

## 3.6 BIG DATA FOR PUBLIC ADMINISTRATIONS (2016.03)

### 3.6.1 IDENTIFICATION OF THE ACTION

Type of Activity	Study and reusable generic tools
Service in charge	DIGIT B4
Associated Services	DIGIT.B2, DIGIT.B6, DG RTD.E2, DG CNECT.R3, DG CNECT.F4, CNECT.02

### 3.6.2 EXECUTIVE SUMMARY

The amount of data generated worldwide increases at an astounding pace - by 40% annually and will increase 30 fold between 2010-2020. Due to the segmentation of public services and due to using non-interoperable means to describe data, most of the data generated in the public sector is not available for re-use.

This action, titled 'Big Data for Public Administrations', aims at addressing the use of big data within public administrations' as a means for enabling wiser decision making. With recent technologies such as big data, data mining, social media, cloud etc. organisations have greater potentials in generating, creating and storing data, information and knowledge thus providing greater opportunities for wisdom.

Consequently, the main objectives of this action are to identify concrete big data opportunities and requirements in public administrations and in specific policy contexts through which a number of kick-off pilots are to be undertaken as a means for creating or compiling tool sets that can be generalised and further extended in order to be used in different contexts. Finally, the action will continuously work towards launching new cooperation with policy DGs and the MSs through the execution of new pilots in order to accelerate the data driven transformation.

Phase 1 of this action was funded by the ISA programme and was executed in 2015. It aimed at carrying out a landscape analysis in order to identify (i) the requirements and challenges of public administrations in Europe and the Commission in the context of big data (ii) on-going initiatives and best practices in these areas including an assessment of the tools and solutions that these initiatives have implemented (iii) synergies and areas of cooperation with the policy DGs and the MSs in this domain. Furthermore, phase 1 also intends to execute some pilots that showcase the usefulness and policy benefit that big data can bring.

This action will continue to build upon the results of phase 1, focusing on the following activities:

- Track 1: continue with the identification of further opportunities and areas of interest whereby the use of big data could help improve working methods as well as ensure better policy making for policy DGs as well as Member States' public administrations.
- Track 2: continue the implementation of already identified pilots through generalising the developed functionalities and thus extending its use to policy agnostic contexts in order to maximise the benefit and return on investment of the proposed solution. As an example, the National Statistics Office of the Netherlands has already shown great interest in collaborating on one of the identified pilots, with a good potential of reuse of the generated outputs (more information in the Technical Approach section).

- Track 3: launch a new wave of pilots in specific domains which hold a potential of later being generalised and scaled-up to be made available to different services agnostic of their specific policy area.

### 3.6.3 OBJECTIVES

The **digital economy, innovation and services** feature high on the priority list of the EU as they are considered as driving elements for the growth of the economy in particular providing opportunities for growth and jobs. In an effort to push further these priorities the European Commission is currently in the process of identifying several actions that are considered as key features for a data-driven economy. Commission has presented a Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions titled '**Towards a thriving data-driven economy**'<sup>39</sup>. This communication focuses on the need for framework conditions that enable a single market for **big data and cloud computing**. The ability to handle and analyse big data is elementary in different areas particularly **scientific research, public sector and innovation**.

In the European Commission and in the Member States' contexts (as elsewhere) the purpose for **capturing data, information, and knowledge** is a vital element in being able to **make wise decisions**. More specifically, big data can help in the challenge of making sense of large data sets thus making sure that data that is most relevant to the organisation is **optimised and used** when taking critical decisions.

The **cross fertilization** of data, information and knowledge with big data technologies provides opportunities for the **exploitation of resources** that would otherwise not be available for influencing decision making processes within the organization and which would thus lead to inefficiencies or poor decisions. The use of big data allows for making **best use of new means of acquiring and spreading knowledge**. Traditional means of data collections are considered to be ineffective and expensive when dealing with a large population of participants, such as the EU citizens.

As a consequence to all this, the **main objectives** of this action are:

- 1) To further identify concrete big data opportunities and requirements in public administrations and in specific policy contexts. The action shall explore the requirements in at least 3 specific areas that shall be chosen based on the best practices identified through phase 1.**
- 2) To generalise and extend further the solutions produced through the already launched pilots, thus ensuring that they can be used by different policy areas and offer at least 25% of new functionality.**
- 3) To launch new co-operations with policy DGs and the MSs through the execution of at least 3 new pilots in order to accelerate the data driven transformation.**

Potential areas of action, where big data could have a real and direct impact are:

- 1) Foresight and agenda setting.** Making best use of available but unexploited data as a means for setting agendas and strategic foresight is possible through the use of big data technologies. The vast amount of data that is available to consider when developing a plan or

<sup>39</sup> <https://ec.europa.eu/digital-agenda/en/news/communication-data-driven-economy>

taking an action is massive and in most instances, the limited human resources available do not provide enough manpower to execute such resource intensive work.

**2) Improve evidence for evaluation assessments.** Current evidence data used in evaluation assessments are based on small data sources (desk research, surveys, interviews, expert groups, case studies etc.). These methods increase the risks of inaccurate estimations, inadequate extrapolations and false perceptions. The use of big data could bring more accuracy by detecting complex and subtle patterns in bigger, all-encompassing datasets and by increasing the capability to analyse short term patterns.

**3) Monitoring of legislative transpositions.** Whenever a new regulation or directive is adopted it is the obligation of Member States to ensure the correct transposition through appropriate implementation measures. The validation of the timely and correct transpositions into national legislation is a resource-intensive procedure for the European Commission and it is therefore foreseen that big data technologies, particularly text analytics, could considerably facilitate the monitoring process.

### 3.6.4 SCOPE

The details of the activities to be performed through this action are provided in the "Technical approach" section of this proposal. The following is a brief overview detailing the scope of the main activities to be performed:

- To **analyse the impact of big data on the policy making life-cycle**;
- To **identify best practices and lessons learnt** in the area of public administrations, more specifically in the areas of policy life-cycle and big data, including an assessment of the **benefits, risks and impacts** of these initiatives;
- To analyse the technical **tools and solutions** that these initiatives have implemented;
- To **identify the requirements and challenges** of public administrations in Europe and the Commission in the context of big data;
- To **identify synergies and areas of cooperation** with the **policy DGs** and the **MSs** in the this domain;
- To **execute some pilots** that showcase the usefulness and policy benefit that big data can bring;
- To **identify areas of interests for the ISA<sup>2</sup> program** to lead and fund future **initiatives** that will allow **practical implementations** that will answer the requirements of the public administrations in Europe.

### 3.6.5 PROBLEM STATEMENT

The amount of data generated worldwide increases at an astounding pace - by 40% annually and will increase 30 fold between 2010-2020. Due to the segmentation of public services and due to not using interoperable means to describe data, most of the data generated in the public sector is not available

for re-use. According to one report<sup>40</sup>, making use of big data techniques could lead to a potential annual savings of 150-300 billion EUR by means of creating more transparency, enabling experimentation to discover needs, segmenting target audience for customised actions, replacing or supporting human decision making and lastly by new, innovative business models and products. The Communication from the Commission titled 'Towards a thriving data-driven economy'<sup>41</sup>, highlights that the annual **growth of the big data sector is equivalent to 40%**, making it one of the strongest assets for economy growth, posing substantial opportunities that have so far not been reaped by the European market players. It identifies **a number of key actions for the EU** to help seize the opportunities of this sector amongst which is the **provision of the enabling technologies and underlying infrastructures and skills** as well as the **provisioning of public data resources and research data infrastructures**.

The **opportunity for public service transformation is real**. Whereas one aspect of big data is the ability to handle the data itself, another aspect is the ability to analyse the data allowing for better results, processes and decisions. Big data analytics can be an immensely powerful tool for helping organizations to learn about how they work. Traditionally, managers and public sector leaders have looked at a relatively small set of key performance indicators to assess the health and efficiency of their organisations. Digitisation has massively increased the quantity of management information available, the resolution and frequency at which it is captured, and the speed at which it can be processed. Data on inputs, outputs, productivity and processes can all be captured and recalled in more comprehensive detail than ever before. So far most organisations have been unable to consume such large data sets, despite being at their disposal either because consuming the data without the appropriate tools is too time consuming or in certain instances even not possible. One particular instance is the difficulty for organisations to be able to analyse research results coming from different domains related to a common topic. In this case the amounts of data to be analysed can be considerably large and in most instances not feasible to be done manually through human resources. In an attempt to tackle such challenges this action is executing a proof of concept in order to demonstrate the use of text mining techniques used on large amounts of unstructured research papers as a means of identifying areas of interest that a particular research area should consider prior to launching calls for new grants.

Furthermore, **big data opens up the realm of reliable predictive analytics**. By examining the relationships embedded in large datasets it is possible to build a new generation of models describing how things are likely to evolve in the future. This approach can be combined with scenario planning to develop a series of predictions of how a system will respond to different policy choices. The state of the art in **predictive analytics** can deliver forecasts for some domains with a very high degree of precision, providing an auditable, scientific basis for making decisions in complex systems. These techniques could be used to improve evidence in the context of the evaluation assessment process.

### 3.6.6 EXPECTED BENEFICIARIES AND ANTICIPATED BENEFITS

Beneficiaries	Anticipated benefits
---------------	----------------------

<sup>40</sup> McKinsey Global institute - Big data: The next frontier for innovation, competition, and productivity

<p>- EU institutions - Member States public administrations (national, regional, local levels)</p>	<p>The development of an action which will give a background basis for the development of open-government initiatives in a coherent way will provide Member States public administrations, DGs at the European Commission, EU institutions, and European agencies at all levels with the following benefits:</p> <ul style="list-style-type: none"> <li>• A harmonised ways to manage big data resulting in more effective and informed actions by public administrations.</li> <li>• Support public administrations to become more modern, adaptive, responsive, dynamic, flexible organisations and meet better the expectations of their stakeholders.</li> <li>• Reach out to citizens through consideration of opinions and data sources beyond the traditional means. This will ensure a more effective process for open government activities.</li> <li>• An increase of the efficiency, e.g. by: <ul style="list-style-type: none"> <li>○ Reducing time to make the right decisions;</li> <li>○ Getting the right knowledge at the right time; and</li> <li>○ Making use of tacit knowledge to support policy making.</li> </ul> </li> <li>• Building cost-efficient solutions implementing the actions based on “lessons learnt” and “risks to avoid” highlighted on the framework, thus reducing the “time to market”.</li> </ul>
<p>- Citizens and business in Europe</p>	<ul style="list-style-type: none"> <li>• Innovate through the reuse of open source software for knowledge discovery.</li> </ul>

### 3.6.7 RELATED EU ACTIONS / POLICIES

Action / Policy	Description of relation, inputs / outputs
<p><b>Digital Agenda</b></p>	<p>Pillar I: Digital Single Market pillar Within this pillar there is an action that promotes government bodies at all levels: local, regional, national, European and international, to open up and disseminate the vast amounts of information that is collected as part of their tasks in order to allow for the reuse of such information as well as a means for improving transparency of organisations.</p> <p>Pillar II: Interoperability &amp; Standards This pillar addresses the need for interoperability and standards as a means for allowing for data originating from different sources to be interoperable.</p>
<p><b>COM(2014) 442</b></p>	<p>Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.</p> <p>This action addresses the need to provide the right framework conditions for a single market for big data and cloud computing as a means for helping accelerate the transition towards a data-driven economy.</p>

<b>ISA Action 4.2.5.- Sharing and reuse strategy</b>	The reusable practices and guidelines as well as the reusable software solutions delivered through this proposed action can adopt the strategy defined in Action 4.2.5 in order to ensure that the outputs are shared and re-used with public administrations in Europe.
<b>ISA Action 4.2.1.- Integrated collaborative platform – Joinup</b>	The Joinup collaborative platform shall be used as a means for sharing the experiences as well as the deliverables of this action with the Member States' public administrations.
<b>ISA Action 1.1.- Improving semantic interoperability in European eGovernment systems</b>	Reusable interoperability solutions, core vocabularies, the linked data pilots, studies (e.g. on the business value of linked data) and other resources made available through Joinup and which might be relevant to this action shall be explored and reused as much as possible in order to ensure interoperability as well as continuation of existing efforts.

### 3.6.8 REUSE OF SOLUTIONS DEVELOPED BY ISA, ISA<sup>2</sup> OR OTHER EU / NATIONAL INITIATIVES

The final output of phase 1 will greatly impact the choice of solutions that might be reused by this action. Currently, the reuse of the following initiatives can be identified:

- European Classification of Skills/Competences, Qualifications and Occupations (ESCO) developed by the European Commission and linked to ISA Action 1.1 detailed in Section 1.1.7 above. This initiative shall be necessary in the implementation of the pilot within the job vacancies domain defined for phases 2 and 3 in the Technical Approach Section.
- Joinup, developed under ISA Action 4.2.1, shall be used as the platform for sharing the deliverables of this action to the various stakeholders.

### 3.6.9 EXPECTED RE-USABLE OUTPUTS (solutions and instruments)

All outputs generated by each of the pilots shall be provided in a re-usable format. The following are concrete re-usable outputs that can be identified at this stage:

<b>Output name</b>	RTD Pilot Software Components
<b>Description</b>	The software components that shall be developed or re-used as part of the RTD proof of concept to be implemented in phase 1 of this action shall be referenced and/or made available for re-use by other actions.
<b>Reference</b>	This output shall be made available through the Joinup platform.

Target release date / Status	Q2 2016 / currently under development
------------------------------	---------------------------------------

Output name	MOVIP Pilot Software Components
Description	The software components that shall be developed or re-used as part of the MOVIP pilot to be implemented in phases 2 and 3 of this action shall be referenced and/or made available for re-use by other actions.
Reference	This output shall be made available through the Joinup platform.
Target release date / Status	2017 / currently under development

### 3.6.10 ORGANISATIONAL APPROACH

#### 3.6.10.1 Expected stakeholders and their representatives

Stakeholders	Representatives
Interoperability solutions for European public administrations (ISA)	The ISA <sup>2</sup> Committee will oversee the project, with the assistance of the ISA <sup>2</sup> Coordination Group. The ISA unit (DIGIT.B6) is an associated service of this project, participates in the definition of requirements and guarantees continuation and exploitation of work conducted in other ISA Action and particularly Action 1.1 on semantic interoperability.
DIGIT.B4 Digital Business Solutions – Corporate Financial Procurement & Policy Solutions Unit	This unit is the service in charge of this action. It will coordinate the interaction between the different stakeholders within the European Commission and MSs which may express an interest for a more active involvement in the work.
DIGIT.B2 Corporate knowledge and decision making solutions Unit	This unit is the service in charge of Information System development supporting document management and corporate decision making processes. This unit an associated service of this project in the context of the execution of the pilot for briefing preparations, taking part in the definition of requirements, the development and deployment of the solution.
DG RTD.E2	This unit is an associated service of this project and will take part in the definition of the requirements, the performance and the guidance

	of the proof of concept planned to be executed in the context of research paper analysis through text mining within phase 1 of this action.
DG CNECT.R3 Support Systems and Tools Unit	This unit is an associated service of this project and will take part in the definition of the requirements, the performance and the guidance of this action. This unit will liaise with other relevant units within DG CNECT.
DG CNECT.02 Knowledge Sharing Unit	This unit is an associated service of this project and will take part in the definition of the requirements, the performance and the guidance of this action.
DG CNECT.F4 European Semester and Knowledge Base	This unit is an associated service of this project and will take part in the definition of the requirements, the performance and the guidance of the job vacancies pilot project (MOVIP) planned for phases 2 and 3 of this action.

### 3.6.10.2 Communication plan

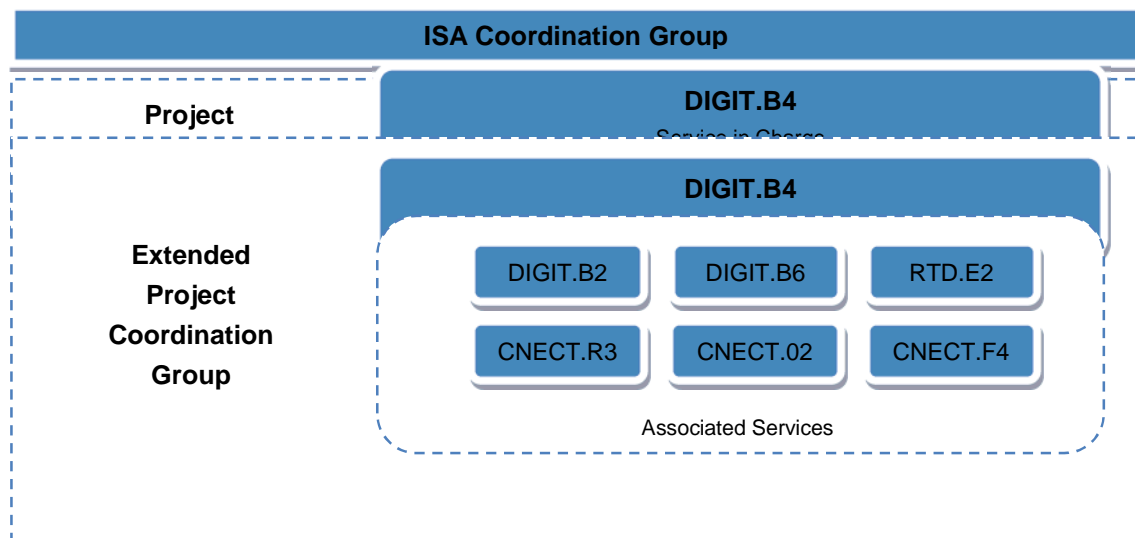
Event	Representatives	Frequency of meetings / absolute dates of meetings?
Bilateral meeting with Member States	DIGIT.B4 representatives, Member States representatives	These are arranged by DIGIT B4 on ad hoc basis.
Bilateral meetings with Policy DGs	DIGIT.B4 representatives, Policy DGs representatives	These are arranged by DIGIT B4 on ad hoc basis.
Relevant conferences and events	DIGIT.B4 representatives with any other project stakeholder	On occasions whereby DIGIT is invited to participate in meetings organised by Member States, policy DGs etc.
ISA Trusted Information Exchange Cluster	DIGIT.B4 team members, Stakeholder's representatives, ISA unit	These meetings are arranged by the ISA unit.
Joinup	(online)	



### 3.6.10.3 Governance approach

This action is part of the ISA programme and therefore it follows the ISA governance structure with the action reporting back to the Member States' representatives pertaining to the ISA working groups.

This action will actively involve all associated services at each of the different stages. The governance structure for this action is depicted in the diagram below.



In order to achieve its objectives this project will work closely together with primarily identified associated DGs to better define the business needs and challenges, bearing in mind the need of potentially other services in order to come up with generic and extendable software solutions.

Particular attention will be given to the coordination activities in order to ensure a successful undertaking of the different activities.

### 3.6.11 TECHNICAL APPROACH

During 2015 this action focused on the execution of a **landscape analysis** in Europe, in the areas of big data. The analysis assesses current good practices including an assessment of tools and solutions used as well as the identification of future plans in Member States and EU institutions and their relations to policy actions.

The purpose of the analysis, which has been defined as **Phase 1** of the action, is to:

- identify the **requirements and challenges** public administrations in Europe are confronted with in the area of big data and identify opportunities;
- identify **best practices** by public administrations and/or organisations which could be used as **lessons learnt** including an assessment of the **tools and solutions** that these best practices have implemented;
- identify **synergies** and **areas of cooperation** with the **policy DGs** and the **MSs** in the big data domain;

- execute a **proof of concept** that showcases the use of big data in the EC research domain, in cooperation with DG RTD, in order to prove the usefulness and policy benefit that big data can bring. This proof of concept shall demonstrate the use of text mining techniques used on large amounts of unstructured research papers as a means of identifying areas of interest overlap that a particular research area should consider prior to launching calls for grants;
- identify **areas of interests** whereby the ISA<sup>2</sup> programme and its proposed successor could have an **active role in launching initiatives** for enabling **practical concrete implementations** that will answer the requirements of the public administrations in Europe.

The work performed during 2015 shall be continued in 2016 and 2017, through phases 2 and 3, in order to ensure that the benefits of using big data in different policy areas are continued and further explored. The following are the planned phases and the respective planned activities:

**Phases 2 and 3** shall be composed of **3 tracks** each:

- **Track 1:** continue with the identification of **further opportunities and areas of interest** whereby the use of big data could help **improve working methods** as well as ensure **better policy making**. These efforts shall be done in close **cooperation** with both policy DGs as well as with Member States' public administrations and shall cover the further **assessment of not only tools and solutions but also standards and vocabularies** that are identified to be mature enough to be considered as potential enablers to be reused by either MSs or Commission Services;
- **Track 2:** continue the implementation of already identified pilots through **generalising the developed functionalities** and thus **extending** its use to **policy agnostic contexts** in order to maximise the benefit and return on investment of the proposed solution. This track shall also ensure that further analysis of the needs derived from these pilots is performed and that further functionality is provided whereby future potential of generalisation and re-use is identified. Such extension of efforts shall be performed on the following pilots:
  - o the proof of concept within the EC research domain executed in cooperation with DG RTD;
  - o the pilot within the job vacancies domain, executed in cooperation with DG CNECT. This pilot aims at using web sources in order to identify information related to ICT vacancies and their related skills in different geographical areas. This information is to be used as a basis for funding training targeting any shortage in specific ICT skills in a specific geographical area. The first phase of the project is being funded by DG CNECT but the inclusion within this action shall ensure that the work performed is generalised beyond the ICT domain and made available for national public administrations, such as National Statistics Offices, to be reused within their local contexts. DIGIT.B4 has presented this action to the National Statistics Offices during the Big Data Taskforce meeting organised by ESTAT and which was held in June in The Hague. The National Statistics Office of the Netherlands has shown great interest in future collaborations on this pilot with a good potential of reuse of the outputs generated through this pilot.
- **Track 3:** launch a **new wave of pilots** in specific domains which hold a **potential** of later being **generalised** and **scaled-up** to be made available to different services **agnostic of their specific policy area**. The following are a number of potential areas of interest that could be explored:
  - o A number of business cases whereby big data could play an important role are to be identified through the work executed by DG CNECT.02 titled 'Big data and other innovative data-driven approaches for evidence-informed policy making'.

- Monitoring of legislative transpositions is an area currently being assessed by DG GROW as part of the Single Market Strategy. This area could greatly benefit from the exploitation of big data and big data technologies such as text analytics as a means of facilitating the work of monitoring the timeliness and correctness of legislative transpositions.
- Currently all briefings preparations within the European Commission are prepared manually with the knowledge that an individual or a service has on the topic. Investigation on related documents needs to be done manually and in most instances this can be relatively time consuming. Consequently, this stream shall look into the option of using big data technologies in assisting individuals preparing briefings by providing the appropriate references to related material and documentation. This activity shall be done in cooperation with DIGIT.B2.
- Analysis of open procurement data using data analytics tools for the early detection of actual or prospective anomalies.
- Finally, further areas of interest may be identified through the work executed in phase 1 of this action, which is further detailed in the first part of this section.

These different sources could provide information on opportunities for pilots either with Member States' public administrations or else with different policy DGs.

### 3.6.12 COSTS AND MILESTONES

#### 3.6.12.1 Breakdown of anticipated costs and related milestones

Phase: Inception Execution Operational	Description of milestones reached or to be reached	Anticipated Allocations (KEUR)	Budget line ISA <sup>2</sup> / others (specify)	Start date (QX/YYYY)	End date (QX/YYYY)
Execution	<u>Phase 2.1</u> D.1.- MS and EC requirements assessment D.2.- Inventory of tools and solutions, standards and vocabularies available for reuse	100	ISA <sup>2</sup>	Q2/2016	Q1/2017
	<u>Phase 2.2</u> D.3.- Specification definition D.4.- Generalisation of core platform and modular components D.5.- Extension of core and modular components in the	250	ISA <sup>2</sup>	Q2/2016	Q1/2017

	context of the pilots				
	<u>Phase 2.3</u> D.6.- Pilot requirements assessment D.7.- Specification definition D.8.- Implementation of core and modular components in the context of the pilots	250	ISA <sup>2</sup>	Q3/2016	Q1/2017
Execution	<u>Phase 3.1</u> D.1.- MS and EC requirements assessment D.2.- Inventory of tools and solutions, standards and vocabularies available for reuse	100	ISA <sup>2</sup>	Q2/2017	Q1/2018
	<u>Phase 3.2</u> D.3.- Specification definition D.4.- Generalisation of core platform and modular components D.5.- Extension of core and modular components in the context of the pilots	400	ISA <sup>2</sup>	Q2/2017	Q1/2018
	<u>Phase 3.3</u> D.6.- Pilot requirements assessment D.7.- Specification definition D.8.- Implementation of core and modular components in the context of the pilots	400	ISA <sup>2</sup>	Q3/2017	Q1/2018
Operational					
	<b>Total</b>	1500			

### 3.6.12.2 Breakdown of ISA<sup>2</sup> funding per budget year

Budget Year	Phase	Anticipated allocations (in KEUR)	Executed budget (in KEUR)
2016	Execution	600	
2017	Execution	900	

### 3.6.13 ANNEX AND REFERENCES

Description	Reference link	Attached document
A vision for Public Services	<a href="http://ec.europa.eu/digital-agenda/en/news/vision-public-services">http://ec.europa.eu/digital-agenda/en/news/vision-public-services</a>	
Data for Policy - A Study for big data and other innovative data-driven approaches for evidence-informed policy making	<a href="http://www.data4policy.eu/">http://www.data4policy.eu/</a>	