



# Research Directions and Recommendations (pending EC approval)

This document is issued within the frame and for the purpose of the *Big Policy Canvas* project. This project has received funding from the European Union's Horizon2020 Framework Programme under Grant Agreement No. 769623. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.

This document and its content are the property of the *Big Policy Canvas* Consortium. All rights relevant to this document are determined by the applicable laws. Access to this document does not grant any right or license on the document or its contents. This document or its contents are not to be used or treated in any manner inconsistent with the rights or interests of the *Big Policy Canvas* Consortium or the Partners detriment and are not to be disclosed externally without prior written consent from the *Big Policy Canvas* Partners.

Each *Big Policy Canvas* Partner may use this document in conformity with the *Big Policy Canvas* Consortium Grant Agreement provisions.

(\*) Dissemination level.-**PU**: Public, fully open, e.g. web; **CO**: Confidential, restricted under conditions set out in Model Grant Agreement; **CI**: Classified, **Int** = Internal Working Document, information as referred to in Commission Decision 2001/844/EC.

---

## Document Information

---

List of Contributors	
Name	Partner
Eleni Kanellou	NTUA
Ourania Markaki	NTUA
Panagiotis Kokkinakos	NTUA

<b>Document name:</b>	Research Directions and Recommendations	<b>Page:</b>	2 of 20				
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b>	Final

# Table of Contents

Document Information .....	2
Table of Contents .....	3
List of Tables.....	<b>Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.</b>
List of Figures .....	5
1 Approach.....	6
2 Roadmap Recommendations.....	7
3 Content-specific Recommendations.....	12
3.1 Cluster 1 - Privacy, Transparency and Trust .....	15
3.1.1 RC 1.1 - Big Data nudging .....	15
3.1.2 RC 1.2 - Algorithmic Bias and Transparency .....	15
3.1.3 RC 1.3 – Open Government Data.....	15
3.1.4 RC 1.4 - Manipulation of statements and misinformation .....	15
3.2 Cluster 2 - Public Governance Framework for Data Driven Policy Making Structures .....	16
3.2.1 RC 2.1 - Forming and monitoring of societal and political will.....	16
3.2.2 RC 2.2 - Stakeholder/Data-producer-oriented Governance approaches.....	16
3.2.3 RC 2.3 - Governance administrative levels and jurisdictional silos.....	16
3.2.4 RC 2.4 - Education and personnel development in data sciences .....	16
3.3 Cluster 3 - Data Acquisition, Cleaning and Representativeness .....	17
3.3.1 RC 3.1 - Real time big data collection and production .....	17
3.3.2 RC 3.2 - Quality assessment, data cleaning and formatting .....	17
3.3.3 RC 3.3 – Representativeness of data collected.....	17
3.4 Cluster 4 - Data Storage, Clustering, and Integration.....	17
3.4.1 RC 4.1 - Big Data Storage and processing .....	17
3.4.2 RC 4.2 - Identification of patterns, trends and relevant observables.....	18
3.4.3 RC 4.3 - Extraction of relevant information and feature extraction .....	18
3.5 Cluster 5 - Modelling and Analysis with Big Data .....	18
3.5.1 RC 5.1 - Identification, acceptance and validation of suitable modelling schemes inferred from existing data.....	18
3.5.2 RC 5.2 – Collaborative model simulations and scenarios generation .....	19
3.5.3 RC 5.3 – Integration and re-use of modelling schemes.....	19
3.6 Cluster 6 - Data Visualisation.....	19

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	3 of 20	
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b>	Final

3.6.1	RC 6.1 – Automated visualisation of dynamic data in real time .....	19
3.6.2	RC 6.2 - Interactive data visualisation.....	19
	References .....	20

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	4 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b> Final

# List of Figures

---

*Figure 3-1: The BigPolicyCanvas Knowledge Base* ..... 7

*Figure 3-2: BigPolicyCanvas Panorama (3-dimensions mapping)* ..... 8

*Figure 3-3: BPC Assessment Framework Axes of Application* ..... 9

*Figure 4: Visual representation of the linkages among the recommendations* ..... 14

*Figure 5: Example* ..... 14

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	5 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b> Final

# 1 Approach

The elaboration of a set of practical recommendations to interested stakeholders constitutes, as stated in deliverable D5.1, the third and last step of the Big Policy Canvas roadmapping exercise, targeting to accelerate the roadmap's take-up and crowning project endeavors to promote (big) data-driven policy making in the public sector. As such, it builds on the outcomes of the preceding roadmapping steps and more specifically on the definition of a set of future research challenges, which makes up the core activity of the roadmap, while also leveraging the experience accumulated by the rest of the activities of the project.

Content-wise the elaborated set of recommendations is divided into two key parts, namely:

- Methodological recommendations, addressed to policy makers and public administration representatives, aiming at supporting and facilitating the adoption of the roadmap for data-driven policy making and the benefits lying therein and
- Content-specific recommendations, addressed, besides the aforementioned groups of stakeholders, to researchers and the industry as well, and targeting to delineate future policy strategy.

Methodological recommendations build, as already suggested, on the experience gathered in the frame of key activities of the project, such as the creation of the Big Policy Canvas Knowledge Base or the establishment of the affine community and are instrumental not only to the adoption of the roadmap itself but also to ensuring sustainability of the Big Policy Canvas outcomes.

Content-specific recommendations on the other hand result from the transformation of the identified research challenges into concrete proposals for action. In this respect, they reflect the needs, trends and problems identified, and take into account the current capacity and future potential of technologies and tools to deal with the former. Content-specific recommendations benefit also from the exchange of ideas with the project community of stakeholders and in particular the comments gathered in the online version of the roadmap, thereby achieving validation by and being representative of the opinions of the latter.

Content-specific recommendations are exposed at two levels of granularity. The first level corresponds to the research clusters and individual challenges identified and further encompasses research and policy recommendations (the latter pertaining to technological but also societal, cultural, political and economic aspects). The second one corresponds to higher level recommendations, touching upon horizontal aspects and framing key guidelines for future policy making.

Following the above distinction, Section 4 presents the methodological recommendations, whereas Section 5 exposes the content-specific recommendations.

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	6 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b> Final

## 2 Roadmap Recommendations

Big Policy Canvas has been guided by the vision to transform the public sector into an effective, efficient, precise, consistent and evidence-based policy making structure. In this respect, it has laid particular emphasis on the potential of Big Data technology for the design, development and implementation of effective, evidence-based and precise policies and has targeted the development of an ever evolving methodological approach, backed by an active community for the introduction of big data related innovations in the policy making process.

BigPolicyCanvas methodological approach is grounded on key project outcomes, including the BPC Needs and Trends Assessment Framework, the BPC Knowledge Base, the BPC Research Roadmap and the BPC Community. The former outcomes make up the basis for the set of methodological recommendations provided in the following paragraphs. These recommendations have to be seen under the prism of the road mapping exercise that has been the ultimate goal of the BigPolicyCanvas project.

During its course BigPolicyCanvas has designed and populated the BPC Knowledge Base, a state-of-the-art, online, dynamic repository, structured along the dimensions of Needs, Trends and Assets. This repository is

- *state-of-the-art* in the sense that it incorporates the project findings on current and emerging needs and trends impacting public administrations and the policy making process in specific, but also on the pool of technological and methodological assets that can be used to accommodate them;
- *online* in the sense that it is integrated into and made accessible through the BPC web site;
- and *dynamic* in the sense that it further provides a comprehensive mapping among the aforementioned three dimensions, thereby the needs, trends and assets.

The information presented for each of the aforementioned items follows a standard structure, as shown in Figure 2-1.

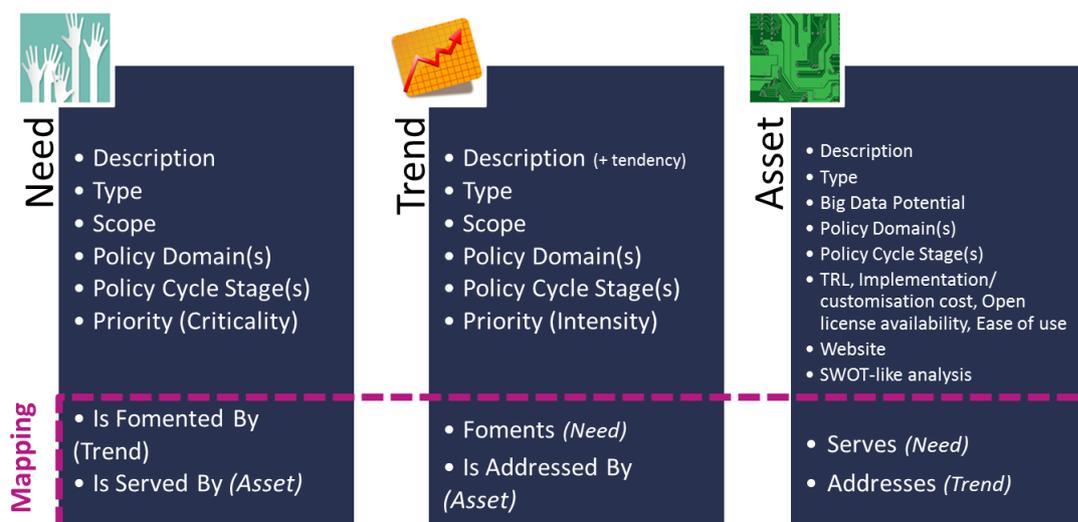


Figure 2-1: The BigPolicyCanvas Knowledge Base

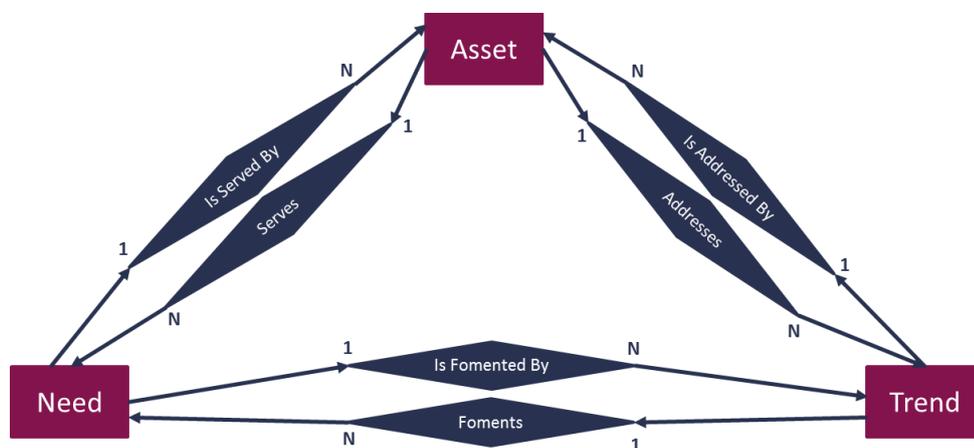
<b>Document name:</b>	Research Directions and Recommendations			<b>Page:</b>	7 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0
				<b>Status:</b>	Final

**Methodological Recommendation 1**

It is recommended to leverage the BigPolicyCanvas Knowledge Base to facilitate the uptake and reuse of the accumulated knowledge by the public sector and the rest of interested stakeholders, namely policy makers, public officials and researchers.

An integral part of the BPC Knowledge Base is the BPC Panorama. The latter is essentially a 3-dimensions mapping of needs, trends and assets, bringing forward potential inter-relations among these elements (see Figure 2-2). In particular

- *needs* are linked to the *trends* they are fomented by and the *assets* by which they can be served,
- *trends* are linked to the *needs* that they foment and the *assets* by which they can be addressed, whereas
- *assets* are accordingly linked to the *needs* that they can serve and the *trends* they address.



**Figure 2-2: BigPolicyCanvas Panorama (3-dimensions mapping)**

**Methodological Recommendation 2**

It is further recommended to leverage the BigPolicyCanvas Panorama to explore relations among identified needs, trends and assets.

From a methodological point of view Big Policy Canvas has designed the Needs and Trends Assessment Framework, a structured approach for mapping public sector needs and related trends to application domains, as well as for assessing the former in terms of their criticality or intensity respectively, with the ultimate goal of prioritising application domains and bringing forward those of greater interest, importance, urgency and capability for innovation.

This approach builds upon the concept of the policy cycle and the individual stages that the former incorporates, as well as upon an assortment of specific and representative policy domains in the sense that it places application domains, and thereby potential areas for intervention at the intersection of

<b>Document name:</b>	Research Directions and Recommendations	<b>Page:</b>	8 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU
		<b>Version:</b>	1.0
		<b>Status:</b>	Final

policy cycle stages with policy domains. Hence, the framework’s rationale, and thereby the rationale for the prioritisation of application domains, lies in comparing for each application domain Demand and Supply, whereby

- Demand is reflected in the *number* of needs in each application domain and their *criticality*, as well as in the *number* of trends and their *intensity*;
- and *Supply* is accordingly reflected in the *number* of assets available in the same application domain and their *score*, a metric calculated based on the criteria of *Technology Readiness Level (TRL)*, *Implementation or Customisation Cost*, *Ease of Use*, *Availability of an Open License* and *Big Data Potential*.

The Big Policy Canvas Assessment Framework is more specifically structured across three axes. These encompass as shown in Figure 2-3:

- the assessment and prioritisation of existing and emerging needs of public administrations,
- the assessment and prioritisation accordingly of existing and emerging trends that may impact the policy making process and
- the prioritisation of specific application fields based on selected criteria.



**Figure 2-3: BPC Assessment Framework Axes of Application**

Axes I and II deal with the assessment of the “Demand” side, whereas Axis III presupposes the assessment of the “Supply” side and targets the identification of areas for intervention. The rationale for the assessment of both “Demand” and “Supply” lies in mapping needs, trends and assets against application domains, i.e. combinations of policy cycle stages and policy domains. The determination of Demand counts more specifically with the determination of the number of needs and trends in each application domain, and the assessment of their criticality or intensity respectively. The determination of Supply on the other hand presupposes the identification of the number of assets in each application and the calculation of their score, as the aggregate of the factors “technology readiness level”, “implementation/customisation cost”, “ease of use”, “open license availability” and “big data potential”.

*Methodological Recommendation 3*

It is recommended to use the Big Policy Canvas Needs and Trends Assessment Framework as an aid for prioritising both public sector needs and related trends, as well as application domains.

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	9 of 20	
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b>	Final

*Methodological Recommendation 4*

It is further recommended to customise the application of Framework Axes I and II by weighting appropriately the different stages of the policy making cycle, as well as by attributing greater importance to specific policy domains. Accordingly, it is recommended to customise the application of Framework axis III in terms of the assessment criteria employed as well as of their relative importance in the calculation of the assets’ overall score.

To design its Roadmap for Future Research Directions in Data-driven Policy Making, and thereby aggregate the different research and innovation directions that should be followed in order to reach the anticipated vision of making the public sector a key player in tackling societal challenges through data-driven policy making approaches, Big Policy Canvas has invested on a policy oriented approach, adopted by previous projects as well, including a foresight element by combining roadmapping with scenario building techniques. The roadmapping exercise has included in particular the following steps:

1. Identification of the gaps that hinder the rapid and effective uptake of data-driven policy-making and policy-implementation solutions and approaches;
2. Elaboration of a set of future research challenges and application scenarios related to the use of big data in policy making;
3. Definition of a set of practical research directions and recommendations for all stakeholders involved.

The gap analysis in specific, has been implemented by comparing the already identified needs of PAs and the potential to be covered through the exploitation of existing Methods, Tools, Technologies and Applications, that is, the assets identified in the context of previous project activities, thus taking into account both demand side or market-pull as well as supply side or technology-push perspectives. Gaps have been considered therefore as the mismatch between what currently can be provided in this universe through the use of existing assets, and what are the current needs in terms of information, organisation, strategy, legislation and technology and accordingly the current conceptual, societal and technological trends. The individual steps of the gap identification process included specifically i. the selection of high priority needs, ii. their broken down into functionalities, iii. the assessment of the related assets against these functionalities and finally iv. the extraction of the gaps themselves from the maximum level of compliancy of the assets against the given functionalities.

*Methodological Recommendation 5*

Bringing together demand and supply perspectives, the BigPolicyCanvas Roadmap development approach is recommended as a methodological aid for the implementation of similar roadmapping exercises.

With the goal of ensuring that the knowledge gathered by the project is diffused and shared as widely as possible, as well as that it has an impact beyond the end of the project, BigPolicyCanvas has pursued and achieved during its course the establishment of an active community in the field of data-driven policy making.

The partners of the BPC consortium and the BPC Experts Committee have been the first members of the BPC Community, which has been enlarged along the project course with several new members,

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	10 of 20	
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b>	Final

attracted using the pull of the Experts Committee and the own contacts of the BPC partners as well as the impact achieved by means of the active presence of the BPC project in various events. To further consolidate the position of this community, the BPC partners soon joined forces with the Big Data Value Association<sup>1</sup>, a fast-growing big data community that at this moment lacked from the Public Sector perspective.

The role of the BigPolicyCanvas Community in the production and validation of the project outcomes has been instrumental in all phases of the project, as it has not only enhanced the quality and credibility of the latter but has also enabled to orient these outcomes towards the real needs of interested stakeholders, who, ultimately are the best-positioned to endorse the generated knowledge, promote its adoption and pursue its further enhancement.

*Methodological Recommendation 6*

Leveraging collective knowledge and wisdom is pivotal for the generation of quality and representative results that address the targeted stakeholder needs; therefore, it is recommended to adopt a community-driven approach that encompasses stakeholder engagement and validation processes.

*Methodological Recommendation 7*

As creating a community from scratch is an arduous task, it is recommended to exploit synergies with related projects and initiatives, as well as join forces with already existing active communities.

<sup>1</sup> <http://bdva.eu/>

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	11 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b> Final

## 3 Content-specific Recommendations

In this section the content-specific recommendations, as they came forward from the roadmapping exercise are presented. The term content-specific refers to recommendations that on the one hand, deal with societal, cultural, political and economic challenges and could assist policy makers to address them, and on the other hand bring forward areas of interest in the research front that researchers and industry representatives can delve into.

Specifically, six areas of interest were brought forward after the roadmapping exercise and the broad (umbrella) recommendations are:

- Impose radical changes in the modus operandi
- Develop or customise and employ the state of the art technologies
- Make changes in the legal framework
- Implement changes that will have a societal impact
- Enhance interoperability among the public sector actors
- Invest in the exploitation and employment of innovative approaches

The aforementioned parent recommendations can be broken down to more specific ones.

### Impose radical changes in the modus operandi

- Employ algorithmic decision making as a supplement and not a substitute of human decision making.
- Embed data collection practices in stakeholders' transactions and interactions with the public administration.
- Ensure adequacy of consent management and provide for symmetric transparency, allowing citizens to see how their data is being used, by whom and for what purpose.
- Enable reusability of government data among public administrations.
- Map HR strategy goals to specific skills and capabilities that are necessary to adopt and apply data-driven policy making.
- Leverage user-generated content from social media as well as sensor data (e.g. mobility, biometric, digital footprint) to support the agenda setting and policy monitoring and evaluation phases of the policy cycle.
- Combine big data with traditional datasets to improve representativeness.
- Use techniques such as principal component analysis, to be able to compare different datasets at a first glance, before delving in for the analysis.
- Ensure common/interoperable databases schemas among public administration organisations.

### Develop or customise and employ state of the art technologies

- Explore the impact big data applications on human behaviour.
- Ensure personal data protection, where applicable (e.g. by employing data anonymisation).
- Leverage blockchain technology to address data source reliability and data integrity issues.

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	12 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b> Final

- Advance research in anonymisation and encryption techniques.
- Develop methods/techniques/tools in order to achieve real-time (big) data (sentiment) analysis.
- Focus research efforts on automating data cleansing processes.
- Develop quality assessment frameworks for Big Data.
- Develop methods and tools to assess the representativeness of data collected.
- Develop shared formats among different actors and ensure the standardisation of datasets to enhance interoperability.

#### Make changes in the legal framework

- Allow local authorities to have access and/or collect data or adjust the legislative framework accordingly to so that local authorities can access and/or collect data.
- Allow government data to be not only accessible but also easily reusable.

#### Implement changes that will have an impact on society

- Adopt a multi-disciplinary approach to enforcing data-driven policy making, one that employs behavioural sciences as well.
- Ensure transparency of machine made decisions.
- Raise awareness around ‘fake news’ and misinformation attempts and combat related phenomena.
- Bring strategic policy making closer to citizens and engage the latter in scenario planning and foresight methodologies.
- Incorporate data science curriculum in civil servants’ training.
- Leverage private sector top-talent to build in-house expertise on (big) data science.
- Develop tools that allow citizens to better control or understand how their data are used.

#### Enhance interoperability among the public sector actors

- Promote standardisation around algorithmic decision making (practices, algorithms, data management).
- Use innovative techniques to achieve standardisation across the already existing datasets.

#### Invest in the exploitation and employment of innovative approaches

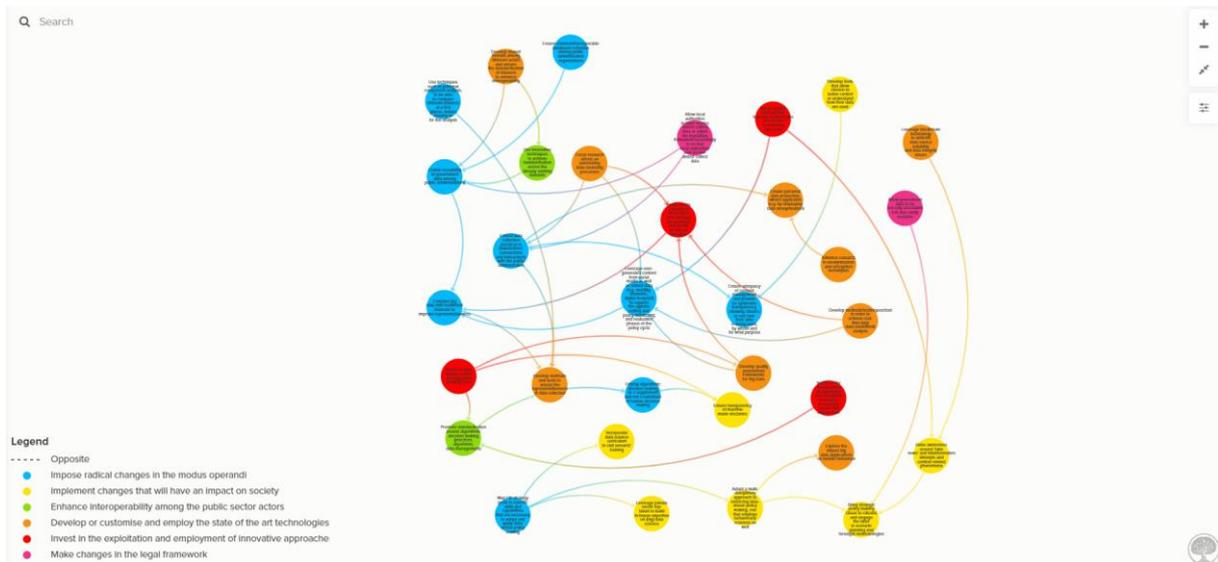
- Invest on data quality control and algorithm auditing tools.
- Invest on the development of algorithmic and encryption techniques to robustify results and ensure trust
- Invest on NLP and machine learning technologies as a means of detecting fake news.
- Invest on the concept of smart cities and leverage the potential and wealth of data of the latter.

The analysis has been conducted in six different areas of interest. These six areas of interest are broad and cover a wide range of recommendations. The first area of interest includes recommendations that if followed they would have a positive impact in society. Another area covers changes in the legislative framework. What is more, there are the recommendations about interoperability, and investments that

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	13 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b> Final

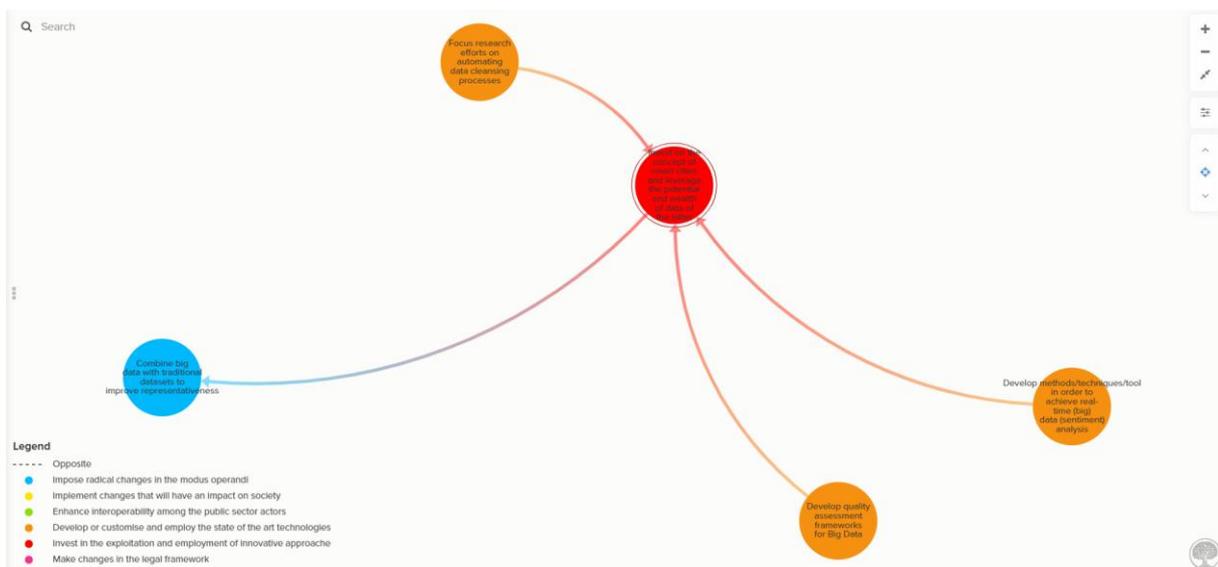
are needed in order for the public sector to become an evidence-based entity. Finally, there are recommendations concerning radical changes in modus operandi and the development and/or utilisation of state of the art technologies.

In the following figure a visual representation of how the recommendations are interconnected is presented. An interactive map showing the linkages of the recommendations can be found [here](#).



**Figure 4: Visual representation of the linkages among the recommendations**

In the figure below, the map is focused on the recommendation “Invest on the concept of smart cities and leverage the potential and wealth of data of the latter” that falls under the category: “invest in the exploitation and employment of innovative approaches”, that also can be seen by the colour code that has been given to the categories. This recommendation is associated with three of the recommendations from the “Develop or customise and employ state of the art technologies” category and also leads to one of the recommendations from the “Impose radical changes to the modus operandi”.



**Figure 5: Example**

<b>Document name:</b>	Research Directions and Recommendations			<b>Page:</b>	14 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0
				<b>Status:</b>	Final

In the section below, the recommendations are broken down to the research cluster they fall under.

### 3.1 Cluster 1 - Privacy, Transparency and Trust

---

Recommendations under cluster one aim at ensuring that the notions of privacy transparency and trust are met when dealing with data in policy making processes.

#### 3.1.1 RC 1.1 - Big Data nudging

##### Research recommendations:

- Explore the impact big data applications on human behaviour.
- Adopt a multi-disciplinary approach to enforcing data-driven policy making, one that employs behavioural sciences as well.
- Develop tools that allow citizens to better control or understand how their data are used.

##### Policy recommendations:

- Allow local authorities to have access and/or collect data or adjust the legislative framework accordingly so that local authorities can access and/or collect data.

#### 3.1.2 RC 1.2 - Algorithmic Bias and Transparency

##### Research recommendations:

- Invest on data quality control and algorithm auditing tools.
- Invest on the development of algorithmic and encryption techniques to robustify results and ensure trust

##### Policy recommendations:

- Employ algorithmic decision making as a supplement and not a substitute of human decision making.
- Promote standardisation around algorithmic decision making (practices, algorithms, data management).
- Ensure transparency of machine made decisions.

#### 3.1.3 RC 1.3 – Open Government Data

##### Policy recommendations:

- Allow government data to be not only accessible but also easily reusable.
- Ensure personal data protection, where applicable (e.g. by employing data anonymisation).

#### 3.1.4 RC 1.4 - Manipulation of statements and misinformation

##### Research recommendations:

- Invest on NLP and machine learning technologies as a means of detecting fake news.
- Leverage blockchain technology to address data source reliability and data integrity issues.

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	15 of 20	
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b>	Final

Policy recommendations:

- Raise awareness around ‘fake news’ and misinformation attempts and combat related phenomena.

## 3.2 Cluster 2 - Public Governance Framework for Data Driven Policy Making Structures

---

Recommendations under Cluster 2 target the establishment of a public governance framework that promotes, supports and takes advantage of data-driven policy making.

### 3.2.1 RC 2.1 - Forming and monitoring of societal and political will

Policy recommendations:

- Bring strategic policy making closer to citizens and engage the latter in scenario planning and foresight methodologies.
- Experiment with policies at a local, restricted level and then scale up.

### 3.2.2 RC 2.2 - Stakeholder/Data-producer-oriented Governance approaches

Research recommendations:

- Advance research in anonymisation and encryption techniques.

Policy recommendations:

- Invest on the concept of smart cities and leverage the potential and wealth of data of the latter.
- Embed data collection practices in stakeholders’ transactions and interactions with the public administration.
- Ensure adequacy of consent management and provide for symmetric transparency, allowing citizens to see how their data is being used, by whom and for what purpose.

### 3.2.3 RC 2.3 - Governance administrative levels and jurisdictional silos

Policy recommendations:

- Enable reusability of government data among public administrations.

### 3.2.4 RC 2.4 - Education and personnel development in data sciences

Policy recommendations:

- Incorporate data science curriculum in civil servants’ training.
- Leverage private sector top-talent to build in-house expertise on (big) data science.
- Map HR strategy goals to specific skills and capabilities that are necessary to adopt and apply data-driven policy making.

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	16 of 20
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b> Final

### 3.3 Cluster 3 - Data Acquisition, Cleaning and Representativeness

---

In this cluster, the recommendations concern actions in the policy making front that would make the acquisition, cleaning and storing of data more efficient and effective as well as bring forward areas in need of further research in order for processes to become more efficient and effective.

#### 3.3.1 RC 3.1 - Real time big data collection and production

Research recommendations:

- Develop methods/techniques/tools in order to achieve real-time (big) data (sentiment) analysis.
- Use innovative techniques to achieve standardisation across the already existing datasets.

Policy recommendations:

- Leverage user-generated content from social media as well as sensor data (e.g. mobility, biometric, digital footprint) to support the agenda setting and policy monitoring and evaluation phases of the policy cycle.

#### 3.3.2 RC 3.2 - Quality assessment, data cleaning and formatting

Research recommendations:

- Focus research efforts on automating data cleansing processes.
- Develop quality assessment frameworks for Big Data.

#### 3.3.3 RC 3.3 – Representativeness of data collected

Research recommendations:

- Develop methods and tools to assess the representativeness of data collected.

Policy recommendations:

- Combine big data with traditional datasets to improve representativeness.

### 3.4 Cluster 4 - Data Storage, Clustering, and Integration

---

Under this cluster, the recommendations referring to data clustering, integration and fusion are presented.

#### 3.4.1 RC 4.1 - Big Data Storage and processing

Research recommendations:

- Develop shared formats among different actors and ensure the standardisation of datasets to enhance interoperability.

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	17 of 20	
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b>	Final

Policy recommendations:

- Understand the storage, encryption, privacy and migration requirements of public administration large amounts of data.
- Promote the utilisation of cloud computing, while taking actions towards the protection of the data (for instance service agreements with third parties).
- Clarify and impose a common regulatory framework within the EU.
- Take advantage of cloud computing third party infrastructures in order to limit costs.

### 3.4.2 RC 4.2 - Identification of patterns, trends and relevant observables

Research recommendations:

- Advance research in machine learning and build public sector oriented applications and tools.
- Develop tools to detect anomalies in the datasets.

Policy recommendations:

- Leverage machine learning, data mining and anomaly detection techniques to respectively identify patterns and detect fraud in public sector big data.
- Decide per different case whether structured or unstructured data are better suited, as text is equally informative with numerical data.
- Use techniques such as principal component analysis, to be able to compare different datasets at a first glance, before delving in for the analysis.
- Ensure common/interoperable databases schemas among public administration organisations.

### 3.4.3 RC 4.3 - Extraction of relevant information and feature extraction

Policy recommendations:

- Ensure transparency of how data are collected and ingested.
- Promote data repurposing for research and analysis goals.
- Bring together both data analysts and domain experts to exploit the power of big data.
- Make use of social interactions when interpreting the data.

## 3.5 Cluster 5 - Modelling and Analysis with Big Data

---

In cluster 5, the recommendations that have to do with the modelling and analysis when handling big data are mentioned.

### 3.5.1 RC 5.1 - Identification, acceptance and validation of suitable modelling schemes inferred from existing data

Research recommendations:

- Develop effective infrastructures, merging the science of data with the design of highly predictive models.

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	18 of 20	
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b>	Final

Policy recommendations:

- Leverage big data to test and validate policy models.
- Allow some non-model-centric approaches in simulation, for instance machine learning based instead of modelling based approaches.

3.5.2 RC 5.2 – Collaborative model simulations and scenarios generation

Research recommendations:

- Develop tools to forecast future impacts.

Policy recommendations:

- Promote several levels of stakeholder involvement (e.g. provision of feedback, model co-creation, etc.)
- Allow collaboration and discussion among not only technical experts but all the stakeholders.
- Perform socio-technical analyses with the existing data, when it is not possible to elicit information from the stakeholders.

3.5.3 RC 5.3 – Integration and re-use of modelling schemes

Policy recommendations:

- Ensure consensus around modelling standards, procedures and methodologies.

## 3.6 Cluster 6 - Data Visualisation

---

Finally, in cluster 6, the recommendations applying to data visualisation are mentioned.

3.6.1 RC 6.1 – Automated visualisation of dynamic data in real time

Research recommendations:

- Develop low-complexity, real-time algorithms for the visual analysis of large-scale data.
- Focus research on data visualisation at the juncture of engineering, statistics, computer science and graphic design.
- Customise the visualisation depending on the end user’s needs.

3.6.2 RC 6.2 - Interactive data visualisation

Research recommendations:

- Treat visualisation as a key and inseparable step of the data analysis process.
- Design visualisation tools, suitable for layman users as well

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	19 of 20	
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b>	Final

---

## References

---

- [1] Big Policy Canvas D5.1 lead author (Mureddu, Francesco): D5.1 Roadmap for future research directions- First Version. Deliverable of the BPC project, [2019].
- [2] Big Policy Canvas D5.2 lead author (Mureddu, Francesco): D5.2 Roadmap for future research directions – Final Version. Deliverable of the BPC project, [2019].

<b>Document name:</b>	Research Directions and Recommendations				<b>Page:</b>	20 of 20	
<b>Reference:</b>	D5.3	<b>Dissemination:</b>	PU	<b>Version:</b>	1.0	<b>Status:</b>	Final