any code segment to any hardware device completely transparently to the user. By using the same parallel and structured patterns such as map and reduce, developers will benefit from the underlying hardware capabilities without having to manually code their applications in a different programming language.	Economic	TRL 3	Prototype	ICT	Incremental	Expected in +2019
E2Data	Health Analytics	Service	In order to improve the predictive capability of the hospital readmission risk prediction algorithm, E2Data will enhance the patient discharge profile with profiles of highly correlated patients in terms of recent hospital activity. E2Data will extract patient correlations from available data matrixes for the purpose of enhancing a readmission risk prediction model through the use of Model-Based Collaborative Filtering	Significant	It has been shown in numerous studies that one of the most important preventive factors against readmissions of recently discharged patients is the correct discharge and follow-up monitoring to ensure that patients adhere to medication and lifestyle guidelines. E2Data will build upon a working system in one of the largest London hospitals that relies on historical admission data to build a model that aims to determine the risk of readmission for every patient that is discharged from the hospital on a daily basis.	Economic and Societal	TRL 3	Prototype	Health, ICT	Incremental	Expected in +2019
E2Data	Natural Language Processing	Service	The Language Processing use case will focus on processing large amounts of messages from social media, such as Twitter, in order to perform semantic information extraction, sentiment analysis, summarization, interpretation and organization of content. Such analysis occurs by extracting from each tweet phrases with specific syntactic forms. The process uses a number of different dictionary types storing a diverse range of information from word lists (vocabularies) to complex networks structures expressing syntactic patterns. These dictionaries provide hints with which each tweet (or arbitrary text) is going to be marked. The execution involves critical and complex algorithms (words prox- imity, fuzzy matching, etc.) that are invoked upon each new text, thus requiring their acceleration in order to become as efficient and scalable as possible.	Medium	Processing of unstructured data (text) is widely used to extract knowledge from articles and messages including social media. It is applied within several business domains to support various types of operations, where sentiment analysis and opinion mining are of significant importance (e.g., tourism, marketing, press, etc.). In the financial sector, the processing of online text and order streams is useful when we need to correlate financial news with trade facts, especially in the domain of fraud detection. Extreme time constraints during the execution of such algorithms makes it challenging for them to achieve their real-time business goals.	Economic	TRL 3	Prototype	ICT, Finance, Politics	Incremental	Expected in +2019

E2Data	Green Buildings	Service	Green Buildings provides a service for achieving more energy efficient buildings based on analytics of real time stream data derived from Internet of Things (IoT) deployed sensors in a number of public schools.	Significant	Affecting the behavioural characteristics of human interaction with buildings can have a tremendous impact on the overall reduction of the energy consumption throughout the EU. IoT technologies can enable real-time feedback on the impact of all actions, while environmental parameter measurements enable the conception and realization of diverse people-centric applications and scenarios.	Economic and Societal	TRL 3	Prototype	ICT, Environment	Incremental	Expected in +2019
E2Data	Biometric Security	Service	E2Data Biometric Security service will be able to meet the dual challenge of intensive computational workloads and differentiated quality of service to deliver a high availability and performant anti-spoofing service. A typical use-case would be that of a third-party secure access controller for high-risk online transactions setting up additional payees on bank accounts. The liveness (anti-spoofing) component verifies that an actual human is present and a real face is being shown to the camera. It is therefore capable of detecting mask attacks (in which the attacker wears a realistic mask of someone else), and replay-attacks (in which the attacker presents a video of someone else to the camera).	Significant	Real-world adoption of such a service has been impeded by the high compute cost required to process video streamed from user sessions. The amount of data and computation required is orders of magnitude greater than previous methods. E2Data Biometric Security service will be able to support it.	Economic	TRL 3	Prototype	ICT, Finance	Incremental	Expected in +2019
EDI	Amigo Climate, EDI-2018-2-UBIMET Quality control of weather observation data	Product	Station-based weather observations serve as a crucial input to different analysis and forecast models. To use these data in a reasonable way, outliers have to be identified. There are two types of outliers: extremes, which represent rare and intense events, and erroneous values. The challenge proposed by UBIMET is to derive an algorithm capable of detecting incorrect values while retaining extremes. AMIGO solution incorporates the best of their expertise into one disruptive algorithm, SOON (Station Observation Outliers fiNder). SOON is a highly advanced classificator based on the team's experience in meteorological data, statistical analysis, and machine learning. It boosts state-of-the-art methods with highly innovative tools, resulting in the perfect solution for UBIMET challenge.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s001/	Economic	TRL8	MVP level progressing into marketable product	Energy & Environment	Incremental	Expected in +2019
EDI	ZYLK, EDI-2018-11-VW_2 Predictive maintenance models for industrial robots in body shop	Product	Zylk is developing Industrial Robots Insights (IRI), a product that creates value from your company's industrial robots. By means of the interconnection of both Industrial and Big Data realms, IRI will be capable of reading, processing and visualizing data from Siemens PLCs connected to either KUKA robots and FANUC devices, including times-series and related statistics visualization in real time. Furthermore, a trained deep learning algorithm will be included as a preset, in order to predict deviations in the robots' behaviour. This will not only allow a successful predictive maintenance of your industrial robots, but also transform you into a data-driven company. Zylk also offers a customized product-based service to adapt the product to your particular data and company goals.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s002/	Economic	TRL8	MVP level progressing into marketable product	Industry 4.0	Incremental	Expected in +2019

EDI	BigDa Solutions, EDI-2018-10-VW_1. Thermal power plant smart management	Product	The project for Volkswagen challenge is to pre-process for data quality and create a dataset that will be used to generate a prediction model for each energy meter and predict the corresponding energy measurements.The workflow process will be: 1- Data preparation: pre-process to standardize all the information from different sources, merge all the internal information with external databases. 2- Database storing system: a stable pipeline process with the company database where all information from energy meters is stored. 3- Modelling: machine learning tools to make forecasts that improve the decision taking on data driven environment. Validation of model accuracy. 4- Improving and reporting: business intelligence platform to improve the user interaction and understanding of the information.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s004/	Economic	TRL8	MVP level progressing into marketable product	Industry 4.0	Incremental	Expected in +2019
EDI	LIS Solutions, EDI-2018-11-VW_2. Predictive maintenance models for industrial robots in body shop	Product	Body shops within automotive production are highly automated and more than thousand industrial robots work remotely and coordinated in a ballet-like choreography producing thousands of car bodies every day. Failures and stops lead to a decreased productivity and an increase of cost. For the VW Plant in Navarra, LIS-Solutions will create a system that can visualize and monitor the robot's performance in real time including the prediction of mechanic wear and failures. Anomalies and problems will be detected before the machine stops, and the maintenance teams will be informed so that corrective actions can be planed and scheduled optimally without affecting the productivity. This system will be based on data created by sensors constantly measuring attributes of each robot.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s005/	Economic	TRL8	MVP level progressing into marketable product	Industry 4.0	Incremental	Expected in +2019
EDI	Statice, EDI-2018-13-MIGROS. Prevent fraud in cashier operations	Product	Statice will solve the Cashier Fraud Detection challenge by Migros. Cashier and customer fraud can amount to large losses, given the huge number of transactions they deal with everyday. Being able to detect fraudulent transactions in real-time allows companies to make significant savings. But this can be tricky, especially in the case of employee or customers sensitive data. With stricter global privacy regulations in place, it is not sufficient anymore for companies to just pseudonymise data. Regulations such as the GDPR explicitly state the need for truly anonymous data. In the past, it has been shown that pseudonymised data is not enough to guarantee the anonymity of the individuals in the datasets. Statice will show that it is possible to build highly performant solutions such as fraud detection on truly anonymous data by using synthetic data.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s006/	Economic	TRL8	MVP level progressing into marketable product	Retail	Incremental	Expected in +2019

EDI	Accelegress, EDI-2018-14-VIAVERDE. New pricing bundles for drivers in Porto	Product	This project aims to develop a big-data application for price optimisations in transport use cases. In the challende Via Verde looks to increase the percentage of tolling and parking transactions before their introduction. Machine learning techniques, such as clustering for segmenting customer consumption patterns, are applied to understand existing data better. An appropriate demand forecasting model and dynamic price optimisation strategy will be developed. This will be accompagnied by an impact analysis to estimate positive or negative impact of potential bundle schemes, simulating future service consumption after applying pricing optimisations. Additionally, the potential to introduce Save-a-Space parking reservations as a value-added end-user service to Via Verde's customers will be evaluated.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s007/	Economic	TRL8	MVP level progressing into marketable product	Smart Cities	Incremental	Expected in +2019
EDI	Pragma-IoT, EDI-2018-12-ELEKLJ. System for detection of anomalies in smart meter	Product	Electricity companies face a lot of anomalies in the energy consumption of final customers. Pragma has already developed an innovative white label (OEM) IoT framework capable to collect and analyse data from different infrastructural nodes. Pragma will facilitate the detection of anomalies and potential patterns in energy consumption by analysing the data in real-time, utilizing advanced algorithms (eg graph analytics, trend analysis, etc). Further processing will indicate if a malfunction is justified (e.g. maintenance) or not (eg fraudulent), while analysis in the spatio-temporal domain through analytics tools will identify categorization patterns (e.g. fraudulent, meters' errors, etc). The information is processed on the fly, producing real-time notifications and parameterizable reports.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s008/	Economic	TRL8	MVP level progressing into marketable product	Energy & Environment	Incremental	Expected in +2019
EDI	Morphl, https://edincubator.eu/edi-explore-s013/	Product	The product development process consists of a few important phases: planning, design, development & launch. To increase engagement and conversion rates, this process undergoes multiple iterations which developers seldom navigate by looking at the data. Usually there's a product owner, marketing or sales person, analyzing it and feeding developers a feature list. This leads to lots of guess-work. MorphL solves this problem by using machine learning to predict user behaviors in mobile & web applications and facilitating a personalized user experience. MorphL started at the beginning of 2018 and it has been funded via Google Digital News Initiative. Through EDIncubator, MorphL is able to add "sentimental analysis in consumer journey (EDI-2018-4-JOT_2)" to their collection of integrations, enabling internet & media companies to AI-enhance their products.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s013/	Economic	TRL8	MVP level progressing into marketable product	Internet & Media	Incremental	Expected in +2019

EDI	Energy Sequence, https://edincubator.eu/edi-explore-s015/	Product	EnergySequence(ES) plan to build, develop and deploy different kinds of building energy models for characterizing and predicting HVAC consumption patterns and design further applications for the identification of optimal operation schedules, waste reduction and improvement of the overall energy efficiency. Energy optimization in commercial buildings is the goal of this proposal. ES platform can forecast energy demand and automatically adjust HVAC sets accordingly. The platform learns and adapts to the building's needs. In a wide range of installations, EnergySequence aims to demonstrate it can reduce HVAC energy consumption and costs by 10-25%, while maintaining occupant comfort levels. ES platform has deep capabilities in data ingestion, aggregation, monitoring, trending, predictive analytics, visualization, automation and optimization of HVAC energy usage and consumption. Human expertise informs ES's unique energy data analytics to provide the basis for identifying operational and equipment anomalies.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s015/	Economic	TRL8	MVP level progressing into marketable product	Energy & Environment	Incremental	Expected in +2019
EDI	Leapcraft, EDI-2018-9-VPS. HVAC usage management and control optimisation	Product	Leapcraft, seea strong demand for good indoor climate, especially air quality, amongst various types of new commercial building customers and have successfully deployed several cases. In this proposal, Leapcraft want to create a live example of a large commercial building that both provides good indoor climate and reduces energy consumption using data mining and advanced demand-control techniques. Leapcraft plans to achieve this using the available data from this challenge, data from their own indoor and outdoor air quality sensors, connecting to energy metering system and building HVAC (ventilation systems) via their artificial intelligence (A.I) algorithms. Leapcraft anticipate their work will in the future make workplaces more healthy, improve productivity and reduce energy consumption	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s016/	Economic	TRL8	MVP level progressing into marketable product	Energy & Environment	Incremental	Expected in +2019
EDI	SmartCat, EDI-2018-9-VPS. HVAC usage management and control optimisation.	Product	SmartCat is proposing a solution to "HVAC usage management and control optimisation" EDI challenge in the form of state of the art machine learning models, generalised to other similar datasets and delivered as a platform for energy redistribution and load balancing. To solve the challenge and develop proposed platform SmartCat will train regression models for HVAC usage that would uncover load profiles as well as support anomaly detection and alerting real-time, and also build a simulation of HVAC usage and tenants' actions in a reinforcement learning setting. Ingestion, (re)training and scoring would serve as backend to web-based platform with models' results visualised on dashboards and exposed as suggestions to energy provider to support automatic load balancing.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s018/	Economic	TRL8	MVP level progressing into marketable product	Energy & Environment	Incremental	Expected in +2019

EDI	Carmetry, EDI-2018-6-RACC. Improving the road safety in Barcelona	Product	This project proposes a new prediction technique based on a probabilistic regression model that allows to predict the severity of a traffic accident for each of its occupants. The model is intended to provide a collective intelligence to the vehicle and also will be useful to improve the user's driving behavior through the provision of collective road intelligence to the vehicle and the devices used in it. A mobile application will be used to guide the driver and it will show alerts identifying possible black spots in the different routes, based on the results obtained by the proposed technique. These alerts that will appear in the mobile application can be verified or denied if traffic has already normalized at that kilometric point.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s022/	Economic	TRL8	MVP level progressing into marketable product	Smart Cities	Incremental	Expected in +2019
EDI	SummarizeBot, EDI-2018-4-JOT_2. Sentimental analysis in consumer journey	Product	SummarizeBot has multilingual aspect based Sentiment Analysis technology . The EDI challenge helps to build and test completely new sentiment analysis based product. It will be sold as application programming interface and chatbot different messaging platforms as Slack. The new Sentiment Analysis product will be able to identify and categorize keywords by the following categories: research, complain, opinion, action, comparison, etc. The product will be built based on our existing prebuilt models, that allow us to deploy new machine learning algorihtms easily.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s024/	Economic	TRL8	MVP level progressing into marketable product	Internet & Media	Incremental	Expected in +2019
EDI	INOVAlabs, EDI-2018-10-VW_1. Thermal power plant smart management	Product	The first step will be the construction of a simulation model of the thermal plant as accurate and realistic as possible. The model will be based on data provided by the customer (and other relevant information like the combustion conditions and properties of the gas) and will be the used to train a Reinforcement Learning algorithm. In Reinforcement Learning, a software agent makes observations and takes actions within an environment, and in return it receives rewards. Its objective is to learn to act in a way that will maximize its expected long-term rewards. The algorithm used by the software agent to determine its actions is called its policy. In the context of this proposal, the policy learnt by the algorithm will determine when to switch on or off the plant.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s026/	Economic	TRL8	MVP level progressing into marketable product	Industry 4.0	Incremental	Expected in +2019

EDI	AIception, EDI-2018-11-VW_2. Predictive maintenance models for industrial robots in body shop	Product	We will: find an easy way to export/ extract the sensor data (velocity, current, temperature) preprocess the data and do feature engineering and analyze the data train unsupervised anomaly detection models using Spark MLlib and scikit learn deploy the models in a Big Data Stack using Apache Spark 2.3 create intuitive visualizations that will help and alert VW engineers The software should: predict mechanical/ electrical/ software failure do predictive maintenance(vs preventative vs corrective maintenance) handle the collisions or crashes do a robot diagnosis We want: to work closely with VW engineers and discover their needs to create a product that will bring VW value (safety, cost savings, uptime)	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s027/	Economic	TRL8	MVP level progressing into marketable product	Industry 4.0	Incremental	Expected in +2019
EDI	Ciclogreen, EDI-2018-7-TUB Optimisation of traffic flow in Berlin	Product	Ciclogreen will develop an analytics and prediction tool that, given the historical data of the traffic flow at a given point in the city, will simulate behaviours that could generate congestion or flow of traffic, public transport, and pedestrians. The solution will include an easy-to-use specific website that will: – Allow Traffic operators to study mobility patterns.. – Allow users to observe the impact of changes to the traffic control infrastructure. Ciclogreen will take advantage of their current solution, specially their Transportation Heatmaps and the possibility to take additional mobility data using their tracking app (www.ciclogreen.com).	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s028/	Economic	TRL8	MVP level progressing into marketable product	Smart Cities	Incremental	Expected in +2019
EDI	Ciclogreen, EDI-2018-7-TUB Optimisation of traffic flow in Berlin	Product	Ciclogreen will develop an analytics and prediction tool that, given the historical data of the traffic flow at a given point in the city, will simulate behaviours that could generate congestion or flow of traffic, public transport, and pedestrians. The solution will include an easy-to-use specific website that will: – Allow Traffic operators to study mobility patterns.. – Allow users to observe the impact of changes to the traffic control infrastructure.	Significant	Incremental innovation, medium. Full details at: https://edincubator.eu/edi-explore-s028/	Economic	TRL8	MVP level progressing into marketable product	Smart Cities	Incremental	Expected in +2019

			Ciclogreen will take advantage of their current solution, specially their Transportation Heatmaps and the possibility to take additional mobility data using their tracking app (www.ciclogreen.com).								
EDI-Statice	EDI: Fraud detection for retail stores	Product	As part of the EDI program, we developed a fraud detection software for grocery retail chains to decrease their stock shrinkage. The value of this software is for retail stores to predcit and detect fraud in order to decrease the value lost sales.	Medium		Economic		Prototype	Retail	Incremental	Delivered in 2018
Fandango	**Multilingual text analytics**	Technology	Service of multilingual text analytics to be used by juournalists, media professionals and fact chekers to detect misleading messages or disinformation claims	Significant	Misinformation challange is a problem of high importance in Europe, but not only; the possibility to have services supporting the human in discovery misinformation is a n high (significant) impact for media professionals and, even, for citizens	Economic and societal	TRL6	first demo (prototype ready in 2019)	Media	Disruptive	Expected in +2019
Fandango	**Spatio-temporal analytics**	Technology and system	Analyzing news posts and finding duplicates or near duplicate posts in the past or referring to other geographic/physical locations or in other contexts.	Medium	A common case of fake news is the re-posting of a real case of the past that it is no longer relevant or was removed from context. Such spatio-temporal or out-of-context correlations can generate strong fakeness markers as soon as they exploit data from various data silos to identify fully or partially copied information.	Economic and societal	TRL6	first demo (prototype ready in 2019)	Media	Incremental	Expected in +2019
Fandango	**Copy-move detection on audio-visual content**	Technology and system	The innovatio is focused in the detection of the intentional manipulation of images and videos to add or remove or in any way modify the visual content.	Medium	Th innovation exploits the power of deep learning architectures to identify manipulated content as well as to identify the pool of near duplicate content and visuals that were used as sources for creating the fake object.	EConomic and societal	TRL5	first demo (prototype ready in 2019)	Media	Incremental	Expected in +2019
Fandango	**Source credibility**	Technology, Services and methods	Profiling of the sources of the news posts and apply graph analytics to detect paths and nodes that tend to produce such fake news and defuse them widely on the public web.	Significant	A critical part of the source credibility profiling is the totally automated pipeline so that no human intervention will be needed. This is important also for the trust of users towards the platform and the mitigation of possible biases especially when it comes in business and economics analytics.	economic and societal	TRL6	first demo (prototype ready in 2019)	Media	Disruptive	Expected in +2019
i-BiDaaS	GPU accelerator technology (FORTH)	technology	A real-time high speed stream processing and pattern matching engine, tailored for continuous stream data, utilizing GPGPUs for the acceleration of computations. Will be integrated with CEP engine within I-BiDaaS.	Medium		Economic	3	Proof of Concept	Banking, Telecommunication, Manufacturing	Incremental	Expected in +2019

i-BiDaaS	Commodity cluster & Privacy enhanced execution (FORTH)	technology	Computational and storage resources with a GPU cluster; enables functionalities of privacy preservation through commodity hardware (Intel SGX). The innovation is on integrating the component within a big data as a self-service platform.	Medium		Economic	3	Proof of Concept	Banking, Telecommunication, Manufacturing	Incremental	Expected in +2019
i-BiDaaS	COMPSs (Programming Model and Runtime) (BSC)	model & system	Data-driven task scheduling and data movement optimization. Innovation in I-BiDaaS is a pool of ML algorithms implemented in this framework.	Medium		Economic	8	First of a kind Commercial System	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Hecuba (BSC)	set of tools		Medium		Economic	5	Large Scale Prototype	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Qbeast (BSC)	system	Indexing system - a Hecuba DB sub-module that provides data-thinning and multidimensional indexing capabilities. Planned to used Qbeast sampling for ML algorithms within I-BiDaaS.	Medium		Economic	5	Large Scale Prototype	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Test Data Fabrication (IBM)	platform		Medium		Economic	6	Prototype System	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Apama Streaming Analytics Platform (SAG)	platform	Complex event processing sub-module. APAMA is built on an in-memory architecture that enables real-time processing of extremely fast, large data volumes—orders of magnitude larger than traditional database-based IT architectures. The innovation is on integrating the component within a big data as a self-service platform.	Medium		Economic	6	Prototype System	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Universal Messaging (SAG)	product	Passive publish/subscribe message oriented-middleware; primarily used to ingest, prepare, and integrate data from various sources and make it available for batch processing and streaming analytics. The innovation is on integrating the component within a big data as a self-service platform.	Medium		Economic	7	Demonstration System	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018

i-BiDaaS	MashZone (SAG)	software	MashZone is a visual analytics software that provides a fast way to explore and analyze streaming data and events. It allows to analyze both real-time and at-rest data. The innovation is on integrating the component within a big data as a self-service platform.	Medium		Economic	4	Small Scale Prototype	Banking, Telecommunication, Manufacturing	Incremental	Expected in +2019
i-BiDaaS	Advanced visualization and monitoring (AEGIS)	toolkit	Extensible platform for data visualization and transformation. The tool will be significantly extended within I-BiDaaS.	Medium		Economic	4	Small Scale Prototype	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Pool of ML alorithms based on structured (non)convex optimization (UNSPMF)	methods	This approach allows the modelling of various data analytics tasks through formulations of (possibly non-convex) optimization problems (e.g., a classification task corresponds to minimizing a logistic or hinge loss function). This is a very flexible and generic approach and is capable to accommodate a broad range of analytics tasks. I-BiDaaS planned innovation is to capitalize on novel theoretical results to implement novel ML algorothms in COMPSs.	Medium		Economic	3	Proof of Concept	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Cloud Application Service management (ATOS)		Computational and storage resources. It is based on Apache Brooklyn allowing deployment and life-cycle management of applications in hererogeneous (private and upublic cloud technologies).	Medium		Economic	4	Small Scale Prototype	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Resource management and orchestration module (ATOS)	module	Performs management and orchestration of resources. The innovation is on integrating the component within a big data as a self-service platform.	Medium		Economic			Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Batch analytics module	module	Contains two sub-modules, Advanced machine learning sub-module (UNSPMF), and COMPSs programming model submodule (BSC). The latter is a sequential programming model able to exploit the inherent parallelism of the applications developed with it; used for implementation of the advanced machine learning submodule. The former is a python implementation (using COMPSs programming model) of a pool of machine learning algorithms for batch analytics. The innovation is on providing an open source pool of ML algorithms that are planned to incorporate Qbeast based sampling.	Medium		Economic		Proof of Concept	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018

i-BiDaaS	Streaming analytics module	module	Consists of two sub-modules, Apama complex event processing sub-module (SAG), and GPU-accelerated streaming analytics and pattern matching sub-module (FORTH). The former is a complex event processing infrastructure for streaming analytics. The latter is a system for optimizing the pattern matching procedures (i.e., string searching and regular expression matching) taking advantage of the benefits that GPU accelerators (FORTH cluster) offer. The innovation is on query partitioning between GPU accelerated analytics and CEP module.	Medium		Economic		Proof of Concept	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
i-BiDaaS	Advanced visualization module	module	Consists of the MashZone sub-module (SAG) and Advanced Visualizations Toolkit – AVT (AEGIS). The innovation is on integration of the two and on significant extension of AVT.	Medium		Economic		Proof of Concept	Banking, Telecommunication, Manufacturing	Incremental	Delivered in 2018
iCARUS	Data Collection	Service	A novel bundle of services to check-in, curate, map, harmonize and upload private and confidential aviation data in the ICARUS platform. Such services are based on the ICARUS aviation core data model and the ICARUS metadata schema, and practically involve the upstream, downstream and indirect collection of data assets from the supply-driven perspective of the data providers. The guidelines for all services are configured on the ICARUS platform but executed in the on-premise environment of the end-user.	Medium		Economic	7	Supported within the ICARUS platform prototype	Aviation	Incremental	Expected in +2019
iCARUS	Data Analytics & Visualisation	Service	A novel service that enables the design, execution and monitoring of the data analytics workflows. This service provides value to the data scientistis, offering an user friendly and interactive UI that supports the whole data analysis life cycle, from the design of the analysis to the execution, with a plethora of data analysis algorithms and visualisation types that can be customised according to their needs and scheduled to be periodically executed. This service is also relevant for business users as they may gain insights on the results of aviation-related analytics through intuitive visualizations.	Medium		Economic	7	Supported within the ICARUS platform prototype	Aviation	Incremental	Expected in +2019
iCARUS	Data Security	Service	A service that protects and ensures the confidentiality and the integrity of the data in the whole lifecycle of the data exploitation by safeguarding the security aspects of data in storage, in transit and in use. This services provides value to the stakeholders of the platform by preventing unauthorized disclosure and any intentional or accidental unauthorized modification to the data assets with an advanced authorisation engine, and by protecting their confidentiality with an end-to-end data encryption approach.	Medium		Economic	7	Supported within the ICARUS platform prototype	Agnostic	Architectural	Expected in +2019

iCARUS	Data Sharing	Service	A DLT-based service for data brokerage that defines in an immutable manner the data license terms according to which an aviation data provider makes available his/her data, and the data contracts that are signed between the rightful aviation data provider and a data consumer (that meets the authorization policies set by the data provider and has been uniquely identified based on a strict KYC approach). The data sharing service manages the smart data contract drafting, the bilateral agreement among the involved parties,the smart data contract validation and the "offline" payment confirmation.	Significant	Data sharing is a very tangible challenge in aviation that still typically operates in silos, with the different aviation stakeholders only having fragmented data at their disposal. Through the ICARUS data sharing service, bilateral trusted and secure sharing of proprietary data assets is enabled, according to the specifications provided by the ICARUS stakeholders and external stakeholders engaged.	Economic	7	Supported within the ICARUS platform prototype	Aviation	Disruptive	Expected in +2019
iCARUS	Resource Orchestration	Service	A service for end-to-end orchestration and provisioning of a secure environment for data analysis execution over a virtual infrastructure. This service provides value to the stakeholders with the dynamic on-the-fly deployment and management of a secure isolated environment where the data analytics jobs can be executed safeguarding the confidentiality of both the used data assets and the results of the analysis.	Medium		Economic	7	Supported within the ICARUS platform prototype	Agnostic	Architectural	Expected in +2019
TheBuyForYou	VIPS	Service	Vendor intelligence solution offering advanced analytics capabilities for supplier risk monitoring, collusive tendering and procurement decision support	Medium		Economic and Societal	6	First prototype in Q3 2019	Public and private	Incremental	Expected in +2019
TheBuyForYou	PITO	System, service	An API feed of global procuremnet data , tender search and alert service, and contract pipeline, with all data in he Open Contracting Data Standard	Significant	The PITO service will likely be the most complete source of procurement data on a global scale	Economic and Societal	7	API is in use	Public and private	Incremental	Delivered in 2018
TheBuyForYou	RS	Service	Reconciliation service that matches references to companies to canonical records or entity resolution.	Medium		Economic	6	Will be used by the TBFY API	Public and private	Incremental	Expected in +2019
TheBuyForYou	SEPPAS	System, service	Procurement data analytics and anomaly detection	Significant	Exploitation of the analytics services will have major impact on procurement intelligence and fraud detection	Economic and Societal	5	Prototype to be presented in 2019	Public	Incremental	Expected in +2019
TheBuyForYou	COPIN	System, service	Platform providing EU tender information and analytics for bidders and public administrations, in order to promote economic development and efficiency in public administration procurement	Medium		Economic	5	Prototype to be presented in 2019	Public and private	Incremental	Expected in +2019
TheBuyForYou	TBFY Ontology	Method	Knowledge Graph, based on the OCDS standard, incorporating the euBusinessGraph ontology for company information	Medium		Economic	5	Supported by prototype	Public and private	Incremental	Delivered in 2018

Typhon	TyphonML Modelling Tools	Technology	These technologies comprise the software that implements the TyphonML modelling language and will consist of the modelling tools supporting the creation of TyphonML models.	Significant	These tools will allow organisations to effeciently and effectively exploit a vast amount of Big Data currently maintained in data silos that are difficult to combine to create new products and services.	Economic and Societal		In development	All sectors that utilise Big Data	Disruptive	Expected in +2019
Typhon	TyphonML Model Analysis and Reasoning Tools	Technology	These technologies will include the algorithms and tools for reasoning about the feasibility and properties of TyphonML models.	Significant	These tools will allow organisations to effeciently and effectively exploit a vast amount of Big Data currently maintained in data silos that are difficult to combine to create new products and services.	Economic and Societal		In development	All sectors that utilise Big Data	Disruptive	Expected in +2019
Typhon	TyphonML to TyphonDL Model Transformation Tools	Technology	These technologies will provide an extensible transformation infrastructure for deriving deployment models expressed in TyphonDL from abstract TyphonML models.	Significant	These tools will allow organisations to effeciently and effectively exploit a vast amount of Big Data currently maintained in data silos that are difficult to combine to create new products and services.	Economic and Societal		In development	All sectors that utilise Big Data	Disruptive	Expected in +2019
Typhon	Optimized Hybrid Polystore VM Assembly Tools	Technology	These technologies will be able to generate configuration scripts to assemble Hybrid polystore VMs from source TyphonDL models.	Significant	These tools will allow organisations to effeciently and effectively exploit a vast amount of Big Data currently maintained in data silos that are difficult to combine to create new products and services.	Economic and Societal		In development	All sectors that utilise Big Data	Disruptive	Expected in +2019
Typhon	TyphonQL Compilers and Interpreters	Technology	These technologies are based the TyphonQL query language, which is a new language for uniformly querying heterogeneous database stores, abstracting from the specificities of underlying technology.	Significant	These tools will allow organisations to effeciently and effectively exploit a vast amount of Big Data currently maintained in data silos that are difficult to combine to create new products and services.	Economic and Societal		In development	All sectors that utilise Big Data	Disruptive	Expected in +2019
Typhon	Event Publishing and Monitoring Framework	Technology	These technologies will provide a high performance architecture for publishing and processing data access and update events in a performant, scalable and fault-tolerant manner.	Significant	These tools will allow organisations to effeciently and effectively exploit a vast amount of Big Data currently maintained in data silos that are difficult to combine to create new products and services.	Economic and Societal		In development	All sectors that utilise Big Data	Disruptive	Expected in +2019
Typhon	Text Processing Pipelines	Technology	These technologies will implement a number of software pipelines, assembled from pre-existing text-processing components (e.g. sentence splitters, parsers), which will process textual data persisted in hybrid polystores.	Significant	These tools will allow organisations to effeciently and effectively exploit a vast amount of Big Data currently maintained in data silos that are difficult to combine to create new products and services.	Economic and Societal		In development	All sectors that utilise Big Data	Disruptive	Expected in +2019

Typhon	Hybrid Polystore Continuous Evolution Tools	Technology	These technologies will provide a consistent set of integrated tools for continuous evolution of the hybrid polystore, for recommending polystore schema reconfigurations when relevant, and for informing the polystore evolution process.	Significant	These tools will allow organisations to effeciently and effectively exploit a vast amount of Big Data currently maintained in data silos that are difficult to combine to create new products and services.	Economic and Societal		In development	All sectors that utilise Big Data	Disruptive	Expected in +2019
Lynx	Multilingual Question Answering on Legal Documents	Service	The system is loaded with documents by legal experts. These documents are analyzed with the help of a knowledge graph, and an automatic Question Answering system is deployed, that answers final-user queries in natural language, pointing them to the correct documents.	Medium	Enables law firms in one member state to consult legislation from another member state, greatly reducing compliance costs.	Economic and Societal	3	Prototype	Compliance, Law	Architectural	Expected in +2019
Lynx	Rental Contract Analysis System	Service	A system that takes a rental contract and finds relevant laws, regulation and case law for each of the sections.	Medium	Reduction in the time that an entity purchasing real state subject to rental contracts needs to invest in verifying compliance.	Economic	3	-	Compliance, Real Estate	Architectural	Expected in +2019
Lynx	Geothermal Project Analysis System	Service	A system that takes a project proposal in the geothermal energy domain and find relevant laws, regulations, permits and projects.	Medium	Reduction in the time that compliance officers in the geothermal energy field need to find the legislation and permits relevant to a project.	Economic	3	-	Compliance, Energy	Architectural	Expected in +2019
Lynx	Semantic Document Manager	System	A system for storing documents and their annotations. These annotations are backed by semantic-web technology, and serve as a basis for several important processing systems.	Significant	The need to keep documents and their annotations syncronzied arises in a variety of applications. By allowing flexibility regarding the nature of the annotations, this system can integrate with many other products to provide unified access to annotated documents within an organization.	Economic and Societal	4	Prototype	Any dealing with documents	Incremental	Expected in +2019
Cross-CPP	Multilingual Question Answering on Legal Documents	Service	The system is loaded with documents by legal experts. These documents are analyzed with the help of a knowledge graph, and an automatic Question Answering system is deployed, that answers final-user queries in natural language, pointing them to the correct documents.	Medium	Enables law firms in one member state to consult legislation from another member state, greatly reducing compliance costs.	Economic and Societal	3	Prototype	Compliance, Law	Architectural	Expected in +2019
Cross-CPP	Rental Contract Analysis System	Service	A system that takes a rental contract and finds relevant laws, regulation and case law for each of the sections.	Medium	Reduction in the time that an entity purchasing real state subject to rental contracts needs to invest in verifying compliance.	Economic	3	-	Compliance, Real Estate	Architectural	Expected in +2019
Cross-CPP	Geothermal Project Analysis System	Service	A system that takes a project proposal in the geothermal energy domain and find relevant laws, regulations, permits and projects.	Medium	Reduction in the time that compliance officers in the geothermal energy field need to find the legislation and permits relevant to a project.	Economic	3	-	Compliance, Energy	Architectural	Expected in +2019

Cross-CPP	Semantic Document Manager	System	A system for storing documents and their annotations. These annotations are backed by semantic-web technology, and serve as a basis for several important processing systems.	Significant	The need to keep documents and their annotations syncronzied arises in a variety of applications. By allowing flexibility regarding the nature of the annotations, this system can integrate with many other products to provide unified access to annotated documents within an organization.	Economic and Societal	4	Prototype	Any dealing with documents	Incremental	Expected in +2019

5.2 New economic viable services of high societal value (KPI II.7)

Projects where asked to quantify, list, describe new economically viable services of high societal value developed or resulting from the project during 2018 and also to forecast (2019+). They were also asked to describe why services qualified.

63 new economically viable services of high societal value were developed during 2018 as result of the projects. 47% (over 30 projects) contributed to this KPI. **Table 11** provides details per project.

Project	#	Description of the services of high societal value
BigDataOcean	1	Poseidon Oil Spill Simulation
Data Pitch	4	4 startup solutions focused on providing a positive social impact within the Health industry
DataBio	12	11 of the DataBio innovations listed in "Innovations to market" have societal value. 3 of the DataBio pilots directly support and improve policy development at EU and national levels, with direct impact in bio-economy and the distribution of government funding: • One pilot in agriculture generates Earth Observation services and data for improving decision-making related to the EU Common Agricultural Policy, in particular improving the process of policy making as well as more accurate development of subsidy policies by improvement of operational subsidy pay out processes of Payment Agencies. • Another pilot is dedicated to providing services for a national agricultural policy agency. • One pilot and service are dedicated to development of web-mapping service for government decision making in forestry. Another 9 DataBio pilots which mainly develop commercial services in agriculture, forestry and fisheries, also provide some significant societal value as they enable saving and efficient use of natural resources of high societal value: • Two forestry pilots that help to early detect forest diseases and to prevent disease spreading. This helps to prevent enormous losses of forest, which is of high societal value that goes beyond private commercial use of forest One pilot that enables a better pest/disease prediction and more efficient use of fertilizers in agriculture, which besides lowering the costs of farmers also provides additional societal value of a more sustainable and environmental friendly agriculture • One pilot that enables more efficient use of water through early detection of irrigation inhomogeneities. Besides lowering the costs of farmers, this pilot also enables a more efficient dealing and saving of the natural resource water • Five pilots that improve productivity in fisheries and fish stock assessment and enable better control of fisheries quotas and environmentally friendly fisheries.
e-SIDES	1	Advice for developers of BDA systems to help them avoid typical ethics and legal pitfalls

euBusinessGraph	7	List provided in "Innovations to Market"
TT	5	a) Improved Routing Service (Connected Trucks): Enhanced component to provide better and more detailed planning possibilities for truck routing and processes of Logistic Service Providers; Category: Software; Main features are described as follows: • Calculation of alternative routes according to travel times and POI analyses • Data analyses of different data sources and comparison to real LSP data • Provision of recommendations for planners and drivers b) Descriptive Analytics Service (Dynamic Supply Networks in e-commerce): it provides interaction with data through the selection of various criteria/dimensions, and as well it detailed analysis of data behaviour in 3PL partners through a set of appropriate visuals, and deduction of patterns and trends. c) Forecasting Analytics Service (Dynamic Supply Networks in e-commerce): A forecasting tool that will allow the potential user to periodically predict the behaviour of different groups of data depending on the selection of various criteria/dimensions, while presenting its forecasts through simple and convenient graphs d) Route Optimization Service (Dynamic Supply Networks in e-commerce): Inventory routing in an omnichannel environment to facilitate the postcode assignment to hubs. e) Customer Reviews Analysis Service (Dynamic Supply Networks in e-commerce): Extract online consumer insights with respect to logistics and delivery processes to perform a sentiment analysis over the reviews.
QROWD	19	Included in Innovations to Market
MHMD	1	Dynamic consent APP
BigDataGrapes	1	BigDataGrapes Platform as a Service
BigMedilytics	5	• Solution for tracking efficiency of emergency departments in hospitals (Version 1) • Solutions for treating prostate cancer patients more effectively (Version 1) • Predicting inpatient treatments of patient with known heart failure with health insurance data • Search engines enabling to exploit healthcare data to improve individual diagnostic quality and corresponding care • A knowledge graph that can be explored and traversed
CLASS	1	Data gathering and generation of a city knowledge base upon which advanced mobility services can be built
DataBench	1	Business KPIs motivating the Use of BDA technologies
EDI	1	In round 1 the following challenge of societal value was proposed https://edincubator.eu/2018/03/26/improving-the-road-safety/, seeking "Improving the road safety"
FANDANGO	4	Multilingual text analytics; Spatio-temporal analytics; Copy-move detection on audio-visual content; Source credibility service.

Table 11: List and description of services of high societal value per project (delivered 2018)

Additionally, projects have forecasted (2019+) additional **61** new economically viable services of high societal value as result of their projects in the upcoming 2 years (2019-2020). 53% of the projects active in 2018 are planning to contribute (9 of them new to provide contributions to these KPI).

Project	#	Description of the services of high societal value
AEGIS	1	The AEGIS service - a public instance of the AEGIS platform.
Data Pitch	10	10 startup solutions will be supported in 2019 to create positive social impact within industries such as Energy, Health and Agriculture
euBusinessGraph	8	List provided in "Innovations to Market"
SPECIAL	2	Tracking how personal data is being used across an enterprise is only possible under the assumption that the controller is aware of its existence, in all its forms, as well as its precise location. But, to know what truly constitutes personal data, one must also be able to tell what contributes to the identification of an individual and, ultimately, what belongs to whom. Within SPECIAL, TenForce is looking to build on these premises to deliver an AI-driven data discovery, classification and annotation solution that would provide a data subject-centric view of the company data landscape. One Service by DT and/or a DT subsidiary - Motionlogic) will be greatly enhanced (almost a "new service")
BigMedilytics	9	• Solution for tracking efficiency of emergency departments in hospitals (Version 2) • Solutions for treating prostate cancer patients more effectively (Version 2) • Multiparty Computation (MPC): a method/tool to analyse sensitive data from multiple parties (who do not wish to share data with each other) without revealing the data itself, but only revealing the outcome of the (data) analysis, i.e. the coefficients of the learned model. • A LASSO based regression that allows for prediction of comorbidities of hospital admissions out of > 4.000 features in a Health Insurance database. • A smart watch-based protocol to increase % of HF patients enrolment for Heart Failure Revalidation • A smart watch-based protocol to increase adherence of HF patients to increased exercise post Heart Failure Revalidation • Search engines enabling to exploit healthcare data to improve individual diagnostic quality and corresponding care • increased productivity of nurses and other caregivers in hospitals. This is an important result because the demand for care is rising while current personnel is scarce, and the deficit will increase the coming years.
BODYPASS	2	• Data interchange using Hyperledger • Virtual Couture Fashion™, is a vision for a new Real Time Fashion System bringing AI driven design, new technologies, a new value chain, a completely new business model and new manufacturing methods to the Fashion industry

Boost 4.0	2	Description is provided in the "Innovation to Market" document with the short name FPT
CLASS	2	Based on the knowledge based generated in 2018: Generation of a pollution model at street granularity level; Smarter advanced driving assistant systems based on information provided by the city
DataBench	1	Benchmarking tool aggregator
E2Data	4	1. Health analytics, 2. Natural Language Processing, 3. Green Buildings, 4. Biometric Security (all services described in "Significant Innovations to Market" questionnaire)
FANDANGO	4	Evolution of the same services started in 2018
ICARUS	1	More accurate and realistic prediction model of epidemics spreading across the globe with the help of airline transportation data (To be implemented by the Computational Epidemiology group of ISI, a private research institution located in Turin, Italy that has developed the global epidemic and mobility model, GLEAM)
TheyBuyForYou	8	Listed In "Innovations to market"
TYPHON	3	1) Polyglot data management services, 2) Polyglot data deployment services and 3) Polyglot data evolution services.
Lynx	1	Citizen portal to access linked legislation from several jurisdictions
Cross-CPP	3	Description is available in "Innovations to market"

Table 12: List and description of services of high societal value per project (additional expected 2019+ by projects active in 2018)

5.3 New Systems and Technologies (KPI II.17)

Projects have reported **204 new systems and technologies** developed during 2018. Many of them are reported already as part of the KPI "Significant innovations to Market". Others are additional technological components (not directly related to an exploitable value in the market).

Table 13 provides details per project.

Project	#	List and description systems and technologies developed 2018
AEGIS	1	AEGIS platform (described in Innovations to Market)
BigDataOcean	1	BigDataOcean Platform - A state-of-the-art big data analytic engine for managing cross-sectoral maritime data of differing velocity, variety and volume in a trusted way.
Data Pitch	18	Innovations to market will be realised by the 18 companies we funded in 2018

DataBio	42	We started with 91 software components, mainly from the technology providers in DataBio. Of these 42 were modelled including their interfaces and refined for use in the trials. 38 of these components were used in the 26 trials during 2018. Furthermore, 13 pipelines of software components were used in more than half (14) of the pilots. In addition, 6 data sets were improved and 16 new datasets created during 2018.
euBusinessGraph	5	Service for semantic multilingual annotation Relation Tracker Grafterizer 2.0 / DataGraft knowledge graph summaries with ABstraction and STATistics Assisted Semantic Interpretation and Annotation of tables
EW-Shopp	14	i. ABSTAT ii. ASIA iii. Big Data workflow tool iv. Weather data API v. Product Data Management Tools vi. Event and Weather Analytics Toolset vii. Keyword Clustering Tool viii. Pragmatic Interoperable Event Management ix. Campaign BOOSTER x. EW-Shopp DataFlow Methodology xi. PERFORMANCE INSIGTHS xii. Contact Center Workforce Optimization Manager xiii. Weather and Event-aware BI Sales Strategy Advisor xiv. Event and Weather Scout Dashboard
FashionBrain	5	1. Flair - A very simple framework for state-of-the-art Natural Language Processing (NLP) 2. MonetDB - The open-source MonetDB software suite wih FashionBrain extensions 3. System TASTY: High performing Named Entity Linkage System linking text to text data based on deep learning methods. Published at COLING 2016 and WWW 2018 4. SYSTEM IDEL: In-database Entity Linkage system bases on MonetDB (and is currently further extended towards Exasol RDBMS). The system executes machine learning in a database with Tensor Flow. It is published at IEEE BigComp 2019 and received a best paper award. 5. FEIDEGGER: A Multi-modal Corpus of Fashion Images and Descriptions in German
SLIPO	6	Answer embedded in the 'Innovations to Market' questionnaire; SLIPO Workbench, TripleGeo, LIMES, FAGI, DEER, SANSA
SODA	3	1) FRESCO MPC framework (see Innovations to Market), 2) MPC technology, 3) MPyC
SPECIAL	5	Resources The SPECIAL Usage Policy Language http://purl.org/specialprivacy/policylanguage The SPECIAL Vocabularies https://www.specialprivacy.eu/vocabs The SPECIAL Policy Log Vocabulary http://purl.org/specialprivacy/splog SPECIAL Ex-Post Compliance Checking demonstrator Demonstrates how usage policies together with event logs can be used to perform ex-post compliance checking https://www.specialprivacy.eu/platform SPECIAL Consent and Transparency Interfaces demonstrators Various consent user interfaces and the transparency dashboard https://www.specialprivacy.eu/platform
TT	11	All these systems and technologies are embedded in "Innovations to Market": - Process - Port and Terminal Productivity Cockpit - Crane and Spreader Sensorization Infrastructure - Service - Emission Reduction System - Simulation model - Traffic micro-simulation models in city centres - Model - Predictive High-Speed Network Maintenance - Predictive & automated diagnostic capabilities in the mainline rail industry - Dashboard to improve probability for trucks to arrive

		at planned time - Dashboard to predict car breakdown - Dashboard to predict and prevent failures in railways - Dashboard to measure cockpit efficiency and analyse truck turnaround time - MAID - Animal intrusion detection system
QROWD	3	QROWD Platform Back-end component that stores sensor streaming data (big data) and linked data. The i-Log (mobile app) requires it to collect data e.g. for the Modal Split calculation. Allows city staff to define and run crowdsourcing tasks and citizen engagement challenges with gamification. Dramatic improvement in engaging citizens and data contributors to solve city issues.
MHMD	1	Blockchain-based data exchange architecture and relevant smart contracts
BigDataGrapes	1	BigDataGrapes Platform as a Service as described in the "Innovations to Market" questionnaire
BigDataStack	3	DATA_SKIP LXS_DB GENERIC_CLEANING
BigMedilytics	8	• Multiparty Computation (MPC): a method/tool to analyse sensitive data from multiple parties (who do not wish to share data with each other) without revealing the data itself, but only revealing the outcome of the (data) analysis, i.e. the coefficients of the learned model. • A LASSO based regression that allows for prediction of comorbidities of hospital admissions out of > 4.000 features in a Health Insurance database. • GDM system • GDM mobile App • GDM medical portal • GDM analytics • Questionnaire • A lung-cancer pilot is under development what intends to contribute in the reduction of costs in hospitals, but also in a reduction of toxicities.
BODYPASS	3	1) Hyperledger 3D data exchange network; 2) Anonymization of face information in full body 3D scans; 3) 3D human body templates enabling 3D data aggregation from different source (Details in Innovation to Market)

Boost 4.0	24	• *Blockchain for Volvo supply chain line. • *Digital twin AGV • * IDSA connected AGV (Autonomous Guided Vehicle) • * IDSA connected Coordinated Measurement Machine (CMM) • * IDSA connected Milling Machine • * Blockchain applied in the supply chain of a manufacturing industry • * FDT • * FPT • * DSS-fusion • *UNINOVA Big Data Architecture • *UNINOVA Data Harmonization / Transformation IDS Big Data App • *ENEO Active Manufacturing Network Monitoring • * TRIMEK M3 Big Data 3D Point Cloud Analytics • * INNOVALIA AlTube Decision Support Platform • * SQS European IDS Certification Laboratory • FILL Cybernetics – IoT platform for FILL milling machine solutions • TTTECH Nerve Edge – Fog computing platform for Industry 4.0 applications • TID-ASTI 5TONIC – open experimentation facility for 5G and Industry 4.0 application connectivity performance and feasibility assessment. • Visual 4.0 Digital Twin simulation platform for internal logistics. • ESI Mineset Hybrid Twin Big Data Analytics Platform • SIEMENS Mindsphere Digital Factory IIoT Platform • One Assistant System in Production to support operators in their daily work with data analytics features. • i2Tracking - I2Tracking is a unique software developed by i2CAT. It is designed to provide seamless integration between positioning technologies like UWB, BLE, GPS... and RFID readings, matching each event to a specific location and determining the different states of the objects that are being moved around different areas providing massive data generation in real time. • Scalable industrial data: Scalable Industrial data is a distributed data solution that creates a modular pipeline of data from its acquisition to its external sharing. The data can be obtained at several points in the chain between the collection and the external distribution (sharing). Using this approach, a data consumer (application) can be seamless moved around the chain and placed closer to the data collection or closer to the data sharing, depending on the requirements.
CLASS	3	COMPSs, EXPRESS, Rotterdam
DataBench		Expected in 2019+
E2Data	5	1. Heterogeneous-aware Big Data framework (E2Data Framework) 2. Health Analytics 3. Natural Language Processing 4.Green Buildings 5. Biometric Security (all descriptions available in the "Significant Innovations to Market")
EDI	16	The solutions produced by the 16 companies in EXPERIMENT stage that took part in Call 1 of EDI are detailed in https://edincubator.eu/startups-2018/
FANDANGO	4	Multilingual text analytics; Spatio-temporal analytics; Copy-move detection on audio-visual content; Source credibility service.
I-BiDaaS	8	COMPSs (Programming Model and Runtime) (BSC), Qbeast (BSC), Test Data Fabrication (IBM), Advanced visualization and monitoring (AEGIS), Pool of ML algorithms based on structured (non)convex optimization (UNSPMF), Batch analytics module, Streaming analytics module, Advanced visualization module
ICARUS	5	Data Collection, Data Analytics & Visualisation, Data Security, Data Sharing, Resource Orchestration
TheyBuyForYou	1	PITO: An API feed of global procurement data, tender search and alert service, and contract pipeline, with all data in the Open Contracting Data Standard

TYPHON	8	See the short names in the innovations to market
Lynx		Expected 2019+
Cross-CPP		Expected 2019+

Table 13: List and description of systems and technologies delivered in 2018 per project

All these new systems and technologies will be promoted in the BDV CPPP Marketplace developed by the BDVe CSA project to spread the knowledge

6 SUPPORT MAJOR SECTORS AND MAJOR DOMAINS BY BIG DATA TECHNOLOGIES AND APPLICATIONS (II.13)

This KPI measures the number of sectors and major domains supported by Big Data technology and applications developed in cPPP projects. As described in KPI I.4 the BDV cPPP projects address a great variety of sectors[22], **20% of the projects are fully cross-sectorial** (their outcomes can be used in any sector or application domain) and **80% of the projects are working in more than 1 sector or application domain** (this explain why the total is superior to 100% in Table 14)**.** In particular the BDV cPPP projects address a great variety of sectors[23] as shown in Table 14.

Sector/application domain	Projects addressing this sector/application domain (% over active projects in 2018)	Innovations to market (delivered in 2018, % over the total)
Public services and Smart Cities	50%	20%
Transport, Mobility and logistics	43%	75%
Retail	37%	20%
Business Services	37%	20%
Health and healthcare	33%	8%
Manufacturing	23%	19%
Media	23%	0%
Finances and banking	23%	14%
Telecom	20%	13%
Energy	20%	0%
Bio-Economy: Agriculture, forestry and fishing	17%	14%
Water and natural resources	17%	0%

22 Grouped with a good level of alignment to the NACE registry These categories are part of the information in the BDV CPPP Marketplace that will be used for promoting all exploitable solutions coming out of the cPPP Projects (if needed new categories can be added)

23 Grouped with a good level of alignment to the NACE registry These categories are part of the information in the BDV CPPP Marketplace that will be used for promoting all exploitable solutions coming out of the cPPP Projects (if needed new categories can be added)

Earth Observation	13%	7%
Others	43%	27%

Table 14: Support major sectors and domains

Others (43% or the projects; this includes sectors such as insurance, public safety, personal security, public tenders, e-commerce, marketing, fashion industry, citizen engagement, ICT/Cloud services, social networks, procurement, legal domain, etc).

The BDV cPPP lighthouse projects[24] active in 2018 focused on the Bio economy (Agriculture, Fisheries, and Forestry) (**DataBio project**), Transport, mobility and logistics (**Transforming Transport project**), Health and Healthcare (**BigMedilytics** project) and Manufacturing (**BOOST4.0**), with a total of **4 major sectors supported** by Lighthouse projects and therefore widely supported with multiple use cases, scenarios and solutions.

Considering the whole project portfolio, the number of sectors supported is larger than **15**, with a solid distribution of use cases, experiments, solutions, and outreach activities among different sectors.

The BDV CPPP Marketplace as a supporting tool will help to further promote the marketable and exploitable results of the cPPP Projects to create a higher impact of the cPPP and contribute to its sustainability.

In addition to the projects, BDVA's Task Force 725 (application) bring together European stakeholders to to identify and act on the needs (technology, skills, etc.) of different industrial sectors as well as the areas of interest applicable to different industrial sectors such as language technologies, HPC, etc. During 2017 seven different application domains were represented in BDVA, organised in different subgroups.

In particular:

- TF7.SG2 Telecom (Lead: Robert Seidl, Nokia)
- TF7.SG3 Healthcare (Lead: Marija Despenic, Philips)
- TF7.SG4 Media (Lead: Mike Matton, VRT)
- TF7.SG5 Earth Observation & Geospatial (Lead: Florin Serban, Terrasigna)
- TF7.SG6 Smart Manufacturing Industry (Lead: David Dalle Carboarne, Engineering, and Sergio Gusmeroli, Engineering)
- TF7.SG7 Mobility and logistics (Lead: Vivian Kiousi, Intrasoft Intl.)
- TF7.SG8 Smart Cities and Public Services (Lead: Lanfranco Marasso, Engineering)
- TF7.SG9 AgriFood (Lead: Nuria de Lama, ATOS)

A specific report on the outcomes of the work done by these subgroups can be found in **Annex 5**.

24 Large-scale data-driven innovation and demonstration projects that aim at creating superior visibility, awareness and impact in specific relevant economic sectors

25 http://bdva.eu/task-force-7

7 EXPERIMENTATION (II.11, II.12, II.14)

7.1 II.11: Large Scale experiments conducted in cPPP projects and i-Spaces involving closed data

Projects reported **82 large scale experiments** (aggregated number of experiments conducted by all projects in 2018), **64** of which involve closed (private) data (78% of the total).

Projects also provided criteria to for experiments to qualify as "large-scale" with following criteria being mentioned:

- Large number of users involved with high TRLs
- Large geographical coverage
- Large number of involved actors and users
- High data value flowing cross-border
- A combination of data volume, data complexity and velocity.
- Complex datasets integrating data in some case from different countries
- Number of data sources
- Significant impact for society and/ or business

Table 15 provides details in the information provided by each of the projects contributing to this KPI.

Project	Number of large scale experiments (2018)	Experiments involving private data	Description of criteria to consider large scale
DataBio	26	26	In agriculture, 9 pilots combine 4 or more technology components to offer solutions which in 2018 reached TRL 7 or 8. Several of them are also run over large areas (thousands of hectares). The remaining 4 pilots use 3 or less technology components but they do use big data (e.g. in genomics) reaching also a TRL of 7 or 8 in 2018. In forestry, all 7 pilots are run over large areas (thousands of hectares) and/or involve thousands of users. In addition, they all reached TRL 7 or higher in 2018. In fishery, all 6 pilots are run over very large areas (5 million square kms each) and involve 5 or more partners each.
EW-Shopp	3	3	850 GB dataset enrichment. The dataset was moved across countries, decompressed, re-formatted, cleaned-up, transformed to a graph format and published in an ArangoDB database in minutes timespan. The workflow ran in a 10-node cluster of host machines that were connected via Ethernet.; widget on Ceneje site could potentially reach million users per month; national level datasets in different countries;

			several million keywords are being clustered.
FashionBrain	21	10	Large scale training of an AI language models, with huge multi-lingual text corpora Size: At least millions of values. Duration: more than 10 years. Users: thousands
SPECIAL	1	0	Performance and Scalability testing with simulated data
TT	13	13	A combination of data volume, data complexity and velocity. Number of users is also a criterium for large scale experiment in some cases.
QROWD	1	1	Size of dataset. We experimented with techniques for RDF-izing a large dataset put on the table by one of our industrial partners (thus, closed)
BigMedilytics	17	11	Large no. of data points, e.g. tracked entities can report location data every 1.5 seconds. Data of 5 million patients monitored over a 5-year period Using real world data (close to 4Ms) searching for clustering or comorbilities and their impact in health care burden and costs Number of potential subjects, 3 million of EHRs and a wide number of stakeholders, around 50 primary care physicians involved. The "big" component of the data in the lung-cancer pilot is related to the heterogeneity of the data. Thus, our experiments involve various data types, i.e. Electronic Health Records (that contain both structured and unstructured information), phone transcripts, unstructured public data (e.g. biomedical articles), and structured public data (e.g. drug-drug interaction databases), biomedical ontologies and a knowledge graph represented with semantic web technologies. Regarding the provenance of the data: we have the Electronic Health Records and phone transcripts that are situated in Spain, a knowledge graph in Germany and open-data processing in Greece.

Table 15: Number of large-scale experiments

BDVA i-Spaces reported **38 large scale experiments**, **28 of them involving close data**. In the same way than in the projects main criteria to qualify as large scale relate to size, impact and cross-border experimentation. 5 BDVA i-Spaces provided input to this KPI and in particular (**Table 16**):

BDVA i-Space	Number of large scale experiments (2018)	Experiments involving private data	Description of criteria to consider large scale
ICE RISE	10	10	cross-border, funding size, physical size, data size
Instituto Tecnológico de Aragón	2	2	European Impact
ITI	4	4	Impact for the company receiving our service, number of users.
CeADAR	16	8	More than 6 companies involved or EU projects or more than 50k in funding.
Big Data CoE	6	4	At Big Data CoE Barcelona we consider "large -scale" experiments those having an impact over an important segment of the population mainly for its social applications. We explain two examples. ATM experiment: ATM stands for Metropolitan Transport Authority. It is the agency that manages public transportation in Barcelona's metropolitan area. We have conducted an experiment involving 100 TB. The experiment consisted of computing the Origin-Destination matrix from Call Data Records and several other data sources in order to infer multi-modal mobility in Barcelona's metropolitan region. This experiment is relevant because it is cross-sector, implies large volume of data analysis involving large scale population (>2M) Deep-Lung experiment: This project/experiment aims at early detection of lung cancer nodules by analysing and classifying radiological images with DL (CNN) algorithms. More than 10 GB data involved-

Table 16: Number of large-scale experiments in BDVA i-Spaces

7.2 II.12: Uptake of BDV use cases and experiments

This KPI intends to assess the year over year increase of the number of Big Data Value use cases and experiments supported in cPPP projects and BDVA i-Spaces.

Projects reported **224 use cases or/and experiments** conducted during 2018 with a contribution from 18 different projects. **Table 17** provide details per project of the experiments and use cases developed during 2018. It is important to notice that the large-scale experiments reported in II.11 are normally a subset of this list (as normal rule).

BDVA i-Spaces have reported additional **165 experiments** with 5 i-Spaces contributing to this KPI. Details can be found in Table 18. In total the cPPP has developed **389 experiments.**

project	Number of data experiments/use cases 2018 (number)	Please provide some qualitative data to support your answer (i.e list, brief description,..)
AEGIS	3	The **AEGIS Automotive Demonstrator** explores how vehicle driving data and other road safety related data including e.g. weather data to name one concrete source can be meshed and modelled, aggregated, and semantically annotated in order to extract meaningful, safety-relevant information. For this, various combinations of vehicle driving datasets and datasets from other domains will be investigated to determine which of them provides the most valuable insights into driving styles and driving behaviour. Beneficiaries including drivers and other stakeholders will enhance their (business) value by using the AEGIS platform to create services for safer driving and safer roads. The automotive demonstrator is developed according to three different scenarios, broken road indicator, safe driving indicator, and regional driving safety risk estimator. These scenarios equal the demonstrator versions, i.e. scenario 1 broken road indicator equals demonstrator version 1, which was implemented in 2018. The **Smart Home and Assisted Living** (SHAL) Demonstrator has the objective to illustrate and implement a services bundle towards advanced monitoring and assisted living management, aiming to improve everyday living and enhance the wellbeing of people belonging to vulnerable groups. The overall demonstrator scenario in summary is the following. A social care service provider, for example a care centre for elderly individuals or a nursing home, desires to exploit big data-driven insights based on data collected from individuals and their surroundings, in order to provide added value services to vulnerable individuals. The services pertain proactive and reactive security and protection through smart notifications and personalised recommendations, as well as indoor comfort and quality preservation. Proactivity and reactivity of the aspired services aim at prolonging self-sufficiency and independence of the at-risk individuals, boosting safety, and facilitating informed decision making, either by the individuals themselves, or by their (in)formal carers. The **Insurance Demonstrator** will be developed in three different scenarios: • Personalised early warning services for asset protection • Financial impact, customer support and services • Marketing strategy and pricing support services The first and the second scenarios (early and medium demonstrator), arisen from the detection of an event that has already happened (close to real time notification) or is expected, will include the analysis of the financial impact for the company, the pricing strategy and support and services to the customer possibly involved. The first scenario is implemented in the reporting period. The third scenario will exploit the analysis related to the previous versions to better understand the HDI market: historical in-house datasets and open data (mainly from Public Italian Registers) will be analysed, providing to the Company models to enhance and ease the business planning and advertising campaigns.

project	Number of data experiments/use cases 2018 (number)	Please provide some qualitative data to support your answer (i.e list, brief description,..)
BigDataOcean	4	4 pilot cases of the project
Data Pitch	18	18 Pilots run with data driven startups working on resolving data related challenges set by large corporate and public sector organisations.
DataBio	26	DataBio runs 26 bioeconomy pilots: 13 in agriculture, 7 in forestry and 6 in fishery. Each pilot may count as one data experiment/use case, although in reality a pilot may execute experiments at more than one location and with more than one technical solutions. The 26 pilots utilize the DataBio platform, which encompasses 91 technology components. 13 major solution pipelines have emerged. The DataBio experiments use 14 existing datasets, have improved 6 datasets and have created 16 new datasets.
euBusinessGraph	5	Project business cases
EW-Shopp	4	• Testing predictability of consumer interest and sales for event-sensitive products (TVs and mobile phones) using historic web-store data enriched with Event Registry event data. Tests performed using EW-Shopp platform's analytics component tools and methodology. • Testing keywork clustering at large scale to identify the right-sizing of clusters in order to increase the scalability of prediction analytics. • Analysis of predictability of calls to a support call centers using a data set enriched with internal events information. Tests performed using EW-Shopp platform's analytics component tools and methodology. • Analysis of predictability of consumer interest using iot and public events data. The integration task was executed using historical data about shops located in 2 different locations. Events data was download from Facebook.
FashionBrain	42	10 different experiments to evaluate the efficiency and accuracy our incremental recovery technique 10 different experiments to evaluate the efficiency and accuracy of our RecovDB system (recovery using MonetDB) and compare it against existing technique 10 different experiments to evaluate the precision of our OpenCrowd technique to find influencers on social networks. 12 experiments for entity linkage and named entity recognition we utilize publicly available datasets, such as WebNLG, ConLL and SEMEVAL.
Transforming Transport (TT)	48	1. Smart highways pilot: Understand better the road traffic and mobility patterns 2. Smart highways pilot: Optimize highway operations 3. Smart highways pilot: Guarantee safer roads and make a better use of these roads 4. Connected cars pilot: Breakdown estimator system.

project	Number of data experiments/use cases 2018 (number)	Please provide some qualitative data to support your answer (i.e list, brief description,..)
		5. Connected cars pilot: Emission reduction system. 6. Connected cars pilot: Traffic jam detector system. 7. Connected trucks pilot: Using Big Data approach on traffic and movement data specifically related to truck fleets. 8. Connected trucks pilot: Generation of truck specific patterns. 9. Connected trucks pilot: Provision of truck specific traffic information (actual and predictive). 10. Connected trucks pilot: Event related information on specific POIs (ramps, ports, toll stations). 11. Connected trucks pilot: Analyses on infrastructure, events, lane-specific traffic conditions for trucks (satellite images). 12. Rail Initial pilot use case 1: Overhead line equipment (OLE) 13. Rail Initial pilot use case 2: Point Machines 14. Rail Initial pilot use case 3: Track Circuits 15. Rail Initial pilot use case 4: Train Track Interface 16. Rail Replication pilot Use Case 1: Prediction of the degradation of point machines 17. Rail Replication pilot Use Case 2: Prediction of track profile degradation 18. Rail Replication Pilot Use Case 3: Use predictive models to optimise railway operation in the Rail Traffic Management System 19. Valencia port pilot: Yard Crane Scheduling Optimization 20. Valencia port pilot: Predictive Maintenance 21. Valencia port pilot: Predictive Web Cockpit 22. DUISPORT pilot: Terminal Productivity Cockpit 23. DUISPORT pilot: Predictive Maintenance System 24. Athens airport pilot: Operation Management Predictive Optimization 25. Athens airport pilot: Descriptive passenger behaviour system 26. Athens airport pilot: Retail model and the consumption ratio per passenger at a flight level 27. Malpensa airport pilot: Improve allocation of STA based on historical operation data analysis

project	Number of data experiments/use cases 2018 (number)	Please provide some qualitative data to support your answer (i.e list, brief description,..)
		28. Malpensa airport pilot: Track and predict Aircraft taxi in and out times 29. Malpensa airport pilot: Track and predict boarding process 30. Malpensa airport pilot: Track and predict Aircraft turn-around timing process 31. Malpensa airport pilot: Track and predict Aircraft sequence before departure 32. Malpensa airport pilot: Feed turnaround prediction data into A-CDM platform 33. Malpensa airport pilot: Feed enhanced ETA into Airport A-CDM platform based on historical operation data analysis 34. Malpensa airport pilot: Monitor performance of the Airport A-CDM platform estimations 35. Malpensa airport pilot: Coordinate prioritization with ATC 36. Malpensa airport pilot: Assess and predict time for passenger to reach the gate 37. Malpensa airport pilot: Track and predict Aircraft de-icing time and sequence before departure 38. Valladolid urban mobility pilot: Generating a traffic model for particular areas in the city where freight transport has more impact. 39. Valladolid urban mobility pilot: Analyse different freight delivery scenarios and make best decision according to the results. 40. Valladolid urban mobility pilot: Creation of a planning tool for delivery fleets and to inform drivers about the optimal route plan 41. Tampere urban mobility pilot: Provision of tools for urban traffic management for diagnosis of traffic status and for selection of alternative solutions for mitigating the impact of roadworks and other events. Addition of new data sources for improved situational awareness, such as social media and traffic cameras. 42. Tampere urban mobility pilot: Provision of tools for drivers and travellers regarding traffic status. 43. Tampere urban mobility pilot: Provision of tools to improve the access of goods delivery vehicles to parking places. 44. Dynamic supply networks pilot: Identification of delivery patterns and problematic issues and forecasting 45. Dynamic supply networks pilot: End to end information sharing 46. Dynamic supply networks pilot: Shared micro-hubs (in terms of space) where customers collect their online order. 47. Dynamic supply networks pilot: Inventory routing in an omnichannel environment 48. Dynamic supply networks pilot: Online consumer insights regarding logistics and delivery processes

project	Number of data experiments/use cases 2018 (number)	Please provide some qualitative data to support your answer (i.e list, brief description,..)
QROWD	7	2 iterations of an experiment regarding Use Case 1 (Modal Split). This includes the integration of three components 2 iterations of experiment regarding collection of mobility data infrastructure, this includes integration of three components 1 experiment regarding dynamic data integration 1 experiment regarding Linked Data Generation Framework (same as large scalr) 1 experiment regarding crowdsourcing monitoring tools
SLIPO	10	Real-world data integration use cases applied by the SLIPO industrial partners and their clients
SPECIAL	3	Consent interface usability testing Transparency dashboard usability testing Compliance backend performance testing
BigDataStack	4	Data skipping experiments CEP experiments ATOS experiments Danaos experiments
BigMedilytics	15	Population Health; Prostate cancer; Stroke workflows; Sepsis workflows; Asset management Analysis of EHRs for detecting comorbidities clusters: this experiment will find using ML techniques how a combination of several relevant diseases could led to a higher risk of hospitalization and mortality. initial pilot experiments to prepare for evaluation campaign assessing value of search in radiologist's workflow. We deployed the prototype of the search engine at the radiology department of MUW/AKH and conducted initial trials to finalize deployment and understand the use case better. Patient characteristics with the reason and frequency of teleconsultation, which is a service established by the HUMPH hospital in Madrid with a view of reducing the unscheduled hospital visits. After developing a framework defining the relevant clinical items, we have deployed an infrastructure to run the image analysis Prepared data usage access and data selection with Curie and IBM for breast cancer pilot.
Boost 4.0	11	TRIAL 1 - VOLKSWAGEN INJECTION MOULDING PLANT: VW will introduce big data analysis in the process to design the moulds for the casting process. TRIAL 2 – FILL: Will work to improve their engineering process with machine data obtained during the operation phase into the design of new machines.

project	Number of data experiments/use cases 2018 (number)	Please provide some qualitative data to support your answer (i.e list, brief description,..)
		TRIAL 3 – VOLKSWAGEN AUTOEUROPA: the main goal is to fully integrate the matiral flow, from the reception at the unloading dock to the point of fit. TRIAL 4 – GF: the ambition is to exploit the opportunities provided by BOOST 4.0 enablers for a zero-defect factory for milling spindles for GF machines. TRIAL 5 – FIAT – Connect automated guided vehicles (AGV) with the laser machine in order to further automize the production process. TRIAL 6 – PHILIPS CONSUMER LINE – Improve the control system for the moulding process providing more data to the operator and using data analysis to predict and prevent the defective parts production. TRIAL 7a – GESTAMP – Improve the navigation information of the AGV to optimize the logistics in the warehouse TRIAL 7b – TRIMEK – Develop a new high-resolution metrology and visualization system to achieve zero-defects manufacturing. TRIAL 8 – VOLVO TRUCKS CORPORATION – Get a real-time tracking of cabs from the plant in Umea to the plant in TUVE, allowing to prevent delays and improve the planning process. TRIAL 9 – WHIRLPOOL – Update their forecasting tool for spare parts distribution and manufacturing incorporating big data analysis tools to process and predict the needs based on the data produce across all the product lifecycle. TRIAL 10 – BENTELER – Develop a machine health monitoring system that will allow to improve the predictive maintenance plan and prevent the defective production.
EDI	16	https://edincubator.eu/challenges-2018/, each data challenge is documented in the page of companies who took part in EXPERIMENT stage
FANDANGO	3	In order to validate and test the project results three main application areas (use cases) have been analysed and selected, these are Climate Change, Migration and European Policies. All of these three use cases present a large amount of Fake News, but a better management and organisation of Data can deeply support journalist together with a advanced use of ML techniques. For this reason, these application domains have been selected.

project	Number of data experiments/use cases 2018 (number)	Please provide some qualitative data to support your answer (i.e list, brief description,..)
I-BiDaaS	1	Use Case: Analysis of relationships through IP address Dataset: IP address synthetic data Partner Owner: CAIXA Bank For the use case of I-BiDaaS MVP, the "Analysis of relationships through IP address" from CAIXA is implemented. The main objective of this use case is to validate and test the I-BiDaaS architecture and tools and verify if the generated synthetic data can provide the same or additional insights. Moreover, it also investigates if the amount of relationships detected in the real data by CAIXA can be also detected by using synthetic data. CAIXA stores information about their customers for security and fraud prevention reasons, verifying if the patterns of the customers interactions with the bank are legit or not (e.g. avoiding fraudulent bank transfers). CAIXA also collects and analyses the information from the operations that customers perform (bank transfer, check their accounts, etc.) using channels such as mobile apps or online banking. That allowed CAIXA to find new kind of relationships between customers using their connection information (e.g. IP address).
TheyBuyForYou	4	Iterations of ingestion of procurement data.

Table 17: Experiments performed by projects during 2018

BDVA i-Space	Number of data experiments/use cases 2018 (number)	Please provide some qualitative data to support your answer (i.e list, brief description,..)
RISE	29	SENDATE EXTEND (a CELTIC+ European project) on holistic automation using 7TB of IT/environmental/equipment data from 2 years of operation (have longer list but cannot paste here)
Cineca	46	23 projects in the IND4.0 domain involving simulation or predictive maintenance / data analytics; 17 projects in the bioinformatics domain involving NGS data processing; 6 projects in other domains (insurance, oil&gas, media, cultural heritage, transport)
Instituto Tecnológico de Aragón	2	
ITI - Instituto Tecnológico de Informática	16	Recruiting problems, Robotic Weld quality in industry, Quality of data in ceramics, automatic labelling / classification in health processes, Ships movement prediction, Digital Twins in industry, political trends.
CeADAR	44	Demonstrators, EU projects, translational research projects, work ready graduate projects, internship projects, research projects with collaborators in both of the two universities where CeADAR is present
BigData CoE	28	This figure corresponds to the number of projects / experiments conducted by the Big Data CoE Barcelona mainly involving private companies willing to adopt big data analytics to their businesses processes. Conducted experiments have multiple purposes, e.g. using data analytics from wearables to determine biometric characteristics, or analysing licence plates data bases of vehicles visiting a town to extract insights on new commercial opportunities for that particular town.

Table 18: Experiments performed by BDVA i-Spaces

7.3 II.14: Amount of data that has been made available for experimentation (cPPP projects and i-Spaces)

This KPI measures the number of Exabytes of data made available in cPPP projects and i-Spaces (including closed data) (access to data for experimentation).

This includes data made available for experimentation (e.g. Data incubators and i-Spaces) and data made available internally in the projects for experimentation purposes (e.g. Lighthouses and other projects).

13 projects provided data for this KPI (Aegis, BigDataOcean, DataBio, euBusinessGraph, EW-Shopp, TT, QROWD, BigDataStack, BigMedilytics, Boost 4.0, CLASS, EDI, TheBuyForYou), reporting a total of 0,08625 Exabytes (**86,25 Petabytes**) for 2018.

Table 19 offers details per project about the amount and type of data used (source, providers, etc). It is important to notice that some of the projects are not only providing internal access to diverse data sets from different sources, but also are improving and creating new valuable datasets (e.g of DataBio project). DataPitch incubator reported that data shared directly between data provider and Startups/SME – and they expect to use the Impact Assessment to determine amount of data shared as a result of the project (to be reported in 2019). Additional projects reported experimentation started in 2019 and therefore they don't contribute to this KPI in 2018.

Project	Amount of **data made available in** the **project in 2018 (including closed/private data)**	**Additional information about the data (source, providers, etc)**
AEGIS	~100 Terabytes	Some examples of data types: automotive sensor data collected by VIF smart home sensor data collected by Hypertech and Konkat private data of the insurance company HDIA weather data social media data open data from European Data Portal.
BigDataOcean	100 Terabytes	AIS data, vessel machinery data, oceanographic data (e.g. data from Copernicus), weather data etc.
DataBio	350 Terabytes	**The DataBio pilots use 14 existing datasets, have improved 6 datasets and have created 16 new datasets**. In agriculture the data include 19 types, with initial volume in the order of 50 TB and velocity in the order of 200 TB/year. In forestry the data include 12 types, with initial volume in the order of 10 TB and velocity in the order of 10 TB/year. In fishery the data include 12 types with initial volume in the order of 10 TB and velocity in the order of 5 TB/year. Note that the PB-scale volumes and velocities of the available and relevant Earth Observation data (e.g. from Sentinel and LandSat) Copernicus data are not included in the above figures.
euBusinessGraph	In the order of Gb.	Company-related data from business case providers.
EW-Shopp	85 Tetabytes + 4 Terabytes from other datasets	ECMWF provided the project access to their weather forecasts (85PT) in GRIB format; EventRegistry made available; JOT has provided data about the performance of digital marketing campaigns of about 3 TBs (but more data for different countries that are not considered in the project would be available) as tabular data; the project has access to the EventRegistry, semantically annotated news sources accounting for about 1TB as JSON data.
TT	156 Terabytes	According to the data received from the pilots, 0.000051393 exabytes (51.39TB) are currently being used or produced and a total amount of 0.00015595 exabytes (155.95TB) have been used during the project. This amount of data is expected to grow as the project progresses. **The description of all the datasets used by the project is publicly available in the following link: http://data.transformingtransport.eu/**

QROWD	500 Terabytes	Geospatial data provided by TomTom (closed) Floating car positions for Trento provided by TomTom (closed) Heterogeneous datasets provided by the Municipality of Trento (open) Sensor data collected from participants in the Trento Use Case (closed, but some of it will be anonymised and opened)
BigDataStack	100 Gigabytes	The Danaos use case provided a few GB of data but we also acquired related whether data which was about 100 GB. In addition, we have 5 GB of ATOS Worldline data
BigMedilytics	>5 Terabytes	The main data source are structured text files describing features from EHRs. These files have been exported from the current Healthcare Information Systems of the Valencian Region. EHR data of more than 5 million patients Some pilots are not about data volume, but focus more on using complex data Imaging data (computed tomography) and connected information (report information) Processing biomedical articles with Natural Language techniques resulted in a 6.5 Gbytes file that is stored in a graph database (Neo4j) Medical images in multiple modalities supported by hospital record data from hundreds of patients.
Boost 4.0	40 Terabytes	The data is mainly provided by factory environments (design data, simulation data, process data, machine data, asset tracking data). The data sets made available so far are focused on the Proof of Concept (PoC) phase and 2019 will extend the use to larger data sets that will scale up the pilots.
CLASS	400 Gigabytes	Extraction of objects and information associated to them (type: pedestrians, vehicles; position; velocity) from images captures by four traffic cameras located in the city of Modena
EDI	86 Gigabytes	From the 15 data providers that are in the consortium, data from 9 Data providers was finally used in the EXPERIMENTATION of EDI. Challenges associated to those 9 Data Providers was supplied
TheyBuyForYou	4 Terabytes	TBFY incorporates procurement data from the EU TED portal with a growth rate of ca 30MB/day, in addition to ca 50 national and regional procurement portals of varying growth rate.

Table 19: Data made available in the projects in 2018 – details per project

Additionally, the BDVA labelled i-Spaces also contribute to this KPI (further information here: http://www.bdva.eu/node/1172). 5 i-Spaces contributed to this KPI reporting additional **20,71 Petabytes** of data for experimentation. In particular:

i-Spaces	**Amount of** data made available in **the** i-Space in 2018 (including closed/private data)	Additional information about the data (source, providers, etc)
ICE RISE	20 Petabytes	The largest size of data is provided from the large-scale test datacentre including IT equipment data i.e server fan power, software platform data i.e open stack instances, environmental sensor data i.e temperature sensors, facility equipment data i.e cooling equipment power
Cineca	280 Terabytes	Cineca has Petabytes of simulated data, but in terms of real data, 277 TB of genomic data coming from NGS and 3 TB of data from the monitoring of the data center (historical data on energy consumption and usage of the resources)
ITI - Instituto Tecnológico de Informática	200 Terabytes	Billing information (millions of clients), Social Networks comments, ...
CeADAR	30 - 100 Terabytes	This include open data and closed data both structured (different nature and format) and unstructured (images, text in 2018) from a variety of verticals, including banking, manufacturing, energy, logistics, electronics supply chain, transportation, health, farming, ...
BigData CoE	130 Terabytes	Most of the data are Call Detail Records from mobile phone operators from one side and medical (radiological imaging and genetic data)

Table 20: Data made available in BDVA i-Spaces

Combining reported information from projects and i-Spaces (members of BDVA) the number of Exabytes of data made available in cPPP projects and i-Spaces is **0,10696 Exabytes** (106,96 Petabytes).

8 SRIA UPDATE AND IMPLEMENTATION (II.9, II.10)

8.1 KPI II.9: Ensure efficiency, transparency and openness of the cPPP's consultation process

The Big Data Value Association is responsible for providing regular (yearly) updates of the BDV SRIA defining and monitoring the priorities as well as metrics of the cPPP. Main foundations of the SRIA update process are:

- **SRIA is a consensus reflecting document**: The update process of the SRIA is a community-driven process
- **It's a data-driven Process**: Survey results gives us the quantitative rationale for changing or keeping the SRIA
- **SRIA document is domain agnostic**: Any technical or non-technical priorities are argued without reference to particular domains
- **Descriptive of the Problem**: SRIA document describes and motivates the problem but does NOT provide details about possible solutions

At the end of 2017 BDVA released the BDV CPPP SRIA v4.0 (detailed process and results reported in the Monitoring Report 2017). This version was the basis to support the H2020 LEIT ICT WP2018-20. During 2018 a minor update towards a version 4.1 was launched in the community, however this did not crystallise in a new full version of the SRIA but individual deliverables in the format of vision, position or discussion papers that are supporting the transition towards next framework programme and the creation of a new Strategic Agenda and roadmap.

Papers released during 2018 (or launched during 2019) basis of a new Strategic agenda and roadmap for the future of Big Data Value and Data-Driven AI are the following:

- Future Challenges for European Leadership in the Global Data Economy and Data-Driven Society: Input to Framework Programme 9 (March 2018)[26]
- Data-Driven Artificial Intelligence for European Economic competitiveness and Societal Progress (November 2018)[27]
- Big Data Challenges in Smart Manufacturing v1 (March 2018)[28]
- The Technology stacks of high-performance computing and Big Data computing: What we can learn from each other (Oct 2018)[29]
- Joint Vision Paper for an Artificial Intelligence Public Private Partnership (AI PPP) (March 2019)[30]
- TOWARDS A EUROPEAN DATA SHARING SPACE: Enabling data exchange and unlocking AI potential (April 2019)[31]
- Strategic Research, Innovation and Deployment Agenda for an AI PPP: A focal point for collaboration on Artificial Intelligence, Data and Robotics. Consultation Release. (June 2019)[32]

General process:

As part of the 2015 update the BDVA defined a general update process that was implemented in 2015 and used in 2016 and 2017. Incorporating the feedback of the 2015 update process, the BDVA modified in 2016 slightly the series of workshops and interactions with BDVA community in the update process to ensure maximal opportunities for efficient interaction and to be in synch with other BDVA meeting and events. Figure 4 gives an overview of the various elements of the general process (with dates for 2016). This well-defined process ensures transparency as requested in KPI II-9.

The process for SRIA v4 update (2017) was also slightly modified due to the short timeframe and the nature of the update that required major interaction and discussion, and it was not focus on prioritisation for the roadmap but the adoption of advances produced by the Task Forces and introduction of cross-domain topics. For this reason update to SRIAv4 did not incorporate a community survey as such. **Table 21:** Overview of activities conducted in 2018 supporting road-mapping provides details of the timeframes and activities for the update of SRIA v4.

26 http://www.bdva.eu/sites/default/files/BDVA%20position%20to%20Fp9_v1.pdf
27 http://www.bdva.eu/sites/default/files/AI-Position-Statement-BDVA-Final-12112018.pdf
28 http://www.bdva.eu/sites/default/files/BDVA_SMI_Discussion_Paper_Web_Version.pdf
29 http://www.bdva.eu/sites/default/files/bigdata_and_hpc_FINAL_16Nov18.pdf
30 http://www.bdva.eu/sites/default/files/VISION%20AI-PPP%20euRobotics-BDVA-Final.pdf
31 http://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpace%20PositionPaper_April2019_V1.pdf
32 http://www.bdva.eu/sites/default/files/AI%20PPP%20SRIDA-Consultation%20Version-June%202019%20-%20Online%20version.pdf

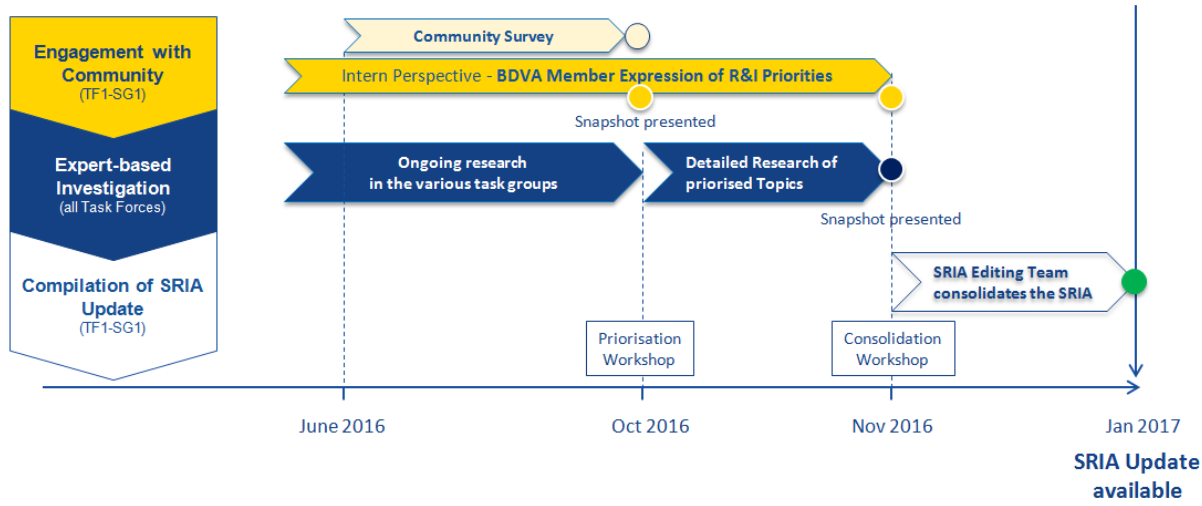


Figure 4: Overview of the annual BDV SRIA update process and its implementation (and details for 2016, same process applies to all versions)

The updated process during 2017 was based on 3 major streams: Engagement with the Community, Expert-based investigation, Compilation of the BDV SRIA Update.

Engagement with the community: Within the update process, the BDVA engages with the a) BDVA members as well as b) the wider community to ensure a comprehensive perspective concerning the technical and business impact of the BDV SRIA technical and non-technical priorities as well as to identify emerging priorities with high potential impact. This was done through face-to-face internal and external workshops, online workshops, and online input and reviews.

Expert-based Investigations: The BDVA has established task groups for all technical and non-technical priorities as well as the task groups for lighthouse projects and I-Spaces. The task groups are continuously working in order to produce related state-of-the-art analysis or working papers.

Compilation of the SRIA Update: In accordance to the outcome of the community engagement process, the particular task groups are consulted to discuss the scope and details of the BDV SRIA update. This part of the update process is introduced and initiated in a dedicated priorisation workshop. The objective of the SRIA priorisation workshop is establish a common understanding about the SRIA update process and how to submit SRIA change requests. In addition, the workshops is an important foundation for the BDVA member to get to know each other in order to establish the basis for further collaboration. At the workshop, the overall SRIA update process has been explained and discussed as well as responsibilities defined. In addition, the task forces could already discuss potential input as well as overlaps in technical priorities.

Based on the ongoing state-of-the-research and by reflecting the quantitative and qualitative feedback of the community engagement process, the BDVA task forces consolidates after the priorisation workshop concrete changes for the SRIA document.

The process described above (for the BDV CPPP SRIA development until v4.0) has been adopted as main process to follow by all the documents released by BDVA contributing to the future roadmaps. Therefore, to calculate this KPI we will use the number of contributions/participants to all the papers described above during 2018.

Table 21 lists events conducted for the update to support road-mapping activities in 2018 (to support the production of the documents listed above)[33], and includes at the end accumulated values from previous exercises. The figures show the number of participants/contributions upon which KPI II.9 is based.

In total there were at least 12 events during 2018 that contributed to input in the BDVA Strategic papers, multiple online meetings with a total of **2085** participants/contributions. This is in alignment with previous

33 List is not exhaustive as more meetings (Physical and online) were put in place to support the road-mapping activities

numbers provided for 2017 (1592), 2016 (1437) and earlier numbers (to 875 participants/contributions that were relevant for the previous version of the BDV SRIA and 433 participants/contributions for the initial version in 2014). Increase reveals increase of participant members and engagement with external communities.

In total since the launched of the BDV cPPP we can count **6422** potential contributions to the Strategic road-mapping activities.

Event	date	Participation/Contribution	Comments
Activity Group workshop 24	01-Feb	55	
Activity Group workshop 25	15-Mar	50	
Activity Group workshop 26	16-May	50	
Activity Group workshop 27	20-Jun	50	
Activity Group workshop 28	12-Sep	50	
Activity Group workshop 29	10-Oct	50	
European Big Data Value Forum 2018, Vienna	12-14 Nov	600	
Big Data Value Meet-up Sofia	14-16 May	250	
Data4AI workshop	18-Sep	50	
Workshop DIHs WG on Big Data and AI	07-Nov	150	
IoT Week 2018 workshops. (specific workshops organised by BDVA	4-7 June	100	
IoT Open Data Access to advance mobility in European cities Workshop (organised with Ertico and AIOTI)	04-Sep	70	
BDVA Task Forces	Ongoing	560	E-mail distribution lists (only Activity Group considered)
Total road-mapping activities 2018		**2085**	
Total SRIA V1 - V4		4337	
Total since the launched of the PPP		**6422**	

Table 21: Overview of activities conducted in 2018 supporting road-mapping

8.2 KPI II.10: Ensure that technology progress is in line with multi-annual roadmap of SRIA

To determine how well the SRIA technical priorities and challenges are covered by ongoing research and innovation activities of the running BDV CPPP projects, the BDVA (supported by BDVe) performed a systematic collection of data. To this end, a common template was shared among the BDV CPPP projects, to collect information about the contributions to the different technical areas of the SRIA.

At the end of the reporting period, 38 projects have provided the detailed mapping of their contributions and technologies to the SRIA technical priorities and challenges.

SRIA Technical Priorities

The BDV SRIA[34] defines the overall goals, main technical and non-technical priorities, and a research and innovation roadmap for the European Public Private Partnership (PPP) on Big Data Value. The SRIA has been developed by the Big Data Value Association (BDVA), an industry-led organisation representing large businesses, small and medium-sized enterprises (SMEs), and research organisations in Europe.

Five *technical priority* areas have been identified for research and innovation:

Data Management as a foundation for higher-level big data techniques;

Optimized **Data Processing Architectures** for data-at-rest and data-in-motion;

Data Analytics to improve the understanding of data;

Mechanisms ensuring **Data Protection** and anonymization, to enable the vast amounts of data which are not (and never can be) open data to be incorporated into the data value chain;

Advanced **Data Visualization and User Interaction**.

Results and Analysis of SRIA Coverage

Based on the data reported by the BDV CPPP projects, Table 22 shows the consolidated and aggregated results. The table depicts the number of technical contributions and big data technologies per SRIA technical priority and challenge, both as totals (red column) and per different phases / years of the PPP (green columns). As part of SRIA Version 4.0, four new technical challenges were introduced and are included in the table. As a result, the 2017 column does not report coverage of these challenges, as they were not part of the previous SRIA.

As can be seen on a high-level of analysis (red column in the table), each of the SRIA priorities is covered, as are all technical challenges.

The major focus of technical contributions lies in the "Data Analytics" priority, followed with some distance by "Data Processing Architectures" and "Data Management". This is a significant change in from MR2017, where "Data Management" was the top priority. As already anticipated as part of the analysis in MR2017, a clear trend to focus on technical contributions in the areas of "Data Analytics" and "Data Processing Architectures" was anticipated, thus supporting our explanation that a solid base of "Data Management" solutions twill enable analytics and processing innovations.

34 S. Zillner, E. Curry, A. Metzger, R. Seidl (Eds.), "European big data value strategic research and innovation agenda (SRIA)," Version 4.0, November, 2017; http://www.bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf

	Number of contributions/technologies				
Technical priorities and challenges	TOTAL	Delivered in 2017	Delivered in 2018	Ongoing (yet not delivered)	Planned (for 2019+)
Priority "Data Management"	**424**	**44**	**122**	**90**	**168**
Semantic Annotation of unstructured and semi- structured data	92	9	31	17	35
Semantic interoperability	51	11	13	9	18
Data quality	63	7	15	19	22
Data Management lifecycle and data governance	48	9	11	12	16
Integration of data and business processes	101	3	44	12	42
Data-as-a service	40	4	4	13	19
Distributed trust infrastructures for data management	29	1	4	8	16
Other (specify)	0	0	0	0	0
Priority: "Data Processing Architectures"	**431**	**9**	**126**	**118**	**178**
Heterogeneity	111	4	33	34	40
Scalability	101	4	28	30	39
Processing of data-in-motion and data-at-rest	55	0	19	15	21
Decentralizatrion	44	0	12	10	22
Performance	68	1	18	19	30
Novel architectures for enabling new types of big data workloads -- new in SRIA4.0	36		13	7	16
Introduction of new hardware capabilities -- new in SRIA4.0	7		1	0	6
Other (specify)	9	0	2	3	4
Priority: "Data Analytics"	**540**	**11**	**183**	**114**	**232**
Semantic and knowledge-based analysis	91	5	21	25	40
Content validation	35	2	6	11	16
Analytics frameworks & processing	96	4	25	26	41
Advanced business analytics and intelligence	106	0	53	8	45
Predictive and prescriptive analytics	120	0	53	21	46
High Performance Data Analytics (HPDA) -- new in SRIA 4.0	33		13	4	16
Data analytics and Artificial Intelligence -- new in SRIA 4.0	56		12	19	25
Other (specify)	3	0	0	0	3
Priority: "Data Protection"	**110**	**15**	**33**	**26**	**36**
Generic and easy to use data protection approaches	34	0	9	10	15
Robust Data privacy ((incl. multi-party computation)	52	1	21	14	16
Risk based approaches	8	0	3	2	3
Other (specify)	16	14	0	0	2
Priority: "Data Visualisation and User Interaction"	**184**	**5**	**56**	**48**	**75**
Visual data discovery	49	3	10	14	22
Interactive visual analytics of multiple scale data	43	0	10	11	22
Collaborative, intuitive and interactive visual interfaces	62	2	31	12	17
Interactive visual data exploration and querying in a multi-device context	28	0	5	10	13
Other (specify)	2	0	0	1	1
Other (specify)	**0**	**0**	**0**	**0**	**0**

Table 22. Coverage of SRIA Technical Priorities and Challenges (Number of Contributions/Technologies)

Achievement of Specific KPI:

For what concerns SRIA coverage, the relevant specific KPI is "Ensure that technology progress is in line with multi-annual roadmap of SRIA" (KPI II-10).

As agreed by the respective BDVA task force, this KPIs is measured as "% of research priorities covered compared to overall scope of research priorities defined in SRIA".

Based on the above table, the KPI is quantified as follows:

- Delivered in 2018: 100% (which is as reported as planned in MR2017)
- Ongoing: 100%
- Planned for 2018+: 100%

Non-technical priorities:

In addition to the technical priorities' implementation is important to assess the non-technical aspects of the SRIA and assess status of implementation.

- **Skills**, is fully covered by the BDVe project and BDVA TF9 and complemented by many activities and contributions from most of the projects as detailed in section 2 of this annex.
- **Societal and Ethical** aspects of Big Data are covered by e-Sides project and BDVA TF5.
- **Policy** aspects are mainly covered by BDVA TF5.
- **Business** in particular regarding the support to SMEs, it is covered by BDVA TF8 and BDVe project and it has been reported under section 3 of this annex.

9 TECHNICAL PROJECTS (II.4, II.15, II.16)

9.1 KPI II.4: Privacy and Security KPI

The analysis of KPI II.4 "Enable advanced privacy and security respecting solutions for data access, processing and analysis" is based on two sources of information. On the one hand, we used the reports from projects about the coverage of the SRIA technical priorities (see section 8 of this report). On the other hand, we collected from the projects specific information on three specific sub-KPIs:

- II.4a) Number of patents filed by cPPP projects that enable advanced privacy and security
- II.4b) Number of publications by cPPP projects that describe advanced privacy and security
- II.4c) Number of OSS contributions / products / SW components resulting from cPPP projects that enable advanced privacy and security

Based on the data reported by the BDV CPPP projects. Table 23 shows the consolidated and aggregated results for the three specific sub-KPIs.

The table also shows results reported in 2017 to show the evolution (2017 was the first year of the PPP projects)

II.4 a) Number of patents filed that enable advanced privacy and security respecting solutions for data access, processing and analysis			II.4 b) Number of publications that describe advanced privacy and security respecting solutions for data access, processing and analysis			II.4 c) Number of OSS contributions / SW / Products that enable advanced privacy and security respecting solutions for data access, processing and analysis		
Year 2017	**Year 2018**	2019+ (planned)	Year 2017	**Year 2018**	2019+ (planned)	Year 2017	**Year 2018**	2019+ (planned)
3	**2**	7	23	**61**	79	**7**	**24**	44

Table 23. Sub-KPIs for KPI II.4 "Enable advanced privacy and security respecting solutions for data access, processing and analysis"

From the PPP projects active during 2018, 3 technical projects are fully dedicated to advance the state of the art in privacy-preserving Big Data technologies: SPECIAL, SODA and MHMD. Specific contributions from SPECIAL, SODA and MHMD can be found in **Table 24**.

	II.4 a) Number of patents filed that enable advanced privacy and security respecting solutions for data access, processing and analysis		II.4 b) Number of publications that describe advanced privacy and security respecting solutions for data access, processing and analysis		II.4 c) Number of OSS contributions / SW / Products that enable advanced privacy and security respecting solutions for data access, processing and analysis	
Project	**Year 2018**	**2019+**	**Year 2018**	**2019+**	**Year 2018**	**2019+**
SODA	0	1	17	10	3	5
SPECIAL	0	1	13	8	3	3
MHMD	1	1	5	5	0	1

Table 24. Sub-KPIs for KPI II.4 "Enable advanced privacy and security respecting solutions for data access, processing and analysis" for ICT-18-2016 projects (RIA)

However many other projects contribute to these KPIs and more expected to do it during 2019 and 2020 (as H2020-ICT-13-2018-19 call for proposals was focused on scaling privacy perseving technologies and close to deployment industrial and data platforms). (Note: values indicated as 2019+ are only considering input from projects running in 2018, so projects from calls 2016 and 2017).

2 projects have reported patents filed that enable advanced privacy and security respecting solutions for data access, processing and analysis: MHMD and BigMedilitycs (both in the Health/Healthcare sector).

The amount of publications that describe advanced privacy and security respecting solutions for data access, processing and analysis has doubled with respect 2017 and the 3 projects focused in this area are the major contributors, with more than half of the publications being part of this projects. Other projects such as BigDataGrapes are strongly contributin in this area.

Finally a large list of projects have contributed delivering OSS, software or products that enable advanced privacy and security respecting solutions for data access, processing and analysis during 2018. In particular AEGIS, EW-Shopp, SPECIAL, SODA, BigDataGrapes, BigMedilytics, Boost 4.0, I-BiDaaS and TheyBuyForYou. Some examples include:

- AEGIS: the AEGIS platform
- EW-Shopp: ASIA - Weather data API - Event and Weather Analytics Toolset
- SPECIAL backend and frontend consent and transparency components
- BOOST: Hyperledger support to IDS & data usage control in IDS connector
- I-BiDaaS: Test Data Fabrication - IBM. Platform for generating user modelled realistic synthetic data for testing and development.
- TheyBuyForYou ontology, data processing pipeline, Procurement data APIs

9.2 KPI II.15: Availability of metrics for measuring the quality, diversity and value of data assets

This KPI refers to the number of metrics developed by the cPPP that allow for assessing quality, diversity and value of data assets. From the running projects in 2018, 15 over 30 (50%) confirmed they are assessing quality, diversity and value of data assets.

Project	**Metrics** used **to quantify** quality, diversity and value of data assets
BigDataOcean	N/A
Data Pitch	DataPitch measures the quality and value of data assets by tracking startup and SME impact through related KPI's. These KPI's consist of the total sales and investment generated by each startups from their related solution.
euBusinessGraph	Based on the raw data received from data providers, euBusinessGraph measures coverage of all attributes based on number of empty cells (null) per attribute / total number of rows in dataset. Coverage 100 equals zero null values in dataset. In the project context, the coverage measure is an important indication of how relevant each attribute is to data consumers. Low coverage results in lower quality of queries since more attribute values are omitted. Low coverage also reduces relevance of specific attributes for data analysis.
EW-Shopp	• Grafterizer implements an effective approach for Visual Data Profiling (VDP) that simplifies the process of preparing tabular data and contributes to improving data quality. In a nutshell, this particular kind of data profiling consists in a column-based statistical assessment of the tabular data set to be transformed adequately and visualized with the goal of helping the user to swiftly identify potential quality threats to the overall data set quality as missing values and valid (non-null) values for the currently selected column. More in details, the visual data profiling service analyzes and assesses the quality of the dataset and returns a statistical profile. This profile is an essential part of the underlying core application logic that suggests transformations and renders profiling charts: i. Count – the total number of values in the selected column. ii. Distinct – the number of unique values. As an example, a column attribute 'week' might count in total 1000 rows and 7 unique values, one for each day. iii. Histogram – an array containing one value for each histogram bin. iv. Quality – an array that contains three different values, one value representing valid entries, one for invalid entries and another one for outliers. v. Boxplot – array that contains all values necessary to render a boxplot chart, i.e. the first, second and third quartiles, and the median. vi. Histogram_labels – labels for the histogram chart visualization. vii. Quality_labels – labels for the quality chart visualization. • It uses a tool named ABSTAT to measure quality of RDF data; this tool provide several indicators that can be used to check quality, although not resulting in a single metrics; indicators are aggregated at the level of schema patterns, i.e., connections between types qualified by a property. For each pattern, we count: frequency, number of relation instances; max/min/avg number of distinct subjects per object, max/min/avg number of distinct objects per subject; outliers in these counts often point to errors in the data supporting their inspection.

FashionBrain	FashionBrain measures the quality of its machine learning models with technical measures such as F1/Precision/ Recall/Mean Average Precision/nDCG, AUC, ROC etc.
SLIPO	Data quality indicators and statistics embedded in the SLIPO Workbench and its individual software components (TripleGeo, LIMES, FAGI, DEER, SANSA)
TT	Data Quality in TT is assessed at pilot level, i.e., each pilot evaluates the quality of their data assets based on data quality indicators (i.e., completeness, uniqueness, timeliness, validity, accuracy, consistency) in order to meet the specific quality requirements of that pilot (e.g., percentage thresholds of completeness or of accuracy). Moreover, the DMT defined a set of Data Quality KPIs aiming to capture further indicators about the Data Quality process performed by the pilots. These Data Quality KPIs cover the following areas: availability and discoverability of data assets (e.g., initial number of available data assets, percentage of available data assets that are open data), comprehension of data assets (e.g., percentage of data assets that have been selected by a pilot based on their metadata, on which that pilot could perform specific quality assessments), and usability of data assets (e.g., percentage of data assets actually used by a pilot, which eventually meets the specific quality criteria of that pilot).
QROWD	• Positional accuracy (for geospatial datasets) • Binary classification (True or false), for some geospatial datasets and for machine learning classification • Confidence interval of classification (For machine learning classifiers) • For diversity, number of different formats. For Semantic datasets, number of different vocabularies used.
BigMedilytics	• Study plan is specifically is designed to ensure that all data is properly collected during the prospective phase of the pilot. This guarantees that the data collected has high enough quality (e.g. in terms of timestamps collected). • Inclusion/Exclusion criteria of studies ensures diversity. Data is collected with pre-defined KPIs in mind. This ensures that the data collected has high enough value. • Statistical approach by using R. • Correctness mainly (avoid repeating codes, missing values, inconsistent dates, etc.). We count the number of data rows/column with one of the aforementioned issues and, usually, such data is discarded. • All the data is collected at Achmea in a separate operational system. The data is made available to the organization with a datawarehouse (Oracle) to different types of application. The quality standard of the datawarehouse is very high, because it is also used for mandatory reports to the government. • Model performance measured by statistical measures such as misclassification rate, sensitivity and specificity. • Metric regarding the speed, quality, completeness, and certainty of radiology reports produced by radiologists using the system. • Data type, e.g.: clinical data, real-time, text, other structured information, etc. Also in terms of variety and velocity. • Completeness and correctness of the data.
BODYPASS	Quality: The project will include at least some quality tags referred to aggregated body data. It is possible that additional quality metrics related to 3D data or body measurement data will be included along with the project life.
DataBench	Extensive research on economic and technical measurement processes See Deliverable D1.1 and D2.1 on www.databench.eu

EDI	A checklist was issued for every Data Provider that it had to be answered before uploading their datasets: EDI Big Data Infrastructure: upload check list In order to obtain the credentials to upload data on EDI's Big Data infrastructure, it is mandatory to answer all the questions in this check list. General Information Name Mail Address Data provider/Challenge Data Legislation In order to upload data on EDI's Big Data Infrastructure, it is mandatory to be compliant with all the aspects of the General Data Protection Regulation. The GDPR regulates how personal data within the European Union should be stored, processed and exported to non-EU countries. In short, data is required to be anonymized or pseudo-anonymized and must be protected by design and by default. More information is available at [1]. Do you confirm that you are fully aware of the GDPR regulation? (Y/N) Personal data Do your datasets contain data that requires anonymization? (Y/N) Anonymization Have you duly anonymized the personal data of the datasets to be shared with the corresponding Subgrantee/s and with personnel of the EDI consortium? (Y/N) Inferring personal data Could personal details be inferred from the full dataset by means of a reasonable effort (i.e. reverse engineering methods, crossing datasets, and so on)? (Y/N) Informed consent Do you count with explicit consent from the data subjects to process and share their data, or with any other legitimate basis? (Y/N) Ownership of the data Is your organisation the owner of the data that will be uploaded? (Y/N) Right of use Do you have the right of use of the datasets you will upload? (Y/N) Contents The only data allowed to be uploaded on EDI's Big Data Infrastructure are the ones strictly connected to the challenges. Any different data with improper contents must not be uploaded and is responsibility of the data provider The datasets that you will upload contain only data useful for the specific challenge(s) you are involved (Y/N) Consent to share the datasets You consent to let the EDI Consortium share your datasets in part or fully to anyone involved with your challenge, meaning the startups and/or SMEs that have signed the Sub-Grant Agreement, and researchers working in the EDI project, in order to leverage the data to reach the goals of the challenge itself (Y/N) Consent to download the dataset Do you agree to the sub-grantee storing/downloading the dataset in its own infrastructure? (Y/N) Dataset Quality & Features I have removed from the dataset all the incomplete or inconsistent data (Y/N) What is the dimension (in Giga or Tera) of the whole dataset?
I-BiDaaS	A Data Quality Assurance Process (DQAP) aiming at ensuring the high quality of the data generated/collected during the lifetime of the project, has been established for the I-BiDaaS project. The data quality dimensions to be measured, determining the quality objectives, are the following: (a) Validity, (b) Reliability, (c) Precision, (d) Integrity, (e) Timeliness, (f) Completeness.
ICARUS	Accuracy, Completeness, Veracity, Timeliness, Variety, Temporal & Spatial Coverage, Mapping to ICARUS data model - metadata to be captured per data asset according to the ICARUS metadata schema
Lynx	No metrics, since the datasets consist of legislation and case law published by official institutions or standardization bodies.

Table 25. Metrics used to quantify quality, diversity and value of data assets

In addition, BigDataStack develops a generic software to detect erroneous IoT records, some projects such as Boost4.0 and Fandango are planning to introduce metrics in 2019.

These results show the intense usage of metrics to measure quality, diversity and value of data assets in projects and some projects have developed specific metrics and methods to ensure quality, diversity and value in the data. However, we cannot talk yet (2018) about "cPPP" developed metric expected for 2019+[35].

9.3 KPI II.16: Increase the speed of data throughput compared to 2014

40% of the projects reported they expect the project to improve data throughput. In particular BigDataOcean, Data Pitch, euBusinessGraph, FashionBrain, SLIPO, MHMD, BigMedilytics, Boost 4.0, CLASS, E2Data, I-BiDaaS and TYPHON reported expected outcomes in this area.

Some projects such as BigDataOcean and FashionBrain have measured improvements over 1000%. Others such as I-BiDaaS have specific objectives to develop data processing tools and techniques applicable in real-world settings and demonstrate a significant increase of speed of data throughput and access, however results are still not available at the end of 2018 (project started in 2018).

Table 26 provides details per project in relation to the metrics used and the specific levels of improvement.

Project	Metrics used for speed of data throughput	improvement (in percent) measured in the project
BigDataOcean	Execution time for specific queries	1500%
Data Pitch	We track the number of data providers sharing data with startups through our programme. The intention will be to change the mentality of data providers in working collaboratively, resulting in an increase in the amount of data being shared	We have no benchmark to compare percentage increases.
euBusinessGraph	Amount, scope, various jurisdictions, how reach/deep the data is.	Not explicitly measured.
FashionBrain	IDEL measures query execution time in milli-seconds, TASTY measures extraction time in milli-seconds	We compared our approach for relation extraction between a disc based data system (Information Systems Journal 2015) with a main memory database system (COLING 2016). The increase is by more than 1000%.
SLIPO	Runtime and scalability of data integration workflows for POIs (Points of Interest) geospatial data assets for each of the SLIPO Workbench components and world-scale data integration workflows as a whole	1-2 orders of magnitude increases in performance in the individual software components; world-scale data integration of POIs in <10mins

35 DataBench project

BigMedilytics	a) Time of care, visits to emergency room, visits to specialist, hospitalization, mortality, costs of health and labour. b) Search speed in large-scale PACS archives	a) 20% b) Not measured yet, part of evaluation campaign in coming project year
Boost 4.0	As this is not a main objective of BOOST 4.0 not all the pilots have defined any metrics, and not even all the pilots are measuring it. Some of the metrics that have been defined and is being measured or will be measured are: - Lot-size-2 engineering lead time - Processed sensor measurements per second - Number of different data sources processed per second - Metrics defined by 3GPP in TS28.554 - Collection, storage and data processing times - Number of samples per second	The pilots have not started measuring this yet.
CLASS	Amount of distributed data sources (mostly images coming from cameras) that the CLASS SW is capable to process in real-time	No available info yet
E2Data	Runtime (ms), Relative performance compared to currently available processes	No available info yet
I-BiDaaS	One of the I-BiDaaS objectives is to Develop data processing tools and techniques applicable in real-world settings and demonstrate a significant increase of speed of data throughput and access. The detailed analysis of the user needs, culminating in the requirements elicitation framework, serves as the baseline for identifying and establishing those concepts that would be applicable to all use cases and for investigating the case for a generic conceptualisation of industrial Big Data analytics as a service solution. To measure data throughput, the evaluation variables will be Operational, Performance.	not measured yet
TYPHON	Query times, data transformation times	Still in development.

Table 26. Increase the speed of data throughput – details per project

10 MACRO-ECONOMIC KPIS (KPI II.1, II.2, II.3, II.5)

The monitoring of the macro-economic KPIs (KPI II.1, II.2, II.3 and II.5) is based on input from the European Data Market Monitoring Tool36 as they are presented in the most recent report[37] by IDC. In their study the authors assess several KPIs related to the data market, among them the number of European Data Professionals, the number of European Data Companies, the value of the data market and the revenues of data companies. The study provides also an open dataset with the historical series of the indicators measured from 2013 to 2018 (actual data), forecasts to 2020 and to 2025. In this monitoring report we present figures for 2016, 2017 and 2018. All multi-annual comparisons are based on the 2013 figures.

The IDC report provides forecast data for 2025 based on 3 alternative scenarios[38]:

- The Baseline scenario is characterised by a healthy growth of data innovation, a moderate concentration of power by dominant data owners with a data governance model protecting personal data rights, and an uneven but rather wide distribution of data innovation benefits in the society;
- The High Growth scenario is characterised by a high level of data innovation, low data power concentration, an open and transparent data governance model with high data sharing, and a wide distribution of the benefits of data innovation in the society;
- The Challenge scenario is characterised by a low level of data innovation, a moderate level of data power concentration due to digital markets fragmentation, and an uneven distribution of data innovation benefits in the society.

10.1 KPI II.1: Market share of the European Union in the global Big Data Market

KPI II.1 assesses development of the market share of the European Union in the global Big Data Market. As an indicator, we will compare the total revenues of EU Data Companies with other economies, i.e. the US, Japan, and Brazil, as they are used as a benchmark in the IDC report[39].

The EU share of the total revenues in these economies the 2013 baseline was 27,7%. This share increased slightly to 27.9% in 2018, which is remarkable because the international indicators grew very fast in this period, but the EU kept pace with them. In absolute terms, the amount of the total revenues of US Data Companies in 2018 was approximately twice the revenues of EU28 Data Companies in the same year (€162 bn vs. €77 bn).

Table 27 presents a detailed comparison for the years 2017-2018 and shows that the EU Data Companies revenues (growth rate: 12%) outperform Japan (10%%) and Brazil (8%) in terms of relative and absolute growth in revenues. The United States show strong growth in revenues of 10% in relative terms but starting from a much higher baseline and therefore by far the highest value of revenues growth in absolute terms (€15 bn).

36 SMART 2016/0063 – Study "Update of the European Data Market Monitoring Tool", IDC and Lisbon Councils

37 Gabriella Cattaneo, Giorgio Micheletti et al. "Update of the European Data Market Tool - Second report on Facts and Figures" April 2019 www.datalandscape.eu

38 Gabriella Cattaneo et al. "idbidem

39 Gabriella Cattaneo et al. Ibidem, chapter 10 pp.129-142

World Region	Revenues (€B)		Share		Growth 2018/ 2017
	2018	2017	2018 % of total	2017 % of total	
EU	77	69	27,9%	27,6%	12%
USA	162	147	58,7%	58,9%	10%
Brazil	8	7	2.7%	2,8%	8%
Japan	29	27	10.6%	10,7%	10%
Total	276	250	100.0%	100,0%	11%

Table 27: Comparison of data companies revenues by country

Finally, Figure 5 shows the increase of the total revenues of Data Companies revenues of the selected economies since 2013. The US and Europe show the highest growth rate compared to 2013, with the US clearly being ahead of the rest of the world in value.

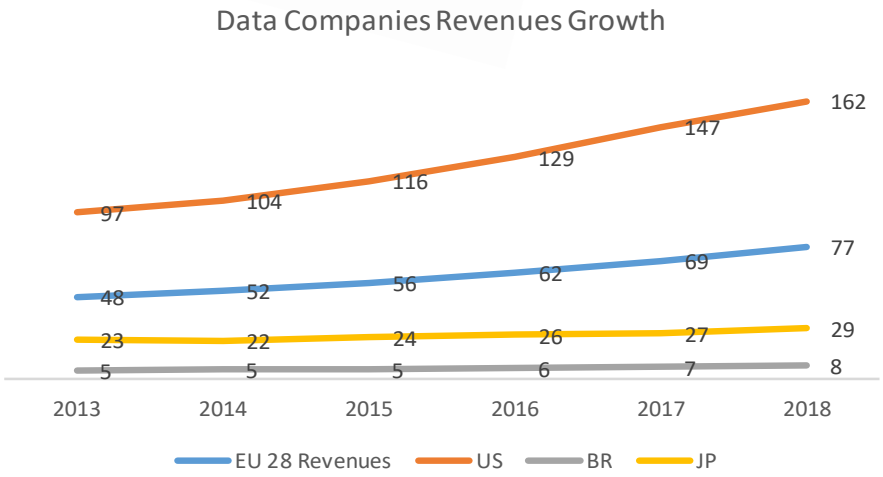


Figure 5: Development of Total Revenues compared to EU, 2013-2017

70% of the cPPP projects active in 2018[40] reported contribution to increase revenue share of EU companies against total of revenue of EU, US, Japan, Brazil. While some projects provide statements in potential impact in market share, others can already share some evidence in real examples: e.g some European companies of euBusinessGraph project reached new markets thanks to the project; in TT project INDRA within the scope of Airports pilot, they are managing to replicate the outcomes achieved in the Retailing services business case trying to sell to other Airport Managers worldwide. The international presence of Indra could be a catalyst to speed up the market uptake of the solutions by other end-users. Other projects, e.g DataBio state that is part of the project objectives to open the possibilities for European ICT and Earth Observation industries (including SMEs) to participate in the European and worldwide bio-economy market.

Some projects contribute to the quick adoption of new technologies, making their user industries more competitive in a world-wide data economy (E.g DataBio increasing the productivity of European bioeconomy

40 Based on number of respondents.

industries (as first adopters of DataBio results), they have the opportunity to better compete with non-European industries, or Boost4.0 whose main objective is to accelerate the adoption, by Industry 4.0.

Other ways projects are supporting this KPI:

- Incubators: Supporting EU data driven companies to build innovative solutions that can be scaled internationally
- Accelerating adoption (mainly in lighthouse projects)
- Developing innovative technologies that can be European companies more competitive, e.g
- adopting data protection approach that can make European data industries more competitive (SPECIAL, SODA, etc or BodyPass that unlocks the potential of analysing large amount of medical data while protecting patient privacy).
- Enabling industries to exploit their big data efficiently and therefore increase their market share and services provided to their customers (e.g I-BiDaaS)

Table 28 provides details of the qualitative contribution of projects to the Market share of the European Union in the global Big Data Market.

Project	Qualitative statements about how the project is contributing to Market share of the European Union in the global Big Data Market
AEGIS	AEGIS has developed innovative technologies contributing to European technological leadership and implicitly to increase the revenue share of EU
Data Pitch	Supporting 47 EU data driven companies to build innovative solutions that can be scaled internationally. These companies are solving real world challenges set by large corporations, creating competition for already established organisations within their respective markets.
DataBio	DataBio objectives include to open the possibilities for European ICT and Earth Observation industries (including SMEs) to participate in the European and worldwide bioeconomy market. In addition, by increasing the productivity of European bioeconomy industries (as first adopters of DataBio results), they have the opportunity to better compete with non-European industries.
euBusinessGraph	The European companies present in the project reached new markets.
EW-Shopp	Creating tools that facilitate the understanding of e-commerce /retail/ marketing dynamics in relation to events (weather and not), will give European companies a significant competitive advantage, generating revenue in terms of savings and improvement of service.
FashionBrain	Flair is being used live in an automized NLP application at Zalando.
SODA	Make Europe leading in advanced crypto solutions like multiparty computations. Knowledge and frameworks resulting from this form the basis of commercial business activity in form of software and services.

SPECIAL	The use cases within the SPECIAL project work may trigger a new strategy of organisations doing business in Europe. This strategy could be to embrace data protection friendly approaches more in future (2018+) as something that can also be a competitive advantage. Therein, the project's approach to support informed consent acquisition easily manageable by the organisation as well as by the data subjects themselves, may contribute to a reconciliation of both economic interests and informational self-determination in the digital era.
TT	Advanced services developed and tested in TT will allow EU companies to commercialise cutting-edge services and products allowing to increase competitiveness against competitors and gaining market share. For instance, INDRA within the scope of Airports pilot, they are managing to replicate the outcomes achieved in the Retailing services business case trying to sell to other Airport Managers worldwide. The international presence of Indra could be a catalyst to speed up the market uptake of the solutions by other end-users. The same situation occurs with CINTRA and their international concessions of tolls. The positive impacts and achievements obtained in their operational efficiency measurements provides a competitive advantage to be replicated and implemented within other similar tools around the world. These two cases (among others) states the contribution of TT to gain market share compared to companies and institutions of EU, US, Japan and Brazil.
QROWD	Reinforcing the creation of services becoming competitive at global scale, to be commercialised by 2 QROWD partners showing strong, consolidated global market positions and wide commercial networks (ATOS and TomTom)
BigMedilytics	Targeting solutions in the Oncology, Population Health and Hospital Workflow Optimization domains. New products exploiting large-scale healthcare data
BODYPASS	With the rise of artificial intelligence algorithms, data has become increasingly valuable. Medical data, especially, can lead to raising standards of life. Our project unlocks the potential of analysing large amount of medical data while protecting patient privacy. Extracting this value will not only increase revenue of EU companies, but also benefit the overall quality of life of EU residents.
E2Data	Technologies developed in the framework of the E2Data project are novel and are expected to increase the revenue of European companies taking advantage of them. These technologies include, but are not limited to, Health Analytics, Green Buildings, Natural Language Processing, and Biometric Security.
I-BiDaaS	I-BiDaaS is expected to provide a significant boost by providing its tools and services in the telecom, banking and manufacturing sector based on the Big Data as a Self Service approach, will enable industries to exploit their big data efficiently and therefore increase their market share and services provided to their customers.

Boost 4.0	One of the main objectives of the BOOST 4.0 project is to accelerate the adoption, by Industry 4.0, of data and advanced analytics platforms, which will have a direct input on the revenue share of EU companies. In general, the creation of the BOOST 4.0 connected smart Factory 4.0 model will help companies deploy big data solutions that will give them many competitive advantages, such as increase the operational efficiency, E2E manufacturing planning and deliver, improved smart product costumer experience, and also foster new business models, like outcome-based and product customization. A few examples of the specific solutions that will prove this impact and that are being deployed within the BOOST 4.0 pilots are: - Using Big Data analysis to simulate the manufacturing process and predicting machine breakdowns will improve the productivity and will reduce the energy consumption and the waste production. - Understanding better the manufacturing process will also help to optimize the general operation of the production line, increasing productivity. - Providing machines with IoT technologies and "smart" capabilities gives them competitive advantages and creating new business opportunities. - By deploying better AGV (automated guided vehicles) technologies based on big data analysis will improve the internal logistics and the productivity. - Improving the information shared among the supply chain and monitoring the delivery network allow the prediction of delays. The project is supporting companies such as CAPVIDIA, TRIMEK, RISC, ATLANTIS and big vendors to build solid cases for international sales.
ICARUS	Optimisation of services delivered, (e.g. improving the readiness of the airports) or creation of new services addressed to passengers as well as businesses (e.g. AIA) signifies minimisation of costs Optimisation of SW solutions (e.g. reducing costs and minimizing the ecological footprint of airliners by PACE) or development of new software solutions (e.g. CELLOCK) or services (e.g. OAG) signifies increased customer base
TheyBuyForYou	Project partners OpenOpps and OpenCorporates are global players in the fields of, respectively, procurement information and company information services
TYPHON	Companies in Europe using the project technologies will have a market advantage in delivering more complex hybrid data-based applications and services.
Lynx	To help SMEs and large companies in compliance processes by saving time and efforts.
Cross-CPP	Key motivation of Cross-CPP project is to give cross-sectorial industries access to the great spectrum of sensor data coming from high volume products from various industrial sectors (vehicles, smart home devices, etc.). With the increasing number of connected sensors and actuators within such mass products, this number will rise in short-term. This enormous amount of data continuously generated by mass products will represent: • a NEW information resource to create new value, allowing the improvement of existing services or the establishment of diverse new cross-sectorial services, by combining data streams from various sources • a major big data-driven business potential, not only for the manufacturers of Cyber Physical Products (CPP), but in particular also for cross-sectorial industries and various organisations with interdisciplinary applications

	Assuming that the Cross-CPP Ecosystem will be implemented as a commercial solution, the Cross-CPP marketplace will represent a disruptive innovation, enabling radical innovations by numerous cross sectorial Service providers. This would help to increase revenue of EU companies against total revenue of US, Japan, and Brazil.
BDVe	BDVe advances the competitiveness of European Big Data Companies by a rapidly evolving Big Data Landscape by supporting the ecosystem in (1) identifying and promoting Data Driven Business Opportunities, (2) supporting the development of CoE in MS and (3) by developing and promoting intuitive, state-of-the-art tooling to help SMEs and start-ups build better Data-Driven value propositions

Table 28: Qualitative contribution of projects to the Market share of the European Union in the global Big Data Market

10.2 KPI II.2: Development of the number of Data Companies in the EU

Data Companies in terms of the IDC study are defined as data suppliers that as their main activity or one of their main activities produce or deliver digital data-related products, services, and technologies. These companies represent the emerging data industry, that is the supply side of the data market.[41]

Regarding the development of the number of Data Companies (KPI II.2), according to the most recent report[42] they increased to 283 100 by 2018, compared to 271 700 in 2017, with a growth rate of 4,2%. It should be noticed that, due to the high concentration of the ICT industry in the UK, almost half of them are based in the UK.

The vast majority of the Data Companies (279 500 in 2018, 98,7% of total) are under 250 employees, reflecting the structure of the European industry, but they accounted for 72% of revenues. The number of data companies of large and medium size in 2018 is 3600. The share of SMEs is more or less unchanged since 2013.

In terms of future expectations, the report forecasts a total number of 338 700 Data Companies in the Baseline Scenario in 2025 representing a Compound Annual Growth Rate (CAGR) of 2.8% in the period 2020-2025. Under the favourable High Growth scenario data companies could increase to 389 100 (CAGR 5,7% on 2020); in the Challenge scenario the number could be lower at 320 700 (CAGR 1.7% on 2020). Figure 6 summarizes these numbers. The share of SMEs remains quite stable in all three scenarios at a level of roughly 99%, reflecting the structure of the EU industry, even though data companies of more than 250 employees are expected to grow twice as fast as smaller companies.

41 Gabriella Cattaneo et al. "ibidem, pages 67-74
42 Gabriella Cattaneo et al. "ibidem

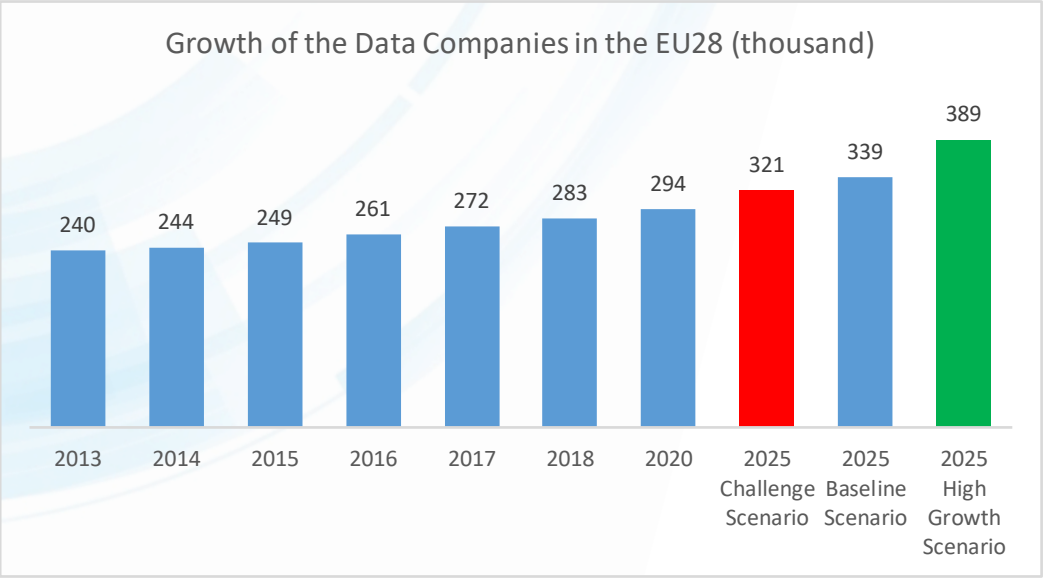


Figure 6: Number of Data Companies in the European Union in 000s

BDVA i-Spaces and **Data Incubators (**ICT 14-b projects**,** so DataPitch and EDI**)** are in particular designed to contribute to this KPI as they support start-ups and entrepreneurs from early ideas to technical and business development till go-to-market[43].

77% of the BDV cPPP projects active in 2018[44] reported contribution from their project to increase the number of European Companies offering data technology, applications. Projects contribute in different ways that can be summarized as follows:

- Creating tools that will stimulate the creation of new companies (e.g EW-Shopp, FashionBrain, E2Data, or Cross-CPP) or lowering costs for data owners and consumers (e.g SLIPO)
- Creating new companies as results of the project (e.g BigDataOcean)
- Supporting EU data driven companies to each build innovative solutions that to solve data related challenges (data incubators such as DataPitch and EDI)
- Supporting companies in complying with the GDPR (e.g SPECIAL)
- Lowering the threshold to create new business in a particular sector:
    - DataBio in databio economy e.g education professionals and workers;
    - ICARUS in the aviation industry at facilitating the development and delivery of added value services over the aviation data.
    - TT project some concrete targets have been set up for assessing how much SW and HW products has been produced in the project that might increase the product offering of the company and create new companies through spin-offs focused on the exploitation of new technology assets or services based on data. The answers provided in question 38 and 39 related on Large Scale Experimentations on Big Data demonstrate that End-users, Technology Providers, Industrial Partners and Research Centres has contributed to extend product systems capabilities by making use of data analytics (descriptive, predictive, etc.) and machine learning capabilities to enhance the current market positioning of the services portfolio of consortium partners.

43 Further information can be found in section 2.1 of this report.
44 Based on number of respondents.

- Boost4.0 provides the EU industry with a model to take advantage of the big-data technologies, the project itself is a contribution to increase the number of European Companies offering data technology and applications.

Table 29 provides details per project to their contribution to the development of the number of Data Companies in the EU.

Project	Qualitative statements about the contribution to the Development of the number of Data Companies
AEGIS	The project delivers a technology enabling companies to collect, process and analyse data.
BigDataOcean	New company created by the project.
Data Pitch	Supporting 47 EU data driven companies to each build innovative solutions that to solve data related challenges.
DataBio	Indirect contribution: DataBio educates several workers (by engaging them in events, hackathons, pilots etc.) and lowers the threshold to create new business in bioeconomy. Combined with the lack of commercial actors that will exploit the new technologies in some domains (e.g. in some fishery applications), new companies can fill the gap. At least one partner considers the option of exploiting the results by creating a spin-off offering data technology.
euBusinessGraph	Through the companies that are part of the project.
EW-Shopp	EW-Shopp works to create a professional open source toolkit that can manage enterprise data in a comprehensive way, helping actors involved in the processes of transformation, cleaning, semantic annotation, enrichment, analytics, and visualization. We believe that having this tool at disposal will stimulate the birth of European companies offering data technology and applications.
FashionBrain	Flair is open source and potentially allows other EU companies to offer NLP data products.
SLIPO	Data integration as a service; lower-cost and higher quality assets for cross-sector geospatial data owners/consumers.
SODA	Indirectly: Make Europe leading in advanced crypto solutions like multiparty computations. Knowledge and frameworks resulting from this form the basis of commercial business activity in form of software and services
SPECIAL	SPECIAL aims to support companies in complying with the GDPR in terms of obtaining consent, providing transparency and demonstrating compliance. As such we envisage that European Companies will use the technology and knowhow developed in SPECIAL in order to give their customers more control and transparency with respect to personal data processing.

TT	TT has set up concrete targets for assessing how much SW and HW products has been produced in the project that might increase the product offering of the company and create new companies through spin-offs focused on the exploitation of new technology assets or services based on data. The answers provided in question 38 and 39 related on Large Scale Experimentations on Big Data demonstrate that End-users, Technology Providers, Industrial Partners and Research Centres has contributed to extend product systems capabilities by making use of data analytics (descriptive, predictive, etc.) and machine learning capabilities to enhance the current market positioning of the services portfolio of consortium partners. By M18 mid-term review, it has been reported a total of 4 new services produced with BD capabilities and 16 services which have been improved by integrating BD capabilities.
QROWD	Progressive take-up of QROWD mobility, transport and crowdsourcing by local Public Sector and/or their data providers will contribute to generate new opportunities for city suppliers to jump in into the urban mobility data services solutions space.
MHMD	The SMEs participating to the project, in particular Gnubila and Digi.me, are already offering data technology to their customers. Through the exploitation of other tools for secure data analytics, new companies might offer similar services.
BigDataGrapes	The project is developing and demonstrating powerful data processing technologies that will increase the efficiency of companies that need to take important business decisions dependent on access to vast and complex amounts of data within the agriculture, food and beauty sectors. Moreover, the project aims to support the creation of a data ecosystem and economy that will increase the competitive advantage of companies that serve with IT solutions these sectors.
BigDataStack	The CEP component will be transferred to Leanxcale.
BigMedilytics	It supports the involved companies and might aid additional companies serving as platforms for healthcare IT solutions.
Boost 4.0	As BOOST 4.0 aims to provide the EU industry with a model to take advantage of the big-data technologies, the project itself is a contribution to increase the number of European Companies offering data technology and applications.
E2Data	Technologies developed in the framework of the E2Data project are expected to foster the creation of European Companies offering data technology applications. These technologies include, but are not limited to, Health Analytics, Green Buildings, Natural Language Processing, and Biometric Security.
EDI	16 data providers have been able to create new Data Products which try to address the challenges of 9 Data Corporations in Europe.
ICARUS	ICARUS aims at facilitating the development and delivery of added value services over the aviation data that will be made available, promoting entrepreneurship and the creation of European companies offering data applications and services.

TYPHON	New field of products and services addressing hybrid data stored will be made possible by the project technologies.
Lynx	We plan to contribute to law firms and standardization bodies in including data technology and services in their processes.
Cross-CPP	Cross CPP project has plans to offer cross domain CPP data in a uniform format via big data marketplace. This marketplace will be equipped with big data analytics toolbox thus, offering European companies to provide data related technology in future.
BDVe	BDVe is supporting SMEs/startups in the area of Big Data in several ways: 1.) Marketplace: all SMEs/startups from the PPP, both projects and association, can present their solutions/products in the marketplace. They will be supported in the promotion and matchmaking. 2.) BDVe is creating a network of Intermediaries (Accelerators, Incubators and Innovation Clusters) that support Big Data SMEs/start-ups in their ecosystems. 3.) BDVe is developing training and tooling for business modeling targeted at start-up intermediaries that work with Data Companies. The tools are designed to be embedded in programs running on large start-up platforms such as F6S. Any data collected on start-up business models will further advance understanding of the emerging European Big Data Economy. In addition to that, BDVe is supporting the startups in the Accelerator Project to get access to investors for additional funding.

Table 29: Qualitative contribution of projects to the Development of the number of Data Companies in the EU

In addition, **25% of the BDVA members** reported their organisation run or support a programme that is specifically targeted at **supporting start-ups or entrepreneurs in the Big Data area**. Some examples include:

- **Eindhoven University of Technology** supports SMEs through EIT Digital.
- **Software AG** is running a start-up program http://software.ag/startup/ to support start ups in the start up phase with free licenses of our Big Data software like APAMA, Zementis, UM.
- The **Inria** programme Horizon StartUp prepares and supports PhD students, Post-Doctoral students, and Inria researcher develop their ideas for launching a Start-Up.
- **Fraunhofer-Gesellschaft**: https://www.fraunhoferventure.de/en.html
- **Instituto Tecnológico de Aragón** is working in a POCTEFA project called Acelestartups (https://acelestartups.com/) related to this topic. Moreover, they are part of the foundation "emprender en Aragón"
- **Universidad de Granada**: The Data Science and Computational Intelligence Innovation Hub (DaSCII Hub) is a Digital Innovation Hub (DIH) focused on performing cutting-edge research, knowledge transference and postgraduate teaching in Artificial Intelligence with outstanding lines of research in Data Science and Computational Intelligence. This innovation hub has been created with the aim of becoming the reference of the Artificial Intelligence and Data Science of the region of Andalusia. It brings together both public and private entities that know that the future lies in the digitalization of companies, but above all, in the incorporation of these new data processing and intelligence technologies to enhance their business possibilities. The DaSCII Hub is composed by some of the most outstanding researchers in our region, seven of whom are included in the Highly Cited

Researchers list (Clarivate Analytics, see 2018 list in https://hcr.clarivate.com/). http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs-tool/-/dih/12691/view

- **Prizztech Robocoast**: -"Make the AI as a success story of Satakunta district" - project, 2 years, 2018-2019. -Digital Cities -Project-AIF (Allied ICT Finland) network.
- **Siemens AG**: Next47; Labs Network Industrie 4.0; MindSphere World.

10.3 KPI II.3: Development of the revenue of Data Companies in the EU

The revenues of Data Companies in the European Union (KPI II.3), according to the IDC report[45], reached €77 €bn in 2018 compared to €69 bn in the year before, with a growth of 12%.

The revenue share of SMEs in 2018 amounts to €55,5 €B (72% of the total revenues), an absolute growth of 5,7 €B on the year before. The growth rate of revenues increases in proportion to company size, with the revenues of large companies over 500 employees growing at 16% in 2018 over 2017.

With respect to future growth IDC projects an increase in total revenues for the Baseline Scenario to EUR 126 €B in 2025 representing a Compound Annual Growth Rate 2025/2020 of 7,1%. For the 2025 High Growth scenario the revenues could instead go up to EUR 179 €B representing a CAGR on 2020 of 14,8%; in the Challenge scenario the revenues would be lower, reaching €101 bn, representing a CAGR 2025/2020 of 2,4%. The revenue share of SMEs remains quite stable at about 63% for all three scenarios with the Baseline Scenario resulting in the highest share (63,5%)[46]. Figure 7 depicts the values from 2013 to 2018 and the projected three growth scenarios as discussed in this paragraph.

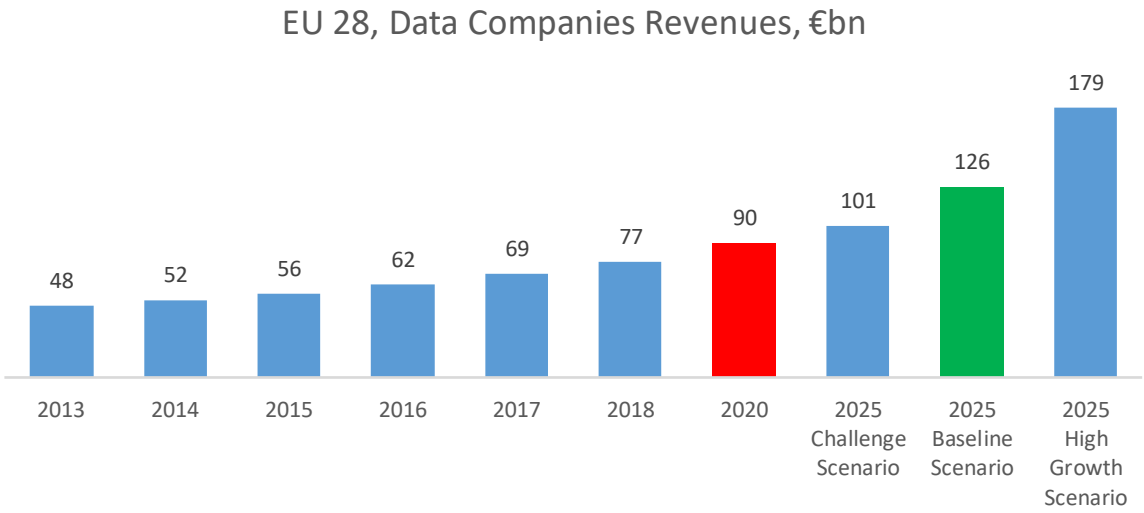


Figure 7: Revenues of Data Companies in the European Union in billion EUR

77% of the cPPP projects active in 2018[47] (BigDataOcean, Data Pitch, DataBio, euBusinessGraph, EW-Shopp, FashionBrain, SLIPO, SODA, SPECIAL, TT, QROWD, BigMedilytics, Boost 4.0, E2Data, EDI, ICARUS, TYPHON, Lynx, Cross-CPP) reported contribution (or plan to contribute) to the revenue generated by European Data Companies. The projects contribution to this KPIs is mainly by:

45 Gabriella Cattaneo et al. "Ibidem pages 89-97
46 Gabriella Cattaneo et al. ibidem
47 Based on number of respondents.

- Opening up sectors to data-intensive companies e.g **BigDataOcean** in maritime sector; **DataBio** project where more technology providers are creating new market and revenue opportunities by piloting their technology in new application areas; or **Boost4.0** in Manufacturing helping the industry to exploit for its benefit the huge amount of data that is already generating or that can be generated, to improve their revenues and reduce their expenses;
- Offering direct support and getting funding for data start-ups to increase revenues.
    - Some examples of DataPitch: https://datapitch.eu/news/data-driven-healthcare-website-expands-into-new-european-markets/; https://datapitch.eu/news/data-pitch-six-month-update/; https://www.digicatapult.org.uk/case-studies/predina/
    - In EDI several start-ups had been able to raise additional funds, from public as well as investment sources, in total adding up to 1,805,000.00€
- Make **data processing easier and cheaper for companies** (e.g EW-Shopp for small e-commerce and retails companies), or **offering technologies or solutions that will make easier for European companies to do more big data analysis thus increasing revenue** e.g BigMedilytitcs in healthcare, ICARUS in the aviation sector optimising the services already offered and at creating new added value services, thus increasing its revenues; or Typhon whose technologies will enable European Data Companies to provide more complex and robust data services.
- **Creating new opportunities through privacy-preserving** analytics solutions or providing tools for companies to become GDPR compliant (e.g SODA, SPECIAL)
- **Through the commercialisation of new services with marketable value** to be developed, demonstrated and later exploited by the project partners (e.g TT in the transport, logistics and mobility market, DataBio in the bioeconomy sector)
- **Creating opportunities for common exploitation** based on joint Big Data technology pipelines (e.g DataBio in the bioeconomy) or developing **simplicity in some business ecosystems** (e,g Lynx will allow law firms, standardization bodies and companies to more efficiently deal with compliance issues).

Table 30 provides additional details on how the project contribute to the development of the revenue of Data Companies in the EU.

Project	Qualitative statements about how your project is contributing to the development of the revenue of Data Companies in the EU
BigDataOcean	The projects further open the maritime activity domain to data-intensive companies.
Data Pitch	Revenue generated by the programme participants is a direct outcome from the support & funding they have received. We have had notable successes within Health and Smart Manufacturing as highlighted by our supporting evidence: https://datapitch.eu/news/data-driven-healthcare-website-expands-into-new-european-markets/ https://datapitch.eu/news/data-pitch-six-month-update/ https://www.digicatapult.org.uk/case-studies/predina/

DataBio	The overall aim of the project is further development and piloting of big data technology for the bioeconomy (in particular: agriculture, fisheries and forestry). Most of the technology providers are creating new market and revenue opportunities by piloting their technology in new application areas. Additional opportunities for revenue increase are created by creating opportunities for common exploitation based on joint Big Data technology pipelines offering advanced bioeconomy services.
euBusinessGraph	Through the companies that are part of the project.
EW-Shopp	The services we are delivering will make data processing easier and cheaper for small e-commerce and retails companies, and for this reason, SMEs that create wealth from data will undoubtedly benefit.
FashionBrain	Flair is being used live in an automized NLP application at Zalando. Flair is open source and potentially allows other EU companies to offer NLP data products.
SLIPO	Provision of SLIPO technologies as OSS; integration of SLIPO into the service offerings of its industrial partners; pre-commercial exploitation of SLIPO.
SODA	Business of European Data Companies is held back by their attitude towards privacy, which privacy-preserving analytics solutions from SODA relaxes and creates opportunities to increase business.
SPECIAL	SPECIAL plans to support companies in developing innovative products and services based on personal data, by providing them with the technology and knowhow required in order to this goal, while at the same time ensuring that said activities are in compliance with the General Data Protection Regulation.
TT	Through the commercialisation of new services with marketable value to be developed, demonstrated and later exploited by the project partners. Indeed, TT has been the opportunity for consortium partners to explore and investigate which impacts occasioned the development and integration of BD capabilities in their service portfolio. The experimentations performance in each pilot involving real end-users which foster the trust and reliability of this solutions when end-users are planning to extend their services or products integration BD capabilities.
QROWD	Reinforcing the creation of services becoming competitive at global scale, to be commercialised by 2 QROWD partners showing strong, consolidated global market positions and wide commercial networks (ATOS and TomTom). Progressive take-up of QROWD mobility, transport and crowdsourcing by local Public Sector and/or their data providers will contribute to generate new opportunities for city suppliers to jump in into the urban mobility data services solutions space.
BigMedilytics	Providing data solutions to hospitals for Oncology, Population Health and Hospital Workflow solutions. It evaluates, establishes and helps deploying new products. By providing the MPC solution, it will be easier for European companies to do more big data analysis thus increasing revenue.

Boost 4.0	Boost 4.0 is about helping the industry to exploit for its benefit the huge amount of data that is already generating or that can be generated, to improve their revenues and reduce their expenses. That means more directly revenue for the tech providers.
E2Data	Technologies and Use cases that will be developed in the E2Data project will contribute to the revenue of European companies. These technologies and their impact is provided in the "Significant Innovations to Market" questionnaire.
EDI	In M15, according to the feedback survey, sent out to the Top 16 startups after the Experiment phase, several start-ups had been able to raise additional funds, from public as well as investment sources, in total adding up to 1,805,000.00€. Additional raised funds are expected from the Top 16 teams after the conclusion of the full incubation round.
ICARUS	OAG, a partner of ICARUS providing big data to the ICARUS platform, aims at optimising the services already offered and at creating new added value services, thus increasing its revenue.
TYPHON	Technologies will enable European Data Companies to provide more complex and robust data services.
Lynx	By allowing law firms, standardization bodies and companies to more efficiently deal with compliance issues.
Cross-CPP	This project aims to provide cross domain CPP data to the service providers, they can be Data Companies providing different services to the end customer. The project also aims to provide big data analytics toolbox which can be helpful especially for SMEs who do not have big data expertise.
BDVe	The BDVe project advances the understanding of Data-driven business opportunities among European companies through insight, analysis and State of the Art tooling. This will allow the community of companies addressed by the cPPPs to drive more revenue from their Data-centric business models.

Table 30: Qualitative contribution of projects to the development of the revenue of Data Companies in the EU

10.4 KPI II.5: Development of the number of Data Professionals in the European Union

Data Professionals according to the IDC study collect, store, manage, and/or analyse, interpret and visualise data as their primary or as a relevant part of their job activities: they should be able to work with massive databases and with emerging data base technology, as well as with structured and unstructured data, to elaborate and visualise data in order to use them for the interpretation of facts and as factors for decision-making processes.[48]

[48] Gabriella Cattaneo, Giorgio Micheletti et al. Updating the European Data Market Tool - Second report on Facts and Figures" pp.57-66

The baseline for Data Professionals in the European Union in 2013 (KPI II.5) amounts to 5,77 million. The number of Data Professionals increased to a total of 7,2 million by 2018, resulting in a total absolute growth of 1 453 thousand professionals since 2013. The rate of growth of data professionals is increasing, with approximately 559 thousand positions added in 2018 and an increase of 8,4% on the year before.[49]

The 2020 projection in the IDC report shows a growth to 8.2 million Data Professionals who could become 11.3 million by 2025 under the baseline scenario, representing an absolute growth of 2 million additional jobs from 2020, or 7,3 million from 2013. The compound annual growth rate (CAGR) between 2013 and 2025 in the baseline scenario is 6,6%. Based on the baseline scenario, the number of data professionals is expected to nearly double between 2013 and 2025.

Depending on framework conditions and the dynamics of the data market, the number of data professionals could reach 13,1 million under the High Growth scenario with a compound annual growth rate of 9,8% from 2013, or be lower (but still relevant) under the Challenge scenario at 10,3 million (CAGR 4.6%).

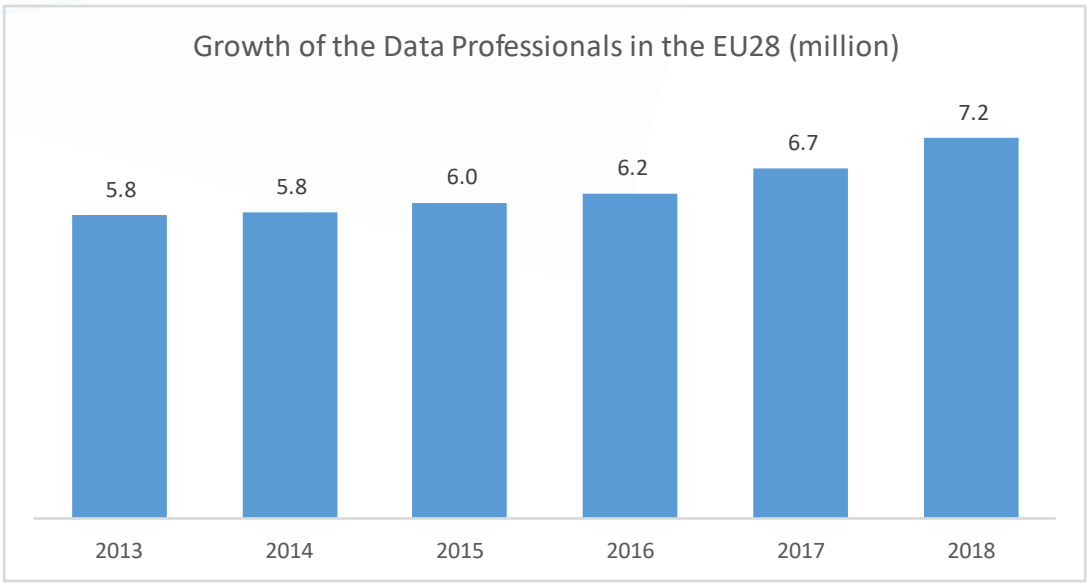


Figure 8: Data Professionals in EU28

87% of the cPPP projects active in 2018[50] reported contribution from their project to increase the number of Data Workers in Europe. Projects contribute to this KPI in different ways:

- **New organisations created as result of the project hiring new data professionals** to exploit project results (e.g BigDataOcean, BigMedilytics) or new jobs for data workers created by the project (e.g DataBio
- **Supporting emerging startups to grow** (e.g companies in DataPitch incubator created more than 40 data related jobs in their companies)
- **More data-driven services** that will **require new data-workers** (e.g EW-Shopp, BigMedlitytics, Boot4.0, E2Data, Icarus, Lynx) or unlocking value of data services **introducing privacy preserving technologies** and therefore attracting more data workers (e.g SODA)
- **Creation of new job profiles** e.g TT project, EDI incubator, Typhon project
- **Adoption of data solutions** in public sector and users in different sectors will increase the amount of data workers (e.g QROWD, BigMedilytics, TT, Boost4.0, DataBio)
- **Education and training** (e.g BigMedilytics, EDI incubator)

49 ibidem

50 Based on number of respondents.

Table 31 provides additional details on how the project contribute to increase the number of Data workers in the EU.

Some projects that don't contribute to this KPI, do e.g contribute to make current professionals more effective (e.g SLIPO increases the productivity of data integrators and GIS specialists).

Project	Qualitative statements about how they project is contributing to increase the number of data workers in Europe (referring to 2018 activities)
BigDataOcean	New workers in the organisation that will exploit the project's results. Data workers in "traditional industries".
Data Pitch	The companies we support are driven by data. They are solving bespoke challenges set by corporates and public sector organisations through the use of shared data. As these startups are quite young/emerging, they require additional resources within a data related capacity to support the needs and expectations of these corporates. The companies we accelerated in 2018 created more than 40 data related roles within their companies.
DataBio	Directly: 34 (2 of them in Norway). See detailed information related to job creation.
euBusinessGraph	By enabling work with company data.
EW-Shopp	The primary objective of EW-Shopp is to develop a Big Data toolkit and methodologies to bridge the gap between companies operating in e-commerce, retail, and marketing industries, and data science with particular reference to techniques for analysis and integration of business data with weather or event information. More data workers will result from the two main outcomes of the project: i. More data-driven services (at least 5) which require data analyst to be sustained ii. Tools that will simplify preprocessing (in particular enrichment, a common task in data analytics), analysis and visualization of data; this technology will require consultancy services from tech partners, which will expand consequently, and by third parties that will advocate the tools.
SODA	The project enables data analytics in a privacy preserving way. Therefore, it enables analytics in situations where privacy, commercial, legislative and ethical arguments prohibit it. The project enables unlocking of this value, which creates jobs for data workers.
TT	Qualitative: Thanks to TT data workers will gain relevance in the transport sector increasing the number and types of data-science related profiles working for Quantitative: currently, TT has no direct metric that allow us to assess the number of data workers produced. Nonetheless, TT has planned to include this metric among the project KPIs in the second Period. The results achieved will directly come from the answers provided by partners in dedicated questionnaires that TT is sharing to feed the calculations of Project KPI levels.

QROWD	Deployment of QROWD solutions will boost activity of third parties e.g. Other city halls and data services providers becoming first adopters outside the QROWD project
BigMedilytics	• Part of the project's results are currently aiding the start-up of a venture within the company. Due to the data-driven proposition of the start-up, the venture will employ more people with a data science background. There will also be other data science workers employed outside the venture to support the activities of the venture. • The following activities developed during 2018 would contribute to increase the number of data workers in Europe: Cluster installation, Anonymization process, Secure channel set up, Data curation and treatment. • Currently in the health domain, there is many potential data to be analysed. Our pilot shows relevant results using such data, fact that encourages the employment of data workers. • All new employees will be data workers, since they will be involved in the curation, analysis, and management of complex medical imaging data and connected information • Education and training • New positions in the different knowledge-driven tasks required to create and make knowledge graphs available are required. • The Lung-cancer pilot application we are developing requires data analytics, machine learning, knowledge engineering, and software development skills • To have validated and integrated hospital data for projects like Breast cancer pilot new data preparation specialists are needed • There is a rapid increase of data scientist job recruitments in Europe • We discussed with key stakeholders the rules for the use of big data technologies in health care and summarized the results in infographics. This work supports the work of data workers in Europe.
Boost 4.0	BOOST 4.0 is about bringing big data to the industrial sector, so we could say that all the activities of the project are addressed to increase the data business in Europe, and thus to increase the data workers. More data shared means more IT workers to install and maintain the digital infrastructure (sensing, communication, storage, processing), more developers to create and deploy new software tools to share data, more data scientists and analysts to obtain conclusions on that data, etc. The challenge is how to customise such capabilities and skills in the context of Industry 4.0 processes, legal frameworks for data sharing and operations of the value chain.
E2Data	The E2Data project is expected to increase the number of data workers, since the main outcome of the project, the E2Data Framework, will enable the deployment of new big data services.
EDI	Some of the companies that took part in EDI already had applications and solutions where data exploitation and analytics was not done in a sophisticated manner. Thanks to EDI training and incentivized by challenges offered by companies, some startups have developed some profiles of Data Analysers within their teams.
ICARUS	Data scientists / analysts either within the consortium organisations or outside the consortium, developing analytics services on top of the ICARUS data.

TYPHON	The project technologies will lead to the creation of a new skills profile of hybrid data management expert having the skills in hybrid data stores, data management, and development of new innovative applications that combine and exploit the diverse big data already collected by enterprise and government systems. The new jobs created will almost be largely data workers.
Lynx	All services being developed in Lynx are intended to deal with data, mostly textual data, so data workers will be needed to manage, maintain and improve those services.
Cross-CPP	Cross-CPP project is to give cross-sectorial industries access to the great spectrum of sensor data coming from high volume products from various industrial sectors.

Table 31: Qualitative contribution of projects to the increase of Data Workers in the EU

11 CONTRIBUTIONS TO ENVIRONMENTAL CHALLENGES (II.6, II.19, II.20, II.21)

11.1 Contribution to the reduction of energy use (KPI II.6)

This KPI intends to measure energy saved in big data analytics using a specified benchmark through solutions provided by specific cPPP projects compared to baseline at beginning of H2020.

Projects were asked: 1) whether they were monitoring Energy Efficiency in your project?; 2) what metrics are you using to measure this parameter? (if monitoring); and 3) improvement (in percent) have you measured in your project.

Only 3 projects (TT, DataBio and E2Data) reported they are monitoring Energy Efficiency in their projects. Except E2Data, actually looking at optimization of data analytics process that can lead to energy efficiencies, responses are not necessarily linked to energy efficiency in Big Data Analytics but to pilot/business scenarios (reported in II.19). Only TT project provides specific results in terms of % of energy reduction and it will be reported in II.19.

E2Data started only in 2018 and we expect the project to provide quantitative results in the future.

Projects provided the following input:

- The **E2Data** framework optimises calculations, leading to decreased use of energy.
- **DataBio**: The fishery pilots focusing on immediate operational choices and fishery planning measure energy consumption and efficiency vs. fisheries operations. Several metrics with base in fuel consumption vs. vessel operation are measured:
  - Specific fuel oil volumetric consumption for propulsion engine per sailed nautical mile in steaming condition (vessel transit between locations).
  - Volumetric consumption of fuel oil per kilogram fish caught.
  - Fuel oil consumed per day of operation.

  Even if these measures are obtainable, they are not a measure of the impact of the project. The reason for this is that the natural variations in factors such as quotas and fish availability are greater than the benefits the project is aiming to provide.

The forestry pilots do not directly monitor energy efficiency. But they are creating the standardized methodology of different forest data integration and its visualization. Therefore, it is expected that as outcome of this project they will define the baseline energy efficiency of big data solutions for forestry.

- **TT**: There is a specific Assessment Category to evaluate Energy Consumption (EC) which measures the energy required by the pilot's activities in both test scenarios (without and with Big Data technologies), therefore potential savings signify improvements in efficiency. The metrics used in TT are implicit in the units in which the KPIs are calculated. The indicators are adapted to assess the specific activities performed by the pilots. In addition, it is important to avoid external factors, such as seasonality or macroeconomic issues that may influence the volume of the activity. In road transport, energy is measured as kWh/100 km, while for airports energy is referred to daily values (kWh/day) and also applicable at passenger level. Also, the scale to which the energy monitoring is applied varies depending on the specific pilot. Since the monitoring of the energy consumption in the global transportation system (eg. Airport energy consumption) would blur the impact achieved, the monitoring unit is scaled to the specific pilot objectives (eg. Airport security area energy consumption). **Current results** show that for the specific monitored items **improvements in efficiency, range between 25% and 51%**. However, demonstrators still have to report their final impact values, which will do at the end of March of 2019. In general, Percentage of KPI Improvements show a growing tendency among the different reporting periods within the project, overtaking most of the predicted target values for energy efficiency, meaning that the range of improvement may move upwards.

11.2 Contribution to Energy use reduction (KPI II.19)

Over 20% of the projects running in 2018 have reported they contribute to the reduction of energy use (BigDataOcean, DataBio, TT, QROWD, BigDataStack, Boost4.0 and E2Data). Only TT project provides quantitative information in Energy reduction with current results show that for the specific monitored items improvements in efficiency, range between 25% and 51%.

Table 32 provides details of the contribution from cPPP projects to energy use reduction.

Project	Qualitative and quantitative statements (2018 and planned)
BigDataOcean	Amelioration of energy produced by waves, due to better location selection for off-shore equipment
QROWD	Future take up of QROWD mobility solutions by cities should render: less traffic across cities, better traffic (less congestion, less traffic jams, etc) promotion of non-polluting modes of transport
E2Data	The E2Data framework optimises calculations, leading to decreased use of energy.
BigDataStack	The Danaos shipping company is planned to optimize ships trajectories thanks to our project

DataBio	The goal of several (4 out of 6) fishery pilots is to increase the efficiency by reducing the time and energy spent on fishery operations. These pilots are providing tools and decision support for optimizing both short- and long-term decisions. The short-term decision support will help the crew choosing the most energy efficient way to operate the vessel. The longer-term decision support aims at helping the crew to find the fishing grounds and fishing times giving the lowest fuel consumption. It is difficult to quantify the reduction of fuel consumption and C02, as this will be different for different vessel types and fisheries, as well as vary with between years because of both fish stocks, fish availability and vessel operational patterns. In forestry, Wuudis (a background solution for activities within forestry pilots) aims to improve operational efficiency and decision making. An improvement in operational efficiency (i.e, ease in daily routines and better communication between stakeholders) indirectly impact the reduction of energy use in the process.
TT	As it has been mentioned, there is a specific Assessment Category to evaluate Energy Consumption (EC) which measures the energy required by the different pilots. Current results show that for the specific monitored items improvements in efficiency, range between 25% and 51%. However, demonstrators still have to report their final impact values, which will do at the end of March of 2019. In general, Percentage of KPI Improvements show a growing tendency along the different reporting periods within the project, overtaking most of the predicted target values for energy efficiency, meaning that the range of improvement may move upwards.
Boost 4.0	Even though is not a direct aim of the project, most of the improvements that BOOST 4.0 brings imply a reduction of the energy use: - Optimizing the production process o Less idle time for machines o Less transportation of parts - Reducing the defective products and breakdowns o Less wasted energy used to produce bad products o Less restart of the machines after breakdown (less warm up waste) - Optimization of the AGV o Better routes imply less distance travelled and less energy consumed

Table 32: Contribution from cPPP projects to energy use reduction

11.3 Contribution to the reduction of CO_2 (KPI II.19)

30% projects have reported contribution to the reduction of CO_2 and in particular BigDataOcean, DataBio, TT, QROWD, BigDataStack, BigDataMedilytics, Boost4.0, CLASS, and ICARUS.

It is important to highlight that mostly all the projects contributing to this KPI are **linked to the transport and mobility sector** (maritime, aviation, terrestrial transport and mobility in general). Even projects focused on other sectors such as healthcare reductions of CO_2 relate to the reduction of needs or mobility.

Only one project, TT, provides evidence in measurements in CO_2 reductions reporting KPIs Improvements concerning **CO_2 emissions, savings reach up to 29% and reaching up to 23% of emission reductions in general (including** PM and NOx)

Table 33 provides details per project in relation to their contribution to the reduction of CO_2.

Project	Qualitative and quantitative statements (2018 and planned)
BigDataOcean	Better selection of locations for off-shore energy production

DataBio	(Input to energy reduction)
TT	In the case of vehicles, carbon dioxide emissions highly depend on energy consumption, therefore reductions are expected in proportion to energy consumption reduction. When energy savings are based on electricity supply, the CO_2 savings depend on the generation mix operating in that specific moment. Current results show that for the specific monitored items improvements in energy consumption per unit of activity range between 25% and 51%, considering contributions belonging to cars, and then standing for CO_2 emissions, and contributions belonging to savings in airport energy consumption, which comes from electric supply. So, taking the average of all KPIs Improvements concerning **CO_2 emissions, savings reach up to 29%**. However, as it has been mentioned, demonstrators will have reported their final impact values by the end of March of 2019. As well, besides CO_2 there are other emissions that are also being assessed, such as PM and NOx ones. Achievements show positive conclusions too, **reaching up to 23% of emission reductions**, taking the average value of all this emissions-related KPIs Improvements, taken from those pilots addressing this kind of indicators (WP4-Highways, WP5-Connected Vehicles, WP6-Railways, WP9-Urban Mobility) In general, Percentage of KPI Improvements show a growing tendency among the different reporting periods within the project, overtaking most of the predicted target values for energy efficiency, meaning that the range of improvement may move upwards.
QROWD	Less traffic and less carbon footprint will render lower levels of CO_2 emissions
BigDataStack	The Danaos shipping company is planned to optimize ships trajectories thanks to our project.
BigMedilytics	Some pilots involve remote monitoring of patients. So, they need to visit hospitals less often. That results in less travel and less usage of resources at the hospital.
Boost 4.0	Reduction of CO_2 emissions isn't either a main objective of BOOST 4.0, but many of the development carried out by the pilots contribute to reduce the CO_2 emissions, mainly by reducing the energy consumption, but also by improving the production process.
CLASS	CLASS will allow to collect information about the traffic of a city in real-time. On top of this, a pollution model is being built to derive the level of pollution of the monitoring area
ICARUS	PACE demonstrator aims at improving fuel consumption, reducing costs and minimizing the ecological footprint of airlines

Table 33: Contribution from cPPP projects to CO_2 reduction

11.4 Contribution to the reduction of waste (KPI II.20)

The 3 or the 4 lighthouse projects running in 2018 (DataBio, Transforming Transport and Boost4.0) have reported contribution to the reduction of waste.

In particular:

- **DataBio:** In forestry, the biomass Manager module of Wuudis solution tracks and provides the material balance of bio-material supply chains. This is a versatile solution and can handle any type of bio-material at any scale. Any material wastage can be easily identified through this solution. In addition, if also offers corrective suggestions to reduce material waste. **Up to 10% reduction in waste** is the experience from our user cases.
- **TT:** In the assessment methodology defined in WP2 there are no specific KPIs that will directly monitor the reduction of waste. However, it can be stated that the project contributes to decrease waste through savings in operational activities and improved management of assets. This will happen especially in those pilots that are implementing predictive maintenance activities (WP4 – Highways, WP6 – Railways, WP7 – Ports) that should extend the life of assets, parts and consumables. Current results concerning these pilots **show a percentage of improvement about 25%** which can fairly well demonstrate a relatively **high-level achievement of reduction waste itself at this final stage of the project**, expecting to be even higher after the final report on KPI values.
- **BOOST 4.0** contributes to the reduction of waste indirectly, in two main ways: reducing the defective products; and improving the part planification.

11.5 Contribution to the reduction in the use of material resources (KPI II.21)

17% projects running in 2018 have reported contribution to the reduction in the use of material resources. The 4 running lighthouses (TT, DataBio, BOOST4.0 and BigMedilytics) contribute to this KPI in addition to TheBuyForYou project. Only one project, BigMedilytics, provides quantitative data in a particular scenario reporting that the Asset Management pilot aims to reduce the no. of unused mobile assets in hospitals by up to 20%. Table 34 provides details per project about their contribution to waste reduction.

Project	Qualitative and quantitative statements (2018 and planned)
DataBio	This is within the scope of the forestry pilots, where we aim to completely remove all paper-based forest management plans, all daily routines over mobile and availability of forest open forest data directly and indirectly impact the reduction in material resources. Some of the agriculture pilots are leading to reduction of chemicals (mainly fertilisers) and also to reduction of fuel consumption. The reduction of chemicals is different from field to field and depends mainly on variability of soil properties in different parts of the field. Some pilots are focused on optimisation of irrigation which leads to reducing the amount of water needed for irrigation.
TT	In the assessment methodology defined in WP2 there are just a few KPIs that will directly monitor the reduction of material resources consumption even though through improved maintenance and operation, savings in that sense are probable. The defined KPIs in this matter are mainly linked to savings in fuel consumption and in predictive maintenance of assets, whose qualitative achievements have been already presented.

BigMedilytics	The Asset Management pilot aims to reduce the no. of unused mobile assets in hospitals **by up to 20%.** Planned to completely replace the current paper-based system for tracking readings by digital solutions.
Boost 4.0	The reduction of the defective products is one of the main contributions of BOOST 4.0 to the reduction of material resources, but not the only one. Also the better planification of the material needed, in case of smart maintenance, where elements (machines, vehicles…) are not subject to a fixed maintenance plan, where parts are changed based on fixed criteria, but to smart maintenance where the conditions of the element will indicate when the part needs to be changed. Scrap rate reduction between 1-5% is common to most cases dealing with defect reduction.
TheyBuyForYou	Improved procurement processes, based on data driven analytics, will ultimately lead to better assessment of material resources.

Table 34: Contribution from cPPP projects to waste reduction

12 STANDARDISATION ACTIVITIES (III.2)

There is a specific Subgroup under TF6 in BDVA (TF6.SG6) looking after coordination of Standardisation and interoperability activities in BDVA and the Big Data Value cPPP. This subgroup has a specific mandate to promote the work of BDVA members in ESBs, but also utilize existing links of BDVA members to the before mentioned international standardization bodies. Standardisation is of high importance for the creation of the EU Single Market. The main goal in the Big Data area is to leverage the European research and industrial capabilities in terms of Data Science. The standard related discussions in the BDVA have shown that it is of paramount importance to foster more industry participation in the development of international standards.

Big Data related sectors have several market needs related to standardisation and interoperability. Addressing them will assure the take-off of the Data Economy market.

On a general level, the subgroup recommends that interoperability and standards related activities should:

- leverage existing common standards (e.g. via open source reference implementations) as the basis for an open and successful Big Data market,
- integrate national efforts on an international (European) level as early as possible,
- ensure availability of experts for all aspects of Big Data in the standardization process,
- provide education and educational material to promote developing standards. Standards must look beyond the 'here and now' of how Big Data is currently being used and instead seek to establish frameworks for dealing with data sets that represent a significant logistical challenge.

During 2017 BDVA and the BDV CPPP set up some foundations defining priorities for the PPP in Big Data Standardisation as follows:

- **Establish an official liaison in between the BDVA Standards Group and the AIOTI WG3**, activity developed through different workshops during 2017 and implemented in 2018 with the signature of a MoU with AIOTI and common activities during the year.
- **Further develop the BDVA Reference Model pursuing alignment with others such as oneM2M, BDE Platform, AIOTI, RAMI 4.0**, etc, implemented through different workshops organised during 2017 and

2018.

- **Open an official dialogue with CEN, CENELEC and ETSI** on standards harmonisation, implemented through different workshops during 2017 and 2018. BDVA is intended to sign an MoU with CEN/CENELEC in 2019 and it is under discussions with ETSI.
- Create BDVA Roadmap for Big Data Standards harmonisation and industry engagement in Global Big Data standards development.

The following activities have been developed during 2018:

- At the end of 2017 BDVA established an official liaison with the ISO/IEC JTC1/WG9 with the main objective of channelling European input (cPPP) into global standards. As WG9 was integrated in a new Sub-Committee SC42 (Artificial Intelligence), BDVA has worked throughout 2018 set up at the beginning of 2019 an **official liaison with ISO/IEC JTC1/SC42.**
- BDVA attended a meeting of ITU-T FG-DPM meeting (Brussels, 20-23 February 2018) and assessed the possibility to establish a liaison with the objective of feeding the FG with use cases from the projects.
- BDV CPPP Big Data Standards session in Sofia (May 15th 2018) with the PPP projects (educational and knowledge sharing session).
- Trustworthy Artificial Intelligence: building a framework with standardization, organised by CEN/CENELEC. BDVA supported organisation and had a promotional stand at the event.
- World Standards Day conference[51] organised in Brussels on October 12th where BDVA participated as panellist in the panel: "Standards meet the future – towards the 5th industrial revolution".
- Session in EBDVF 2018 (Nov 12th 2018, Vienna) focused on setting up conditions in Europe to set a binding framework for international standards on AI. CEN/CENELEC, EC, and representatives from Industry were invited as panellist in the discussion.
- **Data Market Services project**, CSA of the BDV cPPP including as one of its pillars, standardisation support for SMEs, was selected for funding in 2018 and started in 2019.

In relation to the specific KPI, **over 30% of the projects** running in 2018 reported they perform **activities leading to data/Big data Standardisation** (some with a level of uncertainty).

Project	Activities performed by the project during 2018 leading to standardisation
BigDataOcean	Participation in relevant working group.
DataBio	Contributions to: • OGC Testbed 13 – Exploitation Platform Thread • OGC Testbed 14 – EOC Exploitation Platform Thread • OGC DCAT Best Practice Paper OGC 18-001 • OGC Big Data Domain Working Group
euBusinessGraph	Development of an ontology for company information.

51 https://www.cencenelec.eu/News/Brief_News/Pages/TN-2018-068.aspx

FashionBrain	We have standardized the aggregation of open-ended tasks on crowdsourcing platforms using AI models.
SPECIAL	WU have been involved in W3C Open Digital Rights Language (ODRL) working group which released two specifications in February 2018 https://www.w3.org/TR/odrl-model/ and https://www.w3.org/TR/odrl-vocab/. On the 17th and 18th of April 2018 the consortium hosted a W3C Workshop on Privacy and Linked Data at WU Wien, Austria https://www.w3.org/2018/vocabws/. The setup of a W3C Data Privacy Vocabularies and Controls Community Group (DPVCG) in order to address the needs of the broader business community https://www.w3.org/community/dpvcg/, which was launched the day the GDPR came into effect.
BigMedilytics	• Identification of clustering releasing recommendations to relevant scientific Societies • research publication about decentral AI execution in healthcare • We use DICOM and HL7 FHIR and may suggest and enhancement to HL7 FHIR to support medical analytics. • We are defining clinical items to characterise clinical outcome
BODYPASS	Protocols for secure exchange and aggregation of 3D data
Boost 4.0	BOOST 4.0 has a tight bond with the International Data Space Association as one of the main outcomes expected for the project will be a full operational implementation of the IDS ecosystem. To achieve that, we are working closely with the IDSA on the definition and the development of the IDS components. Launch of W3C Web Standardisation for Graph Data Workshop and the agreement to launch a W3C Business Group. Contribution to NIST Quality Information Framework QIF) standard elevation to ISO standard. Start discussion for Pro-STEP collaboration. Proof-of-concept on the application of AI in network security in ETSI ISG ENI.
TYPHON	The following technologies under development will be proposed as new or extensions to existing standards: TyphonML Modelling Language, TyphonDL Hybrid Polystore Deployment Language, TyphonQL Query Language
Cross-CPP	In 2018, Cross CPP has started working towards standardisation of data model for various cross industrial data which is known as CIDM – Cross Industrial Data Model.

Table 35: Activities performed by projects leading to standardisation activities

3 projects have reported contribution to European Standardization Bodies (ESBs) activities. In particular:

DataBio is involved in **6 Working Items** in European Standardization Bodies (ESBs) that are related to the topics of the PPP. In particular:

- OGC 18-001: GeoDCAT-AP Best Practice Report
- OGC 18-050: OGC Testbed 14, ADES & EMS Results and Best Practices ER
- OGC 17-023: OGC Testbed 13, EP Application Package Engineering Report.
- OGC 17-024: OGC Testbed 13, Application Deployment and Execution Service Engineering Report.

- OGC 17-003: GeoJSON(-LD) Metadata for EO Datasets
- OGC 17-084: GeoJSON(-LD) Metadata for EO Collections

It is also involved in **5 of Pre-Normative Research Files** under consultation in European Standardization Bodies (ESBs) that are related to the topics of the cPPP. In particular:

- OGC 18-001: GeoDCAT-AP Best Practice Report
- OGC 17-023: OGC Testbed 13, EP Application Package Engineering Report.
- OGC 17-024: OGC Testbed 13, Application Deployment and Execution Service Engineering Report.
- OGC 17-003: GeoJSON(-LD) Metadata for EO Datasets
- OGC 17-084: GeoJSON(-LD) Metadata for EO Collections

Boost 4.0 is involved in **3 Working Items** in European Standardization Bodies (ESBs) that are related to the topics of the PPP. In particular:

- ETSI GS ZSM002 (ETSI ZSM[52] framework reference architecture)
- ETSI GS ZSM005 (ETSI ZSM means of Automation)
- ETSI GS PDL002

Finally, Lynx project reports being involved in 2 **Working Items** in European Standardization Bodies (ESBs) that are related to the topics of the PPP but does not provide additional information in the specific topics.

20% of BDVA members reported its organisations perform activities leading to data/Big data Standardisation. Some examples include:

- **Institute of Advanced Studies Kőszeg** is involved in standardization of health care patient journey data related to National Heart Registry and payer's patient journey data in cancer diseases.
- **Huawei Technologies Düsseldorf GmbH** is involved and contributing to the new ISO/IEC JTC 1 SC 42 on AI
- **Inria** contributes to TEI (stdfSpec) and in ISO on TC37/SC4, ISO 2473 (Language resource management - Lexical markup framework (LMF)). They are also involved in WG of IETF, IRTF, OpenMP, MPI forum, W3C.
- **Fraunhofer-Gesellschaft** drives to standardise primarily data architectures, interoperability methods at this institute. Other institutes will likely have additional efforts but there are few means of obtaining such statistics.
- **Prizztech Robocoast** participates in ISO/IEC JTC1/SC42 Artificial Intelligence, currently at WG1 and WG3 as members of Finnish Standardization Association SR315 group.
- **Siemens AG** contributes to IEC, DIN DKE and other consensus-base standardisation bodies.

15% of BDVA members reported contribution to ESBs but they did not report specific data about the specific working items of Pre-Normative Research Files they are involved in.

52 Zero-touch network and Service Management

13 QUESTIONNAIRES TO COLLECT INFORMATION

13.1 Questionnaire for PPP projects running in 2018

(Questionnaire fill in through an online tool)

Section 0: Master Data

1. Name of your project:
2. H2020 Call topic addressed by your project:
3. Starting date and duration (in months) of your project:
4. Project coordinator (organization):
5. In case we have any questions to your answers we would like to ask you to provide your name and Email address:

Section I: Success Stories

6. List and describe **success stories** of your project during 2018 (open question)

(Describe relevant project outcomes and achievements here. Add links to media, applications, etc if needed. We appreciate having easy to read / concrete examples that help the reader to understand why this is a success story)

Section 2: job creation, job profiles and skills development

(Contributes to KPI I-2, II-5 and II-8)

7. Does your project contribute to job creation? (Job creation forecasted within a reasonable timeframe in the future, 3 years after H2020 ends)

   (Yes/No)

8. If yes, can you provide an estimation of the impact of your project in job creation forecasted within a reasonable timeframe in the future (or at least until 2022)? Please provide quantitative and qualitative answer if possible
   - Estimation of jobs created until 2022 (estimate a number)
   - Qualitative explanation of how the project will impact job creation (max. 2000 char)
9. Does your project contribute to increase the number of data workers in Europe? (Yes/No?)
10. If yes, please provide some qualitative statements about how your project is contributing to increase the number of data workers in Europe? (Please refer as much as possible to activities developed during year 2018) (Open question. Qualitative answer)
11. Does your project contribute to the creation of new job profiles? (Yes/No?).
12. If yes, how many complete new job profiles have been established by your project in (or by end of) 2018? List and describe the new job profiles created indicating skills required per profile
    - Number of new job profiles established by end of 2018 (number only)
    - List and describe the new job profiles created indicating skills required per profile (open text. 10.000 char)
13. If yes, how many complete new job profiles do you estimate will be created by the end of the project? Please list them (number + description)
    a. Number of job profiles (by the end of the project)

b. List and describe the new job profiles created indicating skills required per profile (open text. 1.000 char)

14. Does your project contribute to the creation/development of new skills? (Yes/No).
15. If yes, explain how your project has contributed to the development/creation of new skills during 2018, and list the new skills developed. (open question)
16. Number of Master students involved in your project in 2018 (number FTEs)
17. Number of PhD students involved in your project in 2018 (number FTEs)
18. Please indicate in which way your project has contributed and will contribute to the Skills Agenda for Europe[53]:
    i. During 2018. _____
    ii. Forecasted 2019+. _____
19. Number of training activities and programmes (such as tutorials, webinars, etc.) organized by your project in 2018 and number of people benefited by those activities (totals). (total number of activities, estimation of people outreached by training activities, description of those activities (name, channel, attendees):
    a. Number of training activities (total number 2018)
    b. Number of people who benefit of those activities (Estimation of a total number 2018)
    c. List and describe the training activities (name, channel, content, audience)
20. Number of European training programs involving 3 different disciplines established by your project by the end of 2018 (include number of programmes, list them and indicate participants per programme). Separate the number and the description

Section 3: Innovations and technical results

(Contributes to I-4, II-4, II-7, II-10, II-13, II-15, II-16, II-17)

Innovations to market

The European Commission has requested all the cPPPs to measure what is called "Number of significant innovations to market". This KPI concerns all developed items that have a marketable value, including products, processes, instruments, methods, technologies, spin offs, patents, ... It should involve all items directly linked to the cPPP projects as developed foreground, as well as any items beyond the scope of the project that is linked to the foreground.

We are asking the projects to list and briefly described all developed foreground, tangible and intangible assets, that have a marketable or at least an exploitable value, including (not only) products, processes, instruments, methods, technologies, spin-offs, patents,....

In 2018 input to this questions is given through a different excel questionnaire to facilitate your input and the assessment. You can find the file here. If you cannot access the file, ask your project coordinator or contact MR2018@core.bdva.eu.
Further explanations about this question and KPI are given in that file.

21. Can you confirm you understood this question is replied via a separate excel questionnaire and that you have access to it? (YES/NO)

[53] http://ec.europa.eu/social/main.jsp?catId=1223

European provision of big data value creation systems and technologies

The Big Data Value Contractual agreement also incorporates a KPI (specific to this Big Data programme) called: "Increased competitive European provision of big data value creation systems and technologies". For this we are asking the project to provide information about the "Number of systems and technologies developed in the relevant sector in cPPP projects (beyond state of the art)"

Theoretically this input should be a subset of your answers to the previous question.

22. Number of systems and technologies developed in the relevant sector in the project during 2018 (number and short description) (if your answer to this question is embedded in "Innovations to Market" questionnaire please indicate so in the description and just referred to the short name)

Support major sectors and major domains by Big Data technologies and applications

23. List the sectors and major domains supported by Big Data technology and applications developed in your project (the following domains listed and more can be added as needed: Health and healthcare, bio-economy, manufacturing, Transport and logistics, Mobility, Telecom, Earth Observation, Media, Retail, Energy, Finances, Public Services, Water and natural resources, Business services, Smart Cities, other)

Enable advanced privacy and security respecting solutions for data access, processing and analysis

24. Number of patents filed by your project that enable advanced privacy and security respecting solutions for data access, processing and analysis:
    a. Year 2018
    b. Forecasted 2019+
25. Number of publications by your project that describe advanced privacy and security respecting solutions for data access, processing and analysis:
    a. Year 2018
    b. Forecasted 2019+
26. Number of OSS contributions/ products/ SW components resulting from your project that enable advanced privacy and security respecting solutions for data access, processing and analysis (number and brief description):
    a. Year 2018
    b. Forecasted 2019+

New economic viable services of high societal value

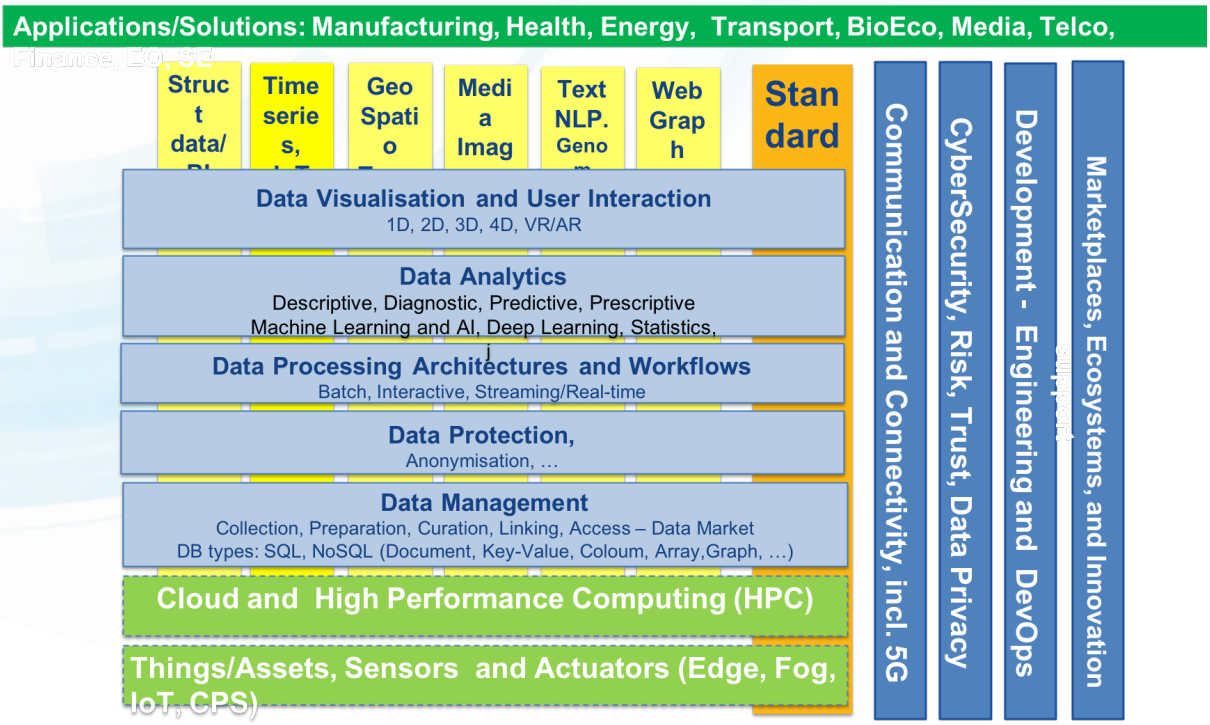
27. (if this is included in Q21, refer to the description with the short name provided in that response):
    a. Year 2018:
        i. Number
        ii. Description per service (describe why something qualifies)
    b. Forecasted 2019+
        i. Number
        ii. Description per service (describe why something qualifies)

    (Note: Criteria for "high societal value" should be up to the project. Project should describe why something qualifies)

Technology progress is in line with multi-annual roadmap of SRIA

With this KPI we intend to measure % of research priorities covered compared to overall scope of research priorities defined in SRIA (differentiate running, upcoming and not covered yet).

To answer this question, you need to use as reference the BDVA SRIA v.4.



Link to SRIAv4: http://www.bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf *Chapter 3 (overview in chapter 2.3, reference model)*

A separate excel file is used to collect information with the appropriate level of details. You can find the excel file here.
Use this question just to confirm you understood you have to submit a separate file and you have access to it.

(Note: This information was initially shared by some of the projects through the BDV Technical Committee. Please contact the person representing your project at the Technical Committee. Chairs of the TC and BDVA TF6 leads will be in charge of assessing results of your input)

28. Can you confirm you understood this question is replied via a separate excel questionnaire and that you have access to it? (YES/NO)

Availability of metrics for measuring the quality, diversity and value of data assets

(Note: Results of this questions will be shared with the BDV CPPP DataBench project)

29. Are you assessing quality, diversity and value of data assets? (YES/NO?)
30. if yes, what metrics are you using to quantify them? (Open Answer)

Increase the speed of data throughput compared to 2014

(Note: Results of this questions will be shared with the BDV CPPP DataBench project)

31. Does your project expect to improve data throughput? (YES/NO?)
32. If yes, what metrics are you using to measure this? (Open Answer)
33. If yes, what improvement (in percent) did you measure in your project? (Numerical Percentage Value).

Section 4: Experiments in Big Data

(Contributes to II-11, II-12 and II-14)

Large Scale experiments conducted in cPPP projects and i-Spaces involving closed data

34. Number of large-scale experiments conducted in your project in 2018 (number)
35. How many of those involved closed data? (number)
36. Please describe your criteria for an experiment to qualify as "large-scale" (e.g. cross-border / Geographical impact, Number of companies involved, Investment, Higher TRLs, number of users, etc) If possible, list the most relevant experiments and explain outcomes (open question)

Uptake of BDV use cases and experiments

37. Number of data experiments/use cases of any kind or size conducted in 2018 (number)
38. Please provide some qualitative data to support your answer (i.e list, brief description,..) (open answer)

Amount of data that has been made available for experimentation (cPPP projects and i-Spaces)

39. Number of Exabytes of data made available in your project in 2018 (including closed/private data)

(This includes data made available for experimentation (e.g. Data incubators and i-Spaces) and data made available internally in the projects for experimentation purposes, so closed data (e.g. Lighthouses and other projects). PLEASE ADD an approximative number and indicate the units (Terabytes, etc). DON't add explanations here.

40. Please add some qualitative information to support your answer (e.g. overview type of data, providers, etc)

Section 5: Contribution to Macro-Economics KPIs

(Contributes to II-1, II-2 and II-3)

41. Does your project contribute (has contributed or plan to contribute) to increase revenue share of EU companies against total of revenue of EU, US, Japan, Brazil? (YES/NO?)
42. If yes, please provide some qualitative statements about how your project is contributing to this. (Open Answer)
43. Does your project contribute (has contributed or plan to contribute) to increase the number of European Companies offering data technology, applications? (YES/NO?)
44. If yes, please provide some qualitative statements about how your project is contributing to this. (Open Answer)
45. Does your project contribute (has contributed or plan to contribute) to the revenue generated by European Data Companies? (YES/NO?)
46. If yes, please provide some qualitative statements about how your project is contributing to this. And examples of success stories (Open Answer)

Section 6: Contributions to environmental challenges

(Contributes to II-6, II-19, II-20, II-21)

Energy saved in big data analytics using a specified benchmark through solutions provided by specific cPPP projects compared to baseline at beginning of H2020

47. Are you monitoring Energy Efficiency in your project? (YES/NO?)
48. If yes, what metrics are you using to measure this parameter? (Open Answer)
49. If yes, what improvement (in percent) have you measured in your project? (Numerical Percentage Value).

Contribution to the reduction of energy use and CO2 emissions

50. Does your project contribute to the reduction of energy use? (YES/NO?)
51. IF yes, please provide some qualitative and quantitative statements (2018 and planned)
52. Does your project contribute to the reduction of CO2 emission? (YES/NO?)
53. IF yes, please provide some qualitative and quantitative statements (2018 and planned)

Contribution to the reduction of waste

54. Does your project contribute to the reduction of waste? (YES/NO?)
55. IF yes, please provide some qualitative and quantitative statements (2018 and planned)

Contribution to the reduction in the use of material resources

56. Does your project contribute to the reduction of material resources? (YES/NO?)
57. IF yes, please provide some qualitative and quantitative statements (2018 and planned)

Section 7: Mobilisation of stakeholders, outreach, success stories

(Contributes to section 2.2 of the monitoring report of the contractual PPPs)

58. Number of dissemination events, seminars, conferences organised by your project in 2018 (number)
59. Estimate the number of people outreached by your dissemination and engagement activities in 2018 (just estimation. Add only NUMBERs) (NEW QUESTION):
    - Total
    - In events (conferences, seminars, training,…)
    - Online (webinars, online training, etc..)
    - Via mass media (online or physical)
    - Others
60. List and describe all the activities performed in 2018 to mobilise and outreach to stakeholders in your project in 2018 (i.e. dissemination and communication activities of any kind, events, workshops, etc). Include in all cases the outreach in numbers of participants and in geographical outreach. (open question)
61. List your project main stakeholders (the ones you are addressing and need to outreach) and if applies tell us a good success story of engagement with your target audience (open question)

Section 8: Contribution to Standards

62. Does your project perform any activities leading to data/Big data Standardisation? (Please include also AI related activities) (yes/no?)
63. If yes, list and describe activities performed by your project during 2018 leading to standardisation
64. Does your project contribute (through one or several project partners) to European Standardization Bodies (ESBs) activities? (yes/no?)

If yes,

65. Indicate the number of **Working Items** in European Standardization Bodies (ESBs) that where your project is involved in and that are related to the topics of the PPP

(A Working Item is a registered topic in one of the ESBs that has been accepted as a matter of standardization activities).

66. List the IDs of the Working Items (list)
67. Indicate the number of **Pre-Normative Research Files** under consultation in European Standardization Bodies (ESBs) that your project is involved in and that are related to the topics of the cPPP. (A Pre-Normative Research File is a draft standard under consultation in one of the ESBs.) (number)
68. List the prEN numbers of these Research Files

Section 9: Contribution to Big Data Ecosystem

69. Does your project run or support a programme that is specifically targeted at supporting start-ups or entrepreneurs in the Big Data area and AI? (yes/no?)
70. If yes, please describe! (You will be contacted for further details, if necessary)
71. Is your project engaged in contributing to any National activities related to Big Data Value? (yes/no?)
72. If yes, provide the country, details of the activity, and an overview about your participation! Focus on 2018 activities Some short statements are sufficient. You will be contacted for further details separately.
73. Is your project present in EU13 Member States? (yes/no?)
74. If yes, provide country(ies), and overview of activities in that country.
75. Is your project linked to another European PPP, ETP or European industrial Association?
76. If yes, provide PPP/ETP/International Association names and briefly describe common/linked activities.
77. Does participation in a larger programme (Big Data Value cPPP) bring any value to your project? (Yes/No) (NEW QUESTION)
78. Please list and describe outcomes of your collaboration with other BDV CPPP projects or your participation in the BDV CPPP programme (e.g data shared with other projects, etc). We are particularly interested in capture success stories that can justify the existence of programme (and not only individual projects). Please focus on 2018 outcomes (Open question) (NEW QUESTION)

Section 10: Additional comments

79. Provide any additional information you may find relevant (if needed)
80. Provide comments to this survey (if needed)

13.2 Technical priorities table for projects - Questionnaire

Survey in excel distributed to all projects running in 2018, and selected ones starting in 2019.

It relates to the Q28 of the MR2018 questionnaire for projects.

Only input in the cells that are relevant for your project.

In "focus" mark with "x" main focus of your project and "o" challenge is addressed in a vertical way given by the main focus.

In the columns "number of" indicate a number (e.g 0, 1, 2, et) when applies.

You can use the column comments to specify more details if you consider so important.

(Add the name of your project here)		Delivered 2017	Delivered 2018	Ongoing (not delivered by initiatied in 2018 or earlier)	Planned (for 2019+)	
Technical priorities and challenges	Focus	Number of contributions/technologies (provided for MR2017)	Number of contributions/technologies 4	Number of contributions/technologies3	Number of contributions/technologies2	Comments
Data Management						
Semantic Annotation of unstructured and semi- structured data						
Semantic interoperability						
Data quality						
Data lifecycle management and data governance						
Integration of data and business processes						
Data-as-a service						
Distributed trust infrastructures for data management						
Other (specify)						
Data Processing Architectures						
Heterogeneity						
Scalability						
Processing of data-in-motion and data-at-rest						
Decentralizatrion						
Performance						
Novel architectures for enabling new types of big data workloads (hybrid Big Data and HPC architecture) -- new in SRIA4.0						
Introduction of new hardware capabilities -- new in SRIA4.0						
Other (specify)						
Data Analytics						
Semantic and knowledge-based analysis						
Content validation						
Analytics frameworks & processing						
Advanced business analytics and intelligence						
Predictive and prescriptive analytics						
High Performance Data Analytics (HPDA) -- new in SRIA 4.0						
Data analytics and Artificial Intelligence -- new in SRIA 4.0						
Other (specify)						
Data Protection						
Generic and easy to use data protection approaches						
Robust Data privacy (incl. multi-party computation)						
Risk based approaches						
Other (specify)						
Data Visualisation and User Interaction						
Visual data discovery						
Interactive visual analytics of multiple scale data						
Collaborative, intuitive and interactive visual interfaces						
Interactive visual data exploration and querying in a multi-device context						
Other (specify)						
Other (specify)						

13.3 Significant Innovations to Market – Questionnaire

Survey in excel distributed to all projects running in 2018.

It relates to the Q21 of the MR2018 questionnaire for projects.

The excel file requested the project list, describe and characterised innovations to market including the following information per innovation:

- Innovation name (short name if exists)
- Type of innovation (technology, system, methods, product, service, methods, instruments, processes, spin-off, patent, other)
- Brief description (Including the marketable/exploitable value)
- Impact (Significant/Medium)
- Impact description (if significant please justify that assessment)
- Impact type (Economic/Societal)
- TRL (if applies)
- Maturity level (e.g whether already supported by a prototype or demonstrator, filed or grant patent, developing or published standard, or spin-off created)
- Sectors
- Type of innovation (details in readme)
- Timing (Already reported in 2017 but progress in 2018 / Already reported in 2017 and no new progress in 2018 / New in 2018 / Expected in +2019)

The following guidelines were provided to support classification of innovations:

The European Commission has requested all the cPPPs to measure what is called "Number of significant innovations to market". This KPI concerns all developed items that have a marketable value, including products, processes,

instruments, methods, technologies, spin offs, patents, … It should involve all items directly linked to the cPPP projects as developed foreground, as well as any items beyond the scope of the project that is linked to the foreground.

We are asking the projects to list and briefly described all developed foreground, tangible and intangible assets, that have a marketable or at least an exploitable value, including (not only) products, processes, instruments, methods, technologies, spin-offs, patents,….

Note 1: It should involve all items directly linked to the cPPP projects as developed foreground, as well as any new foreground beyond the scope of the project that is linked to the project results (most likely through IP). Information on innovations ready to be taken to the market would be an asset given that the EU seeks opportunities to stimulate more interest by investors. Also, other activities could be considered to carry value added, especially at lower technology readiness levels, such as advancement of a certain technology, clustering, and standardisation. However, the measurability must contain some exploitable value. As an example, a (part) contribution to an industrial standard which has an intrinsic value as a method on the market would count as an innovation.

Note 2: Information gathered through this section will be shared with the BDVe projects (BDV CPPP CSA) with the purpose of identifying innovations developed by projects ready to be promoted by the BDV CPPP Market place

Types of Innovation:

- *Incremental Innovation: Improving existing technology to increase customer value (features, design changes, etc) within an existing market*
- *Disruptive Innovation: Applying a new technology or processes in an existing market*
- *Architectural Innovation: taking existing skills and technology and applying it in a different market*
- *Radical Innovation: New technology that creates new industries*

Significant impact: High impact regardless of the type of innovation (e.g. and incremental innovation can have a major impact if scaled reducing CO2 emissions). Justify this significant impact. Rest goes to Medium

13.4 Investments questionnaire

Distributed to all for-profit companies (large and SMEs) part of projects active in 2018 and to all for-profit members.

0	**Company name:**									
1	**Email Contact**									
2	**SME?:**	no								
3	**List all BDV PPP Projects you are involved in (if any)**	LIST OF PPP PROJECTS in "README" page. Please enter the list of projects your are involved in (or none) and remove this message								
4	**Age of the company (years) (for SMEs only)**									
5	**Country of European Headquarters**									
			2014	**2015**	**2016**	**2017**	**2018**	**2019-2023**	**Additional info requested** (and you can use for comments in general. Remove existing text)	**INFO / Clarifications**
6	**Regional cover of the following figures:**	Europe								KPIs refer to private investments within EUROPE (only). Please try to use figures in EUROPE. IF you need to provide World-wide or other regions (please indicate), at least for questions 8, 9, 10 and 11 focus on EUROPE only (and please indicate so)
7	**Turnover**	EUR	Only SMEs	Only SMEs	Only SMEs	Only SMEs			**For SMES**: Explain how your Turnover evolutions links to the Big Data Value PPP (projects or participation in BDVA)	This number is normally a public number published by companies as part of their financial reports
8	**Employment**	FTE	Only SMEs	Only SMEs	Only SMEs	Only SMEs			**For SMES**: If there are links in between your numbero FTEs and BDV PPP realted activities please explain here	**For SMES**: Please enter valus for the last 4 years as it is the evolution what needs to be assessed
9	**R&D expenses**	EUR								This number is normally a public number published by companies
10	*Estimate* of the **total amount of R&D expenses that are related to Big Data and Data-driven AI**	EUR								Linked to the Big Data PPP SRIA roadmap and supporting its objectives: Data storage, management, data tecnologies and platforms, data privacy, analytcs, visualization, Data-Driven AI, etc). Expenses can be CAPEX or OPEX (investments or operational expenses). All sources of funding considered. include private investment mobilised within the entities linked to the beneficiary as well as debt financing, venture capital and other investment proivded by Third parties to these
11	*Estimate* the **amount of R&D expenses that are related to the Big Data PPP (Data, Big Data and Data-Driven AI) but are NOT related to EC funded projects** (B2)	EUR							**ALL**: Please give of examples of those expenses that you can safely communicate	Expenses can be CAPEX or OPEX (investments or operational expenses). Sources can be own provite, private 3rd parties. include private investment mobilised within the entities linked to the beneficiary as well as debt financing, venture capital and other investment proivded by Third parties to these
12	*Estimate* the amount of R&D expenses resulting from follow-up investments of projects funded by the EC that are topic-wise related to the Big Data PPP ((Data, Big Data and Data-Driven AI) however initiated outside the Big Data PPP (in FP7 or in H2020). **Exclude** any expenses that have been funded by the EC! (complements B1)	EUR							List of FP7/H2020 projects related to the follow-up investments.	Expenses can be CAPEX or OPEX (investments or operational expenses). Sources of funding private own or 3rd parties. include private investment mobilised within the entities linked to the beneficiary as well as debt financing, venture capital and other investment proivded by Third parties to these
13	*Estimate* the **amount of R&D expenses resulting from follow-up investments of Big Data PPP projects. Exclude any expenses that are funded by the EC!** (B1)	EUR	0,00	0,00	0,00					Big Data PPP projects started just in 2017 so this figure of follow up investments is probably 0 except for SMEs incubated by DataPitch or simiar. Expenses can be CAPEX or OPEX (investements or operational expenses). Sources of funding private own or 3rd parties Partners part of Lighthouse projects are encorauge to estimate follow up investmente 2019-2023 if they have that information (if possible split by year)
14	*Calculate* **additional investment or resources spent by your company for the execution of cPPP projects, beyond direct contributions (co-funding reported).** (A2)	EUR	0,00	0,00	0,00					**(Only for BDV PPP project partners)** **Expenditure/invesments** to supportproject execution **and not captured in Corda** (EC reporting tools). Any private expenditure in project execusion not included in the 30% private contribition known by the EC and included in your reports. These are typically due to shortfalls in fixed overheads for a project, infrastructure costs not covered by direct funding, consumables, maintanence costs, additional personnel costs, indirect costs higher of 25% reimboursed, other costs) **YOU CAN INCLUDE PRIVATE (own or external), NATIONAL OR REGIONAL INVESTMENT**

Table 36: Investment questionnaire sent to for-profit companies

	2018
EC: EU contribution received for projects under the cPPP	Extracted from Corda
Direct leverage (A = A1 + A2)	
A1: Direct contributions from the industry to the cPPP projects, as predetermined in the projects (co-funding)	Extracted from Corda
A2: Additional investment or resources spent by your company for the execution of cPPP projects, beyond the direct contributions under A1 (*these are typically due to shortfalls in fixed overheads for a project, infrastructure costs not covered by direct funding, consumables and maintenance costs, equipment costs, additional personnel costs, other costs*)	EUR
Indirect leverage (B = B1 + B2)	
B1: Additional investment or resources spent as a follow-up to the activities and results related to cPPP projects (e.g. to exploit or scale-up results) and that are not supporting directly the execution of these projects. This can include additional funding mobilised thanks to the cPPP projects such as venture capital, other public funding, etc.	EUR
B2: Additional investments or resources spent that support the objectives of the cPPP and that have been triggered/encouraged by the existence of the cPPP, but that are not following up activities or results of the cPPP projects as such (hence, not counted in A1, A2 and B1).	EUR

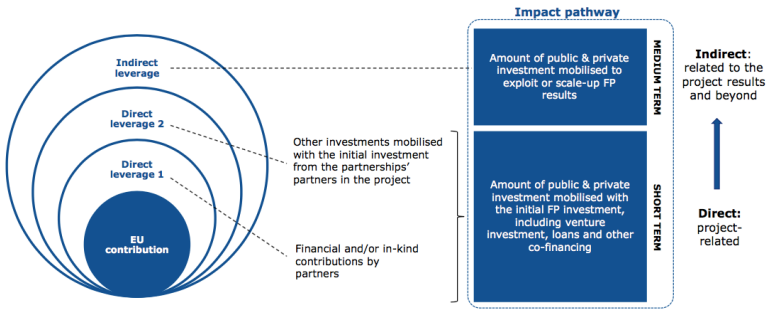


Figure 9: Table and illustration of the subset relationship of the expenses requested in (11) to (14)

13.5 Questionnaire for BDVA members

Distributed to all BDVA members. Responses collected online.

Section 0: Master Data

1. Name of your Organisation*
2. Membership type*
3. My organisation is a University or Non-for-Profit (Research/Innovation) Organization (Yes/No)*
4. My organization is an SME according to EC criteria (1. less than 250 employees, 2. annual turnover less than 50 MEUR or annual balance sheet less than 43 MEUR)*

   In case we have any questions to your answers we would like to ask you to provide your name and Email address*
5. In case we have any questions to your answers we would like to ask you to provide your name and Email address:

Section I: Success Stories

6. Provide any kind of success story of your organization during 2018 in relation to BDVA and the Big Data PPP e.g. where the PPP clearly has proven its value or where the Big Data PPP triggered investments inside your organization, etc. that we are able to publish in the Monitoring Report (open question)*

(Add links to media, applications, etc if needed. We appreciate having easy to read / concrete examples, text that we can "quote" that help the reader to understand why this is a success story)

Section 2: job creation, job profiles and skills development

(Contributes to KPI I-2, II-5 and II-8)

7. Has your participation in the BDVA/BDV CPPP contributed directly or indirectly to job creation? (Yes/No)*
8. If yes, can you provide an estimation of the impact in job creation during 2018 (in your organisation or organisations linked to you)?
   - Estimation of jobs created by end of 2018 (estimate a number)
   - Qualitative explanation (e.g. profiles, etc)
9. Does your organisation contribute to the creation of new job profiles in Big Data? (Yes/No?)*
10. If yes, how many complete new job profiles (in big data/AI) have been established by your organisation in 2018? Please list them:
    - Number of new job profiles established by end of 2018 (number only)
    - List and describe the new job profiles created indicating skills required per profile (open text. 10.000 char)
11. Does your organisation contribute to the creation/development of new skills (scope of Big Data Value)? (Yes/No)*
12. If yes, explain how your organisation has contributed to the development/creation of new skills during 2018, and list the new skills developed (scope of Big Data Value) (open question)

13. Please indicate in which way your organisation has contributed and will contribute to the Skills Agenda for Europe[54]:
    i. During 2018. _____
    ii. Forecasted 2019+. _____

14. Number of training programs in 2018 (such as tutorials, webinars, etc.) conducted by your organisation that have arisen from cPPP activities and number of people benefited by those activities (totals)* (Please don't include training programmes part of BDV CPPP projects and therefore reported as a PPP project activity. Ok if they are linked to topic and developed through other projects and programmes)
    a. Number of training activities (total number 2018)
    b. Number of people who benefit of those activities (Estimation of a total number 2018)
    c. List and describe the training activities (name, channel, content, audience)

15. Number of European training programs in 2018 involving 3 different disciplines conducted by your organisation that have arisen from cPPP activities in 2018 (include number of programmes, list them and indicate participants per programme)* (Please don't include programmes part of BDV CPPP projects)
    a. Total number of programmes involving 3 different disciplines in 2018
    b. List the programmes and number of participants per programme
    c. Any other comments

Section 3: Innovations and technical results

(Contributes to I-4, II-4, II-7, II-10, II-13, II-15, II-16, II-17)

Only for PPP projects

Section 4: Experiments in Big Data

(Contributes to II-11, II-12 and II-14)

(ONLY for projects or BDVA labelled i-Spaces)

Large Scale experiments conducted in i-Spaces involving closed data

16. Number of large-scale experiments conducted in your project in 2018 (number)
17. How many of those involved closed data? (number)
18. Please describe your criteria for an experiment to qualify as "large-scale" (e.g. cross-border / Geographical impact, Number of companies involved, Investment, Higher TRLs, number of users, etc) If possible list the most relevant experiments and explain outcomes (open question)

[54] http://ec.europa.eu/social/main.jsp?catId=1223

Uptake of BDV use cases and experiments

19. Number of data experiments/use cases of any kind or size conducted in 2018 (number)
20. Please provide some qualitative data to support your answer (i.e list, brief description,..) (open answer)

Amount of data that has been made available for experimentation (i-Spaces)

21. Number of Exabytes of data made available in your i-Space in 2018 (including closed/private data)

(This includes data made available for experimentation (e.g. Data incubators and i-Spaces) and data made available internally in the projects for experimentation purposes, so closed data (e.g. Lighthouses and other projects). PLEASE ADD an approximative number and indicate the units (Terabytes, etc). DON't add explanations here

22. Please add some qualitative information to support your answer (e.g. overview type of data, providers, etc)

Section 5: Contribution to Macro-Economics KPIs

(Contributes to II-1, II-2 and II-3)

Only for PPP projects (no need for members to input)

Section 6: Contributions to environmental challenges

(Contributes to II-6, II-19, II-20, II-21)

Only for PPP projects (no need for members to input)

Section 7: Mobilisation of stakeholders, outreach, success stories

(Contributes to section 2.2 of the monitoring report of the contractual PPPs)

Only for PPP projects and TF3/TF4

Section 8: Contribution to Standards

23. Does your organisation perform any activities leading to data/Big data Standardisation? (yes/no?)*
24. If yes, list and describe activities performed by your organisation during 2018 leading to standardisation
25. Does your organisation contribute to European Standardization Bodies (ESBs) activities? (yes/no?)

If yes

26. Indicate the number of **Working Items** in European Standardization Bodies (ESBs) where your organisation is involved in and that are related to the topics of the PPP

(A Working Item is a registered topic in one of the ESBs that has been accepted as a matter of standardization activities).

27. List the IDs of the Working Items (list)
28. Indicate the number of **Pre-Normative Research Files** under consultation in European Standardization Bodies (ESBs) that your organisation is involved in and that are related to the topics of the cPPP. (A Pre-Normative Research File is a draft standard under consultation in one of the ESBs.) (number)
29. List the prEN numbers of these Research Files

30. Does your organisation contribute to International Standardization Bodies (ISBs) activities? (yes/no?)*

If yes,

31. Please describe your activities linked to topics of the BDV CPPP (Big Data/AI)

Section 9: Contribution to Big Data Ecosystem

32. Does your organisation run or support a programme that is specifically targeted at supporting start-ups or entrepreneurs in the Big Data area? (yes/no?)*
33. If yes, please describe! (You will be contacted for further details, if necessary)
34. Does your organisation participate in National activities related to Big Data Value? (yes/no?)*
35. If yes, provide the country, details of the activity, and an overview about your participation! Some short statements are sufficient. You will be contacted for further details separately.
36. Does your organization participate in other European PPPs, ETPs or European industrial Associations?*
37. If yes, provide PPP/ETP/International Association names and briefly describe activities and link with Big Data PPP

Section 10: Additional comments

38. Provide any additional information you may find relevant (if needed)
39. Provide comments to this survey (if needed).

www.bdva.eu