Data Driven Policy Cluster

Co-creating digital tools for better governance

Policy Prediction, the Future of Evidence-Based Policymaking?

How data and converging technologies impact agenda setting and policy design

March 2022

Authors

Claudia Fabó Cartas
is project officer at the European Citizen Science Association (ECSA).

Laura Gavrilut
is project manager at 21c Consultancy, partner in DUET.

Phoebe Koundouri
is professor and director at the Athens University of Economics and Business, partner in IntelComp.

Paresa Markianidou
is principal consultant at Technopolis Group, partner in IntelComp.

Michael Mulquin
is MiMs Ambassador at Open & Agile Smart Cities (OASC).

Ricard Munné
is project manager at Atos Research and Innovation, partner in Policy Cloud.

Francesco Mureddu
is director at the Lisbon Council, partner in DECIDO.

Charlotte van Ooijen
is associate director, digital government and data at the Lisbon Council, partner in DECIDO.

Susie McAller
is founder & managing director at 21c Consultancy, partner in DUET.

Alan Paic
is senior policy analyst, science and technology policies at the Organisation for Economic Co-operation and Development (OECD).

Joseba Sanmartín
is senior policy analyst at the Spanish Foundation for Science and Technology (FECYT), partner in IntelComp.

Nikitas M. Sgouros
is professor at the University of Piraeus, partner in Policy Cloud.

Jurgen Silence
is senior project manager at the Flemish Government, partner in DUET.

Marieke Willems
is project manager at Trust-IT Services, partner in Policy Cloud.

Decido

AI4PublicPolicy

Policy Cloud
Cloud for Data-Driven Policy Management

intelcomp

DUET

Editors

Marieke Willems
Organisation: Trust-IT
Project: Policy Cloud
ORCID: 0000-0001-9664-6225

Zach Smith
Organisation: Trust-IT
Project: Policy Cloud
ORCID: 0000-0002-9984-008X

Design

Lorenzo Calamai
Organisation: Trust-IT
ORCID: 0000-0002-5961-6227

Executive summary

Evidence-based policymaking (EBP) is not new, but its face is changing rapidly in the data-driven public sector, mainly due to the following factors:

1. Evolving meaning of evidence: from scientific opinions and policy evaluations to data-driven insights drawn from modelling and simulations coming from a multitude of data types and sources: from open data to citizen-generated data;.
2. New actors coming into play: EBP is no longer reserved to social scientists, statistical offices and policy evaluators, but now open to data scientists and data savvy public servants;
3. More attention for agenda setting and policy design, while EBP traditionally focuses on policy implementation and evaluation;
4. Different relation between policy and evidence: from reactive and ex-post to proactive and ex-ante; From "What worked?" To "What will work?"

The added value of a more data-driven approach to evidence-based policymaking seems clear, but can it fulfil its promises in practice? And what is needed to scale up the use of policy prediction?

This paper presents evidence from the Data-Driven Policy Cluster[1] to shed light on the questions above. It discusses the critical roles of interoperability and data sharing for the advancement of policy prediction in EBP. Then, concrete applications in different policy domains show what policy prediction looks like in practice and what challenges policy stakeholders encounter. The paper concludes with policy recommendations on how data can be leveraged for policy prediction at a larger scale than what is currently the case:

1. **Start small, experiment, then scale fast through co-creation**
2. **Foster a culture of innovation building on successful experiments and pilots**
3. **Leverage European data sharing infrastructures to increase data availability and accessibility across policy areas and government levels**
4. **Create trusted data spaces to facilitate data sharing by citizens and businesses**
5. **Secure the necessary multidisciplinary resources**
6. **Ensure European building blocks for data-driven policy prediction**

1. **AI4PublicPolicy**, **DUET**, **DECIDO**, **IntelComp** and **Policy Cloud**, five pan-European projects and initiatives dedicated to using cloud for data-driven policy, have joined forces in the Data Driven Policy Cluster, to raise awareness on their cross cutting work on data and cloud-based tools for data-driven policymaking.

1. Data: a game changer for evidence-based policymaking

What Works? This simple question appropriated by the Blair Labour government in the late nineties kicked off a worldwide movement known as evidence-based policy (EBP), which promotes public policy based on rigorous evidence.[2] The rise of EBP embodied a shift from ideologically-driven politics to rational decision-making based on the opinions and judgement of experts.[3] The biggest traditional source of evidence are (social) scientists who can inform/advise policy makers the types of interventions that are likely to produce the best outcomes and why. With the convergence of cloud technology, big data and artificial intelligence, data modelling and visualisations are complementing, and at times replacing scientists, in making predictions about policy outcomes.

Technological convergence has already caused considerable transformation across the Government. With citizens demanding better experiences as their expectations shift towards quick, seamless and personalised services, now is the time for public sector decision makers to embrace digital disruption and new innovative technologies to make more sustainable policy, based on real-time information, predicted impact and citizen input. In that regard, we have the data and tools to forge the future with policy impact analysis and prediction, also called anticipatory governance.[4] The wealth of data at our disposal should not only be used to understand the areas where new policy or change in policy is needed, but also to predict the impact of those policies, giving realistic KPIs on which they can be evaluated. The deluge of data and technologies can be used to forge the future with policy impact analysis and predictions: in a nutshell, to design and implement evidence-based policymaking at all levels of government.

The acknowledgement of data as a core resource for evidence-based policymaking is reflected in policy instruments for evidence-based policymaking.[5] For instance, the 2018 Evidence-Based Policymaking Act of the United States of America, also called OPEN Government Data Act, focuses on the modernisation of data management practices in government and makes government agencies responsible for the collection and analysis of data for evidence-based policymaking. Furthermore, dedicated whole-of-government strategies for the creation of data-driven public sectors are on the rise in countries like Denmark, Korea and the United Kingdom.[7]

The potential of public administrations to generate data-driven evidence from modelling and simulations that complement or potentially replace traditional sources of evidence, such as scientific research and ex-post policy evaluation reports, has implications for the way these organisations function. Evidence-based policymaking is no longer reserved to scientists, statistical offices and government evaluation departments but open to data scientists and data savvy public servants. With the advent of open data and citizen-generated data, a more active role for societal actors may also be envisaged. With regards to the policy lifecycle, more attention can be given to the phases of agenda setting and policy design, largely neglected in traditional evidence-based policymaking. However, policy implementation can also be affected, since dynamic models can account for new outcomes under changing circumstances and the policy can be more easily adapted. Furthermore, evidence-based policymaking becomes more proactive and ex-ante instead of reactive and ex-post. Not only will governments know what worked, but also what will work. That is, if they are careful in selecting the right data and interpreting it in the right context. The danger of cherry-picking "objective" evidence to serve specific policy purposes is ever present.[8]

The concept itself is complex and not yet clearly defined, and it represents the latest in a series of efforts to enhance the efficiency and effectiveness of public policymaking through the application to policy formulation of an evaluative rationality. Evidence-based policymaking represents an effort to reform the process of policy formulation in such a way as to minimise non-design spaces. In that regard, evidence becomes the known relationships between policy instruments and policy goals, e.g. evaluation studies on the results or impacts of specific instruments, or the predictions of the impact of different instruments. The use of evidence in policymaking requires that policy actors have the analytical capacity to carry out such evaluation studies or predictions.

As shown in Figure 1, evidence has different functions throughout the four phases of the policy cycle.

2. Productivity Commission (2010). Strengthening Evidence-based policy in the Australian Federation, Volume 2: Background Paper, Productivity Commission, Canberra, **https://www.pc.gov.au/research/supporting/strengthening-evidence/roundtable-proceedings-volume2.pdf**

3. Sophie Sutcliffe and Julius Court (2005). Evidence-Based Policymaking: What Is It? How Does It Work? What Relevance for Developing Countries? Overseas Development Institute, **https://cdn.odi.org/media/documents/3683.pdf**

4. Charlotte van Ooijen, Barbara Ubaldi and Benjamin Welby (2019), "A Data-Driven Public Sector: Enabling the Strategic Use of Data for Productive, Inclusive and Trustworthy Governance", OECD Working Papers on Public Governance, No. 33, OECD Publishing, Paris, **https://doi.org/10.1787/09ab162c-en**

5. U.S. Chief Information Officers Council (2018). "Foundations for Evidence-Based Policymaking Act of 2018", Policies & Priorities, **https://www.cio.gov/policies-and-priorities/evidence-based-policymaking**

6. OECD (2020a), "Digital Government Index: 2019 results", OECD Public Governance Policy Papers, No. 03, OECD Publishing, Paris, **https://doi.org/10.1787/4de9f5bb-en**

7. Osimo, David (2022). "Context is King: How Correct Data Can Lead to False Conclusions", The Evidence Hub, **https://evidencehub.net/blog/context-is-king-how-correct-data-can-lead-to-false-conclusions/**

8. Howlett, Michael. (2009). Policy Analytical Capacity and Evidence-Based Policy-Making: Lessons from Canada. Canadian Public Administration. 52. 153-175.

1. Data: a game changer for evidence-based policymaking

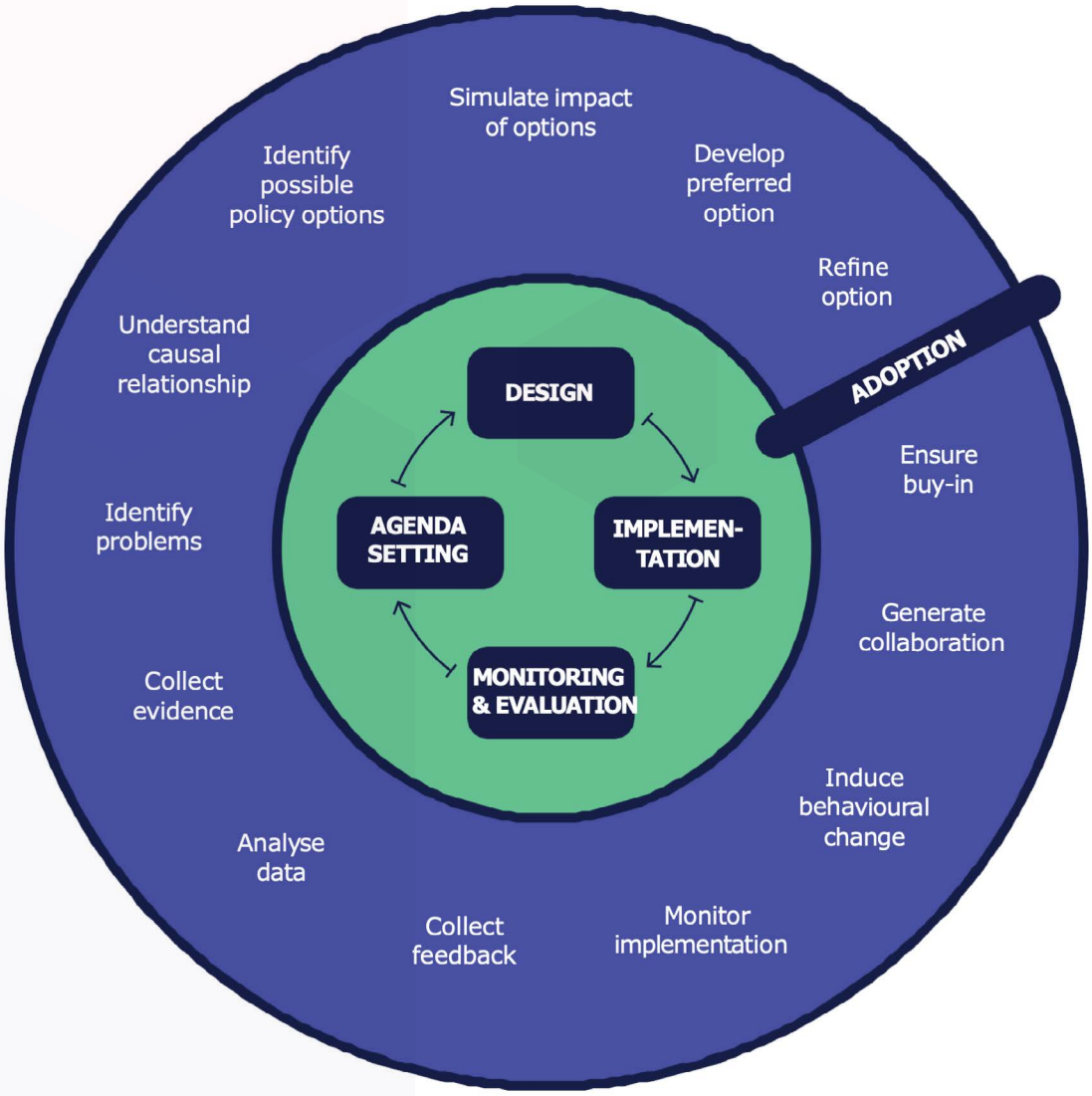


Figure1. Functions of evidence in the policy cycle

1. Data: a game changer for evidence-based policymaking

In the first phase, **agenda setting,** the challenge addressed is to detect (or even predict) problems before they become too costly to face, as well as reaching an agreement of which issues have to be dealt with. In this regard, through data, governments can identify emergent topics early and create relevant agenda points collecting data from social networks with high degrees of participation, and identifying citizens' policy preferences. The second phase, **policy design**, concerns the discussion related to the policy, its formation and acceptance, and the provision of means. Specifically, policy discussion deals with debating the different options on the table, and identifying which is the most important. In this regard, opinion mining and sentiment analysis can help to inform policy makers about the current trend of the political discussion as well as the changes in public opinion in the light of discussed and proposed changes. Policy formation and acceptance deals with the use of big data and data analytics solutions for providing evidence for the ex-ante impact assessment of policy options, by helping to predict possible outcomes of the different options, by making use of advanced predictive analytics methodologies and scenario techniques.

In that regard, robust and transparent predictive modelling and algorithmic techniques can also help in improving policy acceptance. Finally, regarding provision of means, the challenge is to improve the decisions on how to most efficiently provide the required personnel and financial means for the **implementation** of new policies by analysing in detail past experiences. An example is the use of big data in budgeting to increase efficiency and effectiveness while reducing costs. In the third phase, implementation, big data and data analytics can help identify the key stakeholders to be involved in policy or the key areas to be targeted by policies. Another way in which data can influence the implementation phase of the policy process is the real-time production of data. In fact, the execution of new policies immediately produces new data, which can be used to evaluate the effectiveness of policies and improve the future implementation processes.

Finally, in the **policy evaluation** phase, the implementation of the policy provides quantitative and qualitative measures to assess performance and impact on users. Feedback and data on implementation can be collected and analysed by means of disruptive technologies. Clearly, predictions and simulations are deeply related to the **agenda setting** and **policy design** phases. The wealth of data at the disposal of public administrations should not only be used to understand the areas where new policy or change in policy is needed, but also to predict the impact of those policies, giving realistic KPIs on which they can be judged. As such, policy impact prediction and analysis can lead to measurable outcomes and further policy refinement.

From promise to practice

The added value of a more data-driven approach to evidence-based policymaking seems clear, but can it fulfil its promises in practice? And what is needed to scale up the use of policy prediction?

To shed light on these questions, this paper discusses the first insights from five pan-European projects and initiatives dedicated to using cloud technology for data-driven policy, which are united in the data-driven policy cluster (see box 1). In addition, it draws on the results of the 2021 Evidence Based Policymaking in Europe Summit hosted by the data-driven policy cluster, which was held on 9 and 10 December 2021.[9] Besides the members of the data-driven policy cluster, the summit brought together leading change agents from the European Commission and local governments to discuss the new decision-making ecosystems being built by cities and administrations, including the use cases being adopted, and the innovative data and tools being applied for modern policymaking. Section 2 of this paper will focus on the critical roles of interoperability and data sharing for the advancement of policy prediction in EBP. Section 3 presents concrete applications across different policy domains as researched in the projects DUET, AI4PublicPolicy, IntelComp, Policy Cloud and DECIDO that will help to understand what policy prediction looks like in practice. Section 4 discusses the challenges of implementing data-driven policy prediction.

The paper will conclude with policy recommendations on how data can be leveraged for policy prediction at a larger scale than is currently the case.

9. https://Policy Cloud.eu/news-events/_events/evidence-based-policymaking-europe-summit-2021

1. Data: a game changer for evidence-based policymaking

Box 1 - The Data-Driven Policy Cluster

DECIDO aims to demonstrate the ground-breaking impact of the adoption of innovative methodologies, tools and data enabling the effective development of better evidence-based policies by Public Authorities. DECIDO will link Public Administrations to the data and compute infrastructure of the European Open Science Cloud – piloting the access to and exploitation of a great wealth of additional resources.

AI4PublicPolicy will develop cloud computing infrastructures enabling public authorities to harvest the vast amounts of data, and providing High Performance Computing (HPC) capabilities, cost reductions and improved economies of scale, reducing the time needed to develop and roll out new services, allowing the execution of advanced data analytics capabilities over such datasets, as well as the capability to leverage the outcomes of Machine Learning (ML) and Deep Learning (DL) techniques towards holistic and actionable insights.

DUET envisions a world where cities understand the real-world impact of change before making crucial policy decisions. DUET is developing digital twins through 3D interfaces facilitating experimentation and co-creation to deliver policy and services that were previously impossible, to be tested in three city pilots.

Policy Cloud makes use of a tool-based, meta-simulation methodology, to estimate the outcomes of different policy alternatives, providing side-by-side examination and comparative analysis of their assumptions, mechanics and outcomes.

IntelComp provides a Platform able to analyse large volumes of unstructured text data, using Artificial Intelligence tools (natural language processing pipelines, a machine translation system, automatic classifiers, topic modelling or graph analysis) in High Performance Computers to inform Science, Technology and Innovation (STI) policy.

2. Enablers of data-driven policy prediction

As indicated above, evidence-based policymaking increasingly relies on the capacity of public administrations to access and reuse data produced by public sector organisations and societal actors. To tap into the potential of government- and community-generated data for EBP, it is critical that data becomes more widely available in formats that can be used across different information systems. Therefore, tapping into new data sources and ensuring mechanisms for data sharing and interoperability are prerequisites to scale up policy prediction efforts.

Leveraging new data sources

Tapping into new data sources is not just about getting public institutions who previously did not share their data to open it up. It is also about mobilising societal actors as data producers. One way to do this is through citizen science, which is any activity that actively involves the public in scientific research and thus has the potential to bring together science, policy makers, and society as a whole in an impactful way.[10] Through citizen science, all people can participate in many stages of the scientific process, from the design of the research question, to data collection and volunteer mapping, data interpretation and analysis, and to publication and dissemination of results. Citizen science is also an approach of scientific work that may be used as a part of a broader scientific activity. There are several typologies of citizen science. Haklay's 2013 typology aims to explain different levels of participation in citizen science projects:[11]

- Crowdsourcing - citizens as sensors;
- Distributed Intelligence - citizens as basic interpreters;
- Participatory Science - participation in problem definition and data collection;
- Extreme Citizen Science - collaborative science, consisting in problem definition, data collection and analysis.

Citizen science is a great tool for generating data as a source of information for policymaking. In fact, output from citizen science (such as data and knowledge) can be used for the implementation and monitoring of regulation, and citizens can have a big role in agenda-setting and the review of policies. To this extent, "citizen science activities offer an under-used, cost-efficient additional source of knowledge and feedback in the monitoring of the environment and the implementation of environment policies. This includes non-traditional data sources, analytical capacities, opportunities for engaging with citizens and possibilities for knowledge exchange."[12] The development of citizen science in Europe is promoted by the European Citizen Science Association (see box 2 for specific projects). "With the environmental ambitions under the Green Deal, public authorities will be even more driven to tap into citizen science initiatives, communities and outputs, to expand their knowledge base in key areas such as biodiversity, pollution, circular economy, climate change and sustainable food." science initiatives, communities and outputs, to expand their knowledge base in key areas such as biodiversity, pollution, circular economy, climate change and sustainable food."[13]

Box 2 - European Citizen Science Association projects - an extract

- WeObserve - Roadmap for the uptake of the Citizen Observatories' knowledge base (2017-2021)
- LandSense - Citizen Observatory for Land Use and Land Cover monitoring (2016-2020)
- D-NOSES - Distributed Network for Odour Sensing, Empowerment and Sustainability (2018-2021)
- Cos4Cloud - Co-designed citizen observatories for the EOS-Cloud (2019-2023)

Source: https://ecsa.citizen-science.net/

Citizens are not the only sources of new data. In this new era of data-driven policymaking, scientists take on a new role as data producers in addition to being evidence producers. Through open research data, the data gathered by scientists can now also be leveraged by other actors, including data scientists in governmental organisations. Research data is an important input to policymaking. In fact, considering the COVID-19 pandemic, policymaking can be based on epidemiological data (and studies depicting the spread of pandemic), clinical data on lethality and other impacts on public health, mobility data monitoring lockdown measures, consumer behaviour and response of economic agents, and finally sociological data on acceptability of measures. Such data can help in mitigating data-related issues in policymaking, such as quality and reliability (e.g. contested figures on lethality and spread),

10. ECSA (European Citizen Science Association). 2015. Ten Principles of Citizen Science. Berlin. **http://doi.org/10.17605/OSF.IO/XPR2N**
11. M. Haklay (2013), "Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation," In: Sui D., Elwood S., Goodchild M. (eds) Crowdsourcing Geographic Knowledge. Springer, Dordrecht.
12. European Commission (2020). Best Practices in Citizen Science for Environmental Monitoring. European Commission, Brussels, p. 17. Available at: **https://ec.europa.eu/environment/legal/reporting/pdf/best_practices_citizen_science_environmental_monitoring.pdf**
13. Ibid.

2. Enablers of data-driven policy prediction

as well as misuse of data and spread of fake news. The Organisation for Economic Co-operation and Development (OECD) developed several policy instruments and recommendations to foster better access to data and support the advancement of open research data (see box 3). Specifically, in its 2021 revision, the Recommendation for Access to Research Data from Public Funding expands the focus covering research data, metadata, algorithms, workflows, models and software (including codes), and outlines recommendations regarding data governance, technical standards and practises, incentives, responsibility and stewardship, sustainable infrastructures, human capital, and international cooperation for access to research data.

Box 3 - Focus on Open Research Data - OECD

Key initiatives:

- 2020 Policy report Enhanced Access to Publicly Funded Data for Science, Technology and Innovation;[14]
- 2021 revision of OECD Recommendation of the Council concerning Access to Research Data from Public Funding;[15]
- 2021 OECD Recommendation of the Council for Enhancing Access to and Sharing of Data.[16]
- 2017-2018 Digital Science and Innovation Policy project (DSIP)

The Digital Science and Innovation policy (DSIP) Infrastructure is the overarching framework through which governments make intensive use of digital technologies and data resources to support the formulation, delivery and administration of STI policy.[17] Main rationales for DSIP include providing insights at different points in the policy cycles, streamline workflows, reduce administrative burden on researchers and research performing organisations, and support general information discoveries. On the other hand, main challenges include lack of national coordination and strategy, technological backwardness and competencies to use the DSIP systems.

A case study has been carried out on the Norway's DSIP system (OECD, 2020c), highlighting the following drivers:[18]

- A strong legacy of comprehensive administrative records and strong accountability and evaluation culture
- A trust-based societal consensus - citizens and organisations perceive benefits of data use by public authorities
- Statistics Norway (SSB) role extends beyond descriptive statistics, assuming the role of a macro and micro-data clearinghouse
- CRISTIN, the Norwegian Current Research Information System provides a comprehensive list of publication outputs - an essential component of the research assessment system.
- Internal data integration and interoperability of databases - CRISTIN with project databases managed by the Research Council of Norway and other government agencies (e.g. through the usage of individual and organisational IDs)

The study suggests the following areas for improvement:

- Improve skills in digital technologies and access to datasets covering an international dimension of research or innovation activities (e.g. world citation data, research outputs created in other countries).
- Remaining issues of fragmentation and scale, notably due to a sectoral approach of STI policy could be addressed by an inter-agency digital coordination working group.
- Further develop interoperability through compatibility of standards and identifiers across domestic actors and internationally
- Access to data about STI generated in the system (administrative, statistical, commercial, etc.) to be assessed according to data types, purposes and actors, and communicated to potential users.
- Make or buy: consider off the shelf solutions for greater efficiency.

Approaches to manage new data sources

Thus, data availability has taken on a completely new dimension. In fact, more sources and more data are opening new perspectives in management activities and their nature can disrupt the planning and implementation of policy and decision-making measures. Big Data analytics represent one of the main business decision-making tools worldwide, which has the potential to boost productivity and improve the efficiency and effectiveness of governments at all levels. Besides, the advancement in computer power has led to relevant breakthroughs in the AI field and it is

14. OECD (2020b), Enhanced Access to Publicly Funded Data for Science, Technology and Innovation, OECD Publishing, Paris, https://doi.org/10.1787/947717bc-e
15. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347
16. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463
17. OECD (2020c), The Digitalisation of Science, Technology and Innovation: Key Developments and Policies, OECD Publishing, Paris, https://doi.org/10.1787/b9e4a2c0-en
18. OECD (2020d), "OECD Case Study of Norway's Digital Science and Innovation Policy and Governance Landscape", OECD Science, Technology and Industry Policy Papers, No. 89, OECD Publishing, Paris, https://doi.org/10.1787/20f80fa1-en

2. Enablers of data-driven policy prediction

clear the relevant potential of these technologies to transform societies and the economic systems. For instance, at the local government level, smart city ecosystems provide the right environment for the complete data chain by including a variegated set of stakeholders such as universities, private companies, public bodies, and citizens.

Data needed by cities is often not open data. It is difficult to access public, but sensitive data. Furthermore, there are difficulties to access and reuse private data with public interest and current operational models do not scale up. European cities recognise citizen data as a public asset, while safeguarding citizens' digital rights (personal data management), and smart cities strive for portable and affordable, innovative cross-sector services.

The creation of Common European Data Spaces under the European Strategy for data[19] is a key development in this area, which can help to cope with the described challenges. European Data Spaces can boost the sharing and re-use of data. The key elements of Data Spaces include planned iterative implementation strategies, according to which the success of data sharing activities revolves around the basic concept of trust.

The main pillars are:[20]

- Data - Free flow of data from strategies that embed methodologies for data sharing by-design (e.g. interoperability) and clear standard guidelines to determine the market value of data assets;
- Governance - A European-governed data sharing space can inspire trust by adhering to the more advanced European rules, guidelines and regulations and promote European values;
- People - Data sharing needs to guarantee individual privacy and offer fair value or compensation of shared personal data. Reskilling and upskilling are needed to meet the evolving labour market's needs;
- Organisations - More organisations (including business, research and governmental) need to rethink their strategy to fully embrace a data culture that places data at the centre of their value proposition;
- Technology - Safer experimentation environments are needed to catalyse the maturation of relevant technology behind trustworthy data, data access and algorithms (privacy, interoperability, security, and quality), together with standardisation activities.

Minimal Interoperability Mechanisms (MIMs) as developed in the OASC (Open and Agile Smart Cities) project provide a concrete way of supporting the creation of Common European Data Spaces and their use at the city level. MIMs are a set of practical capabilities based on open technical specifications that enable cities and communities to replicate and scale solutions globally. They are the minimal but sufficient capabilities required to achieve interoperability of data, systems, and services. These will facilitate smart communities where bigger volumes of useful data are collected and used by the public administration, businesses, and citizens to help the city function better. Cities are creating local data ecosystems everywhere, but there are some challenges: using data to manage physical assets, managing data analytics, ensuring fair AI, data security management, data quality, ensuring common data models, data discovery, compliance for data sharing, gathering data usage information, etc. Specifically, MIMs aim to figure out all the key building blocks needed for a city to build an effective data sharing ecosystem.

Main characteristics of MIMs are the following:

- Sufficient interoperability to allow good enough integration of systems and development of a viable market;
- Minimal to ensure no unnecessary complexity or time-to-implement and contained costs;
- Clearly defined mechanism to determine if a product or service is compliant as well as the steps to implement.

MIMs are based on existing work. Where existing standards are in place, the MIM will work with cities and standardization bodies to identify the basic requirements of those standards that a city could implement as a first step to immediately start seeing benefits from developing a local data ecosystem. Further, where there are several standards initiatives that cover the same ground, the aim will be to identify the lowest common denominator (or the NIST Pivotal Points of Interoperability) that will make it easy to link products and services that comply with those different sets of standards. Finally, where policy or procurement requirements are being agreed, but with no technical specifications to support these, then the project may fill this gap. Out of the ten planned MIMs (see box 4), so far three are featured in the OASC project: Context Information Management, Common Data Models and Marketplace Enablers (Ecosystem Transaction Management).[21]

19. https://digital-strategy.ec.europa.eu/en/policies/strategy-data
20. https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf
21. https://oascities.org/minimal-interoperability-mechanisms/

2. Enablers of data-driven policy prediction

Box 4 - Envisioned Minimal Interoperability Mechanisms (MIMs)

- MIM1 - Context Information Management
- MIM2 - Shared Data Models
- MIM3 - Ecosystem Transactions Management (Marketplace Enablers)
- MIM4 - Personal Data Management
- MIM5 - Fair Artificial Intelligence
- MIM6 - Security management
- MIM7 - Geospatial information management
- MIM8 - Ecosystem indicator management
- MIM9 - Data Analytics Management
- MIM10 - Resource Impact Assessment

Source: https://mims.oascities.org/

Opportunities of the Common European Data Spaces include achieving wider access to data to realise the full potential of emerging AI technology, as well as a European-governed data space, giving Europe the possibility to assume a prominent position steering international efforts to develop data and AI solutions that reflect and respect European ethical values. For all three main stakeholder groups, business, public sector organisations and citizens, specific opportunities can be identified:

Opportunities for Business

- Open data marketplaces that level the playing field for industrial data sharing.
- Increased availability of vast and heterogeneous data ecosystems for AI.
- Innovative data-driven business models enabled by new value ecosystems.
- Opportunities to tap into 'safe' personal data.

Opportunities for Government and Public Bodies

- Data commons for better government services.
- AI-enhanced digital services.
- Real-time European statistics.
- Lean business environment enabled by access to government services.
- Evidence-based policymaking.
- Data as evidence of Policy compliance.
- Opportunities for Science
- Increasing socio-economic impact of research data across domains and borders.
- Advancing science and open innovation through data availability.
- Monetisation opportunities brought about by emerging data-driven business models.

Opportunities for Citizens

- Full control over personal data.
- Well-being and Quality of Life benefits from personal data sharing in key sectors.
- Access to personalised and cross-sectoral B2C services.
- Increased opportunities of personal data monetisation.
- New professional opportunities.

Specific domain applications of data-driven policy prediction include: city management, healthcare, participatory policy, environmental policy and disaster risk management.

3. Domain applications

3.1 Prediction in City Management

A first application area of data-driven policy prediction is that of city management. The cities of Nicosia (Cyprus), Pilsen (Czech Republic) and Sofia (Bulgaria) are experimenting with ways of using cloud technology and artificial intelligence to improve urban mobility policies.

The city of Nicosia is implementing effective Policies for Holistic Mobility and Accessibility within the scope of the project AI4PublicPolicy. The scope of the pilot is to increase citizens' satisfaction, to improve transportation sustainability, and reduce transport operational costs for the Municipality. Specifically, the aim is to provide citizens the possibility to access all the different transport modalities in a unified way – by APP or by (Transport) Card and allow the Municipality to manage and run a sustainable pool of services. Nicosia Municipality is willing to develop a Holistic Mobility and Accessibility Platform able to optimise transport management within the city, as well as costs and Sustainability. The Holistic Mobility and Accessibility Platform is currently targeting the Nicosia Small Buses Network, Parking Places around the cities, bicycles and E-Cars. The Platform is based on a wealth of data regarding the usage of the listed urban mobility services and which can serve as a basis for the extraction of evidence-based policies for holistic mobility and accessibility (from IT systems of minibuses, parking, bicycles and scooters). Data building the library of the pilot implemented by AI algorithms - AI tools and Opinion Mining Tools (XAI), and AI4PublicPolicy will provide a customised instance of the VPME to enable the Municipality to extract data and start building evidence-based policies. The key performance indicator is to identify and enhance public policies targeted to increase citizens' satisfaction to use public transport, improve transportation sustainability (i.e. reduced CO2 emissions, increased attention to the environment), and reduce transport operational costs for the Municipality. This experience will pave the way towards a new era of Mobility Policies for cities and citizens.

The city of Pilsen is implementing a Digital Twin of Pilsen leveraging on a set of projects. Traffic modelling started with OpenTransportNet (2014-2017), traffic analysis and simulation for policymaking was built by PoliVisu (2017-2020), while DUET (2019-2022) and S4AllCities (2020-2022) are implementing respectively simulations across multiple inter-connected domains and final a digital twin and simulations for crisis management. DUET (and previous projects) provides hands-on experience with new technologies. Specifically, DUET's main added value for the city is the interaction of simulation models in different domains (traffic 2D, air pollution 2D, traffic and air pollution 3D, traffic and noise 3D), as well as a set of traffic visualisations in 2D and 3D. Data helps in analysing the current situation, and setting strategies, milestones and measurable goals (KPIs). Technologies implemented include mobility (e.g. traffic detectors and models), city-operated cameras (traffic management and video analytics). Data include air quality, City GIS, and Digital Technical map. Data selection includes the management of city (open) data, new usage for existing data (e.g. video analytics, traffic detectors), photogrammetric 3D measurements, imagery for 3D modelling and texture mapping, AI to detect city property (traffic lights, sings, posts), measuring road quality and rescue system. Data use examples include cooperative Intelligent Transport Systems pilots, digital Twins of streets, and dynamic traffic management. Challenges are concerned about data interpretation results and its impact on their work routines, as well as skills needed to analyse and interpret data, as well as explain benefits to policy makers to get their support. Expected impacts for the society include improving the current mobility situation, nourishing innovation & startups support, engaging students and data enthusiasts, and supporting informed and data-driven policymaking towards sustainable city.

Through its involvement in Policy Cloud, Sofia Municipality (Bulgaria) will address urban policy as a critical success factor in improving the overall urban environment of the city. Policy design will be adapted through assessment and validation of policies and initiatives, based on big data analysis related to road infrastructure, environment and air quality, waste collection and waste disposal, transport and parking, and cleanliness of public spaces. Specifically, in designing policy, the city will combine the data from both existing sources and from new open data sets that become available. The existing data sources are the Sofia Municipality's Citizens' Contact Centre (CallSofia), facilitating direct

communication from citizens, industry and institutions wanting to signal non-urgent deviations from normal practice within the urban environment; and the Sofia Municipality's air quality measurement stations, analysing the data by category, type, territory and time will enable municipal and district administrations to identify problems, issues, and behaviour trends. Analysing the data by category, type, territory and time will enable municipal and district administrations to identify problems, issues, and behaviour trends according to a set of scenarios. The first scenario concerns road infrastructure, which is among the key and most budget consuming elements affecting citizens' everyday life. Sofia Municipality will be able to carry out a detailed analysis of the territorial distribution of the signals by categories / types, areas, districts, major transport roads, etc. The second scenario deals with air and environmental quality, which are among the key urban topics affecting citizens' everyday life. Sofia Municipality will be able to carry out a detailed analysis based on CallSofia signals and data from Sofia Municipality's air quality measurement stations. The results of the analysis will allow the municipal and district administrations to: identify the problems in the road infrastructure and adjacent urban environment and air quality, adopt or modify adequate policymaking decisions on budget planning and effective use of budget and public resources, and for the long-term improvement of air quality, facilitate better control, monitoring and prevention, validate existing policies and create new ones.

Summarising, Pilsen explained how by focusing on specific use cases rather than just technology they've been on a journey from simple data visualisation to a point where they have a (DUET) digital twin for interactive, real-time traffic simulations and predictions. They can now focus on encouraging green transport policies such as micro-mobility. Sofia showed how the cloud (Policy Cloud) is a gamechanger for them in integrating data from fixed environmental sensors with crowdsourced city contact centre data to identify challenges with road use that can be explored with city managers and policy makers. Nicosia introduced us to the potential of Artificial Intelligence (AI4PublicPolicy) in moving cities to a new era of personalised transport services with their creation of a central mobility platform that will help address individual accessibility and inclusion needs. The north star for all the cities was the execution of their public sector mission to keep improving the current mobility experience for residents and visitors, whilst nurturing innovation for the future.

3.2 Predicting health

Evidence-based medicine existed well before evidence-based policymaking. When dealing with the healthcare sector and policymaking, the two come together again and are both powered by converging data-driven technologies. Covid 19 has highlighted the importance of people's health and social needs and the need to address some long-standing challenges.

AI4PP is fostering data innovations towards establishing a European Healthcare Data Space (EHDS), which will comprise integrated, federated, well-structured, FAIR data (e.g. medical records, laboratory data, real-world data about patients, PROMs/PREMs, alternative data sources, scientific findings). The EHDS will integrate the findings/outcomes of any project's tools, as means of implementing a continuous improvement cycle where past findings are considered in the operation of the tools. Based on EHDS, AI will enable healthcare professionals and policy makers to collaborate effectively towards educated, data-driven, evidence-based, and patient-centric decisions for health prediction and care. The way forward is to build AI Technology innovations for trusted, accurate, and highly personalised clinical decision-making and policymaking, explainable AI (XAI) and causality inference techniques, Federated Machine Learning (FML) solutions, AI-based Healthcare Technology Assessment (HTA) assessments; and powerful Patient Digital Twins. The final output will be a virtualized cloud-based platform to centralise access to AI resources and to enable the integration, consolidation and sharing of assets, including datasets (via EHDS), analytical models for clinical and policy makers' decision-making, AI/ML algorithms and advanced AI tools (e.g. XAI, FML, Patient Digital Twins).

In the specific domain of cancer research, IntelComp provides an application based on the identification of STI policy needs. The developed platform is able to analyse large volumes of unstructured text data, using artificial intelligence tools (natural language processing pipelines, a machine translation system, automatic classifiers, topic modelling or graph analysis) on High Performance Computers.

The first step of the scenario constitutes an analysis of the EU intervention logic on cancer, describing the general vision, operational objectives, targets, policies and roadmap actions, policy initiatives and STI outcomes. The second step entails a stakeholder consultation aimed to develop policy questions, needs and STI priorities. The final step is a prioritisation of the policy questions by a living lab and the identification of use cases. The use case of the IntelComp platform identified so far is an analysis of the impact of funded research projects and the characterization of 'impact pathways' in France. Such use case will be based on data on scientific production ("output") of funded projects in terms of scientific publications, patents and clinical trials; data on medical impact ("outcomes") of research projects in terms of good practises (citations in clinical guidelines), new treatments (pharmaceutical industry) and new diagnostic screening techniques (industrialists / start-ups), building on data from social media, drug use and new diagnostic technologies; and finally on data on social impact ("impacts") of funded projects, in terms of media impact (via the media & social networks), topics of funded projects most often included in position papers; topics of funded projects corresponding to the expectations of patient organisations, positioning of projects in relation to public health data, leveraging on papers and health data.

3.3 Predicting radicalization

Policy Cloud is implementing a pilot on participatory policies to counter radicalization in Italy. The project makes use of a tool-based, meta-simulation methodology, to estimate the outcomes of different policy alternatives, providing side-by-side examination and comparative analysis of their assumptions, mechanics and outcomes. Specifically, the project provides an integrated cloud-based environment for data-driven policy management, technologies for policy optimization across sectors, a policy development toolkit allowing tailored policymaking, interoperable and re-usable models and analytical tools. The integrated cloud-based service environment for data-driven policy management includes a policy development toolkit, a data marketplace, cloud capability, re-usable models and analytical tools, a policy management framework and an ethics framework.

The purpose is to reduce the occurrence of radicalisation by early detection of warning signals and potential risks from social media and other data sources. The project promotes secure access to public spaces for more people by timely adopting cost-effective counter-measures, and encourages citizen engagement and trust in the perceived legitimacy of public authorities. In that regard, the project starts from a set of challenges including retrieving and assessing information from different data sources, presenting the outcomes of the analysis using advanced visualisations, identifying current/future trends and potential risks/threats, keeping track of people moving from mainstream, and the use of coded / hidden language. A first scenario considered deals with monitoring the occurrence of radicalization incidents in a given area leveraging data by the Global Terrorism Database and RAND corporation. A second scenario with identifying the main actors (individuals or groups) involved in violent activities or propaganda spreading through online and offline activities, through data by the Global Terrorism Database and RAND corporation, as well as Twitter, Reddit and RSS feeds. Scenario number three depicts an analysis of the current and future trends of radicalization efforts (keywords detection, new entity recognition, new terms identification), leveraging on data from Twitter, Reddit and RSS feeds. A final scenario depicts the assessment of online propaganda, and it consists in understanding specific events and online activities (sentiment analysis, opinion mining, location surveillance, user monitoring), using data from Twitter, Reddit and RSS feeds.

3.4 Predicting in agro-food markets

The simulations of all alternatives are carried out on a common modelling and environment for execution and analysis, facilitating their side-by-side examination and comparative analysis of their assumptions, mechanics and outcomes. The methodological steps for the simulation consist in creating and executing transparent simulation models for each alternative, enriching these models with meta-simulation structures and processes that make explicit and ground the assumptions behind each alternative, and evaluating alternatives in terms of explicit criteria

supplied by the policy makers. The experimental prototype used is called Politika, whose implementation will be integrated to the actual use cases of the Policy Cloud project, in order to develop a graphical user interface for policy designers with little coding experience.[22] The application of Politika is currently under discussion to support the Aragon region policy to promote their wines in different markets. Specifically, the application would deal with the different motivations for buying a certain wine, be these the price, the buyer's income, the type of wine and social influence, and the promotional campaigns. The next steps involve integrating the Politika implementation to the actual use cases of the Policy Cloud project, developing a UC related to the Aragon pilot, and expanding our methodology to the design and analysis of networks of policies as well as the reasoning capabilities related to evaluation of policy alternatives in Politika. Policy Cloud works in close contact with the European Open Science Cloud, as its data and services will be made available in it.

3.5 Predicting the environment

We are in the midst of a climate crisis which is contributing to public health problems. Governments need to reduce and eliminate emissions fast. Environmental policies exist but have largely failed to make a difference as measures have mainly focused on making emission reductions profitable for business, think carbon credits, energy saving bulbs. For more sustainable outcomes, creative and innovative policies and tools need to be co-created by the whole value chain. Digital enablers such as IoT, Cloud and Big Data ensure stakeholders can work together with a common-view/visualisation of the situation so they can simulate and predict the impact of differing policy choices. Policy predictions can improve environmental policy at both the macro and micro level.

At the macro level, policy prediction can support STI policy for Sustainable Blue Growth under Climate Change. Under the lead of the IntelComp partner ARC (Athena Research Center) in Greece, the project explores the co-development and use of the IntelComp Platform for activities such as a holistic mapping and understanding of the laws, policies and strategies at international, European and national level, as well as an analysis of innovations and new technologies to enhance mitigation and adaptation to climate change. Specifically, the project carried out a set of six thematic preparatory Living Lab Workshops on Climate Change and Blue Growth involving the international cluster for research on sustainability transition. The procedure starts with mapping stakeholders: 1) Enablers, which are organisations enabling the activities of living labs such as public actors, non-governmental organisations (such as towns), municipalities, and regional development organisations; 2) Providers, which are development organisations such as educational institutes, universities, or consultants; 3) Users, representing the citizens or end customers, and they are active or passive actors that participate in living labs in various roles; 4) UHlizers, which are the public or private organisations that will benefit from the results of innovation activities in many ways. The second step involves engaging the stakeholders offline (through co-creation workshops) and online through tools such as MIRO. The goal of the engagement is to discuss risks and challenges in the areas Climate Change and Energy, Climate Change and Industry, Climate Change, Agriculture and Food, Climate Change, Forests and Land Use, Climate Change and Transport, Climate Change, Buildings and Cities. Further engagement aims to understand various stakeholder perspectives, discuss the findings of the challenges, develop a common vision, devise trajectories of change, and co-create a socio-technical roadmap, in the view of devising final evaluation questions and indicators.

At the micro level, policy prediction can be leveraged for city green planning. Within the context of the DUET project, the city of Athens (Greece) is carrying out a process of digital transformation as a smart-hub for city green planning, together with a what-if analysis on green routing and alternative mobility. Specifically, the case starts from a lack of unified and correlated data resources (traffic and environmental) and fosters data-driven policymaking on urban planning with citizens' active involvement as solution proposers, strategy testers and data sensors/providers. Specifically, the case starts with the challenges of digitally transforming Athens, such as lack of open, accessible, usable, interconnected data sources as well as from the typical challenges of a metropolitan city, such as traffic, environmental issues, air-pollution, lack of green spaces/routes. The data sets that are to be leveraged for the project are in the realm of mobility, e.g. public transport itineraries, timetables, stops, traffic data from open sensors; health and environment, e.g. air quality sensor data and maps, air pollution report; and horizontal, e.g. GIS, meteorological and map layers.

22. **https://youtu.be/1Ig3FtncyFM**

The insights coming from the pilot relate to:

- Correlation of traffic, transport and environmental data for green routing;
- Development of DT and dashboard for green routing;
- What-if scenario on pedestrian roads foreseen but not-implemented yet;
- Visualisation of proposed policies for evidence-based decisions;
- Citizens' contribution: feedback, propositions.

An interesting output is the Athens dashboard displaying traffic data from data.gov.gr, measurement of the impact of changes made on traffic, selection of calendar periods for comparison of traffic data and visualisation of data on circulating cars. In that regard, the expected impact includes effective policymaking through evidence-based urban planning, prediction of impact of proposed measures, visualisation of proposed policies and close to real-time response. Further, the use case boosts democratic policymaking via the inclusion of citizens in the decision-making processes as well as having citizens as solution proposers, strategy testers and data sensors/providers, therefore increasing the trust of citizens on city processes through the creation of an open and transparent Athens.

3.6 Predicting emergencies

Policy prediction also has the potential to transform disaster risk management. Through four emergency policy pilots, the DECIDO project explores how policy impact prediction and analysis will lead to measurable outcomes and further policy refinement. Specifically, prediction is used mostly in the agenda setting and policy design phases of the policy cycle, in order to understand causal relationships and simulate the impact of policy options in view of their refinement.

Specifically, there are four applications:

- Pilot on Forest Fires in Kajaani (Finland). Policy objective: prevention and protection against forest fires; Procedures to mitigate damage to nature, infrastructure and life;
- Pilot on Floods in Meisino Park - Turin (Italy). Policy objective: improve design of emergency policies related to floods and weather alerts;
- Pilot on Power Outages in Greek Municipalities (Greece). Policy objective: power outage management of public infrastructure and cultural assets of Greek municipalities via emergency response mechanisms;
- Pilot on Wildfires in the Aragon Region (Spain). Policy objective: improve the design of emergency policies related to wildfires and management of controlled fires.

DECIDO will link Public Administrations to the data and compute infrastructure of the European Open Science Cloud – piloting the access to and exploitation of a great wealth of additional resources. Decido will have three main outputs: an easy to use portal will be released to define, manage and evaluate PA policies in a collaborative manner leveraging services offered by EOSC (Catalogue and Marketplace), external services/tools to EOSC, data made available by EOSC (mainly through services B2Find and EGI DataHUB) and by other data providers (e.g. data.europa.eu), including Public Administrations themselves; involvement of local actors on (1) the methodological side (e.g. co-creation of indicators), (2) the identification of needs and priorities, and (3) the data generation (e.g. through citizen science experiments where applicable); a robust and realistic business plan will be developed backed up by a detailed cost-benefits analysis of ex-ante (not using DECIDO results) and ex-post (using DECIDO results).

Specifically, the project had developed a methodology for using data to implement the decision-making process, according to the following steps:

1. Write the storytelling with pilot needs and challenges
2. Understand what services EOSC can provide
3. Definition of the DECIDO Data Catalogue
4. Data from co-creation activities

3. Domain applications

5. Definition of how to use the selected data
6. Implementation of algorithms to analyse those data
7. Visualisation of Dashboard to take decision
8. Improve policy based on evidence-facts

To do that, DECIDO makes use of data catalogues from data.europa.eu, open data from municipalities (flood areas, waterways, energy consumption and power outages), data from co-creation activities, EOSC Data (climate, earthquake, calendar data, statistical data on the mental health), GIS DATA (flood probability, location and category of municipal buildings), and satellite data from Copernicus or Corine (land uses, vegetation index, satellite images). Further, the project collects needs and challenges of policy makers and maps them with respect to its solutions:

- Employees might not want to be trained in new Technologies (need/challenge) - Co-Creation of a Training Plan with stakeholders (solution);
- Be compliant with GDPR/legislation (need/challenge) - Data Privacy Impact Assessment (solution);
- Lack of data (need/challenge) - DECIDO Data Catalogue (solution);
- Find available resources (need/challenge) - EOSC services & infrastructure (need/challenge).

The project also defines the expected impact for each stakeholder group: Policy makers will be able to reduce the time needed to generate good quality services as well as improve policymaking services, Citizens will actively participate and contribute to public policymaking, have higher data access for better and informed decisions, Businesses will accelerate innovation and business development and reduce costs related to innovation generation, and Research centres and scientists will access a large volume of data for scientific purposes.

4. Challenges to overcome

What are the challenges policy makers face when forging the future with policy impact analysis & prediction? The interactive Evidence Based Policymaking in Europe Summit identified common challenges in the application of evidence for policy prediction across the different application areas. Challenges are related to increasing the capacities of public administrations on the one hand and ensuring the active participation of societal stakeholders on the other. Public authorities need to be enabled to adopt data and cloud technologies (from the PA and research sector) to support evidence-based policies. Their ability to do so is related to three areas of challenges: day-to-day management of operations, construction of a future workforce capable of leveraging data for policy prediction, and ability to balance budgets. Building trust is considered a key challenge when it comes to active participation of societal stakeholders.

4.1 Strengthening the capacities of public administrations

Managing operations

Becoming a data-driven public sector capable of leveraging data from governmental and non-governmental sources for policy prediction demands a structural change in the technical, organisational and legal operations of public administrations. It is not just a case of introducing a new tool or policy, public organisations need to drive continuous innovation to keep up with the people's needs and the market pace. Regarding technical operations, a lack of access to high-powered computing together with insufficient quantity and quality of open, unified and interconnected data are observed. It is not always easy for data producers or users to determine the value of data, due to a lack of data evaluation standards and assessment tools. Furthermore, the integration of technology-related initiatives with traditional policymaking processes suffering from service siloes and legacy systems proves to be difficult, whereas this is essential to move from a testing bed to widespread innovation. This includes the assessment of the transformative impacts, benefits and risks (including ethical) of the deployment of big data tools and methodologies and the use of cloud infrastructure. The difficulty faced by data producers is balancing their data's perceived value (after sharing) against risks exposed (upon its sharing) despite adhering to standard guidelines. Finally, legal operations, such as data security policies and transparency mechanisms must be up to par to ensure compliance with GDPR and national/international legislation.

Building a future workforce

Having a public service workforce with sufficient digital literacy to adequately process, share and reuse data is a challenge for a more widespread use of data for policy prediction. Moreover, policy makers and city managers may be concerned about data interpretation and the impact on their work. Therefore, investments in the development of digital skills and a data-driven culture are essential.

Balancing budgets

Managing operations and building a capable workforce have clear budgetary implications. Adequate resources (e.g. in terms of budget, personnel, infrastructure) should be ensured prior to implementing any initiative. Of course, cost effectiveness needs to be considered and this may be difficult given the often long-term results rendered by EBP.

4.2 Ensuring the participation of societal stakeholders

Several aspects of operational management within governmental organisations have a direct impact on the trust of citizens and businesses and therefore their willingness to both actively participate in EBP and accept the results. These aspects include the implementation of data protection regulation, the transparency in the decision-making process and the demonstration of the impact of EBP. Policy prediction cannot solely depend on data generated or commissioned by public administrations. Especially when it comes to agenda setting, citizen-generated datas is very powerful. However, it needs to be clear to the community, before their participation, how their datas is going to be analysed and used within the policy life cycle and how it will benefit them. The perceived loss of control over data can be a hindrance. More specifically, for business data providers, there may be a fear of losing trade secrets due to unintentional exposure or malicious reverse-engineering. In turn, for public administrations the risk of navigating legal restrictions in the face of possible data policy breaches (including GDPR and exposure of private identities) is significant.

In addition, the lack of suitable data sharing ecosystems hinders both internal (between public organisations) and external (between public organisations and societal actors) data sharing and large-scale participation. More specific concerns include: 1) the lack of robust legal and ethical frameworks, as well as governance models and trusted intermediaries that can guarantee data quality and reliability, as well as its fair use; 2) the lack of widespread adherence to emerging best practices and standards (e.g. interoperability, provenance and quality assurance standards); 3) the lack of technical provisions to better address European concerns like ethics-by-design for democratic AI, and to better address scalability challenges posed by the rapid shift towards decentralised mixed-mode data sharing and processing architectures.

5. Recommendations

What can be done to overcome the identified challenges in the implementation of impact analysis and prediction tools for evidence-based policymaking? The stakeholders present at the Evidence Based Policymaking in Europe Summit 2021 jointly identified the following recommendations:

Start small, experiment, then scale fast through co-creation. Start by focusing on smaller use cases then build up to tackle larger city-wide problems. If you begin with the complex big picture, you may never get started. Keep experimenting with data. Experimentation for joint exploration of market capabilities. Look at data that already exists and see how it can be used and integrated with other data sources for new scenarios and solution development. Unlock innovation through co-creation. Take your stakeholders with you, get them interested by visually showing them examples of the new tools and get them involved in the innovation process. Extend the involvement of the Administration once related to providing access to data. Promote/Disseminate the experience and best practices stemming from successful projects.

Foster a culture of innovation building on successful experiments and pilots. Don't be afraid of data or technology. We are forging a new area of policymaking, so flip your view point and see the technologies and tools as an opportunity to create new possibilities. Leave less room for hesitation on the part of policy makers towards technology, as it is beneficial to their work and not an obstacle. Promote openness to the creation of an ecosystem with science and citizens.

Leverage European data sharing infrastructures to increase data availability and accessibility across policy areas and government levels. Administrations have a certain level of responsibility with the data that can be used in this type of technology. They should make public data as much available as possible, to train artificial intelligence. Variability of data is the fuel for AI. This includes the responsibility to break silos and share data among different players. European data sharing infrastructures should be considered public utility infrastructures that everyone can contribute to and make use of. To do so, alignment and standardisation across initiatives, countries and regions is crucial. Set up an EU-wide data repository as well as national and European data platforms for AI that include all necessary tools for data governance, annotation, and storage, next-generation networks, analytics software and, most importantly, datasets through a structural and investment fund.

Create trusted data spaces to facilitate data sharing by citizens and businesses. European data sharing infrastructures cannot only facilitate data sharing by public administrations but by citizens and businesses as well. For this purpose, it is important to support research on and development of industrial solutions for fast, secure and legally compliant industry data sharing (e.g. encryption) and stimulate sharing of industry data. The creation of trusted data spaces for specific sectors can equally boost the sharing of crowd-sourced data (e.g. healthcare, automotive and agri-food) where actors in a given value chain accept to share data, yet the rights related to the data rest with end users and/or the stages of the value chain where most of the value is being created. Finally, mechanisms should be developed for the protection and control of personal data by individuals, such as a data donor scheme, allowing the donation of data for specific purposes, surrounded by clear governance and transparency, and safeguards for privacy protection. Such mechanisms allow individuals to be empowered by their data, thereby addressing some aspects of the requirements of trustworthy AI.

Secure the necessary multidisciplinary resources (financial, infrastructure, personnel) for effective and evidence-based policymaking by investing in continuous capacity building and using procurement to attract the right teams. Multidisciplinarity is a must when setting up data-driven organisations. Different fields of expertise, such as those of data scientists, policy specialists, legal experts and ethicists will enrich the approach to data analysis. Jointly, they can help AI to take into account all the relevant variables in a policymaking process.

Ensure European building blocks for data-driven policy prediction. Regulatory, strategic, financial and promotional action at the European level is needed to enable the widespread use of data for policy prediction across geographic and sectoral borders. Regarding regulation, consider the introduction of a data access regime on FRAND terms, namely fair, reasonable, and non-discriminatory as well as rules and guidelines to merge data spaces. Furthermore, European regulatory sandboxes can help de-risking. Engagement with ICT Standardisation bodies can help speed up processes and identify common building blocks. At the strategic level, it is advisable to deploy a European-wide Skills Strategy and introduce an EU-wide data governance practices and conformity assessment. Regarding financial support, further investment in R&D and relevant technology is necessary together with funding to test-drive cross-sectoral, cross-border use-cases. Finally, promotional efforts need to be made to encourage widespread adherence to standards, rules and guidelines.

6. Annex - Data-Driven Policy Cluster Projects

Title: AI4PublicPolicy - Automated, Transparent Citizen-Centric Public Policy Making based on Trusted Artificial Intelligence
Duration: 36 months, from 1 March 2021 to 28 February 2024
Cost: EUR 5,080,501
Keywords : #AI-based policymaking #cloud #co-creation #EOSC
Consortium: GFT Italy, EGI Foundation, INTRASOFT International, SIA, Novoville, UNPARALLEL Innovation, ViLabs, Arthur's Legal, Universidad Politécnica de Madrid, City of Athens IT Company, City of Genoa, Nicosia Municipality, Lisboa E-Nova, Reportbrain, Burgas Municipality, EKSO

Abstract: Artificial intelligence (AI) has the potential to revolutionise the citizen-centric development of public policies. This is the aim of the EU-funded AI4PublicPolicy project. This is an ambitious joint effort of policy makers and cloud/AI experts to deliver, validate, demonstrate and promote an open cloud platform for automated, scalable, transparent and citizen-centric policy management. Based on AI technologies, such as machine learning, deep learning, natural language processing and chatbots, the platform will be an Open Virtualised Policy Management Environment. The project will complement public policy development functionalities with the ever-important process of reengineering and organisation transformation activities towards ensuring the effective transition from legacy policy development models to emerging AI-based policymaking.

Title: DECIDO - eviDEnce and Cloud for more InformeD and effective pOlicies
Duration: 36 months, from 1st March 2021 to 28 February 2024
Cost: EUR 4,327,255
Keywords: #cloud #evidence-based #policymaking #EOSC #co-creation
Consortium: Engineering (coordinator), Tecnalia, The Lisbon Council, Fraunhofer, Kpeople, EGI, EY Advisory, National Technical University of Athens, KAJAANIN Kaupunki, KAMK - University of Applied Science, Comune di Torino, Volontariato Torino, Sustainable Cities, Fundacion IBERCIVIS

Abstract: Policymaking is a process that takes place in stages, starting from the initiation of a policy to its implementation and subsequent evaluation. In this context, the EU-funded DECIDO project enables public administrations (PAs) to take full advantage of shared data, analytical tools and methodologies, computational power, cloud services and co-created data for the development of better targeted and more effective evidence-based policies. The aim is to help PAs leverage the capabilities and services offered by the European cloud infrastructure. By developing solid and realistic business plans to ensure the long-term sustainability of the results, the project will contribute to boosting the perceived legitimacy of authorities.

6. Annex - Data-Driven Policy Cluster Projects

Title: Digital Urban European Twins for smarter decision making
Duration: 36 months, from 1st December 2019 to 30 November 2022
Cost: EUR 3,999,282.50
Keywords : #Cloud Infrastructures #Open data #Public sector innovation #Open government #Public sector information
Consortium: Vlaamse Gewest (Coordinator), imec, KUL, ATC, 21c, AEGIS Research, OASC, Grimaldi Studio Legale, Dimos Athinaion Epicheirisi Michanografisis, Virtualcitysystems, The Netherlands Organization for applied scientific research (TNO), Plan4all, City Of Pilsen, Is-Practice, GFOSS

Abstract: DUET envisions a world where cities understand the real-world impact of change before making crucial policy decisions. Convergence of IoT and Cloud Computing creates new possibilities for governance and service delivery using city data. However, a continuous cycle of operational challenges leaves the Government behind. This problem can be solved through the use of digital twins, which are digital replicas or representations of a system, process or place which mimics real world behaviour. They are basically real-time updated collections of data, models and algorithms allowing for better real-time analysis of assets. In that regard, DUET is developing 3D interfaces facilitating experimentation and co-creation to deliver policy and services that were previously impossible, to be tested in three pilots:

- Pilsen case: what-if analysis of the Patton bridge road closure;
- Antwerp case: visualisation of motorised traffic per street segment, based on sensor/mobile data;
- Athens case: mapping shady areas in the city.

Title: IntelComp - A Competitive Intelligence Cloud/HPC Platform for AI-based STI Policy Making
Duration: 36 months, from 1st January 2021 to 31st December 2023
Cost: EUR 4,362,936
Keywords : #evidence-based #policymaking #co-creation #NLP #Text- Mining #HPC #STI
Consortium: FECYT (coordinator), Barcelona Supercomputing Center, Universidad Carlos III de Madrid, Athena Research Center, TILDE, CITE, Technopolis Group Belgium, OPENAIRE, Spanish Secretary of State for Digitalization and Artificial Intelligence, HFRI, ZSI and HCERES

Abstract:.IntelComp sets out to build an innovative Cloud Platform that will offer Artificial Intelligence based services to public administrators and policy makers across Europe for data- and evidence-driven policy design and implementation in the field of Science, Technology and Innovation (STI) policy. Large STI datasets are processed on High Performance Computing (HPC) environment part of the European Open Science Cloud (EOSC) imitative. Public administration at all geographical and organizational levels, STI stakeholders and civil society produce a great amount of dynamic, multilingual and heterogeneous data (i.e. national STI strategies, plans and work programmes, calls, projects, reports, scientific publications, patents, dissemination articles, etc.), so understanding and analysing this data is crucial for evidence-based policymaking. The objective of IntelComp is to deliver a platform that provides tools for assisting the whole spectrum of STI policy, i.e., agenda setting, modelling design, implementation, monitoring and evaluation. It will do so by involving multi-disciplinary teams to co-develop innovative analytics services, Natural Language Processing pipelines and Artificial Intelligence workflows and by exploiting open data, services and computational resources from the EOSC, HPC environments and federated distributed operations at the European Union, national and regional level.

6. Annex - Data-Driven Policy Cluster Projects

Policy Cloud

Cloud for Data-Driven Policy Management

Title: Policy Cloud - Policy Management through technologies across the complete data lifecycle on cloud environments.
Duration: 36 months, from 1 January 2020 to 31 December 2023
Cost: EUR 5,047,884
Keywords: #cloud #evidence-based #policymaking #EOSC #co-creation
Consortium: Atos Spain (coordinator), IBM Israel, EGI, ICCS-National Technical University of Athens, Maggioli, Leanxcale, Ubitech, Okys, Trust-IT, University of Piraeus, Instituto Tecnológico de Aragón, Sociedad Aragonesa de Gestión Ambiental, Stolichna Obshinta, London Borough of Camden, ICT LC.

Abstract: Policy Cloud delivers a unique integrated environment addressing the full lifecycle of policy management: modelling, monitoring, enforcing, simulation, analysis and compliance. The environment utilises the capabilities offered by the European Cloud Initiative, with an emphasis on data analysis to facilitate evidence-based policymaking. Policy Cloud introduces a pioneering approach for the development of policies "collections" to exploit collective knowledge towards policy co-creation and cross-sector optimization. The co-creation of multi-modal policies will follow an innovative modelling process for structural representation of schema-based policies. Targeting high impact, Policy Cloud delivers a data marketplace enabling the creation of an entire ecosystem of stakeholders contributing, producing, processing and using policy-related data assets. The reusable models and tools and this marketplace provide the ground for the proposed dual-business plan, which has been compiled by the consortium to ensure the long-term sustainability and take-up of the Policy Cloud results.

7. Further reading

European Commission (2020). Best Practices in Citizen Science for Environmental Monitoring. European Commission, Brussels, **https://ec.europa.eu/environment/legal/reporting/pdf/best_practices_citizen_science_environmental_monitoring.pdf**

Haklay M. (2013). "Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation." In: Sui D., Elwood S., Goodchild M. (eds) Crowdsourcing Geographic Knowledge. Springer, Dordrecht

Howlett, Michael. (2009). Policy Analytical Capacity and Evidence-Based Policy-Making: Lessons from Canada. Canadian Public Administration. 52. 153 - 175.

OECD (2020a), "Digital Government Index: 2019 results", OECD Public Governance Policy Papers, No. 03, OECD Publishing, Paris, **https://doi.org/10.1787/4de9f5bb-en**

OECD (2020b), Enhanced Access to Publicly Funded Data for Science, Technology and Innovation, OECD Publishing, Paris, **https://doi.org/10.1787/947717bc-en**

OECD (2020c), The Digitalisation of Science, Technology and Innovation: Key Developments and Policies, OECD Publishing, Paris, **https://doi.org/10.1787/b9e4a2c0-en**

OECD (2020d), "OECD Case Study of Norway's Digital Science and Innovation Policy and Governance Landscape", OECD Science, Technology and Industry Policy Papers, No. 89, OECD Publishing, Paris, **https://doi.org/10.1787/20f80fa1-en**.

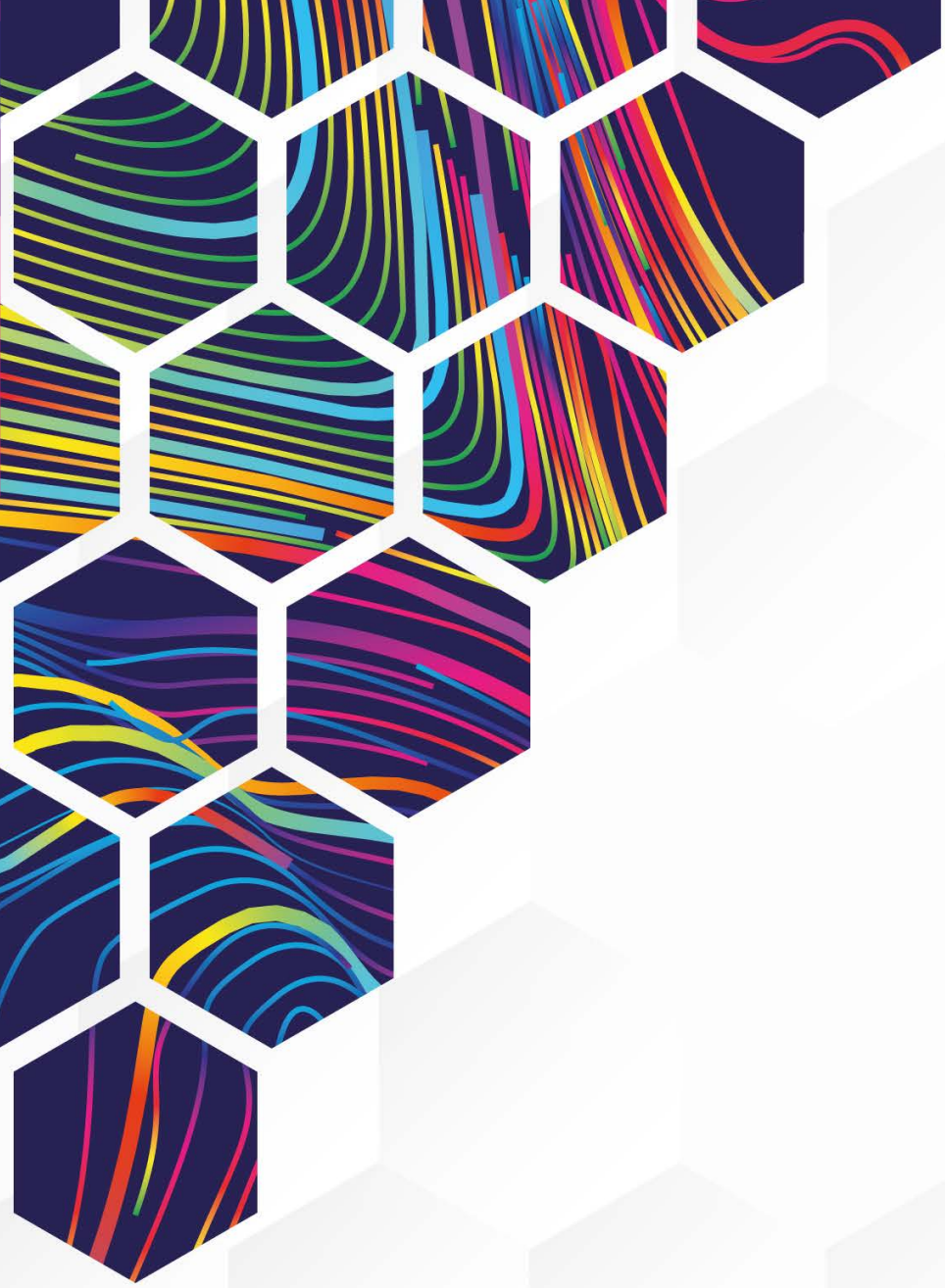
Ooijen, Charlotte. van, Barbara Ubaldi and Benjamin Welby (2019), "A Data-Driven Public Sector: Enabling the Strategic Use of Data for Productive, Inclusive and Trustworthy Governance", OECD Working Papers on Public Governance, No. 33, OECD Publishing, Paris, **https://doi.org/10.1787/09ab162c-en**

Osimo, David (2022). "Context is King: How Correct Data Can Lead to False Conclusions", The Evidence Hub, **https://evidencehub.net/blog/context-is-king-how-correct-data-can-lead-to-false-conclusions/**

Productivity Commission (2010). Strengthening Evidence-based policy in the Australian Federation, Volume 2: Background Paper, Productivity Commission, Canberra, **https://www.pc.gov.au/research/supporting/strengthening-evidence/roundtable-proceedings-volume2.pdf**

Sutcliffe, Sophie and Julius Court (2005). Evidence-Based Policymaking:
What is it? How does it work? What Relevance for Developing Countries? Overseas Development Institute, **https://cdn.odi.org/media/documents/3683.pdf**

U.S. Chief Information Officers Council (2018). "Foundations for Evidence-Based Policymaking Act of 2018", Policies & Priorities, **https://www.cio.gov/policies-and-priorities/evidence-based-policymaking**



Data Driven Policy Cluster

Co-creating digital tools for better governance

These projects have received funding from the European Union's Horizon 2020 Research and Innovation program. Under the following grant agreement numbers:
Policy Cloud No.870675, DUET No.870697, AI4PublicPolicy No.101004480, Intelcomp No.101004870, Decido No.101004605.