Designing an AI compatible open government data ecosystem for public governance

Evrim Tan, KU Leuven Public Governance Institute, evrim.tan@kuleuven.be

Abstract

AI solutions can significantly leverage the use of OGD ecosystems in public governance. For that, it is important to design effective and transparent governance mechanisms that create value in an OGD ecosystem through AI solutions. By analyzing governance challenges associated with OGD and AI solutions in public governance, this article presents a conceptual framework to design an OGD governance model, which adopts a platform governance approach and integrates the governance needs derived from the use of AI.

1. Introduction

The data-driven transformation in public administration is hinged on the theoretical premises of open government data (OGD) [1]. OGD refers to open data that are produced or commissioned by public bodies [2]. Open data is defined as the data or the content that can be freely used, reused, and distributed by anyone, only subject to the requirement that users attribute the data and that they make their work available to be shared as well. Governments, civil society organizations, and private sector representatives consider OGD as a building block for open government, as they see it as a key enabler of improved service delivery, transparency, and public engagement and as a result of better relations between governments and citizens [2].

OGD initiatives are expected to foster democratic and economic processes by promoting transparency, participation and collaboration, and provide opportunities for the development of new products and services [3]. However, current OGD practices suffer from technical, social, institutional/organizational, legal/ethical, economic, operational, political/policy/strategic challenges [4]. Additional challenges emerge when considering the actual use of OGD [5]. On the supply side, OGD programs are often designed not for citizens but technical experts and intermediaries, and the lack of institutional processes for dialogue prevents the integration of public feedback into existing strategies and programs [6]. On the demand side, the lack of incentives, interpretive tools, and contextual and technical knowledge among users can prevent meaningful data use [7].

Recently, the adoption challenges of OGD have begun to be assessed in the larger institutional landscape where the government organizations operate. This approach has been conceptualized by the ecosystem metaphor to assess the complex dynamics among different actors and concerns in the public governance domain [8,9,10,11,12,13,14,15,5]. In this sense, we use OGD governance to refer to the formal and informal arrangements that determine how public decisions are made and how public actions are carried out in an OGD ecosystem.

In a similar vein, the advancements in big data and artificial intelligence (AI) solutions call for the reevaluation of effective models for OGD governance. Studies show that the combination of AI with OGD has a huge potential to improve efficiency, innovation, and crime prevention in public governance, but AI is hardly used by the public to create value from open data [16]. Furthermore, risks of data privacy, and arriving at biased or wrong conclusions undermine the usability of AI solutions in OGD ecosystems [16]. Therefore, it is important to design effective and transparent governance and control mechanisms for policymakers to create value in an OGD ecosystem through AI solutions.

A particular challenge for the use of digital technologies in public governance processes is the 'governance of' and 'governance by' configurations [17] 'Governance of' refers to the design choices associated with the digital technologies in accessing and using the underpinning data infrastructure. The 'governance of' dimension focuses on how the data infrastructure affects the usage of the technology and links the data acquisition and data processing mechanisms in overall public policy processes. The rules and standards in acquiring and processing the available data, and how AI affects data processing and exploitation stages fall under the 'governance of' considerations. 'Governance by' refers to the use of AI in policymaking and policy implementation. Unlike the technical emphasis of the former approach, 'governance by' prioritizes the techno-social power dynamics and control mechanisms in the use of digital technology in

H C S S

public policy processes by looking into the role ascriptions of automated systems and human agents in the overall public governance. A model for OGD governance that is compatible with AI solutions should comprise both aspects.

The aim of this paper is threefold. First, by elucidating on the theoretical approaches to the OGD ecosystem to pinpoint the design principles for OGD ecosystem governance. Secondly, based on the theoretical and empirical cases to identify the key challenges associated with the 'governance of' and 'governance by' configurations for the AI-based solutions in the public sector. Thirdly, to develop a systematic decision-making tool to design a governance model for an OGD ecosystem, which integrates the governance needs derived from the use of AI. The purpose of this tool is to systematically identify and analyze the interrelationships among multiple change factors on governance design and to project the available design options based on the managerial, organizational, legal, technological, moral, and institutional variances in the OGD ecosystem.

The structure of this paper is as follows. Section 2 categorically presents the existing theoretical approaches to OGD ecosystem governance. Section 3 and 4 elaborate on the governance challenges associated with AI solutions in public administration following the 'governance of' and 'governance by' dimensions. Section 5 presents the systematic decision-making tool for the OGD ecosystem design in public governance. The conclusion section summarizes the main findings of the paper and shares recommendations for further research.

2. Conceptual approaches to OGD governance

The literature provides different conceptual models and approaches to the OGD governance design. Based on the existing models in the literature, we identify seven different approaches to OGD governance design. The end of the section presents the common denominators of these approaches and highlights the key determinants.

The first approach to OGD governance design is based on policy processes in the OGD ecosystem. Reggi and Dawes [5] identify two policy cycles in the governance of OGD. One policy cycle addresses the innovation potential of OGD, the other addresses how OGD might support democratic values of participation and accountability. The model integrates the diverse goals of actors in the open data ecosystem and allows the assignment of different role definitions for the intermediaries between data providers and the beneficiaries of OGD products for innovation,

participation, and accountability purposes. This model allows the integration of techno-social power dynamics in the assessment of governance design and tracing the influence of actors on the policy processes.

The second approach to OGD governance design is based on contingencies. For instance, Lee et al [12] distinguish the external and internal contingencies for decentralized and centralized governance approaches in terms of the architecture design of OGD governance. External contingencies refer to the environmental context of a public sector organization, such as regulative framework, market structure, social and economic dynamics, political and institutional factors. Internal contingencies refer to the design choices associated with the platform governance conditions such as degree of control, type of control and strategies for governance. The strength of the model is that it creates a dynamic link between the policy goals and the underlying system infrastructure in the data governance. As such, the model allows estimations of policy outcomes based on the changes in the data governance architecture and changes in the external and internal contingencies (e.g. organizational, regulation, or policy changes).

The third approach in the literature is based on system design thinking. Systems thinking is a holistic approach that focuses on the way that a system's constituent parts interrelate and expounds on how systems work over time and within the context of larger systems [18]. Following the systems thinking approach, Millard [12] defines open governance as linking and integrating the worlds inside the government as well as linking and integrating these with the worlds outside government for the specific purpose of creating public value. Accordingly, he identifies three main components under open governance systems: open assets, open services, and open engagement. Being built on the intersections of open assets, open service, and open engagement, open government is expected to act as a platform, where the government can support a range of actors to collaborate with each other, as well as with the government itself, by facilitating and orchestrating engagements, managing assets, and providing tools to generate public value [11].

The fourth governance approach to the OGD ecosystem focuses on the operational processes in the production and reuse of open data. In this approach, open data is linked to an open governance strategy in which the government builds an open system that interacts with its environment [8]. In an exemplary model developed by Maretti et al [8], the governance choices at the macro-level pertain to the operational processes that are the basis of the open data system, including the digital strategy of the public administrations and the problem of protection and use

of data. At the meso-level, governance choices pertain to the administrative processes of digital public administration. At the micro-level, governance choices pertain to the concrete use and reuse of open data in formal and informal teams among organizations in public administration and civic hackers. In this model of OGD governance, the ultimate governance design purpose of the government is to structure a participatory system that creates economic, political, social, operational, and technical benefits for participants.

The fifth approach to governance design of an OGD ecosystem is based on an organizational or institutional perspective. This approach relies on the platform theory from strategic management and information systems in the identification of the governance design constructs and distinct approaches for the implementation of OGD [9]. The conceptual model developed by Bonina and Eaton [9] for the governance of an OGD ecosystem brings together those constructs (i.e., core architecture, peripheral architecture, platform owner, contributors, developers, tools, rules) to analyze the organizational tools and resources in the supply and demand sides of open governance, as well as the institutional factors affecting the overall platform performance. In another study, Safarov [10] summarizes these institutional dimensions as policy and strategy, legislative foundations, organizational arrangements, relevant skills and educational support, and public support and awareness concerning open data.

The sixth possible approach to governance design is based on the management perspective. In an exemplary model for open data assessment framework, Welle Donker and van Loenen [13] identify governance as a framework of policies, processes, and instruments that structure the interaction between public sector entities and/or private sector entities to enable parties to reach their common goals. In their model, they identify five elements (i.e. vision, leadership, communication, self-organizing ability, and long-term financing) for assessing data governance in open data ecosystems. These elements are also influential on the data supply, and along with the user characteristics, they constitute the input to the ecosystem for the successful reuse of OGD. Welle Donker and van Loenen [13] underline those other important aspects such as the legal and policy frameworks and draw a clear demarcation between public tasks and private sector activities that additionally affect the performance of the overall ecosystem.

The last approach to governance design is based on the commons approach. Commons are based on the principles of bottom-up self-organization, the freedom of collective agency, polycentrism (multiple loci of governance), and subsidiarity (management at the lowest feasible level) [19]. According to Ostrom [20],

the world of natural resources is divided through the axes of rivalry and excludability, where the common goods refer to the rivalrous and non-excludable resources. More recently, the governance of online communities has attracted the attention of researchers [21], which treat open data as a common good. Fuster Morell [22] identifies eight critical aspects that define the direction of online creation communities (OCC) governance: collective mission or goal of the process; cultural principles/social norms; design of the platform of participation (where regulation is embedded in the code); self-management of contributions; formal policies applied to community interaction; license; decision-making and conflict resolution systems concerning community interaction; and infrastructure provision. These eight dimensions are linked to each other through the infrastructure provision. According to Morrell, infrastructure provision can be modeled across two axes: open versus closed to community involvement in infrastructure provision, and freedom and autonomy versus dependency on infrastructure. Based on the empirical analysis of fifty statistical cases and four case studies, Morrell clusters four provision models for OCC governance: corporation service (which is the case of Flickr), mission enterprise (wikiHow), autonomous representational foundation (Wikipedia), and assembly or assemblarian self-provision models (openesf.net).

Despite some overlapping dimensions, each conceptual approach emphasizes unique aspects based on the scope of governance activities. Notwithstanding the conceptual differences, we can highlight the following common denominators from the existing models on the design of OGD governance;

(1) OGD governance design depends on contextual factors such as the regulatory framework, organizational capacities, organizational culture, policy domain, ethical principles, public policy objectives, etc. The available design choices for OGD governance are contingent upon these contextual factors.

(2) OGD governance is most suitable to a platform ecosystem model where government and non-governmental actors can share and reuse the data through the platform. Not only the capacities of actors involved in the platform ecosystem but also regulations, institutional design, and the market structure influencing OGD are important for the effectiveness and sustainability of the governance design.

(3) Policy goals, principles, strategies on data governance, as well as the managerial, technical, and administrative capacities in the ecosystem determine the characteristics of the platform ecosystem and the system architecture design for the use of digital technologies.

(4) OGD governance design needs to address the role of the actors in the platform ecosystem separately as individuals and in communities, as well as holistically

assess all the constituent parts of the ecosystem in setting the rules of engagements in the use, reuse, and share of data.

3. Governance of AI

This section expands on the theoretical basis of OGD ecosystem governance with the governance requirements resulting from the configurations of AI-based systems in public administration. Particularly, we will elucidate the governance of AI-based solutions in addressing data acquisition and data processing challenges in the OGD ecosystem.

3.1. Data acquisition

The AI system acquires data by either conducting a primary data ascertainment through sensor systems and human data input, or by accessing available databases (e.g. machine logs, clouds, or Internet databases) in a secondary data ascertainment [23]. A data acquisition system samples the data input and transforms it into machine-readable data, while the software processes the acquired data for storage or presentation. The data feeding the system and the medium technologies and storages integrated into the system are key governance considerations in data acquisition. Challenges concerning the database size, data integration, data quality, and data standards (i.e. how and what data is collected, and what format it is stored in) can undercut the effectiveness of data processing and AI predictions.

For a successful implementation of AI strategies and programs, organizations must have access to a base set of data and must maintain a constant source of relevant data to ensure that AI can be useful in the selected policy domains. The input data can be in a multitude of formats such as text, audio, images, and videos. The wide range of sources to collect and store these data adds to the governance challenges. For successful predictions, all the relevant data must be integrated in a manner that the AI can understand and transform into useful results. A technological challenge for AI-based systems is to analyze unstructured data. For example, in healthcare, medical imaging represents a large share of relevant data, which even the most advanced AI-based systems (e.g. Watson) cannot read directly [24]. This means that depending on the data source, the AI system may need to be complemented with human experience.

Data quality is another core challenge in data acquisition. AI performs best when it has a good amount of quality data available to it. Therefore, AI solutions built on big data augment the performance of AI-based predictions. But big data pools different data from different origins that need contextualizing for analyses

and reports. Greg Hanson from Informatica argues that for well-curated data, enterprises should build a catalog of data assets and use engineering tools with AI built in the backend [25].

In the public sector, another challenge for data quality is the rules and standards employed in publicly available data. For example, GDPR obliges the purpose limitation principle on data acquisition, which may limit the pattern recognition functions of machine-learning systems. Therefore, not only the quality of data, but the variety of available data can affect the performance of AI applications.

The performance of AI is also related to the quality of the training data. Here, bias embedded in the training data is one of the biggest challenges that AI faces [26]. Often, data sources are laced with racial, gender, communal, or ethnic biases [27]. Biases embedded in the training data could easily lead to discriminatory and unjust consequences in policymaking and implementation processes. For example, the "Correctional Offender Management Profiling for Alternative Sanctions" (COMPAS), a system to predict whether a defendant would re-offend, was found to be as successful (65.2% accuracy) as a group of random humans [28] and to produce more false positives and less false negatives for black defendants [29].

Furthermore, data privacy in data acquisition, and how to achieve an appropriate balance between privacy and data acquisition is another pressing issue in AI adoption [30]. Hence, the ethical challenges accompanying data acquisition processes need to be addressed for the wider adoption of AI-based systems in public services.

3.2. Data processing

The acquired data is stored in data servers for data processing. For security or privacy issues, some organizations may need to store data on in-house data servers. Those organizations need to maintain the cost of in-house data servers and a technical support team. Cloud-based alternatives, centralized government data silos, or digital crossroad data centers present more cost-benefit-friendly solutions and effective upscaling of AI-based systems. However, the interoperability considerations and administrative burden in data acquisition can undermine the appeal of these alternative data storage options for organizations.

The acquired data is processed by algorithms and machine-learning techniques. The quality of the human resources and available software solutions in data analysis are some key considerations for public sector organizations. The massive computing power necessary to process big data to build an AI system, and to utilize data-intensive machine learning systems such as deep learning, can also technically and financially challenge

organizations. The effectiveness of AI-based solutions requires high-end processing power, and the cost of large infrastructure requirements are considered as impediments to the adoption of AI technology [31]. In addition, there is a high demand for a limited number of AI experts, which is associated with the increasing cost of education and salaries [32]. Cloud computing environment and outsourcing can mitigate some maintenance costs. Nevertheless, organizations need to plan in advance the cost and technical requirements to maintain higher computational speed requirements along with higher availability of data to scale up their AI-based systems.

The output procured by the machine needs to be presented in an easily interpretable way by the user. A caveat on data visualization is the phrasing of wording and visuals in the analysis results. The confidence intervals of the results may vary depending on the employed algorithm or statistical techniques, but studies on human cognitive bias show that the framing of information and behavioral factor tend to affect the interpretation of results by the users [33]. There are ongoing technical efforts to detect and remove bias from AI systems, but these efforts are considered in the early stages [34]. A mathematical notion of fairness and technological fixes have their limits in overcoming systematic and cognitive biases in each social context [35].

4. Governance by AI

'Governance by' refers to the use of digital technology in policymaking and policy implementation. The following subsections will elucidate the governance challenges associated with AI-based solutions in policymaking and policy implementation.

4.1. Policymaking

Eggers et al [36] presume that AI can change the role of humans in the policymaking processes in four ways. First, AI can relieve public workers by taking over repetitive tasks, allowing public servants to focus on more valuable tasks. Second, AI can help to break up a job into smaller tasks and can take over as many of them as possible, improving the efficiency of policymaking processes. Third, AI can replace a human agent by automating policymaking processes. Fourth, AI can augment the performance of public servants by complementing their skills and improving the effectiveness of policymaking processes. While each scenario infers diverse efficiency gains in policymaking, their disruptive effect on human agents in administrative and policymaking processes varies drastically.

By the same token, according to Scherer [37], AI presents two risks in policymaking, namely the loss of predictability and the loss of control. Loss of predictability is caused by big gaps in processing capability between AI and human agents, where AI can process huge amounts of data at high speeds beyond the capabilities of human agents, making the results no longer comprehensible and verifiable for humans. Thierer et al [38], therefore, define AI-led information processing systems as 'black box' for human end-users, turning their role as data feeder and recipients of results without the ability to control the validity of methods and criteria in policymaking. Loss of control refers to the dislocation of human control in influencing the system operations. The self-learning mechanisms of AI allow it to reprogram itself for process optimization, and the personnel responsible for maintenance and monitoring of the system can partially or completely lose the ability to realign the system with policy objectives.

Furthermore, AI-led policymaking is stranded by the legal, moral, and ethical framework conditions. Human decisions in public policy are political in nature and soft skills such as ethical trade-offs, social rules, empathy, humanity, and conscientiousness have a conclusive influence on the outcome of decision-making processes and their subsequent evaluation [23]. For the moment, AI technology lacks these human cognitive qualities and has limitations to take over human roles in public decision-making processes [39].

The use of AI in policymaking also raises concern about its behavioral impact on human-led decision-making processes. For instance, a study on an automated profiling system for unemployment claims in Poland has found that less than 1 in 100 decisions made by the algorithm have been questioned by the responsible clerk [40]. Behavioral factors such as lack of time to ponder the details, fear of repercussions from the supervisors, and a belief in the objectivity of the process appear as driving influences behind the behavior of clerks, making human-led decision-making processes practically automated systems [41].

By taking into consideration these potential challenges, Janssen and Kuk [42] envisage governance by AI can at best be used only on mundane tasks. Similarly, Eggers et al [36] underline that AI is most suited to handle repetitive, highly structured, and regulated work processes. They recommend organizations issue a work structure and process analysis regularly to draw up the respective areas of application for AI in policymaking. However, this caveat does not necessarily pertain to the capabilities of AI technologies, rather to the task encroachment and accountability risks imposed by advanced AI solutions in public administration, as a growing number of use

cases show that AI solutions are becoming more attuned to handle complex and cognitive tasks [43].

4.2. Policy implementation

AI can replace the role of human agents over the delivery of public services, and engage with end-users, not only in policymaking but also in policy implementation. AI-led bots and AI-powered digital interfaces can especially replace public servants for repetitive and predictable tasks. Many authors even highlight the risk of increasing the replacement of human agents in healthcare, education, logistics, and organizational processes, which raises the threat of mass unemployment in various public sector areas [44].

Nonetheless, the wider usage of AI in policy implementation can also create new jobs and roles in public services. Skills in algorithms and the use of AI in systems have already become high-demanded skills in the job market [45]. Not only the technical staff but also public managers are required to enlarge the working capabilities by AI usage and to develop a better understanding of how AI can supplement the workforce to achieve better results faster [36]. The challenge is, however, the technical and managerial staff or the people working in the jobs at risk do not necessarily have the required skills and formal training to ease the pressure on human resource management in public administration.

The impact on the workforce is only one governance consideration for AI usage in policy implementation. AI safety and end-user behaviors are other concerns in the governance by AI. AI safety refers to assuring the secure performance and impact of AI [46]. These safety concerns not only relate to issues of information security, but also to complex and safety-critical situations resulting from circumstances where the AI may learn negative behavior from its environment and misunderstand its surrounding [47]. Taking the necessary precautions and safety measures are important for the scope of AI applications. Bostrom and Yudkowsky [48] underline the necessity of AI technology to be resilient against adverse manipulation by humans. For AI applications based on reinforcement learning and automated executions, safety measures need to be in place to avoid catastrophic consequences. It is also important to avoid negative side effects to the working environment during the learning process of AI applications [49].

Lastly, studies show that the uptake of new services by AI-bots has not been particularly high, leading the authorities to believe that some form of proactive marketing would be necessary to change citizens' behavior [41]. Kuziemski and Misuraca [41] assert that to pursue such projects further, local authorities need to develop user experience and awareness. However, user experience and awareness cannot be enhanced solely by marketing practices, and trust in public administration is important to facilitate the transition. Especially, recent studies suggest that public service processes have a significant impact on citizens´ trust in public administration, and in particular the absence of corruption is a strong institutional determinant in trusting public administration for the use of digital technologies [50]. Therefore, holistic approaches are needed in service design to leverage AI technology in service provider and citizen interactions.
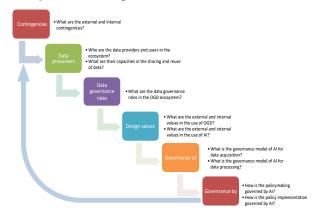
5. A conceptual framework to design AI compatible OGD governance

In this last section, we present a conceptual framework to design an OGD governance model, which adopts a platform governance approach and integrates the governance needs derived from the use of AI.

The purpose of this tool is to systematically identify and analyze the interrelationships among multiple change factors on OGD governance design and to project available AI-based solutions for the OGD ecosystem by assessing the managerial, organizational, legal, technological, moral, and institutional variances. Through the recursive and reflexive analysis of each step, policymakers and system designers can develop reliable strategies in leveraging AI solutions for the use of OGD in public governance.

Figure 1 presents the six steps in the design of an AI-compatible OGD governance. Each of these steps is elaborated on below.

Figure 1. The 6-step model in designing AI compatible OGD governance

Step 1: Identifying contingencies

According to Lee et al. [12], characteristics of platform governance are influenced by external and internal contingencies. External contingencies refer to

the environmental context of a public sector organization, such as regulative framework, market structure, social and economic dynamics, political and institutional factors. For example, in the OGD context, some external contingencies can be related to the legislation and regulations governing the collection and the reuse of publicly available data. External contingencies also pertain to the data feeders in the ecosystem. For instance, the sources of data and its formation (single or multiple sources), the role of public and/or private organizations in the ecosystem, and the data quality and control standards are some aspects of market structure.

External contingencies also influence the choices in internal contingencies. Compared to external contingencies, internal contingencies are the more immediate and direct causes of the characteristics of the platform. Internal contingencies are platform strategy, multi-homing strategy, governance configuration, open strategy, and platform maturity [12]. Platform strategy pertains to the rules on access, control, data provenance, conformance, and monitoring the quality of OGD. Multi-homing and open strategy respectively pertain to the rules of access and the reuse of OGD by the platform users. Governance configuration pertains to how a desirable behavior in the platform is achieved based on authority- or trust-based formations. Lastly, platform maturity pertains to the level of participation and the critical mass of data accumulated in the platform.

Hence, the first step in OGD governance design is the identification of external contingencies that may impede the available design choices in platform governance. Subsequently, the effect of external contingencies on existing strategies, internal and external rules, and structures for the platform governance needs assessment. For example, political scandals and previous cases with fraud may prioritize transparency and privacy considerations in the design of platform governance. It is important to highlight that the contingencies also depend on the outcomes of AI applications in platform governance (see the feedback loop in Figure 1). Therefore, an iterative and dynamic process is called for in the identification of contingencies.

Step 2: Identifying data prosumers

Data prosumers are the organizations, individuals, and automated agents supplying raw and/or processed data in the platform and reusing the data for public, commercial, social, or academic purposes. In a platform ecosystem, the data prosumers can be public or private, and they could be communities or individual users that can access and reuse the data available in the platform. This step is not only about identifying data providers (and beneficiaries of OGD) but also about assessing the

existing capacities among prosumers, and about cultivating the data for the growth of the platform ecosystem. For example, in a system dominated by a few main data providers, the data storage and processing capacities of those organizations would eventually delimit the growth and the big data potential of the ecosystem. Depending on the technological choice in AI-based solutions, the role of data prosumers may vary. For example, in big data-led AI systems, private sector organizations holding large amounts of data may have a more important role in data production.

Step 3: Assigning data governance roles

Lee et al. [12] identify four main data governance roles in a platform ecosystem: data committee, data manager, data owner, and data subject. The data committee is responsible/accountable for clarifying the role of data in the platform ecosystem [51]. It makes decisions about the purpose of data use, desirable behaviors, and appropriate governance mechanisms, aligning business goals. The role is generally taken by the orchestrator of the platform owner. Data manager refers to the role of data management in the platform ecosystem including data collector, data steward, and data custodian. It is responsible for the implementation of data management tasks and verifying the conformance of data governance rules. The role also can be shared with platform users as they can monitor or audit the use of data. The data owner is an individual (or company) who owns data by uploading user content or profile, or by providing the result of analytics jobs to the platforms. The data subjects are the people whose personal data are used in the ecosystem and who are identified or identifiable (Art. 4.1 of the GDPR). The process of assigning data governance roles in an OGD ecosystem requires the interactions between public and private stakeholder organizations. The level of trust vested in public and non-public actors as well ownership of AI technologies may determine the role and power distributions in platform governance. A key challenge for system designers is to create a transparent and inclusive governance mechanism in the distribution of data governance roles that engender legitimacy and trust in public governance processes.

Step 4: Identifying design values

Design values stem from the internal and external value propositions associated with the use of OGD and AI in public governance. Internal values represent the value of OGD and AI-based solutions to a particular organization based on organizational, technical, and managerial investment requirements and expectations. External values stem from the expectations from the platform users and beneficiaries on the (re)use of OGD and AI in public policy processes and service

provisions. For example, improving the efficiency and effectiveness in policymaking processes concerning tax fraud detection by leveraging machine learning algorithms is an example of an internal value proposition. Alternatively, protecting the privacy rights and business interests of platform users are forms of external value propositions. An effective governance design process needs to identify the value propositions associated with the way OGD is used and align the value propositions with data governance goals, in an order of importance.

Indubitably, the values are subjective to the actors involved in data governance and the choice of particular value propositions might incur trade-offs and political costs. Therefore, it is not an easy task for policymakers to create an ordinance among value propositions. It is also possible different prosumers may have contradictive value propositions or expectations in data governance. Therefore, the involvement of the prosumers in this stage is necessary to understand the value hierarchies of prosumers and thereby to design more legitimate platform governance. A possible methodological tool to that end is the means-end chain (MEC). MEC is a formal heuristic tool to define a hierarchy of goals and values that represent potential identities of the actions necessary for the person to reach his or her goal [52]. In a classical MEC, goals are grouped through four stages in a value chain: attributes (concrete and abstract attributes of a goal), the functional consequence of attributes (immediate, tangible effects on the user), the psychological consequence of attributes (the emotional impact on the user), and values (the emotional state-of-mind the user is trying to achieve). According to Gutman [52], MEC provides a flow toward desired ends at successively higher levels of abstraction extending from the product or service to important aspects of consumers' self-concepts. In doing so, MEC creates hierarchically related sets of elements across levels of abstraction and allows differentiating goals from values. As Gutman puts it, goals represent what we want, and values are why we want them. In an exemplary study on the choice of beverage, while studying for a test, first-year marketing students in a New England university replied to a series of 'why' probes to elucidate the instrumentality of an act (e.g. drinking a particular beverage) in goal attainment [52]. The answers of students were processed through a laddering methodology and presented as a hierarchy of goals and value propositions associated with the choice of beverage (e.g. coffee or soft drink).

Step 5: Designing the governance of AI

Preferences in the design values directly influence the choice of AI-based solutions to process the data in the OGD ecosystem. AI-based solutions (e.g. machine learning, advanced analytics techniques) may have different governance attributions in the system architecture choices, with different value propositions and trade-offs influencing the architecture design of the platform ecosystem. Therefore, the design principles and data governance goals set in the previous steps delimit the choice with AI solutions and type of data (e.g closed or private) brought in data acquisition and data processing processes. Additionally, considerations about technology readiness and maturity, associated organizational and human resource capacities, transparency, explainability, and interoperability of AI solutions are other important issues that can delimit the available choices in AI solutions and their inclusion in the OGD ecosystem.

Step 6: Designing the governance by AI

The last step in the governance design of an AI-compatible OGD ecosystem is evaluating the adaptability of policymaking and policy implementation processes with AI solutions. Techno-social power dynamics among data prosumers, control and incentive mechanisms to share and (re)use data in public policy processes, safety, and privacy considerations, and the role distributions among automated and human agents in public governance must be considered. The policy choices on the use of AI solutions in the OGD ecosystem and their policy outcomes are expected to affect, in time, the roles of prosumers and their involvement in platform governance. Therefore, it is important to envisage effective risk and change management mechanisms to adjust changes in user behaviors. For that, there is a need for periodic re-evaluation of the relevance of the existing policy goals and contingencies by strategic planning and management teams, to ensure the sustainability of the OGD governance design.

6. Conclusion

The integration of OGD ecosystems with AI-based solutions can significantly improve the effectiveness of data-driven policies in public governance. In this article, we first identified the existing governance approaches and key design considerations in the governance of an OGD ecosystem, and later elaborated on the governance design challenges derived from the technological properties of AI-based solutions in the public sector domain. Through the synthesis of governance challenges associated with OGD ecosystem design and the use of AI in the public domain, we developed a conceptual framework to act as a decision-making heuristic for policymakers to develop reliable strategies in exploiting AI and OGD in public governance.

With this analytical tool, policymakers and system designers can systematically analyze the compatibility of available AI solutions to the needs and requirements of data prosumers in an OGD ecosystem. Policy choices with values and contingencies are dependent on the area of application, maturity of existing data governance systems, and institutional contexts. To ensure the effectiveness of policy design processes using this tool, the implementation of effective feedback mechanisms and periodic assessments with relevant stakeholder organizations are important. For that public sector organizations need to develop the necessary consulting and deliberation mechanisms.

Several issues remain to deliberate further about how AI can best be used to leverage OGD in public value creation. First, AI adoption in the public sector is influenced by various institutional, legal, cognitive, political, technical factors that create a distinction between available and desirable technological solutions. Furthermore, concerns about transparency and accountability of machine learning systems, task encroachment on human's role in public administration, lack of data analytics skills in the public domain, and biased datasets to supplement learning mechanisms put a strain on the use of advanced AI solutions in the public sector. These concerns inevitably limit the effective integration of available AI technologies within OGD platforms.

Secondly, both OGD and AI governance are contingent upon the (mis)match of value propositions imposed by various actors participating in platform governance. These value propositions are not fixed and are subjected to change in user behaviors and consequential events that can undermine the trust toward the role of public and private sector organizations in data governance. We need further research to understand better the underlying mechanisms between the use of AI and OGD solutions and citizen trust, and public value creation.

Third, for the moment, most data-sharing services are derived from centralized servers in the public sector domain and/or big data repositories controlled by profit-based organizations. This centralized constellation might be beneficial for the implementation of AI technologies to improve the efficiency of public service processes, but in the long run, might undermine the wider adaptability of AI solutions in public governance. The inclusion of other digital solutions such as self-sovereign identity and blockchain technologies might improve the quality of OGD and facilitate the adoption of more advanced AI solutions in public governance. However, for the moment, we lack an empirical and theoretical basis on how best to introduce these various technologies to leverage OGD in public governance. Further theoretical and empirical research is needed to understand the governance implications of these decentralized technologies in the use of OGD and AI in public governance.

10. References

[1] Van Ooijen, B. Ubaldi, and B. Welby, "A Data-Driven Public Sector", OECD Working papers on public governance No.33, 2019.

[2] B. Ubaldi, "Open Government Data: Towards Empirical Analysis of Open Government Data", OECD Working papers on public governance No.22, 2013.

[3] E. Ruijer, and A. Meijer, "Open Government Data as an Innovation Process: Lessons from a Living Lab Experiment", Public Performance & Management Review 43 (3), 2020, pp. 613–35.

[4] M. A. Hossain, Y. K. Dwivedi, and N. P. Rana, "State-of-the-Art in Open Data Research: Insights from Existing Literature and a Research Agenda" Journal of Organizational Computing and Electronic Commerce 26 (1–2), 2016, pp. 14–40.

[5] L. Reggi, and S. Dawes, "Open Government Data Ecosystems: Linking Transparency for Innovation with Transparency for Participation and Accountability", In: H. Scholl et al. (eds) Electronic Government, EGOV 2016, Lecture Notes in Computer Science, vol 9820, Springer, Cham, 2016.

[6] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, Adoption Barriers and Myths of Open Data and Open Government", Information Systems Management, 29 (4), 2012, pp. 258-268.

[7] E. Barry, and F. Bannister, "Barriers to Open Data Release: A View from the Top", Information Polity, 19 (1,2), 2014, pp.129–52.

[8] M. Maretti, V. Russo, and E. del Gobbo, "Open Data Governance: Civic Hacking Movement, Topics and Opinions in Digital Space", Quality & Quantity, 2020.

[9] C. Bonina, and B. Eaton, "Cultivating Open Government Data Platform Ecosystems through Governance: Lessons from Buenos Aires, Mexico City and Montevideo." Government Information Quarterly 37 (3), 2020.

[10] I. Safarov, "Institutional Dimensions of Open Government Data Implementation: Evidence from the Netherlands, Sweden, and the UK", Public Performance & Management Review, 42 (2), 2019, pp. 305–28.

[11] J. Millard, "Open Governance Systems: Doing More with More", Government Information Quarterly, 35 (4), 2018, pp.77–87.

[12] S. U. Lee, L. Zhu, and R. Jeffery, "A Contingency-Based Approach to Data Governance Design for Platform Ecosystems", PACIS 2018 Proceedings, 2018.

[13] F. Welle Donker, and B. van Loenen, "How to Assess the Success of the Open Data Ecosystem?", International Journal of Digital Earth, 10 (3), 2017, pp. 284–306.

[14] F. Van Schalkwyk, M. Willmers, and M. McNaughton, "Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem.", Information Technology for Development, 22, 2016, pp. 68–83.

[15] S. S. Dawes, L. Vidiasova, and O. Parkhimovich. "Planning and Designing Open Government Data

Programs: An Ecosystem Approach." Government Information Quarterly, 33 (1), 2016, pp. 15–27.

[16] Y. Gao, and M. Janssen, "Generating Value from Government Data Using AI: An Exploratory Study", In: Pereira G. Viale et al. (eds) Electronic Government, EGOV 2020, Lecture Notes in Computer Science, vol 12219. Springer, Cham, 2020.

[17] S. Ølnes, J. Ubacht, and M. Janssen, "Blockchain in government: Benefits and implications of distributed ledger technology for information sharing", Government Information Quarterly, 34 (3), 2017, pp. 355–364.

[18] H. Abbas, S. Shaheen, M. Elhoseny, A. K. Singh, and M. Alkhambashi. "Systems Thinking for Developing Sustainable Complex Smart Cities Based on Self-Regulated Agent Systems and Fog Computing", Sustainable Computing: Informatics and Systems, 19, 2018, pp. 204–13.

[19] D. Bollier, "Vernacular Law and Complexity Science: Two Guides for Creating Urban Commons." In: Co-Cities Open Book, 2019.

[20] E. Ostrom, "Governing the Commons", Cambridge University Press, Cambridge, 1990.

[21] Hess, Charlotte, and Elinor Ostrom. 2007. *Understanding Knowledge as a Commons: From Theory to Practice*. MIT press.

[22] M. Fuster Morell, "Governance of Online Creation Communities for the Building of Digital Commons: Viewed through the Framework of the Institutional Analysis and Development." In: B. Frischmann et al. (eds.), Governing Knowledge Commons, Oxford University Press, 2014.

[23] B. W. Wirtz, and W. M. Müller. "An Integrated Artificial Intelligence Framework for Public Management." Public Management Review, 21 (7), 2019, pp. 1076–1100.

[24] T. Q. Sun, and R. Medaglia. "Artificial Intelligence and Public Healthcare Service Innovation: A Service Ecosystem Perspective.", The 40th Information Systems Research Conference in Scandinavia, Halden, 2017.

[25] A. Ross, "The Hardest Part of AI and Analytics Is Not AI, It's Data Management." Information Age, 2019.

[26] C. Rosso, "The Human Bias in the AI Machine." Pscyhology Today, 2018.

[27] J. Manyika, J. Silberg, and B. Presten, "What Do We Do About the Biases in AI?" Harvard Business Review, 2019.

[28] J. Dressel, and H.Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism", Science Advances, 4 (1), 2019.

[29] C. V. Müller, "Ethics of Artificial Intelligence and Robotics." Stanford Encyclopedia of Philosophy, 2020.

[30] R. Begg, "Artificial Intelligence Techniques in Medicine and Health Care", Intelligent Information Technologies, IGI Global, 2009, pp. 1750–57.

[31] A. Roberts, "Five Big Challenges to AI Adoption and Success.", ClickZ, 2017.

[32] J. Bughin, E. Hazan, S. Ramaswamy, M. Chui, et al, "How Artificial Intelligence Can Deliver Real Value to Companies." McKinsey & Co, 2017.

[33] D. Kahneman, "Thinking Fast and Slow", Farrar, Straus and Giroux, 2013.

[34] K. Yeung, and M. Lodge, "Algorithmic Regulation", Oxford University Press, 2019.

[35] A. D. Selbst, D. Boyd, S. A. Friedler, et al, "Fairness and Abstraction in Sociotechnical Systems.", Proceedings of the Conference on Fairness, Accountability, and Transparency, New York, 2019, pp. 59–68.

[36] W. D., Eggers, T. Fishman, and P. Kishnani. "AI-Augmented Human Services: Using Cognitive Technologies to Transform Program Delivery", Deloitte, 2017.

[37] M. U. Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies", SSRN Electronic Journal, 2015.

[38] A. Thierer, A. O'Sullivan Castillo, and R. Russell, "Artificial Intelligence and Public Policy", Mercatus Research, 2017.

[39] A. Andersson, Å. Grönlund, and J. Åström, "'You Can't Make This a Science!' — Analyzing Decision Support Systems in Political Contexts", Government Information Quarterly, 29 (4), 2012, pp. 543–52.

[40] G. Misuraca, and C. Van Noordt, "AI Watch Artificial Intelligence in Public Services: Overview of the Use and Impact of AI in Public Services in the EU", JRC publications, 2020.

[41] M. Kuziemski, and G. Misuraca, "AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings", Telecommunications Policy, 44 (6), 2020.

[42] M. Janssen, and G. Kuk, "The Challenges and Limits of Big Data Algorithms in Technocratic Governance", Government Information Quarterly, 33 (3), 2016, pp. 371–77.

[43] D. Susskind, "A world without work: Technology, automation, and how we should respond", Allen Lane, 2020.

[44] R. Susskind, and D. Susskind, "Technology Will Replace Many Doctors, Lawyers, and Other Professionals", Harvard Business Review, October 2016.

[45] CEDEFOB, "Artificial Intelligence in a Post-Pandemic World of Work and Skills", Cedefop, 2020.

[46] M. Boyd, and N.Wilson, "Rapid Developments in Artificial Intelligence: How Might the New Zealand Government Respond?" Policy Quarterly 13 (4), 2017.

[47] A. Conn, "Artificial Intelligence: The Challenge to Keep It Safe." Future of Life Institute. 2017.

[48] N. Bostrom, and E. Yudkowsky. "The Ethics of Artificial Intelligence." In: K. Frankish, and W. M. Ramsey (eds.), The Cambridge Handbook of Artificial Intelligence, Cambridge University Press, 2014, pp. 316–34.

[49] D. Amodei, C. Olah, J. Steinhardt, et al. "Concrete Problems in AI Safety", Arxiv, June 2016.

[50] K. Migchelbrink, and S. Van de Walle, "D2.3 Final Recommendations to Transform the Public Sector Processes and Services", CITADEL project, 2019.

[51] V. Khatri, and C. V. Brown. "Designing Data Governance", Communications of the ACM, 53 (1), 2010, pp. 148–52.

[52] J. Gutman, "Means-End Chains as Goal Hierarchies." Psychology and Marketing, 14 (6), 1997, pp. 545–60.