Data Strategies, Policies and Agenda

David Osimo*, Francesco Mureddu**, Vassilios Peristeras***, Anita Cioffi****, Cristina Moise*****, Charlotte van Ooijen******

*Lisbon Council for Economic Competitiveness and Social Renewal, david.osimo@lisboncouncil.net
**Lisbon Council for Economic Competitiveness and Social Renewal, francesco.mureddu@lisboncouncil.net
***Lisbon Council for Economic Competitiveness and Social Renewal, vassilios.peristeras@ec.europa.eu
****Deloitte, acioffi@deloitte.it
*****Lisbon Council for Economic Competitiveness and Social Renewal, cristina.moise@lisboncouncil.net
******Lisbon Council for Economic Competitiveness and Social Renewal, charlotte.van-ooijen@lisboncouncil.net

Abstract: Every sector of the economy is affected by the data revolution – and so is government. Data is becoming one of the top strategic priorities of digital government. Many local, national and supra-national organisations have adopted "data strategies": but the actual content of these strategies can vary substantially. In most cases, there is an overarching digital agenda include a digital government policy. Other data strategies are developed in the context of industrial policy, typically covering the overall economy, beyond government, and are designed to grasp the opportunities of artificial intelligence. This paper looks at dedicated self-standing data strategies aiming to maximize data driven innovation in public administration: the Data Agenda Government for the Netherlands, at the New Zealand Data Strategy and Roadmap, and at the initiatives Barcelona Data Commons and Findata. These strategies are typically implemented by those countries or agencies that are at the forefront of data driven public services. Each case includes a high profile, strategic policy initiative. The cases can be considered pioneers in ensuring a strategic approach to data governance in public administration.

Keywords: Data Agenda, Building Blocks, Data Analytics

1. Introduction

1.1. Background of the Paper

The data explosion is affecting all aspects of the society and the economy – and public administration is no exception. Data is a fundamental resource for carrying out all government activities, from regulation to service provision. And governments everywhere and at all levels are looking into the opportunities of data driven innovation, and in many cases experimenting with it. IDC estimates that central government is the fifth largest industry of the of the big data analytics market, covering about 7% of the expenditure, and growing fast. A recent study by Deloitte (2016) identified 103 cases

of big data analytics in government. In that regard, the Communication on "Data, Information and Knowledge management" calls for a more strategic use of data, information and knowledge. In this context, a data strategy (DataStrategy@EC) and a related Action Plan have been set-up in 2018, with the objective of transforming the EC in a data-driven organisation. The eight actions of the Action Plan are centred around 5 different dimensions: data, people, technology, organisation, policy. The data strategy highlights indeed that these dimensions need to mature and evolve harmonically to deliver a real transformation on how data is used in the decision-making processes. In 2019, an operational governance framework has been set up to closely follow-up the implementation and the evolution of the Action Plan. The 2016-2020 ISA² (Interoperability solutions for public administrations, citizens and businesses) programme funded with a budget of 131 million euro, aims to support the development of digital solutions that enable public administrations, businesses and citizens in Europe to benefit from interoperable cross-border and cross-sector public services. All these initiatives foster data-centric public administration. But where do we stand? To understand that the European Commission has commissioned the study Data Analytics for Member States and Citizens, which provides policy Directorate Generals of the European Commission and Member States public administrations with a knowledge base and guidance on the adoption of public sector data strategies, policy modelling and simulation tools and methodologies, and data technologies fostering a data-centric public administration. Specifically, the study covers three domains in relation to data analytics in government:

1) Data strategies, policies and governance: initiatives in the public sector both at the strategic level, such as data strategies, data strategies, data governances and data, management plans; and at organisational level, aimed to create units or departments, and to elaborate new processes and role;

2) Policy modelling and simulation: initiatives to improve policy analysis through new data sources, robust and reliable models to perform "what-if" scenarios, predictive analytics and hypothesis testing, and tools allowing policy makers to carry out scenario analysis through intuitive interfaces;

3) Data technologies: new architectures, frameworks, tools and technologies to be used by public administrations to gather, store, manage, process, get insights and share data. This domain includes the study of how data are governed as well as data collaboratives, and in particular stresses the joint analysis of governance and technologies.

The paper presents the result for domain 1 building on five case studies selected for the in depth analysis, based on the level of ambition and maturity: Barcelona Data Commons, Data Agenda Government in the Netherlands, New Zealand Data Strategy and Roadmap, Secondary use of health and social data (Finland), and Udbetaling Denmark. The cases can be considered pioneers in ensuring a strategic approach to data governance in public administration. These are not economy-wide data strategies – like the digital agendas or data economy strategies, but initiatives focusing on greater adoption of data driven solutions in the public sector. The pape starts off by providing a basic description of the strategies, their objectives and structure. The third section deals with the fundamental components of its governance, in terms of mechanisms in place. The fourth section looks at three necessary enablers of all strategies: the safeguards, the skills and the monitoring mechanisms. The final section looks at the results, the lessons learnt and the policy recommendations and is built on the insight presented in the analysis.

2. What are the Data Strategies For?

A data strategy can be defined as a set of actions aimed to fully leverage on the potential of available data for evidence based policy making. There is an abundance of data strategies nowadays, but they can mean very different things in different contexts. This section describes what are data strategies for, in terms of objectives and key components (building blocks).

2.1. Cases Under Analysis

The paper stems from an analysis on how Member States approach the concept of defining a data strategy, plus the following cases:

- Barcelona Data Commons. Barcelona has been one of the first cities in Europe to establish a Chief Data Officer. the City Council thought it was a priority to open the Oficina Municipal de Dades (OMD - Municipal Data Office), which is now responsible for the management, quality, governance and use of data controlled and/or stored by Barcelona City Council and all of its associated bodies (both public and private). The current volume of information generated by citizens and the City Council itself through its activities, combined with the existing technology and the science of data, allows to make a leap forward and become more efficient at capturing, storing and analysing data, as well as drawing conclusions that help with decision-making for new public measures. This relates also to the creation of data commons and greater control of citizens over their data, as piloted by the Decode project;

- NL DIGITAAL - Data Agenda Government. The Agenda centres on the value of data as a tool to address policy and social challenges. The Dutch Ministry of the Interior and Kingdom Relations leads the implementation of the agenda, but both central and local governments are responsible for implementing it. The agenda also "pays specific attention to the protection of public values and fundamental rights", thus including policy issues related to data ethics and the algorithm transparency. The agenda integrates policy goals oriented to better data management in the public sector and the publication and reuse of open government data. The relevance of the public sector's organisational culture and knowledge sharing for transformation change are also underlined, which is in line with the OECD approach for the digital transformation of the public sector;

- New Zealand Data Strategy and Roadmap. New Zealand's operational Data Governance Framework provides an interesting example where data stewardship is seen more as a skill to be built up among public officials rather than a formal role. This approach aims to embed "data accountability and best practice data management across all data-handling positions, with the goal of evolving beyond the need for traditional data governance roles (e.g. data custodians, data stewards)". The leading agency for government-held data in New Zealand (Stats NZ) developed a new and improved data governance framework for the New Zealand government. The framework is part of the agency's numerous efforts to promote better data management practices across the public sector, and to leverage data as a strategic asset for decision making. One of the central pillars of the framework is the adoption of a so-called "whole-of-data life cycle approach", meaning public bodies and employees are encouraged to think more strategically about the governance, management, quality and accountability of

their data, over the whole data life cycle (i.e. from the design and source of the data to its storing, publication and disposal);

- FINDATA. Finland has a long history of collecting extensive data in registers but making use of the data has been difficult and inefficient. In 2019 a new Act on Secondary Use of Health and Social Data entered in force in Finland. With the new enabling legislation, Finland has become the first country in the world to successfully enact a law on the secondary use of well-being data that meets the requirements of the European General Data Protection Regulation (GDPR). The new legislation enables and expands the use of social and healthcare data from the traditional areas of research and statistics to those of management, development, innovation, education, planning, and steering and supervision work. The new act facilitates the establishment of a new central data permit authority in Finland, known as Findata. Findata has been established and operates within the National Institute for Health and Welfare, but as a separate entity. Findata is the one-stop-shop responsible for streamlining and securing the secondary use of social and health data. It guarantees a flourishing ecosystem around the secondary use of social and health data streamlining the processes for the issuing of research permits and data collection and ensuring that data is being used in secure environments, thereby maintaining the trust that the general public have in authorities and the public sector. This case study reports how the new Act on Secondary Use of Health and Social Data has been developed, the main themes of it and provides an overview of the Finland legal framework around the use of the health and social data. It also described in detail the Findata one-stop-shop, how it works, the governance, and how it guarantees access to the social and health data. A descrption of the users, data sources and data lakes of Findata it is also provided.

2.2. Objectives

The strategies generally pursue the combined goals of fostering data analytics for public value creation and ensuring trust, accountability and citizens agency over how data are used. Both dimensions are important, although arguably with different emphasis.

The Danish, Dutch, New Zealand and Finnish cases are driven by the primary goal to increase data analytics and reuse, and balance this with a strong emphasis on safeguards, ethical aspects and consent as pre-conditions or enablers of data reuse. For instance, in the case of New Zealand the rationale for the data strategy is to address the disconnect between the rhetoric (which focusses on opportunities) and the reality of data-driven policies (which focus on minimizing risks of data misuse). The data strategy is part of the overarching government goal to get more value from data. In Finland, the data reuse strategy is based on the "National health-sector growth strategy" , which aims to make Finland an internationally renowned pioneer in health business and in well-being. In the Danish case, the strategy addresses a clear concrete need: to fight fraud and detect errors at an early stage.

On the other hand, the Barcelona strategy is the only one originally driven by a strong "political" dimension related to data and technological sovereignty – for instance with data provision requirements in the context of public private partnerships.

This difference is related to the specific policy context. In Barcelona, the data strategy is part of the wider city strategy "Transition towards digital sovereignty", while in the other cases it fits under the narrower strategic priorities related to digital government, social affairs or data reuse.

The role of the private sector is also different. In the Finnish case, the private sector is clearly identified as a data reuser, while in the Barcelona and Danish case it is identified mainly as a data provider. In the Dutch case, companies are considered as both data providers and reusers. What is clear across all strategies is that the private sector and citizens are part of the stakeholders that need to be involved in building a data ecosystem for public value creation.

The very existence of a strategy reveals the long-term importance of the topic. In all cases, governments sought via the strategy to ensure a structural commitment to the data priority (at least three years).

2.3. Building Blocks

The analysis of the different priorities of the strategies reveals many commonalities. All strategies include both the element of data exploitation, and the ethical and data protection aspect. Another common aspect is the recognized importance of data sharing as a fundamental pillar of data driven innovation. The New Zealand and Dutch cases also include a strong emphasis on knowledge sharing and skills.

Table 1: Cases and Related Building Blocks

Case	Building Blocks
BCN	1) Understanding data as an urban infrastructure, just as the provision of water and energy are. Data is seen as a meta-utility that will enable the city of Barcelona to support more effective delivery of public services to Barcelona citizens for greater equity, safety and quality of life; 2) Integrating the use of Big Data & Data Analytics to improve public decision-making (Data-driven projects). 3) Treating data as a common asset, and making it available for social and economic innovation processes focused on citizens' needs. This also means that the immense economic value that citizen-produced data represents should be returned back to those that generate that value in the first place: the citizens. 4) Enforcing data & algorithmic transparency (Data Ethics). This not only requires opening up data, but also encouraging the reuse, providing citizens with the tools and knowledge to be able to verify these, and to be informed about automated decisions and their underlying algorithms. 5) Protecting people's privacy and data sovereignty. This is also about shifting agency and control to citizens themselves that have the right to decide what data they want to share, with whom and on what terms.

NL	1) Problem-solving with a data-driven approach: five social challenges have been selected: energy transition, manure issue, infrastructure and spatial bottlenecks, poverty and the issue of debt, and subversive crime. 2) Focusing on legislation and public values to develop new general principles on a responsible way of dealing with data taking into account legal and ethical frameworks. 3) Improving the quality of government data and using it more efficiently to ensure government has the right data and is able to share (open) data at the right time and in the right way in order to foster a service-oriented and transparent public sector. 4) Collecting and sharing knowledge about a data-driven approach (sharing of best practices.) 5) Investing in people, organisations and changes in corporate culture (to address skills needs and cultural change).

NZ	1) Invest in making the right data available at the right time
	   a) To provide visibility of key data sets and proactively address gaps
	   b) To improve accessibility of government held data
	   c) To open up more non-sensitive, non-confidential data to the public
	2) Grow data capability and supporting good practice
	   a) To take a strategic and coordinated approach to uplifting capability across the public sector.
	   b) To make better use of existing data capability.
	3) Build partnerships within and outside government
	   a) To co-design the future data system and work together to maximise use and impact of data.
	   b) To co-design with Māori across the data system
	4) Implement open and transparent practices.
	   a) To establish appropriate accountabilities and protection mechanisms
	   b) To build public knowledge and understanding of how they can benefit from data use.

FI	1) To enable efficient and secure processing of personal data collected during the provision of social and health care as well as personal data collected for the purpose of steering, supervision, researching and collecting statistics on the social and health care sector; 2) to allow the collected personal data to be combined with the personal data held by Social Insurance Institution of Finland, Population Register Centre, Statistics Finland and Finnish Centre for Pensions; 3) to secure the legitimate expectations, rights and freedoms of individuals when processing personal data.

| DK	1) High quality data analysis across registry
	2) Trusted collaboration with sectoral department (national and local)
	3) Collaboration with foreign authoriy
	4) Special initiatives

When looking at the actual implementation of the strategies, other aspects emerge. First, there is limited technological investment. The two larger scale, more ambitious whole of government strategies of the Netherlands and New Zealand do not include the creation of any horizontal "infrastructural platform" for data analytics, but focus on enabling services. It is worth mentioning that the initial case selection included one of the most high-profile national data infrastructural service, the Italian Data Analytics Framework: however, in the course of the study the platform has been quietly scaled down. The Dutch strategy instead includes projects on five specific challenges: energy transition, manure issue, infrastructure and spatial bottlenecks, poverty and the issue of debt, and subversive crime. Barcelona, which does have a centralised data and analytics infrastructure also advances its data analytics activities in an incremental way, based on priority policy challenges. Second, all strategies include data mapping. In the case of Barcelona, it is the Data Office itself which carries out the mapping exercise, while in the others it is part of the distributed data stewardship and process tasks.

3. The Governance of Data Strategies

3.1. Governance Mechanisms

Governments adopt very different approaches to the governance of data strategies. The political positioning of the data strategy and related implementation agencies is varied. First and foremost, the strategies do not entail the promulgation of new legal provisions, but remain at the level of strategic document. Only in the case of Finland, because of the sensitive nature of the data a dedicated law has been approved to clarify the scope of reuse of health data. Only in the case of Barcelona, the strategy sits firmly at the executive level under the mayor's office, which is a fundamental factor in ensuring stability and compliance. In the other cases, the data strategy sits under specific ministries: in NL and NZ, under the department in charge of digital government, while in Finland it is the ministry of social affairs, in Denmark it is the social payments agency (Udbetaling Danmark) under the ministry of employment. On the same line, Barcelona is the only case which presents the figure of a Chief Data Officer, and a central "Municipal Data Office". They have a major role as responsible for the management, quality, governance and use of data controlled and/or stored by Barcelona City Council and all of its associated bodies (both public and private). And the Data Office is not only the coordinator but also the implementer of the Data Commons Strategy. In the other cases, the roles are softer. There is no Chief Data Officer, but the effort is conceived as decentralized and collective: in the New Zealand Case, the role is named "Chief Data Steward" and his role is to foster a culture of data stewardship across government. Data stewardship is intended as "the careful and responsible collection, management, and use of data." In particular, the goal is to spread the role of data stewardship across each agency, as illustrated in the figure below.

Figure 1: Role of Data Stewardship Across Each Agency
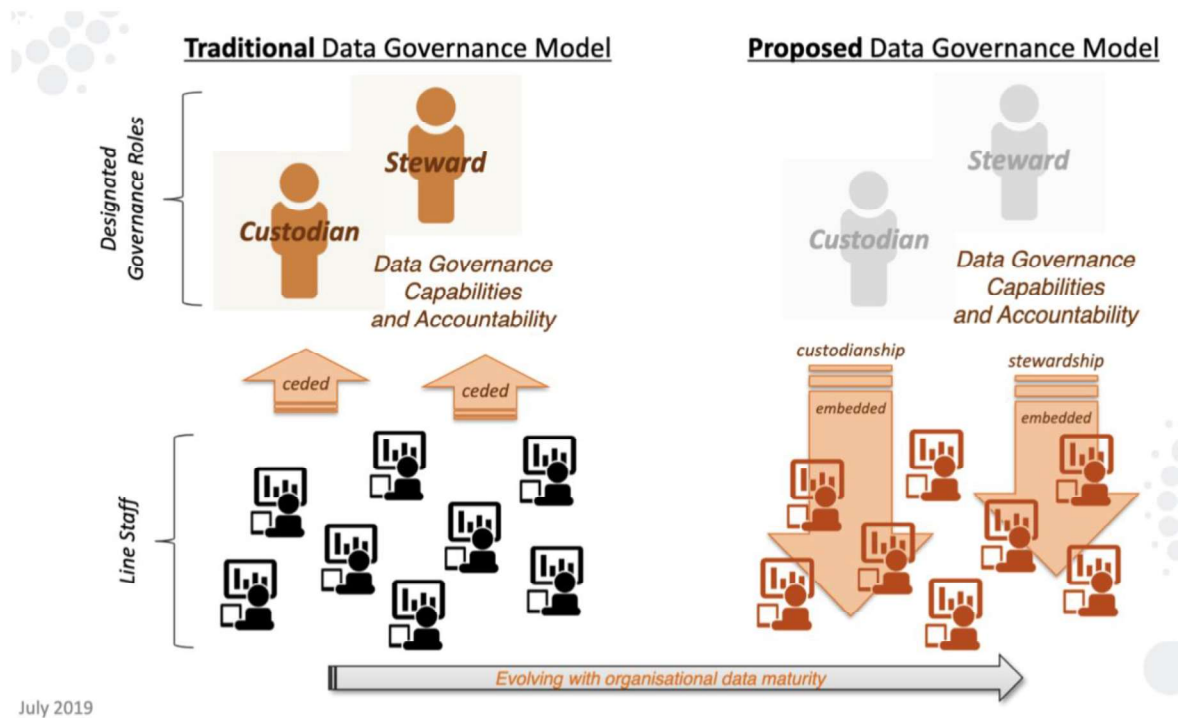
The case of Findata is also different, as its role is not to ensure data quality or sharing but to manage the data and consent flows. As such, it is a dedicated agency with strong enforcement roles but limited to the data and consent flows. This different degree of centralization is related with the broadness of the strategy: the broader the scope, the more decentralized the approach. The Dutch and New Zealand initiatives have a very broad mandate cutting across all departments and levels of government, while the Barcelona and Finland are more focused (respectively on one institutional level and on a specific data type). In any case, whether more or less centralized, all initiatives have an extensive set of boards and steering groups including a wider variety of participants, as outlined in the table below.

Table 2: Sets of Boards and Steering Groups

Case	Building blocks
NL	• Steering with institutional representatives from national govt agencies and local government organisations • Sounding board with technical staff
BCN	• Executive data committee with the central political and executive management • Transversal data coordination board includes representatives of 21 city departments • Data protection board with 14 departments dealing with sensitive data

NZ	• The Digital government partnership includes the public sector chief executives and four cross government working groups
FI	• Steering group includes social welfare ministries, social insurance, central statistical office, representatives of social and health care providers
DK	• Cooperation forum with municipalities and their national association, more operational • The coordination group with the national association of municipalities , more strategic

Consistently with this "inclusive" approach, all initiatives share a strong emphasis on co-creation with all relevant agencies. The extensive range of consultation and collaboration activities in place is repeatedly mentioned as leading edge and unique with respect to traditional processes. As mentioned in the Findata case, "the unique co-operation between public authorities, companies and associations was key to success." It appears that data collaboration not only aims to break data silos as an outcome, but in doing so it requires the adoption of a systematic silos breaking approach as part of its process. For instance, statistics New Zealand "led multiple workshops and interviews, gaining independent technical guidance and expertise on what was important to stakeholders and where help is needed". In the Danish case, the extensive collaboration with municipalities takes place also online through a Sharepoint platform.

This co-creation activity can also extend beyond public administration, towards external stakeholders such as business and civil society. New Zealand and Finland extensively involved these players in the shaping of the strategy, in order to make sure that it includes the perspective of external users. This activity went beyond traditional consultation, as made clear in the case of Findata: "Experts from ministries, authorities, companies and associations from across the private and public sectors worked together to prepare the implementation simultaneously with the legislation process. It was a unique way of working and something carried out for the first time at the national level." However, this co-creation aspect often is too focussed on the process, rather than on the final output: interviewee mention the lack of user orientation of some services as a clear challenge. In data driven innovation just like in digital government, the motto "build it and they will come" is a path for failure.

The budget assignment for the strategies is typically moderate. In the Dutch case, it includes 10 million euros for three years in addition to agencies' contribution. In Barcelona, the overall budget is nearly four millions euros for 2018/19. In Denmark, 3,4 million euros per year, and in New Zealand there is no dedicated funding for cross agency work. The budget for Barcelona, in this case, stands out taking into account that it refers to a single city, and it reflects the strong role of the municipal data office in implementing the strategy. The Danish case actually generates revenues far in excess of its costs (62 million euros in 2019).

3.2. Data Sharing and Quality

Data sharing and improved data quality is a general priority across all strategies. Greater sharing of high quality of data can be considered one of the main goals across the board: for Denmark, New Zealand and the Netherlands, across ministries and levels of government; for Barcelona, across

different municipal agencies; and for all the different player in the value chain. Data sharing has different levels of compliance. In the Netherlands, it is compulsory for public administration to share and reuse data from the base registries, and some of the base registries are open to the public. Six base registries, the BAG, BRK, BRT, BGT, and BRV are in part or completely available as open data. But beside these base registries, there are no strong compliance mechanisms. In the words of one interviewee, "We can't impose: 'thou shalt work in a data-driven way'". Therefore, one of the key policy levers in both the Dutch and Barcelona cases to encourage public organisations to share data is communication on two aspects: why and how. Why data sharing should happen focuses on demonstrating data-driven value creation in best practices. How data sharing can happen focuses on knowledge exchange between organisations on topics such as quality and standards. Similar soft arrangements with regards to data standards are in place in other strategies, as it remains a challenge to ensure compliance. In New Zealand, a Cabinet mandate empowers the GCDS to direct agencies to adopt common capabilities and data standards. Agencies are enabled and supported by the GCDS to voluntarily adopt those standards. Data standards can be of different sort. Dutch base registries rely on "identification keys" that are consistent across registers, as well as x/y coordinates. Yet the adoption of standards is also mentioned as a clear critical aspect looking ahead – it is not a result fully achieved. All the strategies clearly adopt open standards as a priority.

4. Key Enablers

4.1. Safeguards

All strategies have strong emphasis on safeguards, not just in terms of mere compliance, but to create a shared data culture that maximizes analytical power with ethical values. Typically, the data protection competence is separated from the data stewardship or responsibility competence. The notion of safeguards, accordingly, spans well beyond compliance with GDPR, to encompass a full ethical framework. Notably, the concept includes not only data processing, but the ultimate purpose of what is done with the data, with the goal of keeping the interests of citizens first, rather than those of government. This is why in Barcelona the ultimate goal is to empower citizens with data, citizens are involved through experimentation and consultation, and the activities of the strategy include algorithmic accountability and how public decisions are influenced by data. Wherever possible, data-driven projects will be able to check the algorithms using simulations based on city data. Likewise, using open source code or other means, third-party technology suppliers must reveal the underlying logic behind any IT process for (automated) decisions pertaining to any of their systems used by the City Council. By the same token, the Dutch strategy has developed general principles for the responsible use of data, after several municipalities had indicated running into difficulties regarding data sharing with companies. The Dutch case also shows the importance of the purpose of analysis, namely to avoid that data analytics is carried out with punitive purposes. Similarly, in Denmark, the strong data protection provisions go hand in hand with strong citizens' rights when it comes to the investigation, including the need for notification and the impossibility to access sensitive data in other registries (e.g. criminal records). Citizens control over their data is also an important issue. It is one of the leading principles of the Barcelona data sovereignty scheme, defined as "the need for an individual to have control, at all times and in all relevant systems, over the

collection, storage, use, transfer and publication of their data, whether it be of a technical, scientific, economic, social or personal nature." The Dutch government has launched the policy initiative Control over data (Regie op gegevens), which aims to give citizens and businesses more control on what's happening with their data. At the moment, various appointment systems and solutions are being developed to support citizens and businesses in managing their data. Ultimately, this should result in a generic cross-sectoral framework that enables secure, reliable and user-friendly digital exchange of data between governments, private and social organizations But this broad perspective "beyond GDPR" does not mean that GDPR compliance is a done deal.

In the Netherlands, stakeholders find the application of GDPR quite complicated, especially in the social domain. In the social challenge on poverty and debts, for instance, it is not clear which data are allowed to be linked, whether analysis can be done or not, and whether subsequent policy actions can be taken. Municipalities don't always have the same interpretation on how to proceed. The more cautious approach is to not use the data if it's not 100% clear that it is allowed to do so. On a similar note, On 5 February 2020, a Dutch court has ruled the way in which the government uses the fraud detection system SyRI, which links data from different sources, such as the Tax Agency, The Unemployment Agency and municipalities, as an infringement of Article 8 of the European Convention on Human Rights. According to the court, the SyRI legislation doesn't provide a fair balance between preventing and combating fraud in the interest of economic well-being on the one hand and the violation of privacy on the other. This ruling urges the Dutch government to go back to the drawing board.

4.2.   Skills

The lack of adequate skill is a major issue across all strategies, and all strategies include actions to address it.

These actions are basically:
- Training of civil servants. In the Netherlands, data is part of the training provided by the National academy for digitisation (RADIO). In Barcelona, the services provided by the data office are accompanied by training in order to ensure sustainability. In Nez Zealand, GCDS has developed e-learning modules for R coding capabilities.
- Profiles: in Barcelona, a new job profile for data scientists has been created. In New Zealand, GCDS create a data and statistical capability framework to support training and recruitment
- Recruitment: in the Netherlands, a trainee programme was created in the public sector. At the central level, the ministry and the Statistics department have internal expertise and offer support services to other government department. In Finland, Findata has 20 data scientists, while the Danish data mining unit has 25. The issue is then not only recruitment but retaining talent by setting up stimulating working environment that combines IT excellence with public purpose.
- Communities of practice: in Barcelona, an inventory of data enthusiasts has been created to provide training and support, and communities of practice are encouraged in all cases.
- Centre of competence: To deliver support, the strategies create centre of competence that offer services to another department. This is prominent in the Netherlands, with the creation of LED

experts centre and a network of specialised data labs. In New Zealand the data capability hub offers similar services. The municipal data office in Barcelona is also a centre of expertise for other departments

4.3. Monitoring

With the exception of the Danish case, monitoring remains very lightweight across all strategies. There are no KPI in place and no systematic monitoring. In the Netherlands there is a reporting system in place for projects launched under the strategies, which ultimately is presented to parliament. In Denmark, there is strong accountability mechanisms due to market-like relation between the national agency Udbetaling Danmark and the municipalities, with clear Key Performance Indicator, Service Level Agreements and financial accountability.

5. Conclusions and Lessons Learned

5.1. Success, Challenges and Lessons Learned

The first and foremost result of the strategies is putting the topic on the agenda and raising widespread awareness in public administration about the strategic value of data. As the Dutch case shows, the result is that questions like 'Do we use our data fully?' and 'Are we careful enough when using data?' are now on the table. This is also visible in the relation with external stakeholders, for instance in the case of Barcelona it visibly changed the importance of data in the relations with private providers and civil society.

This increased awareness has been reflected in greater access and reuse of data. In the case of the Netherlands and New Zealand, it has enabled greater data sharing between different departments and levels of government. In the case of Barcelona, this has resulted in data clauses in public contracts that allow the municipality to access and use data generated through third party service provision. In the case of Findata, it has cut by an order of magnitude the time needed to access sensitive health data for research purposes – from three years to three months.

It is clear that these strategies led to greater policy awareness and propensity to data sharing. Yet when it comes to visible results of these data sharing, the picture is uneven. Pilots, such as in the Dutch and Barcelona cases, show the potential to use data to improve policy and services, and enable quick delivery on different areas such as renewable energy, public order, housing and poverty. But while pilots are a good start, all strategies also point to the limited uptake and implementation of data analytics projects as a major challenge compared to expectations.

The limited visibility of the results goes hand in hand with the reluctance to share data, in a classical chicken and egg problem. As the New Zealand case shows, there is still a problem in gaining shared ownership of the strategies as many agencies deprioritise work that is very beneficial for the system as a whole, but is not overtly beneficial for their own agency. Despite data sharing being the declared goal of the strategy, it remains one of the most important challenges: this is not only due to technical barriers related to legacy and interoperability, but also to the resistance and lack of know-how to share data as well as the low quality of the data. In fact, one of the issues is that secondary

use of data requires greater data quality and reveals the limitations of data quality, as shown by the Dutch case.

In other words, policy prioritisation is necessary but not sufficient. Just as in other domains such as research data, or open government data, or business data, top down mandates have to be combined with clear benefits and opportunities for reuse: they should be demand driven– whether internal (municipalties in the Danish case) or external (reusers of health data). This is why focussed, thematic intervention such as on health and social data, where the business cases is clear, have more opportunity to grow.

Obviously, one of the expected barriers to data sharing is data protection. Because of GDPR is still relatively new and there is limited knowledge about its practical implementation, several organisations exhibit cautiousness on going ahead with a data project. This continues to be true, but what is clear is that this barriers can be overcome when there is a clear case for it. Again, data protection is a much more powerful barrier when it goes hand in hand with lack of a business case.

The lack of data skills is obviously a problem. But this is not limited to the mere absence of a sufficient number of data scientists in public administration. It refers to the more general problem of data literacy among civil servants, and notably decision makers, which leads to a dangerous cultural gap between the data scientists and policy makers. But this also includes the lack of a policy knowledge by data scientists (policy literacy). This is particularly risky because the ethical implications of data do not concern only the compliance with data protection, but the overall approach to data for policy: for instance, the purpose of data driven policies could spectacularly backfire if it is used with a punitive approach in particular with regard to social issues. The required cultural change concerns therefore all players in the data value chain.

The reason is that data is not a commodity. Extracting value from it requires not only cleaning and preparation, but plenty of assumption and choices in building analytical models. This processing implies that the results of the analysis are to be carefully considered in light of the limitations and assumptions. Data analytics should not be treated as a black box managed by data scientists where data are the input and decisions are the outputs. The recent example of Covid-19, where modelling tools have been at the root of potentially disastrous decisions by policy makers, is a clear demonstration of the need for distributed data competences in every policy domain and of the need of close collaboration between data scientists, data analysts and decision-makers. And the Danish case shows how important the collaboration between data analysts and frontcase workers is.

5.2. Policy Recommendations

The in-depth analysis suggests a set of recommendations for policy makers at EU and national level:

- Start with the problem, not with the technology. Building a data strategy does not necessarily entail an investment in a technological data analytics platform, and certainly it does not start with it. Very few strategies include such investment, and those who do are typically vertically focussed on specific sectors or organisations. On the other hand, there are not many examples

of successful whole of government data analytics platform, but there is room for focussed centralised technological components, as shown by the Reproducible Analytical Pipeline case analysed in task three. A common trait of most advanced horizontal and vertical strategies is a demand driven approach: providing a variety of support mechanisms, from governance to skills to support services, to address real problems, such as health, poverty, urban issues. Focus on the key questions to be answered and the policy problems to be solved. This is important in order to deliver tangible results.

- Analyse permanently user needs. Users include both data holders and data reusers, both internal and external. Too often user needs remain assumed or based on anecdotical evidence. Not only it is necessary to formally analyse them in the first place, but perhaps more importantly to constantly monitor them over time to adapt to how solutions are used. The constant collaboration between the Danish Data Mining Unit and the municipalities frontline case workers is a clear example of this. Iteration of delivery is therefore crucial – no service is designed perfectly the first time.

- Co-creation is a fundamental component of the strategy. Bringing internal and external stakeholders onboard is a necessary (not sufficient) condition of success. But it is equally important to keep stakeholders onboard after the strategy is launched, during the implementation. Other government agencies need to see the benefit to share data and to conform to the required standard and processes, because there are costs in doing so. Of course, there is a shared perception among decision-makers that data is a strategic resource and that investment is needed, but this is only sufficient for kickstarting the process: the difficult part lies ahead.

- It is not sufficient to consult and co-create with stakeholders: what matters is delivering results. There is a lack of business case for data innovation. Existing strategies should focus, as in the case of the Netherlands and New Zealand, on delivering short term results via small scale pilots on topical issues. But pilots should be the beginning of service delivery, as shown by the Findata case, and their results should be well documented and shared. The problem is not only the difficulty in demonstrating impact – the ultimate benefits in terms of quality of public service. It is the actual difficulty to demonstrate deployment and adoption – simple projects that work and deliver. Data strategies should balance long term perspectives to data stewardship with short term delivery of pilots.

- In order to ensure delivery, it is crucial to take a practitioner led approach. The most successful strategies are those were data experts in public administrations are brought together and given a visible role in the process, as in the Netherlands with the creation of a cross department sounding board with data analysts and policy experts. There is a permanent gap between data experts and decision makers, and for data strategies to work, data experts should be empowered. And communities of practices are the fundamental tool to enable mutual learning and empowerment of practitioners.

- Create a data culture across department and institutional level. Data-driven innovation requires cultural change, training and bringing in new resources from the outside. New centre of competences (such as the Dutch labs) have to be created. Data training should be provided to all civil servants, and in particular to decision makers. But it also requires the reinforcement of internal capacity and the creation of effective communities of practice that cut across

government silos, and the creation of knowledge and expertise centre to facilitate knowledge exchange between data champions and novices.

- Because it's a long-term process, expectations need to be managed correctly and hype should be avoided. Delivering data driven innovation is not easy, it's not a low hanging fruit. Data is not a commodity. It requires extensive work for access, preparation and cleaning, but also for processing and reprocessing. There is a constant risk of disappointment that backfires. It is important for data leaders to raise realistic expectations from other stakeholders and to start by focussing on data availability. Pilots should be selected based on two criteria: a genuine need and access to available data. Luckily, the evolution towards a data culture is visible across society and the economy, and it is here to stay – particularly so following the ongoing pandemic crisis. There is no need to overhype the opportunities.

- A robust ethical framework is crucial and can be instrumental to innovation. The results are long term, and it is important to avoid crisis in the short term that would "put back the clock". The safeguards can work hand in hand with more data reuse, by creating a shared data stewardship culture. Actions for data protection compliance should be integrated with those on increased data literacy: in fact, the lack of a data culture is damaging for both data protection and data innovation. But an ethical approach goes beyond compliance with data protection and includes also what is done with the data, for instance to avoid any punitive spirit in the services being put in place to fight poverty based on the data gathered.

- Monitoring should be present and structured but not drive the process. Milestones and KPI should be core part of any strategy – and it is currently very rarely the case. KPI should not concern only outputs, but also the inputs and the process, such as the percentage of datasets in line with the required standards, the access to base registries, and the number of departments taking part in the different activities. In fact, the main compliance mechanism in the case of such soft strategies is monitoring and reporting, as shown by the Dutch case where the most important control mechanism is reporting to Parliament. And they become fundamental in ensuring the long term collaboration of different stakeholders, as in the Danish case. Further, they allow comparison and benchmarking of the strategies for data management developed and adopted in different contexts.

References

Deloitte (2016). Big data analytics for policy making Article title. Report prepared for DG DIGIT.

About the Authors

David Osimo

David Osimo is director of research with 20 years of experience as an advisor on information society and innovation policies, inside and outside government, at local and international level (Open Evidence, Tech4i2 Ltd, European Commission Joint research centre, IPTS, Regione Emilia-Romagna). David is able to combine research and practitioners skills: as a researcher, he is mostly known for his pioneering work on web 2.0 in government and science, on which he advised the European Commission and the United Nations Development Programme. He also published articles on e-government, future science, research and innovation policy and ICT statistics and he is an experienced keynote speaker at high-level events such as Ministerial Conferences and the Digital Agenda Assembly. As a practitioner, he created web- platforms for

policy-making (CommentNeelie.eu, Daeimplementation.eu, and the Grand Coalition Pledge Tracker) and inducement prizes for innovation (iMinds INCA awards and Ideamocracy.it). David is a native Italian speaker and is able to work and interact in English, Spanish and French.

Francesco Mureddu

Francesco Mureddu is an innovation and technology policies analyst and strategist. Francesco has been primarily involved in designing and delivering consultancy and research projects for the European Institutions in a wide variety of topics including Information Society and ICT, e-government, Smart Cities, ICT-enabled social innovation, future science, research and innovation policies. His specific fields of expertise include evaluation of policy initiatives and programmes, impact assessments, policy monitoring and benchmarking, policy development, counterfactual impact evaluation, technology road mapping and scenario development, economic and econometric modelling, cost/benefit and sensitivity analysis, stakeholders' consultation and engagement. Francesco is Director at the Brussel based think tank Lisbon Council, Associated Researcher at the Centre for North South Economic Research (CRENoS) of the University of Cagliari, as well as co-founder of the cyber security startup Intelligence Framework Inc. He also has professional experience in international institutions such as the European Commission and the European Central Bank. Francesco holds an MA in Economics from the Catholic University of Louvain and a PhD in Economics from the University of Cagliari and is able to work and interact in English and Spanish.

Vassilios Peristeras

Vassilios is team leader at the unit data operations, technology and research, directorate-general informatics, European Commission. From 2010 – 2016, he has worked as Programme Manager in the European Commission, Interoperability Solutions for European Public Administration (ISA) Programme being in charge for semantic interoperability and open data. He has also been the coordinator of European Interoperability Framework revision. Vassilios is also Assistant Professor at the International Hellenic University, School for Science and Technology in Thessaloniki, Greece, and he is also Research Fellow at KU Leuven, Public Governance Institute. His teaching and research focus on the areas of eGovernment, eParticipation, interoperability, open and linked data and semantic web technologies

Anita Cioffi

Graduated in Economic in the 2014, Anita has developed consulting experience in leader advisory firms. She is involved in multiple international assignments mainly for European Commission (DG DIGIT, CEF, JRC) and Institutions (EP, EMA) focusing on the development of data strategy, big data and analytics solutions (among others the European Commission data platform, and the CEF Big Data Test Infrastructure building block), design of business models, processes and services. She has collaborated also in several projects and studies within the ISA^2 programme (Programme on Interoperability Solutions for European Public Administrations) and in particular on the EIF (European Interoperability Framework) and NIFO (National Interoperability Framework).

Cristina Moise

Cristina has an extensive expertise in statistical analysis, having worked well at the European Commission on complex DG RTD statistical modeling and representation, including the flagship Innovation Performance Indicators project. In this respect, she provided statistical analyses for the support documents used for various policy analyses and discussions. Cristina has experience in working with data from different sources (EU Industrial R&D investment scoreboard, COMTRADE / COMEXP, Eurostat, OECD, World Bank), as well as with data from patents and bibliometrics, IUC database, Data mining, RIO-PSF website. She has an undergraduate degree in computer science and a master in Economic IT.

Charlotte van Ooijen

Charlotte is a researcher and policy analyst in digital government with a specialisation in data governance and public trust. In January 2020, she joined the Lisbon Council as a Research Fellow Digital Government. In that capacity, she has contributed to the European Commission study 'Data Analytics for Member States and Citizens'. Previously, she worked as a policy analyst at the Organisation for Economic Cooperation and Development (OECD, 2015-2019), leading the analytical work on the Data-Driven Public Sector and contributing to several country reviews on digital government and (open) data, and to the broader OECD work on the digital transformation of the public sector, digital government indicators, open government data, digital welfare, innovation policies and technology trends. Charlotte completed her PhD thesis at the Tilburg Institute for Law, Technology and Society (TILT) at Tilburg University in the Netherlands (2014). Her research focused on analysing the implications of the government use of citizens' mobility data for the government-citizen relationship. She also worked as a lecturer in Public Administration at Tilburg University (2008-2011) and Erasmus University Rotterdam (2013-2014) and was an associate researcher at the Centre of Administrative and Political Sciences (CERSA) of the CNRS and Panthéon-Assas | Paris II University (2014-2015). Besides her mother tongue Dutch, Charlotte speaks English, French, Spanish and German. Additionally, she has taken up the study of Italian.