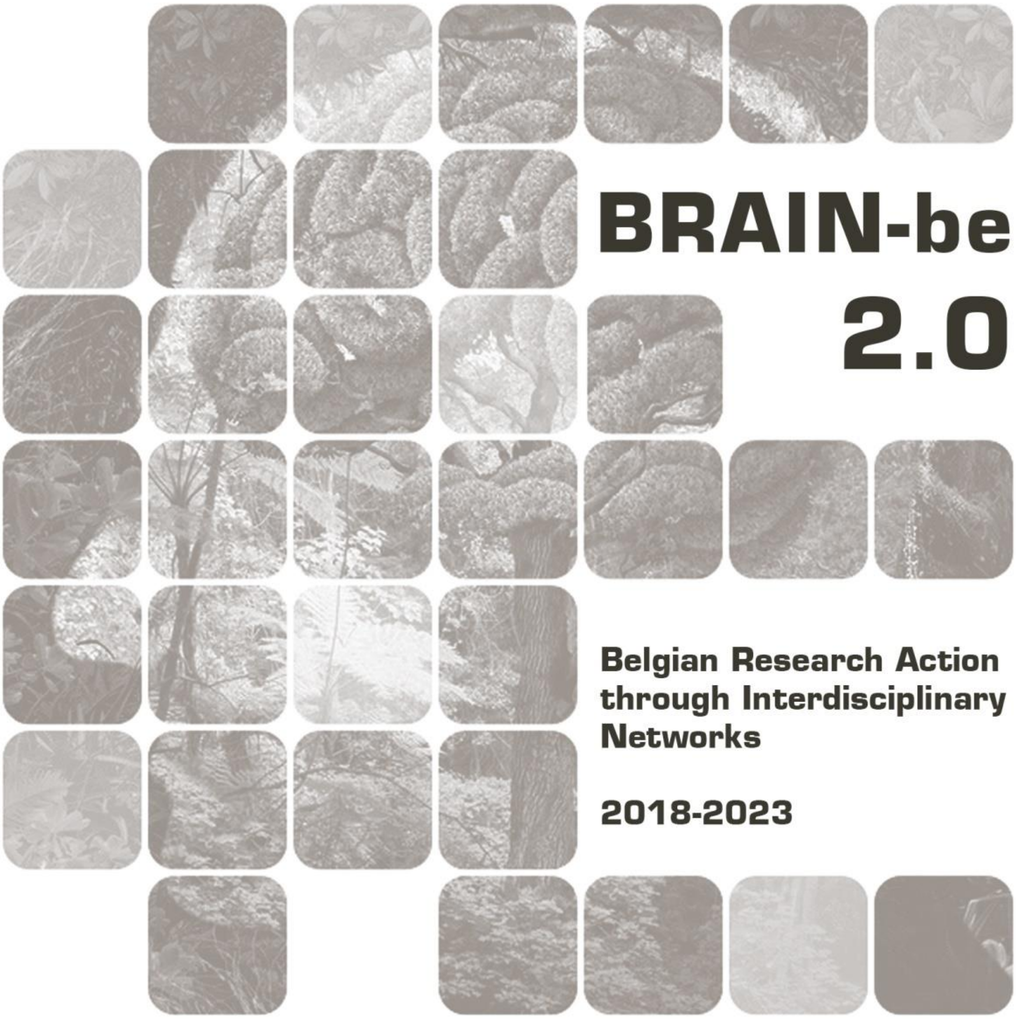


BRAIN-be 2.0

Belgian Research Action through Interdisciplinary Networks

2018-2023

DIGI4FED

Final report

Evrim Tan (KU Leuven) – Bjorn Kleizen (UAntwerpen) –Mathias Sabbe (ULiège) – Anthony Simonofski (UNamur)- Pauline Willem (UNamur)

Pillar 3: Federal societal challenges

BRAIN-be
2.0

Belgian Research Action through Interdisciplinary Networks

2018-2023

NETWORK PROJECT

DIGI4FED

Digital (R)evolution in Belgian Federal Government: An Open Governance Ecosystem for Big Data, Artificial Intelligence, and Blockchain.

Contract - B2/191/P3/DIGI4FED

FINAL REPORT

PROMOTORS:

Joep Crompvoets (KU Leuven)
Wouter Van Dooren (UAntwerpen)
Benoît Vanderose(UNamur)
Cécile de Terwangne (UNamur)
Catherine Fallon (ULiège)

AUTHORS:

Evrim Tan (KU Leuven)
Bjorn Kleizen (UAntwerpen)
Anthony Simonofski (UNamur)
Pauline Willem (UNamur)
Mathias Sabbe (ULiège)

Published in 2022 by the Belgian Science Policy Office
WTCIII
Simon Bolivarlaan 30 bus 7
Boulevard Simon Bolivar 30 bte 7
B-1000 Brussels
Belgium
Tel: +32 (0)2 238 34 11
http://www.belspo.be
http://www.belspo.be/brain-be

Contact person: Emmanuèle Bourgeois
Tel: +32 2 238 34 94

Neither the Belgian Science Policy Office nor any person acting on behalf of the Belgian Science Policy Office is responsible for the use which might be made of the following information. The authors are responsible for the content.

Tan, E, B. Kleizen, A. Simonofski, M. Sabbe, P. Willem. 2022. DIGI4FED Final Report. Brussels: Belgian Science Policy Office 2022 – 83 p. (BRAIN-be 2.0 - (Belgian Research Action through Interdisciplinary Networks))

TABLE OF CONTENTS

ABSTRACT

DIGI4FED aims to understand how (big) data can be used in the Belgian federal administration system to enable better public service provision through new technologies such as artificial intelligence and blockchain. By focusing on the technical, moral, legal and organisational conditions within the internal and external federal decision-making processes, DIGI4FED aims to develop a governance design that serves the administrative and public service processes of the Belgian federal government and makes full use of the potential offered by big data and its application via artificial intelligence and blockchain technology. DIGI4FED focuses on the development of a proof of concept (PoC) of a governance design – the design artefact – in two specific federal policy areas: social security infringements and tax frauds.

Throughout the span of the project, the DIGI4FED team has gathered data through various means (e.g. experiments, interviews, living lab) to design and test the governance modalities for the better use of new digital technologies in the fight against social security infringements and tax frauds. The research has also revealed several types of challenges (e.g. trust, operational, administrative, technical, user acceptance, legal, and policy) to the introduction of new digital technologies in the Belgian federal government. To overcome the identified challenges and to introduce these technologies in the fight against tax and social security fraud, the research findings are compiled into three sets of policy recommendations focusing on legal and operational challenges, trust challenges, governance challenges.

Keywords: Big data, AI, blockchain, data governance, open government data, fraud detection

1. INTRODUCTION

DIGI4FED is a two-years research project funded by BELSPO as part of the BRAIN-be 2.0 Call 2019. DIGI4FED aims to understand how (big) data can be used in the Belgian federal administration system to enable better public service provision through new technologies such as artificial intelligence and blockchain. By focusing on the technical, moral, legal and organisational conditions within the internal and external federal decision-making processes, DIGI4FED aims to develop a governance design that serves the administrative and public service processes of the Belgian federal government and makes full use of the potential offered by open government data and big data, and its application via artificial intelligence and blockchain technology. DIGI4FED focuses on the development of a proof of concept (PoC) of a governance design – the design artefact – in two specific federal policy areas: social security infringements and tax frauds.

Three factors define the context by which DIGI4FED is influenced. The first factor is the growing attention to the potential impact of Big Data (BD) and Artificial Intelligence (AI) on traditional government information processes. The second factor is the growing expectation of society from public administrations, to adopt new technological means to advance efficient and effective governance and public service delivery whilst ensuring the core democratic and moral values are not lost out of sight. The third factor concerns the Belgian federal administration itself. Although several steps were taken toward the digital transformation of the Belgian federal state in the past, challenges remain.

DIGI4FED project has researched the conditions to introduce new digital technologies such as big data, AI, and blockchain to improve fraud detection processes in the taxation and social security domain. Throughout the span of the project, the DIGI4FED team has gathered data through various means (e.g. experiments, interviews, living lab) to identify challenges to overcome in compliance with this aim. The details of the data collection and analysis processes have been reported in various deliverables produced as part of the project. To get a better insight into these processes, we invite the readers to check the deliverables at www.digi4fed.be .

The purpose of this final report is to give an extensive overview of the project's achievements and highlight the main research and dissemination outputs. The structure of the report is as follows. Section 2 gives the state of the art in the use of data-driven technologies in public governance, the factors affecting the trust in the use of these technologies, and the institutional framework of Belgian federal government. Section 3 explains the methodology used in the implementation of the project and elaborates on the key findings from the interviews, experiments, and living lab. Section 4 arrays the policy recommendations on the legal, operational, trust, and governance dimensions in the use of data-driven technologies in fraud detection. Sections 5 and 6 present respectively the dissemination activities and publications produced by the project.

2. STATE OF THE ART AND OBJECTIVES

The use of artificial intelligence (AI) and algorithmic decision-making in government and managerial practices has become ubiquitous in recent years. There is a growing body of scholarship not only in information sciences but also in overall administrative sciences to understand the impact and implications of data-driven technologies such as big data analytics, machine learning, and blockchain on managerial and organizational systems and practices. It appears that there is a consensus among scholars and practitioners that these technologies will overhaul the existing administrative systems and practices into new types of interactions between humans and machines, which is sometimes called algorithmic bureaucracy (Tan & Crompvoets, 2022; Vogl et al, 2020). However, the digital transformation literature also points out that the adoption of new digital technologies is challenging for public sector organizations invoking various value-laden reservations driven by perceived technical, systemic, administrative, and regulative barriers inside and outside of organizations (Tangi et al. 2021, Vogl et al. 2020, Bullock et al. 2020). Public administration research has already begun to investigate challenges associated with the use of AI and algorithmic decision-making on system applications (Exmeyer & Hall, 2022, Neuman et al, 2022), accountability mechanisms (Busuioc, 2020), citizen trust and explainability of decisions (Grimmelijkhuisen, 2022), organizational rearrangements (Meijer et al, 2021), administrative discretion and willingness to implement (Wang et al, 2022, Alshallaqi, 2022), ethical principles and citizen’s privacy (Willems et al, 2022), capacity gaps and knowledge management (Wilson & Broomfield, 2022), and so forth.

This nascent literature in public administration scholarship provides rich but a fragmented picture of diverse drivers that shape the use of AI and algorithmic decision-making in public policy processes. Especially we lack holistic models that explain how different drivers are interrelated to each other and how changes in some can influence system applications (Engvall & Flak, 2022; Dawes, 2009). This lack of understanding complicates developing viable digital transformation strategies for AI and algorithmic decisions in public administration.

One theoretical objective of the DIGI4FED project is to fill this research gap by developing a comprehensive model that can explain the interrelationships of perceived drivers that influence the integration of AI and algorithmic decisions in public administration. Specifically, our research focuses on the adoption of these tools in the taxation and social security domains for fraud detection. Therefore, the first main subsection will focus on elucidating on how these technologies are currently employed in the public sector, what new features and capabilities can be unlocked through their conjoint use, and what opportunities and challenges are associated with these technologies. The second focus of the section will be about the trust dynamics in the use of these technologies in the public sector. Especially, understanding the underlying trust problems through various dimensions influencing the use of these technologies are important to develop a viable governance model to integrate data-driven technologies in public governance.

2.1. The use of data-driven technologies in public governance

Artificial intelligence (AI) and blockchain technologies (BCT) allow organisations to leverage big data (BD) and gain insights in order to make sound public policy decisions. Public sector organisations need to learn about these technologies and develop the necessary skills and competencies to help their organisations stay competitive. One distinguishing characteristic of these emerging digital technologies is that they are not only tools for public sector organisations to enhance existing capabilities in public policy processes, but their technological features demand a complete revisit of the existing administrative systems and processes. Below, we will give an overview of the state of the art about the use of these technologies in public governance.

2.1.1. Big Data

The perception and description of big data (BD) keeps evolving (Alam & Dunny, 2019). The most common usage of the concept relies on the three ‘v’s, namely volume of information, a variety of different data sources and types (structured and unstructured), and the velocity of data (i.e. the speed of creation, storage, and dissemination of data) in real time (Pencheva et al., 2020; Einav & Levin, 2014; Zikopoulos et al., 2012). Some authors even add other ‘v’s to the definition, which correspond to the value of data (Kimble & Milolidakis, 2015) and veracity of data (Demchenko et al., 2014). Kitchin & McArdle (2016) add to these characteristics the dimensions of exhaustivity, resolution, indexicality, relationality, extensionality, and scalability.

Not only its features, but the scope of BD often varies according to the disciplinary backgrounds. In managerial sciences, BD corresponds to unstructured content generated from a plurality of sources such as Internet clicks, mobile transactions, user-generated content, and social media, as well as purposefully generated content through sensor networks and business transactions such as sales queries and purchase transactions (George et al., 2014). In public policy, BD is associated with new formats, quality, and availability of administrative data (Pirog, 2014). In political sciences, BD refers to the technological innovations to gather either new types of data, such as social media data, or vast quantities of traditional data at less expense (Clark & Golder, 2014). In information and technology sciences, BD is defined through the big and openly linked data (Janssen & Van den Hoven, 2015), and by the massive quantities of information produced by and about people, things, and their interactions (Boyd & Crawford, 2012). In computer sciences, BD refers to the second-by-second picture of interactions over extended periods of time, providing information about both the structure and content of relationships (Lazer et al., 2009). Finally, in public administration, Mergel et al. (2016) identify BD as (1) data created by private citizens through their interactions with each other online (such as social media data), (2) data automatically generated by sensors and automatically transmitted online, and (3) data that is automatically collected by public entities in the course of their operations.

BD is expected to bring value creation at almost every stage of policy design processes. First, BD-based solutions can improve agenda setting and policy formulation processes. BD allows a quick reaction to and incorporation of collective information from a variety of sources (Höchtl et al., 2016; Mergel, 2017). It provides greater accuracy, efficiency and speed of the administrative processes (Pencheva et al., 2020), It allows predictive analysis of which policies will work and under what conditions (Clarke & Margetts, 2014; Cook, 2014). It allows for the design of policies better matched to the preferences of stakeholders (Stritch et al., 2017; Taeihagh, 2017). It improves the legitimacy of policy formulation processes by enabling citizens and governments to engage in more meaningful dialogue and to collaborate in policy design (Desouza & Bhagwatwar, 2012; Schintler & Kulkarni, 2014).

Second, BD can improve policy implementation processes. BD ensures compliance with regulations, fraud prevention and detection, and criminal investigation (Janssen & Kuk, 2016). It allows quick responses to emergencies, prevention of damages, and fewer casualties (Hondula et al., 2018). It improves operative efficiency by exposing and eliminating redundancies (Gamage, 2016; Shindelar, 2014), and improves effectiveness in the delivery of public services by better resource allocation modelling and real-time operation optimisation (Daniell et al., 2015). It increases public sector efficiency by delivering saving and boosting productivity (Agostino & Arnaboldi, 2017; Johnes & Ruggiero, 2017). It helps policymakers to better understand and segment users for personalised policies (Pencheva et al., 2020). It facilitates better supervision of the implementation processes through the detection of irregularities (Maciejewski, 2017).

Third, policy research and evaluation can be augmented via BD-based solutions. BD improves policy analysis and helps identify areas for reform (Decker, 2014). Through BD, policymakers can understand the long-term effects of interventions on citizens in policy areas and develop new interventions for hard-to-reach populations (Cook, 2014; Blume et al., 2014). BD provides superior insight and decision-making capabilities (Pencheva et al., 2020). BD enables holistic evaluation of policy outcomes with the capacity to handle time-series data from multiple, diverse sources and the ability to simultaneously observe individual and aggregate variables (Pencheva et al., 2020; Jarmin & O'Hara, 2016). BD allows a rapid and real-time evaluation process (Höchtl et al., 2016). Through BD-based solutions, it is possible to experiment with new business models and organisational performance techniques (Arinder, 2016), and supplement traditional techniques in policy evaluation such as surveys (Bachner & Hill, 2014; Pandey et al., 2017; Zhu et al., 2017).

Fourth, new modes of data-driven public governance can be formulated by improving mutual government-citizen understanding (Clarke & Margetts, 2014). The use of data-smart governance models (Goldsmith & Crawford, 2014) and open data governance (Millard, 2018) models depend in particular on the effective use of BD-based solutions. The adoption of these new modes of digital public governance relies on the development of new policies and programmes describing what type of data will be produced and how it should be analysed (Pirog, 2014). Furthermore, data and time-conscious evaluative frameworks can emphasise evidence-based decision-making and longitudinal cost-benefit analytics as part of policy-making processes, making public policy processes more predictable and transparent, and thereby changing the administrative culture in public sector organisations (Arinder, 2016: 394).

2.1.2. Artificial Intelligence

There are a variety of definitions for artificial intelligence (AI) in the literature. Russell & Norvig (2016) define the term AI to describe systems that mimic cognitive functions generally associated with human attributes such as learning, speech and problem-solving. For Haenlein & Kaplan (2019), AI is a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation. In an even broader conceptualisation, Tegmark (2017) defines AI as the ability of a non-organic, mechanical entity to accomplish complex goals. According to Dwivedi et al. (2019), the common thread amongst various definitions of AI is the increasing capability of machines to perform specific roles and tasks currently performed by humans within the workplace and society in general.

The use of AI technologies in public policy and administration processes is not a new phenomenon but, over the last decade, there have been dramatic advances in core AI technologies like machine learning, natural language processing, deep learning (or neural networks), virtual agents, and computer vision. AI literature distinguishes between 'narrow AI' and 'general AI' (Bullock, 2019). Narrow or weak AI refers to systems capable of carrying out tasks that require single human capabilities, e.g. visual perception, understanding context, probabilistic reasoning and dealing with complexity (Russell & Norvig 2016). In public administration, narrow AI is often used in boosting the capabilities of data analytics and data mining techniques to deliver deeper and better insights beyond what human analysts can do. General or strong AI refers to (hypothetical) systems with human or superhuman intelligence, which simulate the complex human ability to think and execute intelligent tasks such as ethical judgements, symbolic reasoning, managing social situations, and ideation (Brynjolfsson & McAfee, 2014). Despite some noteworthy advancements in unlocking general AI technology (e.g. ALPHAGO Zero developed by DeepMind), most real-world applications of AI fall under the category of narrow AI.

The early promise of AI was largely viewed in terms of providing decision support for public managers (e.g. Hadden, 1986; Hurley & Wallace, 1986; Jahoda, 1986; Masuch & Lapotin, 1989). The latest advances in AI allow computers to learn from past experiences and understand the world through a hierarchy of concepts (Goodfellow et al., 2016) that can lead to automation of tasks (Pencheva et al., 2020; Bailey et al., 2016; Barth & Arnold, 1999). The use of BD has enabled algorithms to deliver excellent performance for 'narrow AI', rather than the more human level 'general AI' where the complexities of human thinking and feelings have yet to be translated effectively (Pencheva et al., 2020; Russell & Norvig, 2016). A key contributing factor to the increasing maturity of AI technologies and the viability of AI application to public policy and administration was the availability of data that can be used in the machine-learning process. At the same time, without the underlying analytical technology, the data-driven public governance can be viewed simply as a shift in the scale of the available data rather than as a transformational change (Pencheva et al., 2020). Therefore, advanced analytics and other AI technologies are essential for the usage of BD in the public sector (Mergel et al., 2016, Pencheva et al., 2020).

There are a multitude of AI applications currently in use in the public sector[1]. First, we see the use of AI in knowledge management. Generation and systemisation of knowledge, development of expert systems supporting the codification of the knowledge, and use of neural networks to analyse, distribute, and share knowledge with others are some examples of AI applications in knowledge management (Wirtz et al., 2019).

Second, AI applications are used in process automation systems. Automation of standard tasks through rule-based assessment, workflow processing, schema-based suggestions, data mining, and case-based reasoning improve the pace and quality of administrative services (Wirtz et al., 2019; Chun, 2007). Automated AI systems can also support complex human action processes or repetitive tasks by leveraging the ability of software robots or AI-driven workers to mimic human interaction with user interfaces of software systems (Wirtz et al.,2019; Jefferies, 2016).

Third, AI-based systems can also create higher-level cognitive functions in building autonomous systems and replacing human agents. For example, the use of electric-powered autonomous vehicles for public transport (Jefferies, 2016) or robot-assisted surgeries (Collier et al., 2017) are some areas in which AI-based agents can replace humans.

Fourth, AI solutions based on speech analytics and natural language processing are used in creating virtual agents (e.g. chatbots, avatars), digital assistants, and forming complex recommendation systems to replace or support human agents in internal and external human relations management (Wirtz et al., 2019; Zheng et al., 2018; Cortés-Cediel et al., 2017; Mehr, 2017).

Fifth, by leveraging BD, AI solutions such as machine learning, deep learning, predictive analytics are used to create complex predictive and prescriptive models in fraud detection (Hemken & Gray, 2016), preventive actions in crime and terror threats (Power, 2016), and forecasting models in natural resource and water management (Kouziokas et al., 2017).

2.1.3. Blockchain Technology

Blockchain and distributed ledger technology (DLT) belong to a class of technologies known as blockchain technologies (BCT), which create a transparent, autonomous and decentralised data governance system that gives users confidence that archived information has not been tampered with (Beck et al., 2018). In a blockchain, the information is archived in a distributed ledger that is shared across a network of users where participants, called nodes, keep a copy of the ledger and record all transactions. By agreeing on a form of consensus mechanism, nodes validate the transactions, provide an immutable (or nearly immutable) record of transactions, and ensure traceability. Each validated transaction is registered in a block, which is time-stamped and includes a digital trace (called hash) of

[1] A list of AI-use cases in the European public sector can be obtained at: https://joinup.ec.europa.eu/collection/elise-european-location-interoperability-solutions-e-government/news/143-ai-cases-public-sector-are-available-open-data

a prior block, together forming a blockchain. Every block in the system continuously synchronises with other blocks. In this way, blockchain ensures nearly unhackable decentralised systems of encrypted data, without the need for a centralised authority to ensure the continuity of the system.

BCT is not a monolith technology, and it covers several underlying technologies (e.g. smart contracts, token technology, decentralised applications, etc.) that can jointly support a decentralised and automated information infrastructure. In the context of public governance, BCT has the potential for facilitating direct interactions between public institutions, citizens, and social and economic agents. At the most basic level, BCT can be used as an information infrastructure for the more efficient management of public information and exchanging information between public administrations. At a more advanced level, BCT can leverage the features of decentralised information management and automated execution of algorithmic decisions through smart contracts in support of BD and advanced analytic techniques to create autonomous organisations that can replace the human agency and traditional forms of public sector organisations in public governance.

There are three categorical areas where BCT is currently in use in the public sector. First, BCT-based solutions are used in identity verification and in supporting self-sovereign digital identity systems for individuals and businesses. For instance, European Blockchain Service Infrastructure (EBSI) adopts the use of self-sovereign electronic identification, authentication and trust services (ESSIF) as one of the use cases (European Commission, 2021). Another example is a blockchain-based identity system that is used for refugees in Finnish camps (Hempel, 2018). Furthermore, BCT-supported identity verification systems are also used to develop more secure e-voting systems. For example, uPort decentralised digital identity in the Swiss city of Zug is used for e-voting and renting e-bikes (Berryhill et al., 2018). In another case, the Digital Democracy and Data Commons pilot in the city of Barcelona developed a DLT-based voting system to support petition mechanisms in Barcelona's e-governance platform called Decidim.Barcelona (López, 2019).

Second, BCT is used in asset registries. Land titling, patents, health data, diplomas are some areas where BCT is currently used to improve the reliability, traceability, and security of registered data in public and private databases, and to reduce the operational costs and verification times (Allessie et al., 2019). Some examples are the notary and diploma use cases for EBSI, Exonum land title registry in Georgia (Berryhill et al., 2018), Equity Platform, to register and validate individual energy transactions between batteries and the grid (Janssen, 2020), and e-health records in Estonia to register health records at the national level (Einaste, 2018).

Third, BCT is used in automating and tracking high-risk transactions and improving the traceability and transparency of supply-chain management systems. Currently, there are various use cases in logistics, food tracking and energy sectors for blockchain-based supply-chain management systems (Yafimava, 2019).

2.1.4. Convergence of Technologies in Public Governance

So far, we have covered the individual implications of BD, AI, and BCT in public governance. However, their true disruptive impact is contingent on their convergence with each other and with other digital technologies such as the Internet of Things (IoT), sensing technologies, 5G systems, cloud computing, robotics, and 3D printing. Here, the convergence refers to the joint applications and seamless interactions of these digital technologies in the systems of public governance. The idea is that with convergence, separate digital technologies will complement each other, and their disruptive impact will be augmented in creating a world where individuals, organisations, and machines can freely interact with one another with little friction and at a fraction of the current costs (Voshmgir, 2020:109). In this new world, public services can be organised more efficiently and effectively without any interference from intermediaries, and new forms of autonomous and self-governing systems of public governance can be envisaged.

Following the work of Hassani et al. (2019), I surmise that the convergence of AI, BD and BCT can bring forth the following augmented features and outcomes in the sectoral area of application.

Table I. Convergence of AI, BD and BCT

Technology Stack	Augmented Features	Outcome
Big Data + Blockchain	▪ BCT makes BD even bigger and contributes towards making BD more secure and valuable for data analytics. ▪ BCT's decentralised data storage and computing can create a more efficient and seamless management of BD. ▪ BCT enables verified, structured, and secured BD for better analysis and prediction of consumer behaviour. ▪ More efficient governance mechanisms and strategic decision-making in BCT-based systems due to real-time BD sharing and analytics capabilities. ▪ Blockchained BD mitigates a single point of failure in data governance and introduces an extra level of security against cybercrimes.	▪ A bigger, more structured, and secure data-sharing ecosystem. ▪ Improved blockchain management and functioning.
Big Data + AI	▪ Through intelligent modelling techniques, raw data can be transformed into business value data.	▪ More intelligent, customised modelling, prediction, and decision-making.

	▪ Through deep-learning systems, raw input data is used to extract higher-level features that can mimic human cognitive abilities. ▪ Smart data administration and data structuralising.	▪ Better data training and more intelligent AI functions.
Blockchain + AI	▪ Automated decentralised autonomous organisations (DAOs) and intelligent smart contract management systems can be created on blockchain. ▪ Data-sharing in trustless, decentralised blockchain networks can augment the abilities of machine-learning techniques. ▪ AI can promote the efficiency of mining in blockchain and significantly reduce energy consumption whilst increasing the sustainability of BCT.	▪ Secured AI marketplace, crowd-sourcing AI models, techniques and algorithms, and expanded structured data resource for training. ▪ More efficient mining, improved blockchain maintenance, management, and intelligent smart contract initiating.
AI + Blockchain + Big Data	▪ A decentralised infrastructure for intelligent data analytics can be created. ▪ Well-trained intelligent automation can largely boost the processing capacity and accuracy of blockchain-based systems. ▪ Blockchained big data can provide transparency and potentially a degree of traceability of data in a machine learning model and could contribute to the explainability of AI decisions and outcomes.	▪ Transparent, secure, automated data governance ecosystems that can generate useful insights that satisfy the diverse interests of the different parties.

In a more specific look at the public sector, the convergence of these technologies is expected to unlock new features and capabilities in digital governance and in overall organisational and institutional rearrangements in public governance. Nevertheless, any disruptive influence brings with it challenges and obstacles to be overcome during the implementation of these technologies. While these obstacles are expected to vary according to the specific institutional and contextual conditions, we can draw on certain opportunities and challenges for the public sector resulting from the convergence of these technologies.

Opportunities for public governance

(a) **More data availability, better big data management & analytics:** public blockchains are expected to improve AI algorithms and market prediction solutions since data will be available via a public ledger. This will allow scalable and more accurate solutions and better AI models (McConaghy, 2016) within multiple contexts, enhancing the possibilities of data analytics (Casino et al., 2019). The secure and verifiable blockchain structure may be used to ease big data management (Karafiloski & Mishev, 2017) and for efficient risk management and strategic decision-making owing to real-time BD sharing and analytics capabilities (Hassani et al., 2019).

(b) **Higher transparency and reliability of AI solutions in data analytics**: BCT can be used to train and test AI systems and to ensure protection against misuses by creating a safe and socially desirable AI solution (Dwivedi et al., 2019). Dwivedi et al. (2019) propose an AI certification transparency and scorecard blockchain (ACTS-B) to integrate the information about the training dataset used for an AI system, thereby tracking whether the training dataset meet certain criteria such as diversity, equity, etc. For the ACTS-B to engender trust in the AI-based decision-making, it should be a universal publicly viewable blockchain. This would create a transparent mechanism for rating and understanding AI solutions before putting them into use. Blockchain can also provide a tamperproof record of the changes made, along with the authorisation details, making all changes traceable.

(c) **Better auditability**: the adoption of deep learning in conjunction with faster machines and larger storage spaces have paved the way for modern auditing, which is already being enhanced by blockchain (Appelbaum et al., 2017; Issa et al., 2016). However, machine-learning algorithms are characterised by their opaque features. Their opacity most commonly stems from the large number of possible features included in a classifier which prevent us from understanding and explaining decisions made by AI (Burrell, 2016). Justifying why a specific choice was made by AI creates many complications due to the enforcement of GDPR and its derived requirement of explainability[2]. Additionally, concerning automated decision-making, a data subject has the right to be provided with meaningful information about the logic involved[3]. Blockchains can overcome some of these regulative challenges by providing auditable trails to prove why a particular decision was made by an AI system and resolve the discrepancies raised by the non-linear use of numerous factors and use of randomisation (Casino et al., 2019). Furthermore, having a clear audit trail will also eventually increase every

[2] Art. 22, GDPR.

[3] Art 13, ibid

machine-to-machine interaction through IoT and transaction, providing a secure way to share data and coordinate decisions (Corea, 2019).

(d) **Customer-centric services:** AI and human collaborative automation through BCT can improve working efficiency and prioritisation of human-agent workflow, as well as support human-agents with intelligent advice and information (Hassani et al., 2019). BCT provides greater control of data shared for public service processes and decentralised data processing. The convergence of AI, BD and BCT-based applications can enable real-time data processing and the formation of customer-specific public services.

(e) **Automated decentralised autonomous organisations**: AI technology can improve the effectiveness of BCT-based systems by providing higher efficiency in mining processes and governance of blockchain, better scalability solutions, increased security of applications, and coding of more complicated applications for autonomous agents and DAOs (Corea, 2019). AI supported DAOs can autonomously or semi-autonomously operate without a centralised control or third-party intervention (Wang et al., 2019). These new forms of organisations can redefine the mechanisms of control and coordination in public governance. For example, units specialised in streamlining, regulatory and network governance can be replaced by DAOs that are (semi)controlled by other public sector organisations. Furthermore, new forms of management structures with non-governmental organisations can be established as DAOs replacing traditional forms of public-private partnerships and collaborative governance mechanisms.

Challenges for public governance

(a) **Difficulty of applying risk management approach and challenges with change management:** Applying risk management to digital security and other digital risks is challenging for most organisations, in particular where the rights of third parties are involved (e.g. the privacy rights of individuals and the intellectual property rights of organisations and individuals) (OECD, 2019). Especially, insufficient budget and a lack of qualified personnel have been identified as major obstacles to the effective use of risk management to address trust issues (OECD, 2017). Not only the lack of organisational capacities but also the technological features of AI and BCT complicate the application of effective risk management and change management mechanisms. For example, in decentralised blockchains, the absence of a centre to enforce governance policies may hinder the use of effective sanctions against malevolent behaviours or the reparations of harm due to malevolent behaviours. Similarly, the black box features of self-learning AI applications carry the risk of biased decisions in public policies, undercutting effective risk management in data exploitation. For the moment, our knowledge base lacks the determinants of effective risk and change management in the use of advanced AI and BCT solutions.

(b) **Increased operating costs:** Data analytics using blockchain structure incur too many overheads (Casino et al., 2019). One of the identifying features of blockchains is that one can neither delete nor purge specific records/files in blockchains. The ever-increasing number of records can in the long run overburden the cost of storage, operation, and archiving (Liu, 2016). Nevertheless, blockchain-based architectures for BD storage already exist (Kumar & Abdul Rahman, 2017), and intermediate or efficient auxiliary structures may be implemented via cloud solutions, thereby increasing the overall efficiency. Furthermore, the environmental impact of high energy-consuming digital solutions (e.g. PoW-based blockchains) may delimit the scalability of solutions or lead to additional operational costs for compliance with environmental standards (De Vries, 2018). Yet again, AI-based solutions and/or alternative BCT consensus models (e.g. PoS, DPoS) can improve the energy efficiency of system infrastructures limiting the negative externalities of high energy consumption. For a better estimate of the impact of these technologies on the operational costs in public administration, experimentations with new technologies and empirical data from use cases are necessary.

(c) **Need for new interoperability standards and public governance mechanisms:** While AI algorithms are more adaptable in existing information management systems, the use of BCT solutions requires new interoperability standards between different information management systems, and the development of legitimate governance and administrative mechanisms concerning the reading, editing and use of data. In particular, the technological and sectoral challenges of blockchain interoperability still need to be resolved (Janssen et al., 2020). Furthermore, AI and BCT may bring increasing levels of effectiveness and efficiency, but political and administrative legitimacy of these technologies remain key concerns (Salamon, 2002), and the impact of these technologies on equity, accountability, and democracy is far from clear at the moment (Bullock, 2019; Barth & Arnold, 1999). Even if a completely unbiased and trustworthy AI solution can be developed aided by BCT, a balance needs to be found between the amount, type and variety of data processed and the goals involved in public governance processes. Not only the technical interoperability standards, but also legal interoperability challenges (e.g. different transparency standards, different standards in data access and sharing) create barriers in cross-border data flows (OECD, 2019) and the scaling up of AI and BCT solutions for cross-border services. Advancements in the related legislation and interoperability policies, developments with cross-border data infrastructures (e.g. European Blockchain Services Infrastructure), and advancements in technical solutions and governance models will shape the future usage of these technologies in public governance.

(d) **Cultural transformation in administrations:** Data verification and implementing effective BCT and AI solutions often require a cultural transformation in administrative processes. This cultural transformation includes changing working habits in system operations and in managerial processes (Bean, 2020). The effective usage of digital solutions proposed by AI, BCT and BD depends on the openness of data and exchange of information inside and outside administrations (European Commission, 2020). Transformation toward openness can be more challenging in siloed organisations. Another challenge is that such transformations may inherently be resisted in some administrations due to incumbent civil servants distrusting the new techniques or political distrust of citizens, which may contribute to low alignment. Furthermore, it is not entirely clear what the implications of these technologies will be on public values. For instance, the operations of BCT-based systems often require the re-allocation of certain prerogatives in public governance concerning accountability, leadership in change management, and decision-making with organisations outside the administrative hierarchy. AI-based systems, on the other hand, are output-oriented, easy to implement and user-friendly to interpret the results, yet machine-discretion may undercut the accountability mechanisms and significantly affect the trust vested in public sector organisations. There is a need for further research to better understand the implications of these technologies on public values.

(e) **Automation vs human discretion:** A challenge faced by public managers is deciding on the tasks to be automated in AI and BCT-supported systems. The difference between two technologies is that, while AI automates more decision-making processes, BCT automates execution processes (e.g. smart contracts). Bullock (2019) anticipates that tasks that are high in complexity (more deviations from the norm) and high in uncertainty (less analysable) are likely to remain as a discretionary task to be completed by humans. Nevertheless, Bullock acknowledges the limitations of human weaknesses in situations of uncertainty and admits AI's relative advantage for tasks that are high in uncertainty but low in complexity. Although Bullock anticipates that humans will retain a relative advantage in identifying highly complex and abstract patterns across task sets, not all AI experts share the same sentiment that domains requiring creative and abstract thinking will remain in the hands of humans. For example, McAfee & Brynjolfsson (2017) emphasise the creative abilities of AI applications by giving examples of designing novel industry models, composing music, and advancing scientific hypotheses. Their expectation is that machines will generate initial proposals that people can extend and improve with their better attunement to human conditions. Certainly, the decisions to introduce automation are not only managerial or technological decisions. Regulative, social, economic, behavioural, and political factors concerning the involvement of human agents in administrative processes will determine to what extent technologically available automated solutions are implemented in public governance. The sharing of responsibilities among machines and humans in the future of public governance, and the determinants on the role division decisions are other unexplored research areas with which we can expand our current knowledge base.

2.2. Trust in digital governance

As data analytics technologies such as AI and big data become increasingly ubiquitous in the public sector, so too does the importance of trust in these technologies and the ways in which they are governed. This is hardly a new or surprising insight: user trust and the ethical application of emerging technologies have become important topics of inquiry in data science, while citizen-government trust relationships are frequently featured in both social science and academic legal debates (Ahonen and Erkkilä, 2020; Busuioc, 2020; Ananny and Crawford, 2018). The topic has even reached mass media reporting, as is shown by major events such as the US Facebook-Cambridge Analytica scandal or the Dutch childcare benefits scandal, both of which featured major society-wide trust breaches that stemmed in part from the use of large-scale databases and AI (Isaak and Hanna, 2018; Busuioc, 2020; La Fors, 2020).[4] In Cambridge Analytica, the world was introduced to the prospect of AI influencing elections, while one component of the Dutch childcare benefits scandal featured an AI model that intentionally used nationality for fraud risk detection among childcare benefit applicants (Dutch Data Protection Authority, 2020; Dutch Parliamentary Investigation Commission, 2020). Given the salience that some governmental data analytics projects are thus gaining both within academia and in society, it is remarkable that only limited empirical research seems to exist on how governments are dealing with emerging societal trust issues regarding the use of data analytics and the degree to which these endeavours may be expected to be successful (Ahonen and Erkkilä, 2020; Sun and Medaglia, 2019).

2.2.1. Trust Problems in Data Analytics and Government

As algorithms in general and AI more specifically begin to pervade society, so do the concerns over their potential misuse (Sun and Medaglia, 2019). With the private sector leading in the development of AI and big data, it is perhaps unsurprising that many frequently cited examples of issues pertain to major private-sector innovators such as Facebook (e.g. Cambridge Analytica's use of Facebook profile data (Isaak and Hanna, 2018)) and Tesla (the reliability of image recognition to prevent accidents with self-driving cars (Stilgoe, 2018)). However, the public sector is less of a laggard than is often assumed. Governments around the globe are turning to data analytics to better analyse their – frequently substantial – data flows (Ananny and Crawford, 2018). In some cases, AI is not even required, with eligibility for certain permits, tax deductions or social benefits being possible through 'simple' matching algorithms. AI methods such as machine learning and deep learning become more useful

[4] Cambridge Analytica revolved around an AI model using unethically obtained data to predict the profiles of potential voters, with the aim of improving targeted political advertisements. The Dutch childcare benefits scandal revolved around the (illegal) use of ethnicity to determine the risk of fraud among childcare benefits applicants. Although broader because it also encompassed non-AI elements, the childcare benefits scandal partially concerned an AI risk-classification model which incorporated ethnicity in its predictions on eligibility for childcare benefits.

where probabilistic predictions need to be made over large and complex datasets. Governments in the US and Western Europe (including Belgium and the Netherlands) are actively developing and using AI for problems ranging from the automatic interpretation of text (such as chat messages and applications from citizens or jurisprudence analysis) to the prediction of traffic flows, and from personalised recommendations to risk-detection in policing and administrative supervision.

These developments generate important new trust problems for society, a few of which will briefly be discussed here. Many non-AI and AI algorithms have in common that they require access to sufficiently large datasets, which in the public sector frequently incorporate sensitive personal data or pseudonymised data (McDermott, 2017). The use of such data is also changing, as algorithms concurrently analyse thousands or even millions of observations to generate predictions. By contrast, traditionally analysed databases (such as medical records examined by a doctor or criminal records used by an inspector) may incorporate high numbers of observations, but these observations are only analysed in small numbers and normally only with prior cause (e.g., the immediate relevance of the medical record of a patient that is currently under treatment). Depending on the goal of the data analytics project and the sensitiveness of the data involved, the large-scale use of such data may at least be uncomfortable for some members of society (Sun and Medaglia, 2019). This issue may be compounded where data is shared among multiple public entities or obtained from private entities, as more entities, developers and users gain access to more extensive (combinations of) data (Meuwese, 2020). Concerns that large-scale data processing may lead to privacy violations or discriminatory practices are no longer entirely theoretical either, given examples such as the Dutch tax service’s intentional use of nationality data in its social benefits fraud detection algorithms (Dutch Data Protection Authority, 2020).

Where AI becomes involved, concerns regarding black boxes, biases and model drift are also introduced. The black box refers to the potential opaqueness of machine learning and deep learning algorithms regarding the factors that contribute most to a certain prediction. For instance, in a black box recidivism model, it may be unclear whether high predicted risk of committing further crimes is based on admissible criteria (e.g. prior record) or on potentially unlawful and/or discriminatory criteria (e.g. ethnicity) (Sandviq et al., 2016; Chander, 2017). Both human biases in training the data and subsequent model drift caused by the model’s self-learning features may however lead to unlawful and/or unethical criteria being used as a factor in making a certain prediction (Sandvig et al., 2016; Busuioc, 2020). These biases need not even come from a variable in the dataset with direct information on ethnicity but could be implicitly introduced by incorporating data on a person’s name or neighbourhood of residence (Sandviq et al., 2016). Moreover, it must be mentioned that some controversial biases are introduced by way of policy choice rather than as part of the algorithm itself, as occurred for instance in the Dutch SyRi case, where a network of governments chose to focus fraud detection analytics on specific problem neighbourhoods (Meuwese, 2020). Where such biases eventually become public, they may give rise to major trust breaches among groups that the model was biased against, or even wider society. Moreover, where an algorithm is a true black box, it may be difficult even for developers to ascertain what factors are leading to certain predictors, which may make it difficult to take away suspicions among sceptical external actors, causing tension between such algorithms and the principle of transparency (Ananny and Crawford, 2018; Ahonen and Erkkilä, 2020). It should be noted, however, that 'the black box' presents a problem that data- and computer scientists are generally well aware of, and multiple techniques have been and are being developed to counter it (Sandvig et al., 2016).

Beyond these technical and data-related issues, we may also discern a number of organisational and governance challenges giving rise to potential trust problems. Several authors have noted the risk of automation bias occurring when algorithms support civil servants' decision-making, as civil servants may lack the data literacy, time or autonomy required to assess the limitations or design of a model (Busuioc, 2020; Giest and Grimmelikhuijsen, 2021). This may lead to an overreliance on AI predictions and a lack of internal signalling of issues with models (Busuioc, 2020). It should be noted, however, that no evidence was found of automation bias in a recent study (Alon-Barkat and Busuioc, 2021). An overreliance on black box AI, furthermore, may be an issue from the viewpoint of good governance principles, as it becomes increasingly difficult for civil servants to transparently motivate a decision towards citizens (Ahonen and Erkkilä, 2020). Existing accountability channels may also become ineffective, as actors in a position to provide oversight (e.g. senior management, political superiors and/or the judiciary) lack the required expertise to accurately assess potential issues with data analytics projects (Busuioc, 2020).

All these reasons may cause citizens and civil society organisations to reduce their level of trust in governmental data analytics processes. Lowered levels of trust, in turn, may impact the legitimacy of utilising data analytics and thus hinder governments from utilising advances in data analytics to their fullest benefit.

2.2.2. Trust and Trustworthiness

Before continuing with the various strategies that governments may employ to enhance trustworthiness, it is useful to briefly elaborate on the notions of trust and trustworthiness. Trust has frequently been seen as a willingness to take risks, based on the expectation that the trustee will not behave opportunistically (Levin et al., 2006; Hardin, 1996; Grimmelikhuijsen and Meijer, 2012). Trust is thus a relational state presupposing an interaction between trustor and trustee, which is not always the case when discussing trust in public sector actors. Trustworthiness, on the other hand, refers to the qualities that a trustee has or lacks in terms of their ability to deliver, their benevolence and their integrity. A trustee, such as a government, may actively attempt to foster such trustworthiness by emphasising activities and choices that display ability, benevolence and/or integrity (Hardin, 1996; Levin et al., 2006).

Perceived trustworthiness then refers to the beliefs and expectations that a trustor has regarding the desirable qualities that a trustee possesses (Levin et al., 2006; Grimmelikhuijsen and Meijer, 2012). This distinction between trustworthiness and perceived trustworthiness is essential here, as we emphasise the potential distinctions between governmental strategies to engender trustworthiness and civil society perceptions of such trustworthiness. Such discrepancies may arise due to a trustee lacking several desirable qualities (it is at least hypothetically possible that a government has malicious intent, for instance) or due to the characteristics of the trustor (e.g., different predispositions to trust (Grimmelikhuijsen and Meijer, 2012)). However, they may also emerge due to the bounded rationality of both trustor and trustee, with the former not always being capable of signalling trustworthiness perfectly, while the latter will usually not have the capacity to perfectly predict the trustworthiness of an actor (Hardin, 1996; Bitektine, 2011). Although such mismatches inherently arise in the social interactions between actors and are thus likely to never be resolved completely, matching strategies to enhance trustworthiness to the criteria on which the trustee evaluates trustworthiness may aid governments in enhancing the perceived trustworthiness in their data analytics projects.

2.2.3. Resolving Trust Problems by Focusing on Trustworthiness?

Given the many dilemmas associated with the rise of data analytics in the public sector, it is no wonder that much effort has gone into the way that data analytics projects can be designed to be trustworthy. We begin with a brief approach of measures that can be taken at the technical level, although this discussion will stay relatively general by discussing categories of measures. Subsequently, we discuss organisational aspects of designing trustworthy analytics. Finally, we discuss a recently emerging perspective in public administration, which more broadly examines the place that data analytics should take within public sectors. Although we discuss these as separate dimensions of trustworthy data analytics governance, it must be noted that dimensions are frequently complementary and – for specific measures – may sometimes overlap.

The technical dimension

Perhaps the most prevalent approach to attain trustworthiness for a data analytics project is to focus on technical features. Generally speaking, data scientists and computer scientists are well aware of the many technical problems associated with creating large databases of sensitive data, selecting data, training models and/or examining the robustness of results. As was mentioned earlier, one of the classic criticisms of AI models in particular is that they can function as black boxes, in which the relative weight of factors predicting a certain outcome is unclear even to developers (Stilgoe, 2018). The black box nature of many AI systems has led to a substantial thrust in research on methods to counter the black box problem (Gunning and Aha, 2018; Emmert-Streib et al., 2020). For instance, even deep-learning systems – which are noted for their complexity and opaqueness due to their use of multiple layers of data transformations – may nowadays be accompanied by explainer AI (so-called XAI algorithms (Gunning and Aha, 2018; Emmert-Streib et al., 2020)), which are specifically designed to reveal the contributions of certain variables in the model. It should be noted, however, that XAI is a relatively recent area of inquiry, and that current models frequently face challenges in terms of e.g., accuracy/interpretability trade-offs or excessive computing power requirements (e.g., Gunning et al., 2019). Another frequently used method is the generation of local explanations (i.e. explanations of a single or small group of predictions) to see whether these local predictions depend on factors that are deemed undesirable. Explaining which factors lead to certain predictions also aids in correcting emerging biases and model drift, suggesting ways in which datasets can be altered to prevent 'garbage-in-garbage-out' problems or suggesting a need to retrain the algorithm. For instance, should a natural language processing algorithm become less capable of analysing input from certain dialects, it may be possible to retrain the model with additional data from that dialect. Data scientists argue that XAI will be an essential tool to generate user trust in the future (Gunner and Aha, 2018), although it remains somewhat unclear how explainability should be used towards broader audiences (such as citizens or civil society organisations that are not immediate users).

Data security, data selection and anonymisation are other measures that can be taken to ensure the technical robustness of an algorithm. For instance, to reduce the impact on privacy, a public actor could choose to focus only on publicly available information (for instance, Google maps data instead of data gathered on-the-ground through photographs). Where more personal data is necessary, developers can train models on anonymised versions of a dataset. Even if de-anonymisation is necessary, as might be the case for risk-profiling algorithms in policing or supervision, this can be done only for observations of interest. It should be noted that preventing model drift and biases and ensuring technical robustness and data safety are also heavily emphasised by influential policy documents, the foremost of which is perhaps the EU High Level Expert Group ethics guidelines for trustworthy AI (High-level Expert Group on Aim 2019).

The legal dimension

Another dimension to ensuring the trustworthiness of data analytics is ensuring it does not violate legal requirements and good governance principles. Although elaborated in greater detail in Chapter 5, it is useful to denote here briefly that ensuring compliance entails not only implementing requirements from the GDPR and future AI regulations, but also respect for various fundamental rights and requirements in administrative law (Freeman, 2016; Meuwese, 2020; Donahoe and Metzger, 2019). In practice, complying with legal norms frequently involves balancing different requirements, with a moderate breach of the right to privacy, for instance, being justifiable if it suitably serves for another goal (e.g. fraud detection) and does so in a proportionate manner (Meuwese, 2020). Similarly, a government aiming to engage in fraud detection may have to balance the effectiveness of its data analytics project with transparency requirements, as too much transparency could potentially aid violators in gaming the system (Chander, 2017). Malicious companies may, for instance, attempt to exhibit characteristics that make them unlikely to be seen as high risk by a risk-prediction AI, thus helping them evade detection (Chander, 2017). At the same time, although this is a frequently used explanation for limiting transparency in certain AI projects, one may question to what degree gaming the system will actually occur in settings such as social welfare fraud, and to what degree this can justify limiting the application of the transparency principle. It is therefore worth noting that a Dutch court was unconvinced by a similar reasoning in a case on geographically-focused tax and benefits fraud algorithms. In this case, the court noted that using highly sophisticated and new techniques with a substantial impact on citizens' legal positions places a heightened responsibility on governments to be transparent regarding the implementation of such techniques (Meuwese, 2020) – a relevant conclusion from a trustworthiness perspective.

The ethical dimension

Related to the technical and legal dimensions is an area of inquiry that has seen substantial attention in recent years. Ethical AI is a broad term covering approaches and frameworks specifying that AI and associated data analytics projects must be employed according to norms of 'good' behaviour (Mittelstadt, 2019). Frequently, ethical AI incorporates principles and values such as the idea that data analytics projects may do no harm, should be fair, should promote well-being, should promote privacy and that humans retain their decision-making autonomy (Winfield, 2017). These principles are for instance visible in the EU High Level Expert Group's ethical AI guidelines, which aim to serve as a general framework for both private and public sector AI. Ethical AI as a field borrows from other professional codes, such as those employed in the medical sector (Mittelstadt, 2019), and have recently begun to see active implementation in governments (as will be seen in the results section). Despite its intentions, the ethical AI movement has not been without criticism. Principles are often relatively vague (Mittelstadt, 2019) and some may even be seen as somewhat arbitrary (e.g. is explainability always necessary, given the additional costs and burden that e.g. XAI may introduce?).

The organisational dimension

Although the aforementioned technical, ethical and legal measures seem to be the main focus of recent research into creating trustworthy data analytics initiatives, they provide a useful but arguably incomplete set of tools to address trust in governmental AI systems. An already somewhat more comprehensive view is gained when also taking into account organisational factors, in particular organisational values, culture and management processes (Brendel, 2021). Organisations already possessing strong cultures focused on client-orientation, transparency and openness of decision-making may for instance have an easier time transitioning towards a transparent and explainable implementation of data analytics than entities traditionally seeing too much openness as a risk. Similarly, organisational cultures focusing more on values such as privacy and other fundamental rights could be better predisposed to proactively implementing GDPR requirements (such as comprehensive data protection impact assessments (DPIAs)) than counterparts focused more strongly on the value that algorithms may bring in terms of efficiency or effectiveness (Chua et al., 2017). Given the legitimacy risks of inappropriately managing trust issues in the context of data analytics, fostering a culture that takes ethical, technical, and legal safeguards seriously is highly important (Brendel, 2021). At the same time, the accompanying critical note should be that transitioning an entity towards such a culture may take years, with many aspects of cultures being extremely difficult to manage (Schein, 1990). This transition depends in part on the management structures in place at the entity. Agile management tools or similar management styles – if set up properly instead of being fitted onto existing structures as window dressing (see e.g. Molenveld et al., 2020) – may be particularly suited to the open development with internal stakeholders. Such management styles allow the business side to provide feedback on missing elements and problems in the data or the algorithm and may foster internal trust in data analytics development. Moreover, codes of conduct and leadership styles may matter, providing employees with a fixed frame of reference (Brendel, 2021).

The accountability dimension

Ever since the major incidents such as the COMPAS Scandal and the Dutch benefits affair, social science scholars have also begun to realise the ramifications that data analytics broadly and AI more specifically may have for public governance. Recently, focus has been placed on the blurring of accountability lines due to the rise of data analytics (Busuioc, 2020; Freeman, 2016). Essential components of accountability relationships, including transparency and the provision of information, debates on this information and justifications of the functioning of algorithms is often stymied due to technical complexities and information asymmetries (Ahonen and Erkkilä, 2020; Freeman, 2016). Accordingly, calls have been raised to avoid public sector-wide oversight gaps by introducing an entity with the capacity and competences to investigate data analytics projects (similar to existing data protection authorities), by ensuring that political representatives or their aides have sufficient proficiency in data and AI to evaluate information, and/or by introducing judges specialised in data, IT and data science matters. Others call for regulatory solutions, such as the EU's upcoming AI regulation or requirements for greater sector-wide transparency. It is certainly true that steps can be taken in this area, as the development and implementation of data analytics processes, including inherently value-based choices, remain fractured in most public sectors. Accordingly, while some entities may be proactively engaging with ethics and trustworthiness, other entities may remain less inclined to do so, for instance expecting citizens to 'game the system' if a government is too transparent regarding its algorithms. Such gaming involves knowing which data categories contribute to AI predictions and using this knowledge to window-dress your characteristics in such a way as not to raise suspicion. Fears of gaming may lead some public entities to engage in what has been called fuzzy transparency, in which general aspects of activities are divulged but with low levels of detail (Ananny and Crawford, 2018).

A final and frequently overlooked dimension? Communicating trustworthiness through murky lenses

It is argued here, however, that ensuring trustworthiness through technical robustness, legality, organisational procedures and accountability links may be insufficient – despite the substantial efforts that are sometimes invested in these dimensions. Although ensuring trustworthiness along these dimensions is certainly necessary, it is unlikely that civil society organisations, let alone other citizens, will directly evaluate how a data analytics project operationalises these dimensions (Kolkman, 2020). Instead, a small subset of easily interpretable project features will be filtered through a media lens, which will then be heuristically interpreted in light of existing public perceptions of data and AI, as well with regard to perceptions of the public entity involved (Bitektine, 2011). A similar argument has previously been made in the context of algorithmic transparency, where it was noted that even completely transparent data science projects often lack critical audiences with either the ability or the interest to closely scrutinise algorithmic decision-making (Kemper and Kolkman, 2019; see also Stilgoe, 2018 and Burrell, 2016). Even data scientists themselves will frequently have difficulty interpreting the full scope of a peer's model, regardless of the original developer's transparency (Kemper and Kolkman, 2019).

Perceived trustworthiness is thus best seen as a complex and dynamic social construct, influenced by a wide variety of social and cultural factors beyond the immediate control of public sector organisations. Bitektine (2011) argues that in such situations of incomplete knowledge regarding the organisation, project and algorithms involved, evaluators will fall back on status judgments, in which attributes of an entity are predicted on the basis of its membership of a broader group of similar organisations. The point is that specific policies or actions designed to enhance trustworthiness, especially when these policies are technically complex, may fail to generate perceived trustworthiness among non-expert audiences (see e.g. Kolkman, 2020) – in particular when these audiences are somewhat distanced from a policy, such as civil society actors. Moreover, when some public organisations have engaged in salient unethical data analytics practices, the social judgment process implies that governments as a whole may be evaluated based on these practices. To complicate matters further, cultural associations of algorithms may also affect social judgments. The analogy with nuclear energy is striking here, with the latter having lost significant legitimacy in part due to cultural fears and associations with disasters and danger (Koerner, 2014). Similarly, terms such as big data and AI sometimes seem to invoke cultural associations with a police state in which all actions are monitored, or with sentient 'general AI's' capable of sentience and intelligent action. Given such social interpretation processes, there may be a substantial disconnect between the policies designed to secure data analytics processes and the trustworthiness that external actors attribute to such processes externally.

Translating a project's trustworthiness into perceived trustworthiness among civil society and citizens is therefore in part also a communication challenge. A combination of organisational transparency, project-level motivation procedures and algorithmic explainability can help here but are unlikely to resolve the issue on their own. Perhaps greater promise lies in active communication by signalling moral choices, and trustworthiness measures, such as implementing ethical principles or the applications of data analytics that are precluded by the policy. At the same time, it must be realised that some segments of society may remain averse to even the most well-designed data analytics processes.

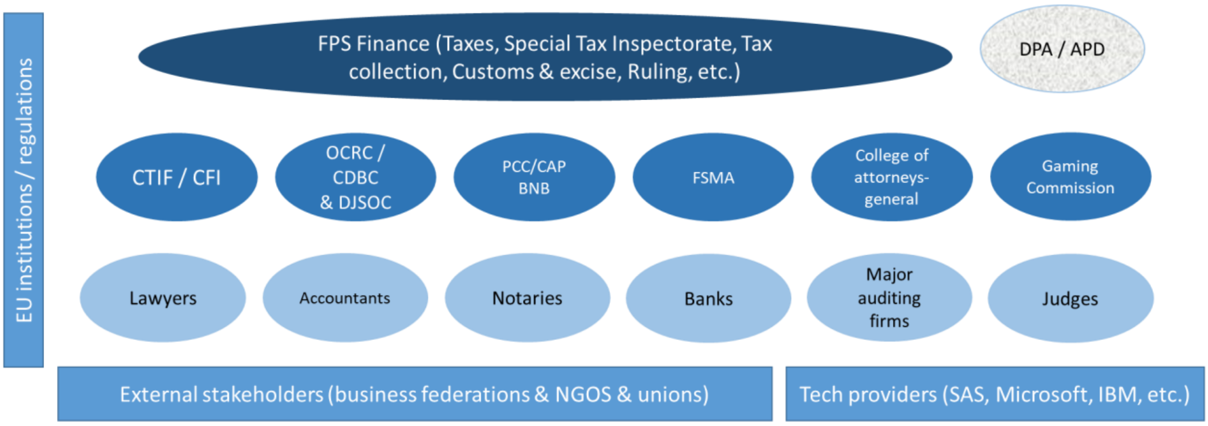
2.3. Belgian federal government context

DIGI4FED focuses on two policy domains: the fight against tax fraud and the fight against social security infringements. In these policy domains, the stakeholders have shown sustained interest in the use of data for public policy. These policy domains also display quite distinctive configurations in terms of stakeholders.

First, in the taxation field, the fight against tax fraud is part of the federal tax policy. Federal tax policy concerns both direct and indirect taxation at the federal level. Direct taxation notably includes personal and corporate income tax, while indirect taxation comprises among others value-added tax, and import and excise duties. Tax policy is a rather centralized policy domain, organized around the FPS Finance who is in charge of both policy design and part of the policy implementation.

Specifically, for the fight against tax fraud, the FPS Finance is the main actor: several directorates-general play a role, including tax collection, fraud detection, customs, and excise. A second layer of actors includes the public sector organizations that can provide information about potential infractions to the FPS Finance. These actors notably include the financial intelligence processing unit (CTIF/CFI), several services of the police forces in charge of corruption (OCRC / CDBC) and of major crimes (DJSOC), the national bank (that has a point of contact – PCC/CAP – to alert it about frauds), the financial services and markets authority (FSMA), the college of attorneys-general and the gaming commission. A third layer of actors is composed of people and public and private organizations that have a mandate to provide information about fraud they encounter to the FPS Finance: lawyers, accountants, notaries, banks, major auditing firms and judges. In addition, there are transversal actors that are present in this field: actors at the EU levels that enforce specific tax policy provisions and actors at the national level that act as external stakeholders (technology providers, unions, business federations and NGOs in charge of fair taxation and/or privacy rights). Figure 1 synthetizes this institutional landscape.

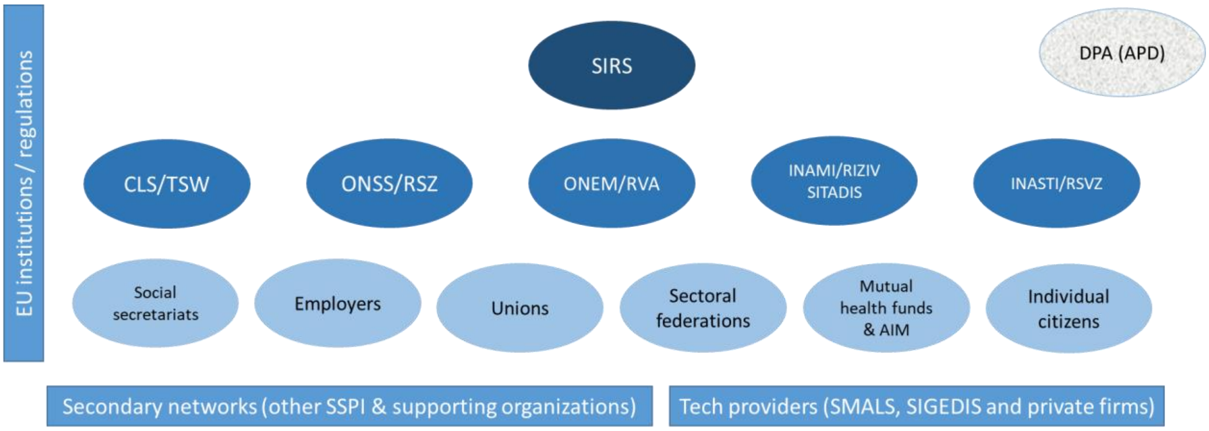
Figure 1. Institutional landscape of tax fraud policy


FPS Finance (Taxes, Special Tax Inspectorate, Tax collection, Customs & excise, Ruling, etc.)
DPA / APD
EU institutions / regulations
CTIF / CFI
OCRC / CDBC & DJSOC
PCC/CAP BNB
FSMA
College of attorneys-general
Gaming Commission
Lawyers
Accountants
Notaries
Banks
Major auditing firms
Judges
External stakeholders (business federations & NGOS & unions)
Tech providers (SAS, Microsoft, IBM, etc.)

Moving to the second policy area, the fight against social security infringements is part of the federal social security policy. Federal social security policy provides several social services to citizens about welfare benefits (e.g. unemployment, retirement, children, maternity leave) and healthcare reimbursement. This policy domain is rather fragmented. The FPS Social Security oversees part of the policy design, but most of the policy implementation, including the fight against social security infringements, is led by a multitude of public bodies. These are called social security public institutions (SSPI). Each of these SSPI has its own specific responsibility: providing employment benefits, providing healthcare reimbursements, collecting social security contributions from contractual workers or from self-employed workers, providing retirement benefits, providing holidays benefits, and facilitating datasharing between SSPI.

Among these topics, there are five areas in which specific actions are undertook to combat what is broadly defined as social security infringements. These are led by five distinct organizations: employment (ONEM / RVA), healthcare insurance (INAMI / RIZIV), social security for contractual workers (ONSS / RSZ), social security for self-employed workers (INASTI / RSVZ), and control of social legislation (FPS Employment). These five organizations are working together under the coordination of the SIRS / SIOD, that is in charge of developing a vision of the fight against social security infringements. In addition, peripheral actors can be identified in this policy landscape: SMALS, a non-profit that acts as the federal IT support service, and the Crossroads Bank for Social Security that organizes datasharing among SSPI. There are also non-governmental actors such as social partners, social secretariats, and mutual health funds, that play a role in the organization and management of social security. Eventually, the EU institutions also play a role in providing specific regulations in terms of social fraud (especially in the matter of social dumping). Figure 2 illustrates this institutional landscape.

Figure 2. Institutional landscape of social fraud policy


EU institutions / regulations
SIRS
DPA (APD)
CLS/TSW
ONSS/RSZ
ONEM/RVA
INAMI/RIZIV SITADIS
INASTI/RSVZ
Social secretariats
Employers
Unions
Sectoral federations
Mutual health funds & AIM
Individual citizens
Secondary networks (other SSPI & supporting organizations)
Tech providers (SMALS, SIGEDIS and private firms)

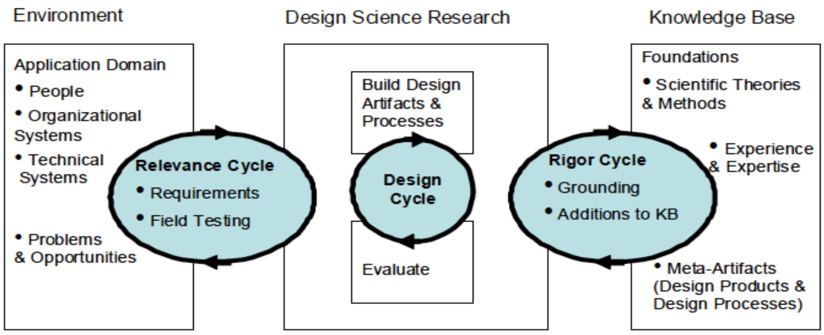
Lastly, there are some overarching political and administrative actors that play a role in both policy domains, such as the federal government, the federal parliament, the data protection authority (DPA / APD) and the Court of Audit.

3. METHODOLOGY

The methodological choices of DIGI4FED are selected according to the compatibility of the research objectives with the feasibility of completing the research tasks within a 2-years implementation period.

As a general methodological approach, DIGI4FED was based on **Hevner's '*three cycle view* of design science research' methodology** (Hevner, 2007). The *three-cycle view* identifies three interrelated cycles of activities in design science research. The *Relevance Cycle* inputs requirements from the contextual environment into the research and introduces the research artefacts into the environmental field-testing. This cycle ensures the connection between the design artefacts and the application domain. The *Rigor Cycle* provides grounding theories and methods along with domain experience and expertise from the foundations' knowledge base into the research. This cycle adds the new knowledge generated by the research to the growing knowledge base of the project. The *Design Cycle* supports the research activity for the construction and evaluation of design artefacts and processes. Figure 3 illustrates the interactions and interrelations of each research cycle.

Figure 3. Hevner's three cycle view of design science research



This methodological approach has been complemented by various methodological choices in the data collection and analysis phases. In a nutshell, DIGI4FED has used expert and focus group interviews, living lab approach, and experiments with citizen panels during research. Below we summarize each methodological process.

3.1. Expert interviews and interpretative structural modelling:

DIGI4FED research team conducted a series of semi-structured interviews in 2020 and 2021 with 66 public officials and technical, business and policy experts from the public sector and stakeholder organizations in the taxation and social security domains in Belgium[5]. The purpose of the interviews was to understand the perceptions of the key players and the issues that may affect the adoption of technologies such as big data, AI, and blockchain in the fight against fraud. The collected data has been thematically analysed to construct the key drivers in technology adoption and later through

[5] The list of interviewees and their organisations can be found in the Annex.

interpretive structural modelling to explore the underlying the relationships among identified drivers. The adoption models have provided the conceptual framework to categorise the identified themes and subthemes into key derivers. Table II gives the categorisation of constructs, their definition and categorisation according to the adoption models, and the elements identified through thematic analysis. The subthemes that are used for the categorisation of elements are included in the Annex.

Table II. Key drivers in the technology adoption

Drivers	Codes	Constructs	Elements	Definitions
Performance Expectancy	1	Technological maturity	Bias and noise Technology convergence Blockchain/DLT AI/ machine learning Fraud detection technologies	This variable captures the maturity of new digital technologies that are used in the fight against fraud.
	2	Perceived usefulness	Automation Improved social security and taxation Better data collection & analysis Past experiences Indirect added value of new digital technologies	This variable captures the perception of stakeholders about the usefulness of new digital technologies in improving the fight against fraud.
Self Efficacy	3	Capacities, skills and competencies	Resources Digital skills Training	This variable captures the resources, digital skills, and training of the administrations concerning the use of new digital technologies in the fight against fraud.
	4	Management/ operational systems	Guidelines Rules & standards Principles Processes Strategies	This variable captures management systems and means in the administrations concerning the use of new digital technologies in the fight against fraud.
Perceived Risk	5	Perceived risk	Legal challenge Control of data	This variable captures the perception of stakeholders

			Democratic challenge Administrative challenges Societal challenges	about the risk of using new digital technologies in the fight against fraud.
Effort Expectancy	6	Governance system	Data governance Open governance Multi-level governance Network governance	This variable captures the modes of governance in relation to new digital technologies that influence the fight against fraud.
	7	Technical infrastructure	Security Quality of database Data collection & analysis Softwares Computer maturity Reliance/dependence on external actors	This variable captures the technical capacity of the system infrastructure that influence the use of new digital technologies in the fight against fraud.
Social Influence	8	Public values	Appropriateness of technology Respecting privacy Tax fairness	This variable captures the public values in relation to the use of new digital technologies in the fight against fraud.
	9	Trust	Trust in administration Trust in society Trust in technology Trust in system Trust in tech providers/private sector	This variable captures the trust dimensions in relation to the use of new digital technologies in the fight against fraud.
	10	Socio-cultural elements	Digital culture Digital divide Willigness to share data	This variable captures the socio-cultural conditions in relation to the use of new digital technologies in the fight against fraud.
Facilitating Conditions	11	Interoperability	Technical interoperability Semantic interoperability	This variable captures the interoperability conditions in relation to the use of new

			Organizational interoperability	digital technologies in the fight against fraud.
			Regulative interoperability	
	12	Policy priorities	EU-level policy priorities	This variable captures the national and international policy priorities in relation to the use of new digital technologies in the fight against fraud.
			Fight against fraud	
			Political support	
			Geo-political aspects	
	13	Regulations	Data	This variable captures the national and supranational regulations in relation to the use of new digital technologies in the fight against fraud.
			Taxation/Social Security	
			Transcending laws	
			Justification of decisions	

A thematic analysis across interview data has revealed 13 variables to keep into account to understand the dynamics in the adoption of new digital technologies in the fight against fraud. These variables correspond largely to the theoretical assumptions of technology adoption models. Nonetheless, technology adoption models provide a limited insight about the interrelationship among these drivers. Understanding how changes in certain variables influence changes in other variables are important to develop effective technology adoption strategies. To address this issue, an analysis based interpretive structural modelling (ISM) and MICMAC has been conducted.

ISM is a mathematically derived, methodical, and cooperative method that allows researchers to examine contextual relationships among factors identified through expert opinion and establish hierarchical levels of challenges (Warfield 1974). ISM analysis involves the development of a directed graph through a hierarchical configuration of the relationships as interpreted by the researchers. MICMAC analysis complements ISM in the development of a graph that classifies factors based on driving power and dependence power (Ahmad et al. 2019). ISM have been used in the e-government literature to explore critical success factors in e-service delivery (Lal and Haleem 2009), citizen's perceptions of e-government's trustworthiness (Janssen et al. 2018), challenges for implementing the Internet of Things (IoT) in smart cities (Janssen et al. 2019). We use ISM method to explore and describe the dependencies between the variables (i.e. constructs) identified by the thematic analysis.

The development of the model goes through three levels of analysis. The first level of analysis is about the development of a structural self-interaction matric (SSIM). In developing SSIM, there are four possible ways of relating variables to each other, that are represented by 'V', 'A', 'X', and 'O' symbols (Hughes et al. 2020). The symbols are to be interpreted as: V = Variable i influences variable j; A =

Variable j influences variable i; X = Both variables i and j are influenced by each other; O = Variables i and j are not related to each other or do not influence each other. The relationships among these variables are interpreted by the research team based on the thematic relationships revealed through the thematic analysis. More specifically, if an underlying concept for a construct is related with another driver, this relationship is used to interpret the nature of the relationship.

Following SSIM, first an initial reachability matrix (IRM), and later the final reachability matrix (FRM) has been developed. IRM illustrates the relationships described by SSIM in a binary way, where '1' shows an existence of the relationship and '0' shows a non-relationship. In the next step, the IRM is converted into an FRM, where the transitive relations are included. Transitive relations correspond to if a variable X influences variable Y, and variable Y influences Z, then variable X should influence variable Z even if no mutual relationship is interpreted between variables X and Z. When such a relationship is found, an initial 'no relationship'-i.e. "0"- has been recoded as "1". FRM also shows the driving and dependence power of each variable. The driving power for each variable is the total number of variables, including itself, which it may help to achieve. On the other hand, dependence power is the total number of variables, including itself, which may help in achieving it. These driving powers and dependence powers are used later in the classification of variables as part of the MICMAC analysis.

The third step in the development of the ISM relies on the level partitioning. For that, the FRM is used to develop reachability and antecedent sets for each of the variables in the matrix. The reachability set, R(Pi), for a particular variable includes the variable itself and other variables that might help to achieve it (i.e. the corresponding value is 1). Similarly, the antecedent set, A(Pi) consists of the variable itself and other elements that might help in achieving it. The variables for which the interaction of these sets, $R(Pi) \cap A(Pi) = R(Pi)$, and the reachability set match, are considered as the top-level variables of the ISM hierarchy. Each iteration of the level partition matrix identifies the hierarchy of variables in achieving other variables. The top-level variables would not help to achieve other variable above their own level of hierarchy. Once the top-levels are identified, they are separated, and the same process is repeated. The iteration process continues until all levels of partitions are established. Accordingly, five levels of partitions are identified in the ISM diagraph.

- Level 1 is the highest level in the hierarchy of ISM diagraph and contains the following variables: technological maturity, capacities, skills and competencies, perceived risk, public value, policy priorities, regulation.

- Level 2 is the second level in the hierarchy of ISM diagraph and contains the following variables: management/operational systems, governance, technical infrastructure, socio-cultural elements

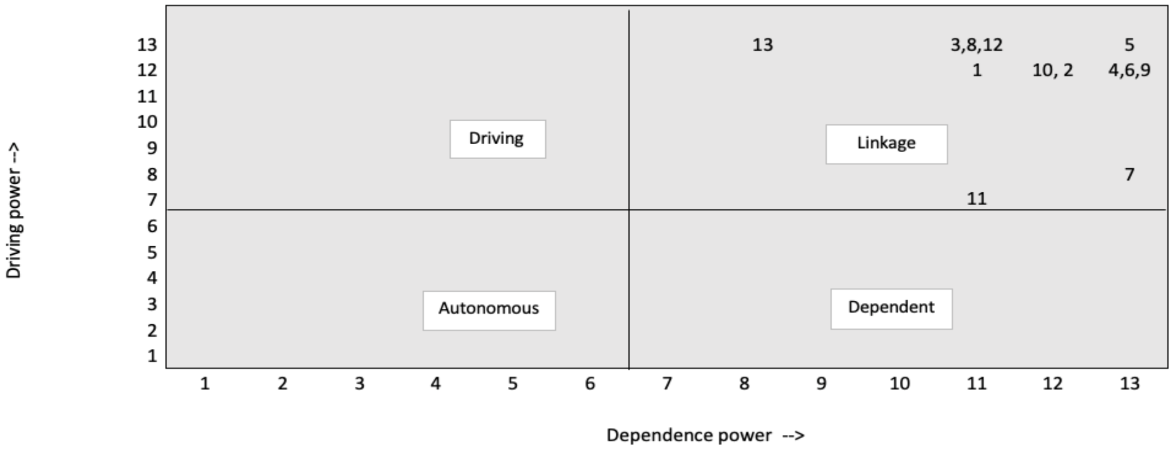
- Level 3 is the third level in the hierarchy of ISM diagraph and contains only perceived usefulness.

- Level 4 is the fourth level in the hierarchy of ISM diagraph and contains only interoperability.

- Level 5 is the lowest level in the hierarchy of ISM diagraph and contains only trust.

Before completing the ISM analysis, we visually categorized the driving and dependency powers among the variables through a MICMAC diagramme. The MICMAC diagramme contains four categories, namely autonomous, driving, dependent, and linkage. Autonomous shows variables with weak dependency and driving powers, showing variables that are mostly disconnected from the system. Driving shows variables with higher driving power and weak dependency power, mostly signifying variables that determines the changes in other variables without necessarily being dependent on the changes in other variables. Dependency shows variables with higher dependency power and with weak driving power, mostly containing variables that vary with changes in other variables without necessarily affecting the changes in other variables. Linkage shows the variables that have both higher dependency and driving power. The variables in the 'linkage' quadrant are the most influential variables in the system. The cut-off point for each quadrant is arbitrarily designated by the number of variables. Since there are 13 variables, cut-off point is designated at 6,5 in both axes.

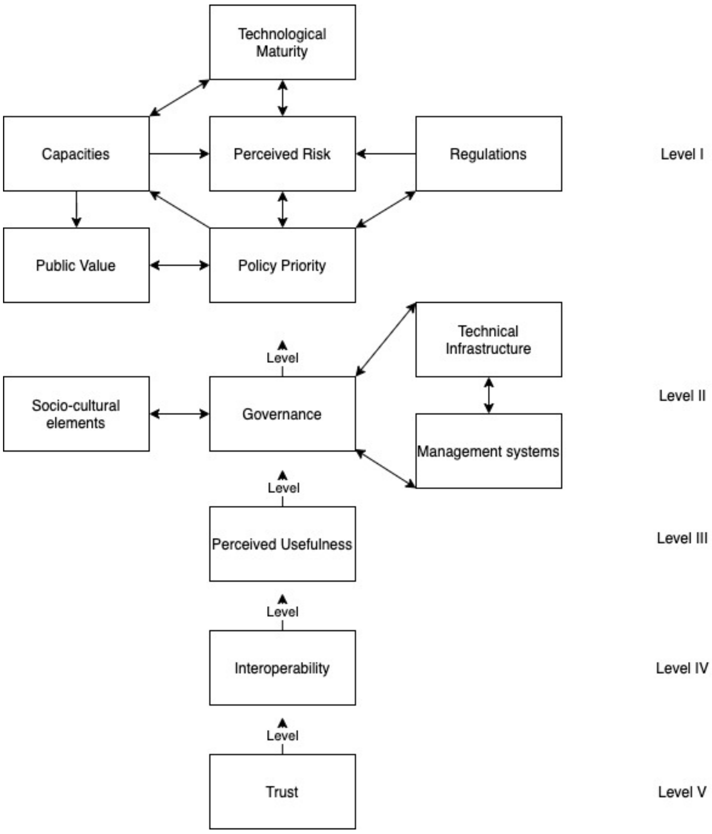
Figure 4. MICMAC diagramme



The MICMAC diagram in Figure 4 shows that all variables fall in the linkage category, which means that all identified constructs are highly influential in the adoption of new digital technologies in the administrative systems focusing on the fight against fraud in taxation and social security. Most variables show higher driving and dependency power suggesting that changes in any of them can lead to drastic changes in the adoption strategies of the new digital technologies in the fight against fraud.

Lastly, based on the level partitioning, and the direction of relationships identified in SSIM, we have developed the ISM model for the adoption of new digital technologies in the fight against fraud in taxation and social security systems. The ISM model in Figure 5 shows the direct and indirect relationships among variables. The direction of relationships is shown with an arrow. The ISM model suggests that trust (variable 9) plays a very significant role in driving the rest of the variables and is positioned at the base of the ISM hierarchy. The level of trust influences interoperability (variable 11) of information systems. Interoperability conditions in return directly influences the perceived usefulness (variable 2) of new digital technologies in the fight against fraud. Perceived usefulness influences the interrelationships between socio-cultural elements (variable 10), governance conditions (variable 6), technical infrastructure (variable 7), and management/operational systems (variable 4) at place. In this level, we observe that governance conditions, technical infrastructure, and management systems influence each other and as such should be assessed together. The model also shows that socio-cultural elements are closely interlinked with governance conditions. The changes and conditions in level II affects the interrelationships in the highest level of hierarchy. In the highest level, we observe direct mutual influences between policy priorities (variable 12) and regulation (variable 13), policy priorities and perceived risk from the use of new digital technologies in the fight against fraud (variable 5), and policy priorities and public values (variable 8). Thence, policy priorities are not only directly influenced by these variables but also perceived risk, public values, and regulations are directly influenced by the priorities set by the policymakers. Furthermore, policy priorities and the maturity of technologies (variable 1) are influential on the capacity conditions (variable 3) which in return influences public values concerning the appropriateness of technology, privacy, and tax fairness. Capacity conditions are also directly influential in perceived risks associated with new digital technologies along with regulations, technological maturity, and policy priorities. Finally, technological maturity is influenced directly by the perceived risk and capacity conditions.

Figure 5. ISM Model



3.2. Experiments with citizen panels

The main focus of WP2 was understanding the trust dynamics in the use of advanced fraud analytics techniques. Quantitatively, 4 survey experiments were held with Belgian citizens. The first two survey experiments (1269 and 756 respondents) focused on how governments might be capable of communicating trustworthiness towards citizens and whether such communications might enhance citizen trust. Different groups of respondents were shown (combinations of) various hypothetical trustworthiness measures taken by the federal government (such as implementing ethical AI policies, improving data security, reducing bias in data, retaining a human-in-the-loop, etc.). We aimed to make the measures shown to respondents as realistic as possible. The selected measures were based on insights from the interviews with government officials, the AI ethics literature, data science literature, and existing guidelines on trustworthy AI already employed by some governments.

Results showed that it is relatively difficult to gain citizen trust by being transparent on the various measures a government has taken to maintain the trustworthiness of AI projects, with no measure or combination of measures exhibiting robust positive effects on trust in government or policy support. This finding challenges the prevailing opinion in data science and AI ethics that increasing the trustworthiness of a data or AI project may be sufficient to solve the primary trust issues pertaining to such projects (e.g. Gunning & Aha, 2019). In other words, creating a well-designed, safe and ethical data analytics project does not guarantee an increase in citizen trust by itself.

Survey experiment 3 consisted of two sub-experiments. The first, experiment 3a (402 observations), was based on comments by respondents from the open fields incorporated into experiments 1 and 2, which showed that many respondents would like to retain some degree of control over their data and how it is used before being willing to share their data with governments voluntarily. Sub-experiment 3a integrates these insights with advances in areas such as Self-Sovereign Identity (SSI), as well as also the increasing degree to which online data submission choices are prevalent. We analysed whether simulating some degree of control over how governments share data might be able to enhance trust. To study the impact of the degree of control, we provided respondents with information on a hypothetical data exchange platform. Some respondents were then also given information on an additional project in which they could exercise some control over such data exchanges, and a list of data they would share, while a final group received information on said additional project as well as the option to indicate what data they would like to share if such a project were implemented.

When giving positive information and a list of data types that are available for exchange (without the option to choose which data type to share), trust among respondents is reduced. Seeing lists of data available for exchange therefore seems to induce perceptions of threat. In some cases, transparency may therefore come at the cost of lower trust. However, when instead of a simple list of data types, we also provide a choice of which data types citizens would like to share, we observe that trust is maintained on the same level as the control group (i.e., the group only receiving general information on the data exchange project). It thus seems that the threat perception can be offset by providing control.

Sub-experiment 3b (266 respondents) analysed how major crises abroad may influence perceptions of and attitudes towards data projects and AI in government in Belgium. We based this sub-experiment on the Dutch SyRI case, in which a court annulled Dutch legislation underpinning a data exchange and fraud analytics program due to it containing too few privacy and transparency safeguards (Rechtspraak, 2020). Respondents in the control group were only shown general information on a data exchange project (identical to sub-experiment 3a). Respondents in the treatment group were shown said information on a data exchange project, complemented with information on the Dutch case. As expected, results show that even foreign crises may have consequences for trust in data exchange and data analytics in Belgium, as respondents in the treatment group reported significantly reduced levels of trust and policy support.

Finally, experiment 4 (402 respondents) sought to build on experiment 3b, examining whether the framing of news articles containing information on privacy and discrimination crises could affect citizen attitudes. A first intervention was again based on the SyRI case, while a second intervention was based on the US COMPAS case (in which an algorithm used by US courts to predict recidivism showed signs of bias against minority groups) (ProPublica, 2016). In this experiment, control groups received relatively neutrally framed news articles on respectively privacy issues and discrimination issues (pointing out that there are issues, but that these are limited and are being addressed), while intervention groups received articles that were strongly negative about the effects of AI in government. Contrary to our expectations, there were no significant differences among respondents receiving relatively neutral information compared to respondents receiving strongly negative information on AI in government. Thus, results from experiment 3b and experiment 4 seem in tension with one another. Based on current data, we unfortunately cannot analyse why these experiments produced different results, although we suspect that the relatively subtle differences between control groups and intervention groups in experiment 4 may have played a role (after all, control groups still received an indication – albeit more subtle than intervention groups – that AI in government can produce privacy issues and biases). Further research is needed to definitively resolve questions regarding the impact of AI crises on public perceptions and attitudes.

The empirical conclusions and their implications are summarized in Table III.

Table III. Overview of experiments

ANALYSIS	OBJECTIVE	FINDINGS	IMPLICATION OF FINDINGS
EXPERIMENTS 1 & 2	Does transparent communication on measures taken to improve AI systems positively affect attitudes (trust and support) towards AI systems used in the federal government? How do pre-existing attitudes relate to attitudes towards AI systems used in the federal government?	No evidence that information akin to press release or website on measures taken to improve AI systems will also improve citizen attitudes. Substantial evidence that pre-existing attitudes (general trust in government, privacy concerns, trust in AI, self-reported discrimination) affect attitudes towards an AI project.	At least in the short-term, citizens seem to interpret brief texts with new information on the trustworthiness of AI in government in line with their pre-existing characteristics & attitudes (trust in government, privacy concerns, trust in AI, self-reported discrimination). For non-negligible sections of society, these pre-existing characteristics & attitudes are moderately to highly negative. This implies that transparent communication will not be a silver bullet to address trust issues in society.
EXPERIMENT 3a	Do contextual cues (positive information) and direct cues (cognitive exercise) on the control that citizens have over their data improve citizen attitudes towards governmental data exchange systems?	Early evidence that combining information on a project that would improve citizen control over data with actual control allows governments to maintain trust, while being more transparent. Simultaneously, also unexpected evidence that simply being transparent on data categories processed without offering citizens some degree of control may be perceived as threatening, in particular of these data categories sound threatening.	Whereas experiments 1 and 2 suggest that providing information alone will not allow governments to build trust and support, experiment 3a suggests that perceived control may act as a safeguard to maintain trust, providing a cue that a government will not or cannot behave opportunistically. At the same time, in the context of higher transparency but low perceived control over what types of sensitive personal data are exchanged, citizens may they update their attitudes to incorporate that data processing by the government may constitute a privacy threat towards them. Thus, we not only reconfirm that transparency is no silver bullet, but also see that it may work adversely.
EXPERIMENTS 3B AND 4	Experiment 3b: Does information on a privacy crisis abroad influence attitudes towards a data	Mixed evidence. Experiment 3b suggests that negative information on a policy crisis abroad	Mixed evidence makes definitive interpretation difficult. Cautiously, we may say that there is some

	exchange system in the Belgian federal government? Experiment 4: Does negative media attention in terms of privacy and the potential of AI systems to discriminate affect attitudes towards AI systems in Belgian governments?	adversely affects attitudes (trust, support, concerns). Simultaneously, experiment 4 does not manage to replicate these findings for negative media attention.	evidence that negative information from abroad may influence attitudes towards AI systems in Belgium, providing a risk factor that needs to be managed. At the same time, these findings need further scrutiny to verify whether there is an effect or not.
QUALITATIVE ANALYSES (CIVIL SOCIETY AND GOVERNMENT)	What are the differences in narratives between civil society actors (e.g. NGO's) and governments on trust in big data and AI?	Both public sector entities and civil society actors emphasize the importance of applying AI in a safe and ethical way. At the same time, narratives on how to achieve safe and ethical AI differ greatly between both types of actors. Public sector entities stress organizational and legal measures, while most civil society actors interviewed would advocate proactively limiting applications, preventing sliding scales, increasing accountability, etc.	Although public sector entities are often concerned about applying AI in a way that is trustworthy for citizens and civil society, their internal paradigm may cause them to inadvertently overlook elements found important by external actors (e.g. fears of sliding scales, even if current application of a data source is relatively non-invasive). Additionally, purely internal organizational measures are often difficult to pick up as a cue of trustworthiness by external actors, making a more comprehensive strategy to address concerns necessary.
QUALITATIVE ANALYSES (TRUST WITHIN GOVERNMENT)	How do internal trust dynamics on AI, big data and other new technologies function?	Interview respondents suggest that quantitative job insecurity (fear of losing one's job) and qualitative job insecurity (fear of losing a valued aspect of a job, such as a valued task) are potential drivers of mistrust. Cultural divides (e.g. between DPO's, development teams and frontline civil servants) may lead to relatively low trust and lower uptake of each other's suggestions. Strategies promoting participation, face-to-face contact, clear communication, co-creation and a reduction of uncertainty are reported as effective to counteract mistrust from user audiences towards new data analytics solutions.	As with other forms of organizational change, the development and integration of new data analytics tools can be experienced as threatening by some groups in the organization, such as frontline civil servants. This is not a one way street, however, as the introduction of new ethics or privacy procedures may similarly be experienced as imposed or threatening by development teams. Change management and/or technology acceptance techniques may aid in reducing uncertainty, perceived threat and perceptions of imposed change.

Table 1: Summary of analyses and findings of WP2

3.3. Living Lab

We conceptualize the living lab as a user-centric innovation space in which stakeholders are engaged in a long-term bottom-up collaboration in (or close to) real-life settings. Such innovation space can take the form of one or several workshops, of an open space, or of any space that allows for dialogue between stakeholders (Alavi, Lalanne, & Rogers, 2020).

In methodological terms, living labs aim at fostering open innovation through knowledge exchange, co-creation techniques and participatory methods, bridging the gap between stakeholders that do not usually meet. They thus act as an intermediary for innovation, connecting actors and supporting the knowledge exchange. In that regard, the process matters often as much as the end-result (Gascó, 2016).

The DIGI4FED Living Lab approach is designed in three phases: exploring, co-creating, and testing and evaluating (Zwetkoff, Elsen, Vigneron, & Pardo, 2018).

1. The first phase is about exploring problems and opportunities to be addressed regarding the integration of BD, AI and BCT in Belgian federal public organizations, as well as identifying emerging ideas and breakthrough scenarios.
2. The second phase aims at supporting the co-creation of an open governance model (or open governance modalities) through deliberative processes including several diversified stakeholders involved either in the policies concerning social security infringements or tax fraud. Stakeholders include actors from public sector organizations (federal and other levels of government, including social security public institutions), from policy-involved non-governmental organizations (as social partners and mutual health funds), and from other profit and not-for-profit organizations (technology providers, sectoral federations, NGOs).
3. The third phase (testing and evaluating) gives to stakeholders the opportunity to test the open governance model in a real-life context, in terms of performance and potential adoption, to identify incentives, possible risks and the discretionary space of professionals and users (Brandsen, Steen, & Verschuere, 2018).

In terms of participants, we work with a closed collaborative network, rather than with an open one. This means participants were invited because they are members of an organization that is linked (to a certain extent) to one of the two policy domains (either tax policy or social security policy). In that regard, they all have an interest to participate as well as a particular expertise to bring to the debate. In an open approach, we would have launched an open invitation to anyone interested in the topic.

Regarding the methods, we combined several qualitative and participatory methods in the exploration and co-creation phases. In the exploration phase, we relied on a literature review and semi-structured interviews to gain insights about emerging problems and opportunities associated with the integration of BD, AI and BCT in the fight against tax fraud and social security infringements. These insights were used in the creation of four scenarios presenting possible futures of digital governance in public organizations. In the co-creation phase, we used these scenarios to facilitate the debates and discussions among stakeholders in two scenarios-workshops. In turn, we relied on the results of these two workshops to support the facilitation of a third "solution-oriented" workshop.

The co-creation phase relied on a series of workshops, organized in an iterative manner:

1. Two scenario workshops - one on tax fraud (TF) and another one on social security infringements (SSI) - were organized in a very short period of time. Scenarios were used to stimulate the debates among stakeholders.
2. The insights provided by the participants were analyzed and directly fed to the set-up of a third solution-oriented workshop.
3. A third workshop was conducted, gathering participants from the first two workshops. Solution-oriented debates where held on the basis of the results of the two previous scenario workshops.

These workshops were analyzed in two successive steps, each time using a thematic analysis. On the one hand, the analysis of the two scenario workshops mostly focused on the identification of the main challenges associated with the integration of new technologies in the fight against tax fraud and social security infringements. On the other hand, the analysis of the third, "solution-oriented", workshop focused on the identification of the main solutions to the challenges that were highlighted in the analysis of the scenario workshops.

The following subsections present the main steps that were undertaken as well as the key results stemming from the analysis of both the scenario workshops and the solution-oriented workshop.

3.3.1. The Scenarios Workshops

The scenario workshops were centered around a short animated movie representing four possible scenarios for the future of digital governance by 2030. These four scenarios were set up to stimulate debates about what the future could look like in terms of digital governance and what are the associated challenges and opportunities.

These scenarios were constructed using the insights brought by the interviews and the literature review that were conducted during the exploration phase of the living lab. More specifically, a matrix approach (Cairns & Wright, 2018) in eight stages was used to construct the scenarios as follows: (1) setting the agenda; (2) determining the driving forces; (3) clustering the driving forces; (4) defining the cluster outcomes; (5) determining the key scenarios factors; (6) framing the scenarios; (7) scoping the scenarios; (8) developing the scenarios. Using a Pestel analysis, the researchers settled on a final selection of eight driving force clusters: (1) Political support in using of AI and BCT in public administration; (2) Availability of skilled IT profiles for public organizations; (3) Trust in public sector using new IT technologies; (4) Technological progress in artificial intelligence and BCT applications for public sector; (5) Belgian federalization process; (6) Environmental regulation regarding new digital technologies; (7) Development, compliance and enforcement of a regulatory framework; (8) Solidarity within society.

These clusters were then positioned in an uncertainty-impact matrix according to their (low to high) degree of impact and their (low to high) degree of uncertainty regarding the integration of BD, AI and BCT in federal public organizations. The most uncertain and impactful clusters - the "political support in using AI and BCT technologies in public organizations" and the "trust in public sector using new IT technologies" - were selected as the scenario macro-factors. After identifying the extreme outcomes for both macro-factors and building a set of descriptors for each scenario, the scenarios were labelled and given an orientation. The following figure presents the four scenarios in relation with the aforementioned macro-factors. Further details on the construction of the four scenarios are presented in deliverable 3.2.2.

Figure 6. The four scenarios

	High political support in using AI and BCT in public organizations		
Low trust in public sector using new technologies	The efficient government	The participatory government	**High trust in public sector using new technologies**
	The privatized government	The transparent government	
	Low political support in using AI and BCT in public organizations		

Each of these scenarios can be further detailed with descriptors as follows:

Table IV. Institutional landscape of tax fraud policy

	The efficient government	**The participatory government**	**The privatized government**	**The transparent government**
Political support in using AI and BCT in public organizations	High through major R&D funding towards private actors to develop these technologies	High with several public initiatives towards these technologies	Low, with the use of AI / BCT being low on the agenda	Low, with focus on open data and controllable technologies and rule-based algorithms
Trust in public sector using new technologies	Low among parts of the population	High through participatory deliberation	Low, as outsourcing has decreased technological expertise in the public sector	High, due to clear political orientations in using technologies in policy (security and transparency)
Public private relationships regarding technologies	Strong partnership with private actors as technology providers	Co-creation and participation with high public expertise	High presence of private actors in policy implementation	Public regulation vis-à-vis private actors to support open data and open source applications
Impact on policy	High efficiency and budget increases but opacity of algorithms and privacy issues	High legitimacy but time-consuming and costly processes	Outsourcing of implementation, with high financial returns but less control and socio-economic disparities	Continuity of current processes with a crossroad bank for policy

Each of these four scenarios were then synthetized and edited under the form of a short video in collaboration with a professional studio that specializes in video animation (Whoosh studio)[6].

The two scenario-workshops were organized in June 2021 in meeting halls provided by the FPS BOSA. The objective of these workshops was to elicit future expectations (including opportunities and risks) of participants regarding a model of governance aiming at integrating big data, AI and BCT in federal public organizations. In these workshops, the four scenarios served as a catalyst of discussion, to

[6] This video is available online at the following link: https://www.youtube.com/watch?v=lsBZwy4w5RA&t=29s

engage participants with a plurality of possible futures and to broaden their perspectives to a wider set than currently existing (Cairns & Wright, 2018).

Both workshops were designed for a physical setting to fit in a 4-hour period. The facilitation protocol was designed to maximize speech time for all participants, combining a plenary with sub-group discussions. As Belgian federal public organizations work in both Dutch and French, no constraints were imposed regarding the choice of a language of expression during the workshops. This choice was made to favor free speech from participants in a setting and a language in which they are the most comfortable. Conceptualizations, complex thoughts and impressions can be difficult to express and convey with nuance in a language that is less familiar. As such, we followed the advices of Rice and Ezzy (1999)**,** who recommend allowing participants to express their thoughts in their own language. The team of researchers also prepared some visual support material to facilitate the presentation of the workshop agenda and objectives to participants. PowerPoint presentations were distributed in French and Dutch versions to ensure participants' best possible comprehension. The animation video presenting the four scenarios was projected in English with subtitles.

Workshop participants were identified from the stakeholder analysis in the exploration phase of the living lab process. We favored the formation of heterogeneous groups of participants for each policy domain, thus fostering a plurality of points of view. More specifically, the following criteria were followed during the selection process of participants: (1) the institutional affiliation of the participant to one of the identified stakeholder groups; (2) his/her expertise regarding key aspects of the policy domain; (3) his/her interest in the topic and his/her level of seniority. About the last aspect, we mostly targeted participants that were not in a senior management position. This choice was guided by our need to get specific insights from operational civil servants while reducing dominance effect within the group. The following table presents the profiles of the stakeholders who participated to the SSI and TF scenario workshops.

Table V. Institutional landscape of tax fraud policy

Participants to the SSI scenario workshop	Participants to the TF scenario workshop
Data analyst, ONSS	Dataminer, FPS Finance, customs
Advisor expert in social fraud, INAMI	Datamining project coordinator, FPS Finance, Special tax inspectorate
Head of datamining, INASTI	E-auditor, FPS Finance, Special tax inspectorate
Data analyst, ONEM	Data protection officer, FPS Finance
Civil servant in charge of controlling social policy legislation, FPS Employment	Enterprise architect, FPS Finance
IT specialist, BCSS	Data analyst, CTIF
Research consultant, SMALS	Business Group Leader Digital, Agoria
Innovation manager, SPF BOSA	Director Tax & Public Affairs, FEBELFIN
National Technology Officer, Microsoft (x2)	Founder, Sagacify

Scenario workshops were set up as follows. After a brief introduction presenting the DIGI4FED project, the participants, the agenda, and the workshop objectives; the team of researchers presented the video containing the four scenarios. After this first screening, participants were given the opportunity to ask for precisions to ensure their best comprehension. Workshop participants were then separated in two sub-groups, each moderated by a researcher. These two groups re-watched the video before engaging discussions on the opportunities and challenges regarding the integration of BD, AI, and BCT in the fight against TF or SSIs. The results of these sub-group discussions were then synthetized and presented in a plenary session.

The first scenario workshop was held on the 24th of June 2021 and gathered private and public actors involved in the use or integration of new IT solutions in the fight against SSIs. This first workshop lasted a bit over four and a half hours. The second scenario workshop was held on the 29th of June 2021 and gathered public and private actors involved in the use or integration of new IT solutions in the fight tax fraud (TF). This second workshop lasted almost 4 hours. These two workshops were audio-recorded and then fully transcribed with the help of the students who were present during both activities.

4. SCIENTIFIC RESULTS AND RECOMMENDATIONS

The scientific results and policy recommendations of DIGI4FED have been compiled under three deliverables (D.1.2, D.2.3, D.3.4). Below, we will highlight the key findings and recommendations on the governance model focusing on the organizational, legal, trust, and technological dimensions.

4.1. Policy recommendations on the legal framework for the use of Big Data and AI in the Belgian federal administrations

Public administrations consistently use more and more data and process a larger amount of citizens' personal data to deliver their public services. Yet, when processing these personal data, they have to comply with the GDPR. The adoption of the GDPR presents several challenges for the administrations as they might have to revise their former way of processing personal data in order to be compliant with this new Regulation. Indeed, the principles of accountability and of data protection by design and by default, which are at the core of the GDPR, were not explicit in the former Directive and in the Belgian Law of 8 December 1992 that transposed it, and the administrations had to adapt their practices in order to meet the new standards set by the GDPR. These privacy and personal data protection rules are especially important in the advent of new technologies, such as Big Data and Artificial Intelligence, which increase the public administrations' capability to process greater amounts of data in order to provide public services and to support their decision-making. Furthermore, some core principles of administrative law have to be considered as well when using these technologies, as they are used by administrations in the context of the pursuit of their public service missions. As the use of such technologies could have a strong impact on the lives of the citizens, it is fundamental to understand the limits that this legal framework puts on their use. This is especially important for public policies and decision-making linked to social security infringements and tax fraud, as these could have a significant impact on the citizens' finances, in particular, and lives, in general.

Deliverable 1.2 provides a deep legal analysis of the implications of the GDPR in order to draw the legal framework within which data-driven technologies can be used in fraud analytics. We invite the readers to check this deliverable for further elaboration. In short, public sector organizations need to address the following principles/conditions: 'Purpose limitation', 'Data minimisation', 'Special categories of data', 'Right to information', 'Right of access', 'Right to erasure', 'Right to non-solely automated decision-making + Right to human public services', 'Equal access to public services', 'Explainability'. In the light of these legal and ethical conditions, below we highlight the key recommendations for public sector organizations that wish to use personal data for fraud analytics processes.

Recommendation 1: A public administration wishing to use personal data for policy- and decision-making should avoid relying on consent as the lawful basis for the processing of personal data.

Recommendation 2: A public administration wishing to use personal data for policy- and decision-making should rely on law, meeting the standards of Article 6.3 of the GDPR, as the lawful basis for the processing of personal data.

Recommendation 3: A public administration wishing to use personal data for policy- and decision-making should be mindful of the specific rules applicable to "special categories of data" (e.g. health data) and to data relating to criminal convictions and offences.

Recommendation 4: A public administration wishing to use personal data for policy- and decision-making can only do so for specified, explicit and legitimate purposes. These purposes should be explicitly mentioned in the law that serves as the lawful basis of processing.

Recommendation 5: A public administration wishing to use personal data for policy- and decision-making will have to ensure that the data will be kept in a form which permits the identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. It may only store the data as long as it needs it to fulfil the specific purpose for which it has been collected. To verify this, continuous data relevance checks should be performed. It is also advised to establish, in advance, maximal retention periods. Once this purpose is achieved, the data needs to be deleted or needs to be anonymised (and not merely pseudonymised).

Recommendation 6: A public administration wishing to use personal data for policy- and decision-making has to clearly define in advance the concrete categories of data that will be collected, stored and analysed, and why those data are necessary for the specific purposes of processing. This should be explicitly mentioned in the law that serves as the lawful basis of processing, and the number of data sources should be as limited as possible.

Recommendation 7: A public administration wishing to use personal data for policy- and decision-making has to make sure that the data is accurate and that it is kept up to date. To do so, it is advised to rely on a network structure, such as the CBSS. If it appears that this is not the case, the public administration will have to take every reasonable step in order to ensure that the data is either rectified or erased.

Recommendation 8: A public administration wishing to use personal data for policy- and decision-making should, to the extent possible, rely on pseudonymised data. If the data needs to be de-pseudonymised, this should be done following a risk analysis.

Recommendation 9: A public administration wishing to use personal data for policy- and decision-making will have to limit its accessibility to the minimum necessary for the purpose of processing. This means that, by default, personal data should only be made accessible to a limited number of people within the public administration. Access should be limited to those for which the access is necessary for the purpose of the processing, and this should be verifiable through access logs, notably in the perspective of third-party auditing (e.g. by the Data Protection Authority) of the process.

Recommendation 10: A public administration wishing to use data for policy- and decisionmaking has to be totally transparent and inform the citizens about the purposes for which their data are being collected and combined, about how these data are being stored and about how this data will be analysed. This information should be provided in a concise, transparent, intelligible and easily accessible form, using clear and plain language.

Recommendation 11: To increase the transparency of data uses by public administrations, for policy- and decision-making, the data transfer protocols and the deliberations of the Information Security Committee should be published on a single source, such as the Data Protection Authority's website, where it should be possible to search through them on the basis of several criteria, such as the types of purposes or of data concerned.

Recommendation 12: A public administration wishing to use personal data for policy- and decision-making, and to restrict the data subjects' right in order to guarantee public interest objectives – including monetary, budgetary and taxation matters, public health and social security –, must keep in mind that such a restriction must be provided by law. Moreover, this law must respect the essence of the restricted fundamental rights and freedoms, must be necessary and proportionate in a democratic society, and must safeguard an important objective of the general public interest.

Recommendation 13: If a public administration wishes to rely on automated processing for policy- and decision-making, it should evaluate the level of risk entailed by this processing prior to it, and especially whether it should be considered as implying "high risks" for individuals' fundamental rights. Indeed, in that case, the public administration will have to conduct a data protection impact assessment (the results of which should be published to increase transparency) and mandatory requirements would have to be respected in order to deploy these processing. These requirements pertain to the need to establish, implement, document and maintain a risk management system, to produce technical documentation, and to ensure high quality data and data governance, record-keeping – i.e. logs –, transparency and the provision of information to users, human oversight, and a certain level of accuracy, robustness, and cybersecurity.

Recommendation 14: Public administrations must refrain from relying on automated processing for policy- and decision-making which create unacceptable risks for the citizens, such as AI systems that deploy subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour; systems that exploit any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability; or social scoring systems.

Recommendation 15: Public administrations should refrain from using "real-time" remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for the targeted search for specific potential victims of crime; for the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack; or for the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence.

Recommendation 16: A public administration wishing to use data for individual decisionmaking, based "solely" on automated processes, has to inform the citizen about the existence of such processes. In that case, the citizen should, at least, receive meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for her. To ensure that the public administration can comply with this legal obligation, the use of rules-based AI systems should be privileged, and, if the algorithm is developed by a private entity, the administration should clearly specify in the public procurement either that it will hold the rights on the algorithm, or that it has the right to receive information about the logic involved behind the algorithm. Moreover, algorithmic impact assessments (AIAs) could be included in the public procurement procedures, in order to enable better communication with the general public, increase the in-house expertise of public agencies, increase the levels of accountability of automated decision-making systems, and offer a meaningful way for the public to question them.

Recommendation 17: A public administration wishing to use personal data for policy- and decision-making will have to set in place the measures allowing it to respond to the exercise of the data subjects' rights, notably the rights of access. The public administration will have to answer the data subjects' requests without undue delay and, in any case, within one month of the receipt of the request, and the exercise of the rights should be free.

Recommendation 18: A public administration wishing to use personal data for policy- and decision-making should anticipate potential erasure requests in the way it builds its AI system. More specifically, it should define, in advance, the moment at which data will be considered as being deleted and justify, in light of the accountability principle, why this complies with Article 17 of the GDPR. Moreover, in light of the iterative way of working of AI systems, that rely on the previous "learned" results for the next training iterations, it should reflect, from the outset, on the impact that the right of erasure will have on these "learned" results: is it possible to erase the data subjects' data not only from the training data but also any trace of it in the "learned" results.

Recommendation 19 a: If a public administration wishes to rely on automated processing for policy and decisionmaking, it should ensure that there is critical and fictitious /nonnegligible human supervision of the process. Light of the constant budget cuts and personnel risk that the few critical thinking avoided.

Recommendation 19 b: If a public administration wishes to rely on automated processing for policy and decisionmaking, it must inform the data subject in case the "flag" decision taken by the system significantly affects the data subject or produces legal effects concerning him or her.

Recommendation 20: If a public administration wishes to rely "solely" on automated processing for policy- and decision-making, this will have to be provided by law. This law will have to contain suitable measures to safeguard citizens' rights and freedoms and legitimate interests, and these measures shall be implemented in the automated decisionmaking process. This means that it has to be built in such a way that the citizen is, at least, provided with the right to obtain human intervention from the public administration, to express her point of view and contest the decision.

Recommendation 21: If a public administration wishes to rely "solely" on the automated processing of health data for policy- and decision-making, it will have to demonstrate that this "solely" automated processing is necessary for reasons of substantial public interest, and it will have to be based on a law, which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

Recommendation 22: A public administration relying on Artificial Intelligence for policy- and decision-making should always be able to explain how such a policy or decision was reached, which implies that the analytics that has been used need to be interpretable. This is a matter of accountability for the decisions taken by the public administrations, and the citizens should be able to understand this explanation. To ensure that the public administration can provide such an explanation, the use of simpler and effective AI systems (such as rule-based systems) should be privileged, and, if the algorithm is developed by a private entity, the administration should clearly specify in the public procurement either that it will hold the rights on the algorithm, or that it has the right to receive information about the functioning of the algorithm and about how specific decisions are taken, in order to be able to explain this to the citizens.

Recommendation 23: To ensure that it meets its explanation requirements, the public administration should use the "Explainability checklist" provided in the "Ethics Guidelines for Trustworthy AI" of the High-Level Expert Group on AI: "Did you assess: - to what extent the decisions and hence the outcome made by the AI system can be understood? – to what degree the system's decision influences the organisation's decision-making processes? – why this particular system was deployed in this specific area? – what the system's business model is (for example, how does it create value for the organisation)? Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can under–tand? Did you design the AI system with interpretability in mind from the start? – Did you research and try to use the simplest and most interpretable model possible for the application in question? – Did you assess whether you can analyse your training and testing data? Can you change and update this over time? – Did you assess whether you can examine interpretability after the model's training and development, or whether you have access to the internal workflow of the model?".

Recommendation 24: A public administration wishing to rely on Artificial Intelligence for policy and decisionmaking must ensure that all citizens are treated equally by the algorithm, which shall not be biased and shall not entail discriminations against some categories of the population. To ensure this, public administrations shall audit their AI systems on a regular basis and shall perform continuous data quality and biases checks.

Recommendation 25: A public administration wishing to rely on Artificial Intelligence for policy and decisionmaking must ensure that the functioning of the algorithm is not too complex, and that the citizen can understand, at least at a high the algorithm works.

Recommendation 26: Public administrations wishing to rely on AI for policy- and decisionmaking will have to be transparent about these objectives, about the role played by AI and by humans, and about the logic behind the decisions (active publicity) and will have to conduct a balance of interests in order to determine whether they need to provide a copy of this information to citizens that would request it (passive publicity). In this regard, they should publish the rules defining the main algorithmic treatments used in the performance of their tasks, and they should communicate to the person who is the subject of an individual decision taken in whole or in part on the basis of algorithmic processing, at the latter’s request, the characteristics of the algorithm in an intelligible form (i.e. the degree and type of contribution of the algorithmic processing to the decision-making; the data processed and their sources; the processing parameters and, where appropriate, their weighting applied to the individual’s situation; and the operations carried out through the processing).

Recommendation 27: Public administrations should leverage Big Data and AI technologies in order to ensure the automatic allocation of social rights and benefits, by streamlining and simplifying their processes, in full compliance with the personal data protection and administrative law requirements. This can be done by further building on the “only-once principle” of data collection from individuals and on the infrastructure of the Crossroad Bank for Social Security (CBSS) in order to facilitate an efficient exchange of data between the SSIs to the benefit of the individuals. At the very least, proactive and personalised information is needed from the administrations towards these individuals, in order to bring to their attention that they are entitled to a specific allocation, and to explain clearly what is required of them in order to benefit from it.

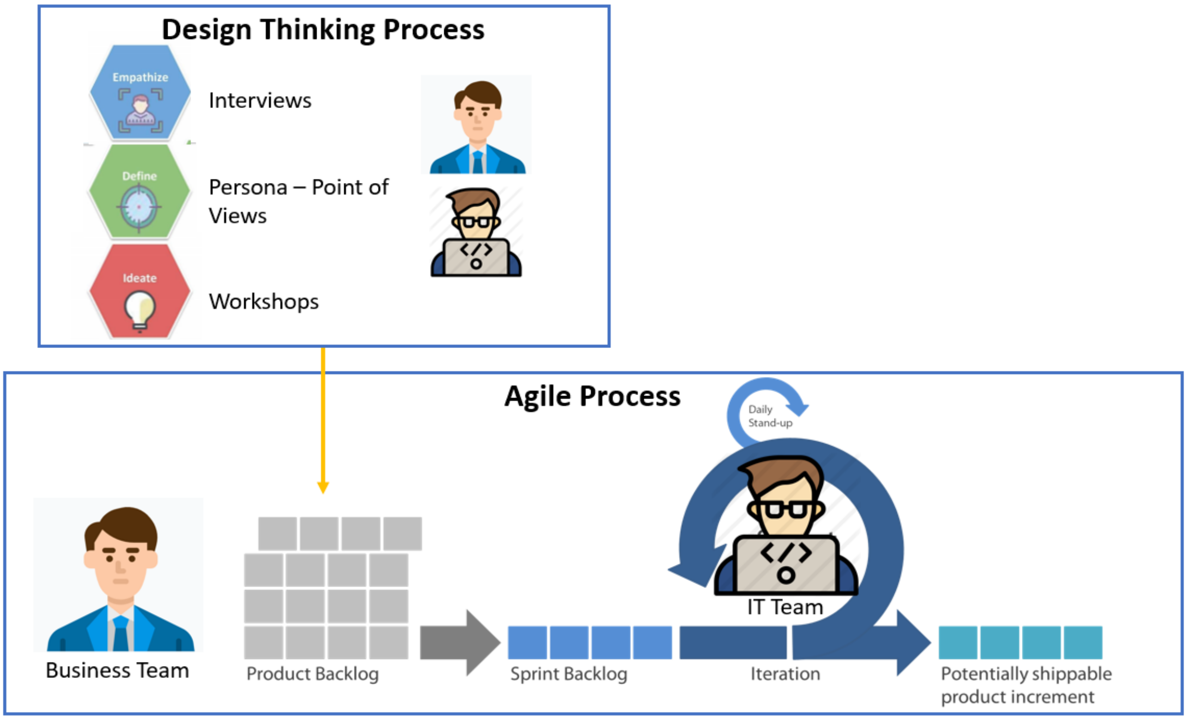
Recommendation 28: While Big Data and AI technologies should be leveraged in order to ensure the automatic allocation of social rights and benefits, the risk of the digital divide must be kept in mind. Accordingly, it is important to keep the possibility, for the most vulnerable individuals, to have physical contact with social workers in order to assist them in a personalised way, despite the existence of automaticity in the back-office.

4.2. Policy recommendations on the implementation of legal requirements in fraud analytics processes

The policy recommendations on ensuring legal and ethical compliance in fraud analytics show that the fraud analytics process is not only performed by IT teams (data scientists, data miners, etc.) but also by stakeholders knowledgeable in the business domain (fraud investigators, legal specialists, etc.). A close collaboration between legal specialists (the business side) and the data scientists (the IT-side) should thus be ensured so that the numerous legal requirements are effectively translated into technical procedures within the fraud analytics process.

Drawing from the insights gathered from the agile methods, participation methods and design thinking literatures, we suggest a tentative methodology to increase Business-IT alignment in Fraud Analytics. This methodology by enforcing collaboration between several profiles in the administrations, should ensure proper collaboration between legal experts and data scientists as well. Figure 7 represents this methodology visually.

Figure 7. Suggested Methodology for Business-IT Alignment in Fraud Analytics



Drawing from the SCRUM agile methodology (Schwaber and Sutherland 2017), the business representatives (fraud investigators, legal experts, etc.) can be considered as the product data scientists owners of the analytics process and the IT representatives (IT managers, data miners, etc.), can be considered as the development team in charge of the iterative provision of analytics output. First, the methodology starts with the Design Thinking process where Business and IT work together. The IT-side interviews the Business side (“Empathize” stage, “Interview” participation method) so clear requirements are defined (“Define” stage). Furthermore, the organization of ideation workshops allow thinking creatively about analytics that leads to solutions to address the requirements (“Ideate” stage, “Workshop” participation method).

Second, the refined leads for solutions are bundled into a product backlog that the business team is in charge of. This backlog summarizes the requirements to be addressed by the IT team in a hierarchical manner. Then, the IT team self-organizes and selects which elements from the backlog they will tackle in a sprint (time-boxed period of work, often 2-4 weeks). The team works iteratively

("Prototype" stage, "Prototyping" participation method), with daily stand-ups to reflect on the work done. After the sprint, they produce potentially shippable product increments submitted for feedback to the business team ("Test" stage). Based on the feedback received through several iterations, the analytics will be more aligned to the needs of the fraud investigators and this will foster collaboration.

4.3. Policy recommendations on building citizen trust for AI and advanced fraud analytics

Big data, AI, and blockchain can all function as useful tools in government. At the same time, the integration of major data projects and AI, in particular, is often seen as threatening by (sub-groups) of citizens. Maintaining societal and civil servant trust in new AI projects is key to maintaining momentum but also a severely underestimated challenge. Major crises based on issues such as algorithmic discrimination or privacy infringements (such as the SyRI case and the benefits affair that have occurred in The Netherlands over the past two years) may disrupt momentum and create spillover effects in citizen attitudes towards government in general. Our research suggests that technical and project-level fixes to ethical and legal concerns are not sufficient at that stage to restore trust quickly.

To prevent such a future of mistrust, the federal government – and in particular entities in social and tax affairs, which deal specifically with personal data and policies which may adversely impact individuals – should act proactively. Integrating major data exchange programs and AI models should be done with clear attention to consequences in terms of good governance, proportionality, privacy, safety, and non-discrimination, recognizing the more stringent obligations governments have compared to private actors. Relatedly, it must be kept in mind that major portions of society may not find far-reaching data-gathering -storage, and exchange programs desirable. In their eyes, data-driven policies should always be heavily circumscribed; the goal-binding principle should be adhered to stringently, citizens must be informed and given some degree of control over their data, and fears of sliding scales must be taken seriously. A reasonable amount of the members of this group are already highly mistrustful of AI and government or may fear discrimination (perhaps based on prior experience), making them particularly susceptible to further alienation through trust breaches. Finding the right balance between new technologies and new applications with broad legitimacy among the Belgian general populace will therefore be a major challenge going forward. At the same time, neglecting to engage with this challenge may entrench opponents and create new opponents, i.e., a far more perilous route.

Below, we summarize the recommendations and their rationale behind building citizen trust for AI and advanced fraud analytics

Number	Policy recommendation	Rationale underlying recommendation
1	Develop and incorporate a common ethical data and AI framework throughout government, e.g. based on guidelines	Results from both the experiments and the interviews suggest that a comprehensive framework is necessary to ensure that there are no ethical blind spots that can lead to issues. Although legally and technically Belgian entities often already have strong policies in place, the application of ethical principles currently seems to be *ad hoc* and determined on the organizational-level. Improvements here, for instance on subjects such as explainability and non-discrimination, could prevent future trust breaches.
2	Leave operational implementation of ethical guidelines to discretion individual governmental organizations, but stimulate them to do so and provide best practices	Some degree of autonomy in implementing guidelines is necessary to adapt to local needs, ethical concerns, data and AI applications and available data sources, likely making overly specific frameworks undesirable. At the same time, organizations often have difficulty translating abstract ethics guidelines into concrete and feasible work-processes. Exchanging best practices and providing support in the implementation of guidelines may aid in improving their implementation.
3	Ethical guidelines should be adapted to the public sector context, taking into account legal and moral requirements that are sometimes more stringent for public sector organizations than for organizations in other sectors Ethics guidelines' and protocols may require ongoing management to ensure they gradually become ingrained in the organization's culture	Ethics guidelines currently available, such as the EU's HLEG Trustworthy AI Guidelines, provide a valuable resource. However, public organizations have to deal with specific concerns not always present, or less present, in other sectors. One example is how to deal with the administrative law principle of proportionality. I.e. how should we test whether an AI system or big data system produces such little added value above another, potentially less invasive

	Specific factors which adapted guidelines could emphasize more strongly are the – frequently more stringent – transparency requirements, the proportionality principle and the prevention of sliding scales.	analysis system, that the application of AI or big data may be considered disproportionate? Another example is how to deal with concerns on gradually sliding scales in society. Analysing proportionality and the risk of sliding scales also relates to assessing the legitimacy of an AI/big data solution, as mentioned under recommendation 3. Other examples include the potentially more stringent transparency requirements and addressing societal fears of sliding scales and function creep.
4	Examine the legitimacy of new AI and/or data projects. Use tools such as surveys and focus groups to determine legitimacy risks among specific sub-groups in society.	Citizen trust in, support for and concerns on data and AI projects vary greatly across society. Even well-designed projects may be mistrusted by some sub-groups in society, e.g. due to negative pre-existing attitudes within this group. This leads to legitimacy risks, in particular when an AI or data system adversely affects (or is perceived to adversely affect) members of such a sub-group. As many of these risks are difficult to anticipate for a government, surveys and focus groups may aid in determining citizen responses.
5	Investigate ethical procurement to address ethical and legal issues in the relationship with private developers	Ethical procurement protocols may stipulate requirements such as the features of a model explaining a given phenomenon always being explainable, limitations on data access, positive obligations to prevent discriminatory effects, access to source code, the relationship between proprietary work of the private developer and administrative law requirements facing public sector organizations (e.g. obligation to motivate decisions), etc. Although ethical procurement has its own downsides (potentially more cumbersome and may limit the amount of available private

		parties), its usage may prevent issues with relatively sensitive data analytics applications.
6	Create a single point of contact for AI and big data questions, for instance within FOD BOSA.	As noted under recommendation 3, many organizations have difficulties implementing big data and/or AI technologies, finding the correct partners within government (e.g. to exchange data), developing ideas regarding the ethical application of AI, etc. A single point of contact or knowledge platform may mitigate such issues and stimulate the coherent and long-term application of legal requirements and ethics guidelines.
7	Incorporate (change) management strategies focusing on co-creation, participation, clear and consistent information (for instance Agile work methods). Identify cultural assumptions underlying another department’s values and actions to prevent imposing illegitimate systems	Interview analyses on trust dynamics within governmental organizations suggest trust is hampered by issues such as quantitative job insecurity (fear of losing one’s job), qualitative job insecurity (fear of losing valued aspects of the job, such as a task considered an important part of a professional identity), cultural misunderstandings on the task to be executed (broad face-to-face coaching of clients versus a single recommendation predicted by an AI, for instance), and imposed change (if employees cannot co-create the system, they may feel it is being imposed unilaterally on them, reducing their willingness to work with the system). Both the change management and technology acceptance literature suggest using participation, co-creation, information-provision and face-to-face contact to identify cultural assumptions in a consistent way. It should be noted that the same strategies may be applied when integrating new ethical or privacy-related protocols that may be experienced as

		imposed, stifling or unnecessary by development teams.
5	Prevent automation bias by (1) training civil servants and (2) preventing excessive downsizing of frontline civil servant workforce	Automation bias, i.e. the overreliance on conclusions generated by a data analytics system, may emerge from multiple sources. Two important potential sources are insufficient data literacy (i.e. civil servants may not see the flaws or limitations of a model) or from excessive reductions in the amount of employees that have knowledge of the task an AI is supporting.

4.4. Policy recommendations on designing a governance system for the use of advanced analytics in the fight against fraud

A core objective of the DIGI4FED project has been to design and test a governance system for the Belgian federal government that can leverage advanced analytics in fraud detection. We have addressed this objective, on the one side by identifying the relationships between various drivers of successful technology adoption inside the administration, and on the other side, by corroborating the viability of different policy actions through the experiment and living lab results.

Our final list of policy recommendations for the Belgian government focuses on how to address various governance challenges in the widespread introduction of advanced analytics in the fight against fraud. The justification and elaboration of these policy recommendations are provided in Deliverable 3.4. Below we give the overview of the governance issues and the corresponding policy recommendations.

Governance Issues	Policy recommendation
Trust	**Policy recommendation 1:** Pursue a transparent data exchange system that leverages self-sovereign identity (SSI) solutions and allows users to track data transactions
	Policy recommendation 2: Combine technological solutions with pedagogical / communication efforts toward citizens. But beware communication itself is not magic bullet.
Interoperability	**Policy recommendation 3:** Establish a single gateway to different departments, administrations, businesses, and stakeholder organizations to explore type of datasets held by different administrations.
	Policy recommendation 4: Create standardized documents and rules to facilitate data exchange through the single gateway.

	Policy recommendation 5: Integrate SSI solutions in the data exchange system in compliance with EBSI and ESSIF solutions.
Data Governance	**Policy recommendation 6:** Don't build a new data exchange platform to fight against fraud but address the system integration challenges among the relevant public and private organizations and existing data-sharing systems.
	Policy recommendation 7: Don't create a new entity to coordinate the data exchange for the fight against fraud. BOSA appears as the ideal candidate to coordinate other actors for the data exchange platform.
	Policy recommendation 8: Establish a data governance hub led by BOSA but participated by the representatives of regional and local government, as well as business and civil society organizations. needs to set the rules and guidelines in data extraction, sharing, and usage of platform data. A modular approach on data share and user rules is advisable.
	Policy recommendation 9: Manage data processing in the fight against fraud in a decentralized way, where data scientists are employed in respective units of social security and taxation domains but supported by the platform management in terms of legal and technical dimensions in data governance.
Digital Skills & Expertise	**Policy recommendation 10:** Collaborate with private sector organizations in developing technology solution in advanced analytics but pursue open-source technologies and build internal competencies in data analytics and advanced technologies to successfully engage with tech vendors.
	Policy recommendation 11: Invest in training programs and activities that supply advanced digital skills in public sector. Interdisciplinary trainings are necessary that combine computer skills with legal and social sciences dimensions.
	Policy recommendation 12: Provide frontline civil servants training and support to achieve a sufficient degree of data literacy and knowledge about advanced analytics. This training and policy support can be provided by the central management of the platform.
Value-based Design	**Policy recommendation 13:** Combine citizen-controlled data sharing (i.e. SSI) with transparency measures in data usage to maintain citizens' trust and engagement in digital governance.
	Policy recommendation 14: When suspecting that a particular data analytics tool may be perceived as illegitimate or discriminatory by some

	societal sub-groups, conduct research on how such a system would be interpreted by members of these sub-groups.
	Policy recommendation 15: Use legal, technical, and ethical requirements to design ethical procurement procedures for private sector service providers in advanced analytics.
	Policy recommendation 16: Engage in clear communication on the contents of digital change with end-users inside administration that are most affected by advanced techniques in fraud analytics. Avoid combining the integration of AI or data-driven support tools with excessive reductions in a specific category of frontline civil servants.
Policy Priorities	**Policy recommendation 17:** Make sure national and regional projects on SSI and digital wallets are compatible with EU-solutions.
	Policy recommendation 18: Participate in the EBSI use cases on taxation and social security to support the adoption of these solutions in the Belgian ecosystem.
Risk Management	**Policy recommendation 19:** Conduct in vitro-experiments to test the feasibility and understandability of XAI solutions for predictive analytics.
	Policy recommendation 20: Run controlled experiments to assess different technological configurations in data governance in finding a balance between performance of predictive analytics and understandability and transparency of algorithmic decisions.
Legal Compliance	**Policy recommendation 21:** Policy recommendation 21: Develop standardized guidelines for DPOs and DPA in managing open data policies for advanced analytics. IT solutions developed at the EU level have a clear advantage in facilitating DPOs' tasks
	Policy recommendation 22: Improve the reliability of IT solutions through regulatory sandboxes and benchmarking
	Policy recommendation 23: Pilot with regulatory sandboxes to test different regulative systems for the use of digital technologies in the fight against fraud.

5. DISSEMINATION AND VALORISATION

During the course of two-years various dissemination and valorisation activities took place. Some of the planned activities, however, had to be postponed or cancelled due to Covid measured. Below we give an overview of the dissemination activities that was organized as part of the DIGI4FED project.

Table VI. Dissemination activities

Event name	Activity	Date	Notes
DIGI4FED & DIGITAX webinar: Big data and artificial intelligence: the challenges for (tax) authorities	Webinar. Presentations and panel debate.	27/11/2020	For presentations: https://www.uantwerpen.be/en/research-groups/digitax/news-and-activities/webinar/ About 80 participants.
HackYourCity Hackathon	Coaching and Jury Participation	01/02/2021	Providing insights about AI use of Open Government Data
The 15 international RCIS 2021	International conference. Poster presentation	12/05/2021	Presentation of DIGI4FED project
DIGI4FED scenarios	Animation videos for the scenario workshops	05/09/2021	The video is accessible at the link
DigiTax interdisciplinary conference on computational taxation	Interdisciplinary conference on computational taxation	23/09/2021	Presentation by B. Kleizen 'Trust is government in the data analytics age'. About 60 participants.
Digitax/World bank series: How is blockchain technology is affecting taxation?	Presentation and public debate	12/10/2021	About 110 followers. For the recording and slides: link
Guest lecture AI in government UAntwerpen	Guest lecture at the Faculty of Social Science, UAntwerpen	1/11/2021	Presentation of trust components of the DIGI4FED project to students, in particular survey experiments

Bachelor's course Leeronderzoek			
Guest lecture AI in government UAntwerpen Master's course on Persuasive Technologies	Guest lecture at the Faculty of Social Science, UAntwerpen	14/03/2022	Presentation of trust components of the DIGI4FED project to students, in particular survey experiments
AI4GOV panel-European AI Week	Presentation of project findings in panel session	18/03/2022	For the presentation of Bjorn Kleizen: link
DIGI4FED closing ceremony	Presentations of project findings. Keynote speakers Public debate	20/05/2022	For the recording: DIGI4FED closing ceremony
Vlaamse Staten-Generaal over AI	A joint public event with Flemish AI Academy (VAIA)	17/06/2022	The event link for the presentations. An event to bring together politicians and decisionmakers in Flanders to talk about AI and government. Findings from DIGI4FED project has been used as part of the panel debate.
Lecture UAntwerpen GOVTRUST Summer School 2022	Lecture at the Faculty of Social Science, UAntwerpen	1/09/2022	Presentation of trust components of the DIGI4FED project to an international audience attending the GOVTRUST summer school
AI4GOV Hackathon	Hackathon	postponed	The proposal by Digi4Fed on the automatisation of social rights was accepted for the hackathon. Initially, the hackathon should have taken place in March 2021, but due to Covid measures, it has been postponed indefinitely.

6. PUBLICATIONS

The list of publications produced during the DIGI4FED project can be found in Table VII

Table VII. Publication List

Title	Type	Author(s)	Venue/Publisher	Status
The New Digital Era Governance	Book (edited volume)	E.Tan, J.Crompvoets	Wageningen Academic Publishers	Published in 2022.
Chapter 1: A new era of digital governance	Book chapter	E.Tan, J. Crompvoets,	The New Digital Era Governance	Published in 2022 as open access.
Chapter 2: The role of big data, AI and blockchain technology in digital public governance	Book chapter	E.Tan	The New Digital Era Governance	Published in 2022
Chapter 3: Business-IT alignment in fraud analytics: fostering collaboration between domain experts and data scientists	Book chapter	A. Simonofski, B. Vanderose and B. Frenay	The New Digital Era Governance	Published in 2022
Chapter 4: Legal framework for the use of Big Data and blockchain in public governance	Book chapter	T. Tombal, P. Willem and C. De Terwangne	The New Digital Era Governance	Published in 2022
Chapter 5: Legal framework for the use of artificial intelligence and automated decision-making in public governance	Book chapter	T. Tombal, P. Willem and C. De Terwangne	The New Digital Era Governance	Published in 2022
Chapter 6: Trustworthiness in an era of data analytics: what are governments dealing with and how is civil society responding?	Book chapter	B. Kleizen, W. van Dooren, and K. Verhoest	The New Digital Era Governance	Published in 2022

Chapter 7: Supporting public sector innovation through a living lab approach: the use of new technologies in administrations	Book chapter	M. Petit Jean, L. Bechoux, C. Fallon and M. Sabbe	The New Digital Era Governance	Published in 2022 as open access
Chapter 8: Conclusion – strategies and policy decisions in the new digital-era governance	Book chapter	E.Tan, J. Crompvoets	The New Digital Era Governance	Published in 2022
Balancing fraud analytics with legal requirements: Governance practices and trade-offs in public administrations	Article (peer reviewed)	A. Simonofski, T. Tombal, et al.	Data and Policy	Published 02/05/2022
The use of new digital technologies in the fight against fraud: An interpretive structural model about the key drivers in digital transformation.	Article (peer reviewed)	E. Tan, et al.	Public Administration Review	Revised and resubmit.
Designing an AI compatible open government data ecosystem for public governance	Article (peer reviewed)	E. Tan	Information Polity	Under review
The living-lab – A strategic tool to support collaborative innovation on digital transformation processes in public administration?	Article (peer reviewed)	M. Sabbe, et al.	Public policy and Administration	Abstract is accepted for the special issue. Manuscript is under review
Is everything under control? An exploratory study on how control over data influences citizen attitudes towards	Article (peer reviewed)	B. Kleizen, W. Van Dooren	In consideration for special issue of Information Polity.	Will be submitted for peer review in

major governmental data exchange projects				September 2022
Can we foster citizen trust in governmental AI projects? Experimental evidence on the limits of project-level measures	Article (peer reviewed)	B. Kleizen, W. Van Dooren, K. Verhoest	Government Information Quarterly	Under Review
Digital (R)evolution in Belgian Federal Government: An Open Governance Ecosystem for Big Data, Artificial Intelligence, and Blockchain	Conference proceeding	E.Tan, J. Crompvoets	The Fifteenth International Conference on Research Challenges in Information Science (RCIS) Proceedings, pp 699-702. Springer.	Published in 2021
Artificial Intelligence and Big Data in Fraud Analytics: Identifying the Main Data Protection Challenges for Public Administrations	Conference paper	T. Tombal, A. Simonofski	Data for Policy 2021 Conference (University College London)	Published in 2021
Designing an AI compatible open government data ecosystem for public governance	Conference proceeding	E. Tan	Hawaii International Conference on System Sciences (HICSS) 55	Published in January 2022
Deliverable 1.2 - Policy report on Big Data policy of the Belgian federal administrations	Report	A. Simonofski, et al.	DIGI4FED repository	Published in June 2022
Deliverable 1.3 - Activity report from the exploration and co-creation phases of the living lab process	Report	M. Sabbe, et al.	DIGI4FED repository	Published in November 2021
Deliverable 1.4 - Activity report and results of the testing phase of the	Report	M. Sabbe and C. Fallon	DIGI4FED repository	Published in June 2022

living lab process (Delphi survey)				
Deliverable 2.2.1 - Big data, AI-based governance and trust in government: developing an analytical framework and a conceptual model	Report	B. Kleizen	DIGI4FED repository	Published in October 2020
Deliverable 2.3- Empirical report: Qualitative and experimental results on trust, data and AI in and within governments	Report	B. Kleizen and W. Van Dooren	DIGI4FED repository	Published in August 2022
Deliverable 2.4. Policy report: Trust in and within government in the context of big data and AI	Report	B. Kleizen and W. Van Dooren	DIGI4FED repository	Published in August 2022
Deliverable 3.2.1 - A conceptual model of the use of AI and Blockchain for open government data governance in the public sector	Report	E. Tan	DIGI4FED repository	Published in February 2021
Deliverable 3.2.2 - Living lab and scenario development as a design methodology to build governance modalities for big data, AI and blockchain solutions	Report	M. Petit Jean and E. Tan	DIGI4FED repository	Published in September 2021
Deliverable 3.3 - Design artefact and evaluation criteria for testing	Report	E. Tan, et al.	DIGI4FED repository	Published in March 2022
Deliverable 3.4 - Designing a governance system for the use of	Report	E. Tan, et al.	DIGI4FED repository	Published in August 2022

advanced analytics in the fight against fraud				

7. ACKNOWLEDGEMENTS

DIGI4FED is funded under the BRAIN-be.2.0 program of the Federal Science Policy (BELSPO) with the grant number B2/191/P3/DIGI4FED.

We would like to acknowledge the support of all those who, in various ways, have contributed to the successful completion of this project. We also deeply appreciate the participation and feedback of colleagues in the follow-up and steering committees of DIGI4FED. Their guidance and suggestions helped us significantly improve our research.

REFERENCES

Agostino, Deborah, and Michela Arnaboldi. 2017. 'Social Media Data Used in the Measurement of Public Services Effectiveness: Empirical Evidence from Twitter in Higher Education Institutions.' *Public Policy and Administration* 32 (4): 296–322. https://doi.org/10.1177/0952076716682369.

Ahmad, M., X.-W. Tang, J.-N. Qiu, and F. Ahmad. 2019. "Interpretive Structural Modeling and MICMAC Analysis for Identifying and Benchmarking Significant Factors of Seismic Soil Liquefaction" Applied Sciences 9 (2): 233. DOI: https://doi.org/10.3390/app9020233

Ahonen, P., & Erkkilä, T. 2020. Transparency in algorithmic decision-making: Ideational tensions and conceptual shifts in Finland. Information Polity, (Preprint), 1-14., DOI: 10.3233/IP-200259

Alavi, H. S., Lalanne, D. and Rogers, Y., 2020. The Five Strands of Living Lab. ACM Transactions on Computer-Human Interaction 27: 1-26. https://doi.org/10.1145/3380958

Allam, Zaheer, and Zaynah A. Dhunny. 2019. 'On Big Data, Artificial Intelligence and Smart Cities.' *Cities* 89 (June): 80–91. https://doi.org/10.1016/j.cities.2019.01.032.

Allessie, David, Maciej Sobolewski, Lorenzino Vaccari, and Francesco Pignatelli. 2019. 'Blockchain for Digital Government.' *Joint Research Center*. Vol. 29677. https://doi.org/10.2760/93808.

Alshallaqi, M. 2022. "The complexities of digitization and street-level discretion: a socio-materiality perspective". Public Management Review, 1–23. https://doi.org/10.1080/14719037.2022.2042726

Ananny, M., & Crawford, K. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New media & society, 20(3), 973-989. https://doi.org/10.1177%2F1461444816676645

Appelbaum, Deniz, Alexander Kogan, and Miklos A. Vasarhelyi. 2017. 'Big Data and Analytics in the Modern Audit Engagement: Research Needs.' *AUDITING: A Journal of Practice & Theory* 36 (4): 1–27. https://doi.org/10.2308/ajpt-51684.

Arinder, Max K. 2016. 'Bridging the Divide between Evidence and Policy in Public Sector Decision Making: A Practitioner's Perspective.' *Public Administration Review* 76 (3): 394–98. https://doi.org/10.1111/puar.12572.

Bachner, Jennifer, and Kathy Wagner Hill. 2014. 'Advances in Public Opinion and Policy Attitudes Research.' *Policy Studies Journal* 42 (April): S51–70. https://doi.org/10.1111/psj.12052.

Bailey, Stephen J., Ari-Veikko Anttiroiko, and Pekka Valkama. 2016. 'Application of Baumol's Cost Disease to Public Sector Services: Conceptual, Theoretical and Empirical Falsities.' *Public Management Review* 18 (1): 91–109. https://doi.org/10.1080/14719037.2014.958092.

Barth, Thomas J., and Eddy Arnold. 1999. 'Artificial Intelligence and Administrative Discretion.' *The American Review of Public Administration* 29 (4): 332–51. https://doi.org/10.1177/02750749922064463.

Bean, Randy. 2020. 'Why Culture Is the Greatest Barrier to Data Success.' *MIT Sloan Management Review*, September 2020.

Beck, Roman, Christoph; Müller-Bloch, and John Leslie King. 2018. 'Governance in the Blockchain Economy: A Framework and Research Agenda.' *Journal of the Association for Information Systems* 19 (10).

Berryhill, Jamie, Théo Bourgery, and Angela Hanson. 2018. 'Blockchains Unchained: Blockchain Technology and Its Use in the Public Sector.'

Bitektine, A. 2011. Toward a theory of social judgments of organizations: The case of legitimacy, reputation, and status. Academy of management review, 36(1), 151-179. https://doi.org/10.5465/amr.2009.0382

Blume, Grant, Tyler Scott, and Maureen Pirog. 2014. 'Empirical Innovations in Policy Analysis.' *Policy Studies Journal* 42 (April): S33–50. https://doi.org/10.1111/psj.12050.

Boyd, Danah, and Kate Crawford. 2012. 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon.' *Information Communication and Society* 15 (5): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Brandsen, T., Steen, T. and Verschuere, B., 2018. Co-creation and co-production in public services: Urgent issues in practice and research. In: Brandsen, T., Steen, T. and Verschuere B. (eds.) Co-Production and Co-Creation: Engaging Citizens in Public Services. Routledge, New York, USA, pp. 3-8.

Brendel, A. B., Mirbabaie, M., Lembcke, T. B., & Hofeditz, L. 2021. Ethical Management of Artificial Intelligence. Sustainability, 13(4), 1974. https://doi.org/10.3390/su13041974

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.

Bullock, Justin B. 2019. 'Artificial Intelligence, Discretion, and Bureaucracy.' *The American Review of Public Administration* 49 (7): 751–61. https://doi.org/10.1177/0275074019856123.

Bullock, J., M. Young, and Y. F. Wang. 2020. "Artificial intelligence, bureaucratic form, and discretion in public service". Information Polity 25(4): 491-506.

Burrell, Jenna. 2016. 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms.' *Big Data & Society* 3 (1): 205395171562251. https://doi.org/10.1177/2053951715622512.

Busuioc, M. 2020. Accountable Artificial Intelligence: Holding Algorithms to Account. Public Administration Review. https://doi.org/10.1111/puar.13293

Cairns, G. and Wright, G., 2018. Scenario Thinking. Preparing your organization for the future in an unpredictable world. Palgrave Macmillan, Basingstoke, UK.

Casino, Fran, Thomas K. Dasaklis, and Constantinos Patsakis. 2019. 'A Systematic Literature Review of Blockchain-Based Applications: Current Status, Classification and Open Issues.' *Telematics and Informatics* 36 (March): 55–81. https://doi.org/10.1016/j.tele.2018.11.006.

Chander, A. 2017. The racist algorithm? Michigan Law Review, 115(6), 1023-1045.

Chua, H. N., Herbland, A., Wong, S. F., & Chang, Y. 2017. Compliance to personal data protection principles: A study of how organizations frame privacy policy notices. Telematics and Informatics, 34(4), 157-170. https://doi.org/10.1016/j.tele.2017.01.008

Chun, A.H.W. 2007. 'Using AI for E-Government Automatic Assessment of Immigration Application Forms.' In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*. Vancouver.

Clark, William Roberts, and Matt Golder. 2015. 'Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?' *PS: Political Science & Politics* 48 (01): 65–70. https://doi.org/10.1017/S1049096514001759.

Clarke, Amanda, and Helen Margetts. 2014. 'Governments and Citizens Getting to Know Each Other? Open, Closed, and Big Data in Public Management Reform.' *Policy & Internet* 6 (4): 393–417. https://doi.org/10.1002/1944-2866.POI377.

Collier, Matt, Richard Fu, Lucy Yin, and Philip Christiansen. 2017. 'Artificial Intelligence: Healthcare's New Nervous System.'

Cook, Thomas D. 2014. ''Big Data' in research on social policy.' *Journal of Policy Analysis and Management* 33 (2): 544–47. https://doi.org/10.1002/pam.21751.

Corea, Francesco. 2019. *An Introduction to Data*. Vol. 50. Studies in Big Data. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-04468-8.

Cortés-Cediel, María E., Iván Cantador, and Olga Gil. 2017. 'Recommender Systems for E-Governance in Smart Cities.' In *Proceedings of the International Workshop on Recommender Systems for Citizens*, 1–6. New York, NY, USA: ACM. https://doi.org/10.1145/3127325.3128331.

Daniell, Katherine A., Alec Morton, and David Ríos Insua. 2016. 'Policy Analysis and Policy Analytics.' *Annals of Operations Research* 236 (1): 1–13. https://doi.org/10.1007/s10479-015-1902-9.

De Vries, Alex. 2018. 'Bitcoin's Growing Energy Problem.' *Joule* 2 (5): 801–5. https://doi.org/10.1016/j.joule.2018.04.016.

Decker, Paul T. 2014. 'Presidential Address: False Choices, Policy Framing, and the Promise of 'Big Data.'' *Journal of Policy Analysis and Management* 33 (2): 252–62. https://doi.org/10.1002/pam.21755.

Demchenko, Yuri, Cees de Laat, and Peter Membrey. 2014. 'Defining Architecture Components of the Big Data Ecosystem.' In *2014 International Conference on Collaboration Technologies and Systems (CTS)*, 104–12. IEEE. https://doi.org/10.1109/CTS.2014.6867550.

Desouza, Kevin C., and Akshay Bhagwatwar. 2012. 'Leveraging Technologies in Public Agencies: The Case of the U.S. Census Bureau and the 2010 Census.' *Public Administration Review* 72 (4): 605–14. https://doi.org/10.1111/j.1540-6210.2012.02592.x.

Dwivedi, Yogesh K., Laurie Hughes, Elvira Ismagilova, Gert Aarts, Crispin Coombs, Tom Crick, Yanqing Duan, *et al.* 2021. 'Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy.' *International Journal of Information Management* 57 (April): 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

Dutch Data Protection Authority (Autoriteit Persoonsgegevens) 2020. 'Belastingdienst/Toeslagen: De verwerking van de nationaliteit van aanvragers kinderopvangtoeslag'. Available at: https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoek_belastingdienst_kinderopvangtoeslag.pdf
Dutch Parliamentary Investigation Commission 2020. 'Ongekend Onrecht'. Available at: https://www.tweedekamer.nl/kamerleden_en_commissies/commissies/pok

Einaste, Taavi. 2018. 'Blockchain and Healthcare: The Estonian Experience.' E-Estonia. 2018. https://e-estonia.com/blockchain-healthcare-estonian-experience/#:~:text=Estonia%2C home to one of,healthcare on a national scale.&text='We are using blockchain as,the integrity of health records.

Einav, Liran, and Jonathan Levin. 2014. 'The Data Revolution and Economic Analysis.' *Innovation Policy and the Economy* 14 (January): 1–24. https://doi.org/10.1086/674019.

Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. 2020. Explainable artificial intelligence and machine learning: A reality rooted perspective. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(6), e1368. https://doi.org/10.1002/widm.1368

Engvall, T., and L.S. Flak. 2022. Digital Governance as a Scientific Concept. In Scientific Foundations of Digital Governance and Transformation. Public Administration and Information Technology edited by Charalabidis, Y., Flak, L.S., Viale Pereira, G., vol 38. Springer, Cham. https://doi.org/10.1007/978-3-030-92945-9_2

European Commission (EC). 2021. 'EBSI.' Accessed May 28, 2021. https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/EBSI.

European Commission (EC). 2020. 'Shaping Europe's Digital Future: Commission Presents Strategies for Data and Artificial Intelligence.' 2020.

Exmeyer, P. C., and J. L. Hall. 2022. High Time for a Higher-Level look at High Technology: Plotting a Course for Managing Government Information in an Age of Governance. Public Administration Review. https://doi.org/10.1111/puar.13513

Freeman, K. 2016. Algorithmic injustice: How the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis. North Carolina Journal of Law & Technology, 18(5), 75.

Gamage, Pandula. 2016. 'New Development: Leveraging 'Big Data' Analytics in the Public Sector.' *Public Money & Management* 36 (5): 385–90. https://doi.org/10.1080/09540962.2016.1194087.

Gascó, M., 2017. Living labs: Implementing open innovation in the public sector. Government Information Quarterly 34: 90-89. https://doi.org/10.1016/j.giq.2016.09.003

George, Gerard, Martine R. Haas, and Alex Pentland. 2014. 'Big Data and Management.' *Academy of Management Journal* 57 (2): 321–26. https://doi.org/10.5465/amj.2014.4002.

Giest, S., & Grimmelikhuijsen, S 2020. Introduction to special issue algorithmic transparency in government: Towards a multi-level perspective. Information Polity, (Preprint), 1-9. DOI: 10.3233/IP-200010

Goldsmith, Stephen, and Susan Crawford. 2014. *The Responsive City: Engaging Communities Through Data-Smart Governance*. Jossey-Bass.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT press.

Grimmelikhuijsen, S. G., & Meijer, A. J. 2014. Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment. Journal of Public Administration Research and Theory, 24(1), 137-157. https://doi.org/10.1093/jopart/mus048

Gunning, D., & Aha, D. 2019. DARPA's explainable artificial intelligence (XAI) program. AI Magazine, 40(2), 44-58. https://doi.org/10.1609/aimag.v40i2.2850

Hadden, Susan G. 1986. 'Intelligent Advisory Systems for Managing and Disseminating Information.' *Public Administration Review* 46 (November): 572. https://doi.org/10.2307/975579.

Haenlein, Michael, and Andreas Kaplan. 2019. 'A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence.' *California Management Review* 61 (4): 5–14. https://doi.org/10.1177/0008125619864925.

Hardin, R. 1996. Trustworthiness. Ethics, 107(1), 26-42.

Hassani, Hossein, Xu Huang, and Emmanuel Sirimal Silva. 2019. *Fusing Big Data, Blockchain and Cryptocurrency*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-31391-3.

Hemken, Terry, and Chris Gray. 2016. 'Smart Move: Technologies Make Their Mark on Public Service.'. Accessed at: https://www.slideshare.net/accenture/smart-move-emergent-technologies-make-their-mark-on-public-service

Hempel, Jessi. 2018. 'How Refugees Are Helping Create Blockchain's Brand New World.' *Wired*, 2018.

Hevner, A.R. (2007). 'A Three Cycle View of Design Science Research'. *Scandinavian Journal of Information Systems,* 19, 4.

High-level Expert Group on AI 2019. Ethics guidelines for trustworthy AI. Retrieved on 02-06-2021 from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Höchtl, Johann, Peter Parycek, and Ralph Schöllhammer. 2016. 'Big Data in the Policy Cycle: Policy Decision Making in the Digital Era.' *Journal of Organizational Computing and Electronic Commerce* 26 (1–2): 147–69. https://doi.org/10.1080/10919392.2015.1125187.

Hondula, David M., Evan R. Kuras, Justin Longo, and Erik W. Johnston. 2018. 'Toward Precision Governance: Infusing Data into Public Management of Environmental Hazards.' *Public Management Review* 20 (5): 746–65. https://doi.org/10.1080/14719037.2017.1320043.

Hughes, D.L., N.P. Rana, and Y.K. Dwivedi. 2020. “Elucidation of IS project success factors: an interpretive structural modeling approach”. Annals of Operations Research 285: 35–66. DOI: https://doi.org/10.1007/s10479-019-03146-w

Hurley, Michael W., and William A. Wallace. 1986. ‘Expert Systems as Decision Aids for Public Managers: An Assessment of the Technology and Prototyping as a Design Strategy.’ *Public Administration Review* 46 (November): 563. https://doi.org/10.2307/975578.

Isaak, J., & Hanna, M. J. 2018. User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer, 51(8), 56-59. DOI: 10.1109/MC.2018.3191268

Issa, Hussein, Ting Sun, and Miklos A. Vasarhelyi. 2016. ‘Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation.’ *Journal of Emerging Technologies in Accounting* 13 (2): 1–20. https://doi.org/10.2308/jeta-10511.

Jahoda, Marie. 1986. ‘Artificial Intelligence: An Outsider’s Perspective.’ *Science and Public Policy*, December. https://doi.org/10.1093/spp/13.6.333.

Janssen, Davine. 2020. ‘Power Grid Operators Launch Blockchain for Home and Car Batteries.’ *Euractiv*, May 1, 2020. https://www.euractiv.com/section/energy/news/power-grid-operators-launch-blockchain-for-home-and-car-batteries/.

Janssen, Marijn, Vishanth Weerakkody, Elvira Ismagilova, Uthayasankar Sivarajah, and Zahir Irani. 2020. ‘A Framework for Analysing Blockchain Technology Adoption: Integrating Institutional, Market and Technical Factors.’ *International Journal of Information Management* 50: 302–9. https://doi.org/10.1016/j.ijinfomgt.2019.08.012.

Janssen, M., S. Luthra, S. Mangla, N. P. Rana, and Y. K. Dwivedi. 2019. “Challenges for adopting and implementing IoT in smart cities”. Internet Research 29(6): 1589-1616.

Janssen, M., N. P. Rana, E. L. Slade, and Y. K. Dwivedi. 2018. “Trustworthiness of digital government services: deriving a comprehensive theory through interpretive structural modeling”, Public Management Review 20(5): 647-671. DOI: 10.1080/14719037.2017.1305689

Janssen, Marijn, and George Kuk. 2016. ‘The Challenges and Limits of Big Data Algorithms in Technocratic Governance.’ *Government Information Quarterly* 33 (3): 371–77. https://doi.org/10.1016/j.giq.2016.08.011.

Janssen, Marijn, and Jeroen van den Hoven. 2015. ‘Big and Open Linked Data (BOLD) in Government: A Challenge to Transparency and Privacy?’ *Government Information Quarterly* 32 (4): 363–68. https://doi.org/10.1016/j.giq.2015.11.007.

Jarmin, Ron S., and Amy B. O’Hara. 2016. ‘Big Data and the Transformation of Public Policy Analysis.’ *Journal of Policy Analysis and Management* 35 (3): 715–21. https://doi.org/10.1002/pam.21925.

Jefferies, Duncan. 2016. ‘The Automated City: Do We Still Need Humans to Run Public Services?’ *Guardian*, September 20, 2016.

Johnes, Geraint, and John Ruggiero. 2017. ‘Revenue Efficiency in Higher Education Institutions under Imperfect Competition.’ *Public Policy and Administration* 32 (4): 282–95. https://doi.org/10.1177/0952076716652935.

Karafiloski, Elena, and Anastas Mishev. 2017. 'Blockchain Solutions for Big Data Challenges: A Literature Review.' In *IEEE EUROCON 2017 -17th International Conference on Smart Technologies*, 763–68. IEEE. https://doi.org/10.1109/EUROCON.2017.8011213.

Kemper, J., & Kolkman, D. 2019. Transparent to whom? No algorithmic accountability without a critical audience. Information, Communication & Society, 22(14), 2081-2096. https://doi.org/10.1080/1369118X.2018.1477967

Kimble, Chris, and Giannis Milolidakis. 2015. 'Big Data and Business Intelligence: Debunking the Myths.' *Global Business and Organizational Excellence* 35 (1): 23–34. https://doi.org/10.1002/joe.21642.

Kitchin, Rob, and Gavin McArdle. 2016. 'What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets.' *Big Data & Society* 3 (1): 205395171663113. https://doi.org/10.1177/2053951716631130.

Koerner, C. L. 2014. Media, fear, and nuclear energy: A case study. The Social Science Journal, 51(2), 240-249. https://doi.org/10.1016/j.soscij.2013.07.011

Kolkman, D. 2020. The (in) credibility of algorithmic models to non-experts. Information, Communication & Society, 1-17. https://doi.org/10.1080/1369118X.2020.1761860

Kouziokas, Georgios N., A. Chatzigeorgiou, and K. Perakis. 2017. 'Artificial Intelligence and Regression Analysis in Predicting Ground Water Levels in Public Administration.' *European Water* 57: 361–66.

Kumar, Deepa S, and M. Abdul Rahman. 2017. 'Simplified HDFS Architecture with Blockchain Distribution of Metadata.' *International Journal of Applied Engineering Research* 12 (21): 11374–82.

La Fors, K. 2020. Legal Remedies for a Forgiving Society: Children's rights, data protection rights and the value of forgiveness in AI-mediated risk profiling of children by Dutch authorities. Computer Law & Security Review, 38, 105430. https://doi.org/10.1016/j.clsr.2020.105430

Lal, R., and A. Haleem. 2009. "A Structural Modelling for E-Governance Service Delivery in Rural India." International Journal of Electronic Governance 2(1): 3–21. doi:10.1504/IJEG.2009.024962.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabasi, Devon Brewer, Nicholas Christakis. 2009. 'Social Science: Computational Social Science.' *Science* 323 (5915): 721–23. https://doi.org/10.1126/science.1167742.

Levin, D. Z., Whitener, E. M., & Cross, R. 2006. Perceived trustworthiness of knowledge sources: The moderating impact of relationship length. Journal of applied psychology, 91(5), 1163. https://psycnet.apa.org/doi/10.1037/0021-9010.91.5.1163

Liu, Paul Tak Shing. 2016. 'Medical Record System Using Blockchain, Big Data and Tokenization. In: Lam KY., Chi CH., Qing S. (eds) Information and Communications Security. ICICS 2016. Lecture Notes in Computer Science, vol 9977. Springer, Cham. https://doi.org/10.1007/978-3-319-50011-9_20

López, Antonio Calleja. 2019. 'Building Alternatives: The Digital Democracy and Data Commons Pilot.' DECODE Project. 2019.

Maciejewski, Mariusz. 2017. 'To Do More, Better, Faster and More Cheaply: Using Big Data in Public Administration.' *International Review of Administrative Sciences* 83 (1): 120–35. https://doi.org/10.1177/0020852316640058.

Masuch, Michael, and Perry LaPotin. 1989. 'Beyond Garbage Cans: An AI Model of Organizational Choice.' *Administrative Science Quarterly* 34 (1): 38. https://doi.org/10.2307/2392985.

McAfee, Andrew, and Erik Brynjolfsson. 2017. *Machine, Platform, Crowd: Harnessing Our Digital Future*. W. W. Norton & Company.

McConaghy, Trent. 2016. 'How Blockchains Could Transform Artificial Intelligence.' Dataconomy. Accessed at: http://dataconomy.com/2016/12/blockchains-for-artificial-intelligence/.

McDermott, Y. 2017. Conceptualising the right to data protection in an era of Big Data. Big Data & Society, 4(1). https://doi.org/10.1177%2F2053951716686994

Mehr, Hila. 2017. 'Artificial Intelligence for Citizen Services and Government.' Harvard Kennedy School, Ash Center for Democractic Governance and Innovation.

Meijer, A., Lorenz, L., and M. Wessels. 2021. "Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems". Public Administration Review, 81(5). https://doi.org/10.1111/puar.13391

Mergel, Ines. 2017. 'Building Holistic Evidence for Social Media Impact.' *Public Administration Review* 77 (4): 489–95. https://doi.org/10.1111/puar.12780.

Mergel, Ines, R. Karl Rethemeyer, and Kimberley Isett. 2016. 'Big Data in Public Affairs.' *Public Administration Review* 76 (6): 928–37. https://doi.org/10.1111/puar.12625.

Meuwese, A. 2020. Regulating Algorithmic Decision-making One Case at the Time: A Note on the Dutch 'SyRI' Judgment. European review of digital administration & law, 1(1), 209-212.

Millard, Jeremy. 2018. 'Open Governance Systems: Doing More with More.' *Government Information Quarterly* 35 (4): S77–87. https://doi.org/10.1016/j.giq.2015.08.003.

Mittelstadt, B. 2019. Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501-507.

Molenveld, A., Verhoest, K., Voets, J., & Steen, T. 2020. Images of coordination: How implementing organizations perceive coordination arrangements. Public Administration Review, 80(1), 9-22. https://doi.org/10.1111/puar.13136

OECD. 2021. Tax Administration 2021: Comparative Information on OECD and other Advanced and Emerging Economies. Paris: OECD Publishing. DOI: https://doi.org/10.1787/cef472b9-en.

OECD. 2019. *Enhancing Access to and Sharing of Data*. OECD. https://doi.org/10.1787/276aaca8-en.

OECD. 2017. *OECD Digital Economy Outlook 2017*. OECD. https://doi.org/10.1787/9789264276284-en.

Pandey, Sheela, Sanjay K. Pandey, and Larry Miller. 2017. 'Measuring Innovativeness of Public Organizations: Using Natural Language Processing Techniques in Computer-Aided Textual Analysis.' *International Public Management Journal* 20 (1): 78–107. https://doi.org/10.1080/10967494.2016.1143424.

Pencheva, Irina, Marc Esteve, and Slava Jankin Mikhaylov. 2020. 'Big Data and AI – A Transformational Shift for Government: So, What next for Research?' *Public Policy and Administration* 35 (1): 24–44. https://doi.org/10.1177/0952076718780537.

Pirog, Maureen A. 2014. 'Data Will Drive Innovation in Public Policy and Management Research in the Next Decade.' *Journal of Policy Analysis and Management* 33 (2): 537–43. https://doi.org/10.1002/pam.21752.

Power, Daniel J. 2016. ''Big Brother' Can Watch Us.' *Journal of Decision Systems* 25 (1): 578–88. https://doi.org/10.1080/12460125.2016.1187420.

Rechtspraak (2020), SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens, available at: https://www.rechtspraak.nl/Organisatie-en contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-wetgeving-in-strijd-met-het-Europees-Verdrag-voor-de-Rechten-voor-de-Mens.aspx

Rice, P. L. and Ezzy, D., 1999. Qualitative research methods: A health focus. Oxford University Press, Oxford.

Russell, Stuart, and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach*. 3th ed. Pearson.

Salamon, Lester M. 2002. *The Tools of Government: A Guide to the New Governance*. Oxford University Press.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. 2016. Automation, algorithms, and politics| when the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software. International Journal of Communication, 10, 19.

Schein, E. H. 1990. Organizational culture. American Psychologist, 45(2), 109–119. https://doi.org/10.1037/0003-066X.45.2.109

Scherer, Matthew U. 2015. 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies.' *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2609777.

Schintler, Laurie A., and Rajendra Kulkarni. 2014. 'Big Data for Policy Analysis: The Good, The Bad, and The Ugly.' *Review of Policy Research* 31 (4): 343–48. https://doi.org/10.1111/ropr.12079.

Schwaber, Ken, and Jeff Sutherland. 2017. *The Scrum Guide: The Definitive The Rules of the Game*.

Shindelar, Stacey. 2014. 'Big Data and the Government Agency.' The Public Manager. 2014. https://www.td.org/magazines/the-public-manager/big-data-and-the-government-agency.

Stilgoe, J. 2018. Machine learning, social learning and the governance of self-driving cars. Social studies of science, 48(1), 25-56. https://doi.org/10.1177%2F0306312717741687

Stritch, Justin M., Mogens Jin Pedersen, and Gabel Taggart. 2017. 'The Opportunities and Limitations of Using Mechanical Turk (MTURK) in Public Administration and Management Scholarship.' *International Public Management Journal* 20 (3): 489–511. https://doi.org/10.1080/10967494.2016.1276493.

Sun, T. Q., and R. Medaglia. 2019. "Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence from Public Healthcare." *Government Information Quarterly* 36 (2): 368–83. https://doi.org/https://doi.org/10.1016/j.giq.2018.09.008.

Taeihagh, Araz. 2017. 'Crowdsourcing: A New Tool for Policy-Making?' *Policy Sciences* 50 (4): 629–47. https://doi.org/10.1007/s11077-017-9303-3.

Tan, E., and J. Crompvoets. 2022. *The new digital era governance: How new digital technologies are shaping public governance.* Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-930-5

Tan, E., S. Mahula, and J. Crompvoets. 2022. "Blockchain Governance in the Public Sector: A Conceptual Framework for Public Management." *Government Information Quarterly* 39 (1): 101625. https://doi.org/https://doi.org/10.1016/j.giq.2021.101625.

Tangi, L., M. Janssen, M. Benedetti, and G. Noci. 2021. "Digital government transformation: A structural equation modeling analysis of driving and impeding factors". *International Journal of Information Management* 60, 102356. https://doi.org/10.1016/j.ijinfomgt.2021.102356

Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Alfred A. Knopf.

Vogl, T. M., C. Seidelin, B. Ganesh, and J. Bright. 2020. "Smart technology and the emergence of algorithmic bureaucracy: Artificial intelligence in UK local authorities". *Public Administration Review* 80(6): 946-961. https://doi.org/10.1111/puar.13286

Voshmgir, Shermin. 2020. *Token Economy: How the Web3 reinvents the Internet*. 2nd Edition. Token Kitchen: Berlin.

Wang, G., Xie, S., and X. Li. 2022. "Artificial intelligence, types of decisions, and street-level bureaucrats: Evidence from a survey experiment". *Public Management Review*, 1–23. https://doi.org/10.1080/14719037.2022.2070243

Wang, Shuai, Wenwen Ding, Juanjuan Li, Yong Yuan, Liwei Ouyang, and Fei-Yue Wang. 2019. 'Decentralized Autonomous Organizations: Concept, Model, and Applications.' *IEEE Transactions on Computational Social Systems* 6 (5): 870–78. https://doi.org/10.1109/TCSS.2019.2938190.

Warfield, J. N. 1974. "Developing interconnection matrices in structural modeling". *IEEE Transactions on Systems, Man, & Cybernetics* (1): 81-87.

Willems, J., Schmid, M. J., Vanderelst, D., Vogel, D., and F. Ebinger. 2022. AI-driven public services and the privacy paradox: do citizens really care about their privacy? *Public Management Review*, 1–19. https://doi.org/10.1080/14719037.2022.2063934

Wilson, C., and H. Broomfield. 2022. Learning how to do AI: managing organizational boundaries in an intergovernmental learning forum. *Public Management Review*, 1–20. https://doi.org/10.1080/14719037.2022.2055119

Winfield, A. 2019. Ethical standards in robotics and AI. Nature Electronics, 2(2), 46-48.

Wirtz, Bernd W., Jan C. Weyerer, and Carolin Geyer. 2019. 'Artificial Intelligence and the Public Sector—Applications and Challenges.' *International Journal of Public Administration* 42 (7): 596–615. https://doi.org/10.1080/01900692.2018.1498103.

Yafimava, Darya. 2019. 'Blockchain In the Supply Chain: 10 Real-Life Use Cases and Examples.' Openledger. Accessed at: https://openledger.info/insights/blockchain-in-the-supply-chain-use-cases-examples/

Zheng, Yongqing, Han Yu, Lizhen Cui, Chunyan Miao, Cyril Leung, and Qiang Yang. 2018. 'SmartHS: An AI Platform for Improving Government Service Provision.' In *AAAI Conference on Artificial Intelligence Thirty-Second AAAI Conference on Artificial Intelligence*.

Zhu, Jiangnan, Huang Huang, and Dong Zhang. 2019. ''Big Tigers, Big Data': Learning Social Reactions to China's Anticorruption Campaign through Online Feedback.' *Public Administration Review* 79 (4): 500–513. https://doi.org/10.1111/puar.12866.

Zikopoulos, Paul C., Chris Eaton, Dirk DeRoos, Thomas Deutsch, and George Lapis. 2012. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw Hill.

Zwetkoff, C., Elsen, C., Vigneron, L. and Pardo, J., 2018. L'utilisateur au coeur du processus d'innovation. Livre blanc méthodologique. WeLL Living Lab, Liege, Belgium, 86 pp.

ANNEXES

Table A. Interview List

Position	Name and type of organization
Data Miner	FPS Finance (Federal)
Service Manager (Data Warehouse)	FPS Finance (Federal)
Head of Data Miners	FPS Finance (Federal)
e-Auditor	FPS Finance (Federal)
Data Miner	FPS Finance (Federal)
Fiscal Coordinator	FPS Finance (Federal)
Regional Director	FPS Finance (Federal)
Operation Manager	FPS Finance (Federal)
Advisor - Administration générale Expertise et Support stratégiques | Centre des connaissances	FPS Finance (Federal)
Research Consultant	Smals (non-profit private org.)
Research Consultant	Smals (non-profit private org.)
Research Consultant	Smals (non-profit private org.)
Administrator	Crossroad Bank for Social Security
Vice-administrator	Crossroad Bank for Social Security
Deputy general administrator	INAMI (federal)
Head of Data Management	INAMI (federal)
Chief Data Officer	INAMI (federal)
Deputy general administrator	INAMI (federal)
Social Inspector	INAMI (federal)
Director of the Social Inspection Direction	INAMI (federal)
General Administrator	ONEM (federal)
IT researcher in data mining	ONSS (federal)
Data scientist	ONSS (federal)
Director service taxation and valuation	VLABEL (regional)
Employee competence centre reporting	VLABEL (regional)
Data protection officer	A Flemish agency in the social affairs domain, participated on the condition of anonymizing both name and organization (regional)
General advisor	Agoria (private-profit)
Professor	UCL/ICTEAM (research institute)
Director	HEC Digital Lab (research institute)
Assistant Professor	UMons (research institute)
Advisor	FGTB (union)
Program Manager	NRB (PPP)
Partner	PwC (private-profit)
Partner	PwC (private-profit)
General Manager	Febetra (business federation)
Director Social Department	Confederatie Bouw (business federation)
Advisor	Confederatie Bouw (business federation)
General Director	Constructiv (service provider org.)

Former Secretary-General	MC (mutuality)
Head of Study Department	MC (mutuality)
Advisor	MC (mutuality)
Information & Technology Manager	AdN (regional)
Expert	AdN (regional)
Head of Study Department	UCM (employer org.)
Secretary-General	UCM (employer org.)
Social Affairs Advisor	UCM (employer org.)
Social Security Advisor	UCM (employer org.)
Associate Professor	HEC (research inst.)
Customer Solution Manager	SAS (private-profit)
Head of Task Department	FEB (employer org.)
Social Security Advisor	FEB (employer org.)
Data Analyst	Ecolo (political party)
Digital Advisor	Centre Jean Gol (political party)
Tax Advisor	Centre Jean Gol (political party)
Secretary-General	CNCD 11 11 11 (NGO)
Head of Study Department	CSC (union)
Advisor	IEV (political party)
Head of Study Department	PTB (political party)
National Technology Officer	Microsoft (private-profit)
Secretary-General	UPTR (federation of companies)
CEO	Skwarel (start-up)
Director Task & Public Affairs	Febelfin (federation of companies)
Professor	VUB AI Lab (research inst.)
President	BATL (Professional association)
Advisor	BATL (Professional association)
Analysis department director	CTIF (federal)

Table B. Codification list

Constructs	Elements	Topics
Regulations	Data	GDPR, national laws, data combination, data anonymization (difficulty of doing, and viable use of it)
	Taxation/Social Security	national laws, EU laws, sanctions
	Transcending laws	environmental laws
	Justification of decisions	the challenge of explainability with deep learning
Trust	Trust in administration	tax authorities, inability to develop in-house solutions, regulative agencies, trust in centralized entity vs the necessity of blockchain

	Trust in society	no adherence to tax, easy acceptance of new technologies (not critical thinking), lack of skills regarding new technologies
	Trust in technology	Cloud systems, social acceptability, algorithms, blockchain, political distrust, digitalization improves trusts, social distrust in new digital technologies (related to Fake news, conspiracy theories, anti-science speeches),
	Trust in system	Surveilance capitalism, undeclared purposes (e.g. commercial, political)
	Trust in tech providers/private sector	implementation of ethical standards, GAFAM (big tech), higher efficiency and expertise than public sector, not adapted to local challenges (better in-house), distrust to private sector in civil servants, distrust to banks to share data
Management/operational systems	Guidelines	Legal guidelines, a precise framework to declare fraud suspicion (for banks, casinos etc.)
	Rules & standards	data confidentiality, ethical rules and standards, lack of standards to assess overall quality of taxation system (e.g. ISO)
	Principles	only-once, trade-offs (weak/strong AI vs ethics, digitalization vs reduction in human resources, efficiency vs transparency, cost reduction vs quality of services, decentralization of data vs difficulty of analysis, privacy vs the cost of more advanced cytopgraphic solutions, cost of investment vs return of investment, openness vs not sharing too much information about business), lack of flexibility in project requirements
	Processes	risk assessment, human in/on the loop, safeguards against AI bias, AI-led data management, streamlining, process automation, reducing administrative burden, administrative simplification (e.g. public procurement), automation of rights, elimination of unnecessary tasks, reporting obligations for companies, treating big companies same as small companies, follow-up after the detection of tax fraud, data sorting, data protection, very long specific investigations vs agile way of working (thanks to data insights), data matching and cross-checking across institutions, purpose delienation, making proportionate cross-checks, administrative burden on authorization of projects (delays timely detection of new types of fraud), public registry for used algorithms (e.g. Amsterdam algorithm register)
	Strategies	data strategy, 5-10 year digital investment plans, a central data management political vision, a common data collection and exploitation strategy, horizontal control, use of AI and data on most problematic cases, big fish vs small fish, incentives to entrepreneurs, pre-

		filling of tax returns, specifities of SMEs, improving targeting of controls (focus on fraudsters leave others alone), early detection of frauds, creation of technology watchgroups in administrations (to improve anticipation), subscription-based IT systems, new business models for media to inform society, operational strategy on data literacy, starting small with a few tenants on big data and AI solutions and incremental increase in time with trusted tenants
Perceived risk	Legal challenge	protection of citizen rights, implications to tax lawyers, rules on data confidentiality, evolution of identity
	Control of data	tech vs government
	Democratic challenge	autoritarian drift, development of a public GAFAM to preserve democracy and equality, avoiding Chinese system
	Administrative challenges	disrupting administrative practices and culture, the need of reinventing itself, difficult to keep public services relevant, budget cuts and personal reduction will human inspectors will not have time to verify AI's decisions
	Societal challenges	job loss, decrease in working time, disruption in labour market (rapidly changing tech skills), environmental impact, increasing digital divide
Governance system	Data governance	sensitive data, reliability of databases (e.g. updating, info about foreign workers), secure networks to exchange data with mutualiteit, collaboration between data centres, exchanges between databases (e.g. social security data+financial data), siloed organizations, automatic collection of transaction data at transaction level, data sharing with banks, access and control of data, inclusivity, frequency of checks on data quality, internal audit processes (e.g. data anonymization)
	Open governance	open data, open sources, better control for experienced inviduals and groups, giving citizens access, reduced control and access to software, access to data by the private sector, use of open data to better understand the tool, user friendliness of the interface
	Multi-level governance	global data platform for tax authorities, international coordination at the EU level, social dumping, collaboration between Belgian and foreign supervisory bodies, administrative fragmentation (local-regional-federal), institutional complexity, EU as a normative actor (e.g. green deal, free movement of workers), EU as a coordinator
	Network governance	institutional rivalry, coordination between services, collaboration with companies and citizens, expertise centres, collaboration with other public actors (police, social security, justice), understanding the challenges of entrepreneurs, need for listening many actors

		(including private actors and NGOs), control of compliance, finding innovation champions
Technical infrastructure	Security	data security, network security, data privacy
	Quality of database	social security data about foreign workers, up-to-dateness of data
	Data collection & analysis	Use of AI with other tools (e.g. IOT), blockchain/DLT, data hubs (across sector, across countries), automation, predictive analysis, setting algorithms properly (parameter choices, big vs small fish), data mining, sensitivity of data, classification of data, probabilistic vs deterministic models, open data, big data platform
	Softwares	subscription-based IT systems, supply model of software, open vs proprietary softwares, entanglement with obsolete legacy software
	Computer maturity	
	Reliance/dependence on external actors	GAFAM, SAS, in-house solution vs outsourcing
Public values	Appropriateness of technology	why we use AI, avoiding discrimination, asking ethical/moral questions, avoid nervousness of an association with “Big Brother”.
	Respecting privacy	profiling fraudsters
	Tax fairness	
Technological maturity	Bias and noise	risk of bias with AI algorithm, misinterpretation, possible errors in probabilistic AI, systemic biases due to algorithm, bias in training data, using too many algorithms
	Technology convergence	blockchain and AI, AI and IoT, machine learning for risk assessment of machine learning, BCT as a means of transparency and trust
	Blockchain/DLT	not reached to maturity, disruptive potential, difficult to find good use cases, GDPR is an obstacle
	AI/ machine learning	absence of AI use cases that are effective as current data-matching operations
	Fraud detection technologies	fraud analytics, predictive models, proactive/anticipatory use of ML, tools for better targeting and sorting data, nowcasting tools
Interoperability	Technical interoperability	lack of harmonization of IT systems in administration, compatibility of systems and means of data exchange between administrations
	Semantic interoperability	homogenization of data, lack of standardization to ensure data quality, cross-border differences in the use of data and metadata (e.g. easier to do with some countries (Netherlands) than others)
	Organizational interoperability	coordination between services, a harmonized social security database at the European level, aligning data

		strategies of multiple actors involved in social security, different applications used by insurance companies (a need for a common application for coordination with the RIZIV)
	Regulative interoperability	data exchange between EU countries about workers, EU as a coordinator (e.g. European Labour Authority)
Capacities, skills and competencies	Resources	human resources (not enough manpower to process data, growing need to IT skills), financial resources, lack of resources at SMEs for digital solutions/sharing faster and better data with tax authorities
	Digital skills	knowledge of algorithms and BCT, data management, critical thinking, data literacy, data storage and sharing, lack of digital skills in society, lack of digital skills in unions, lack of digital skills for new technologies in companies/SMEs
	Training	problem of initial training of civil servants, lack of competence and poor training in tax administration, technological developments, foresight on technological changes, access to appropriate IT training, rapid change of technological skills, development of civic and professional skills in public, inclusivity of training (including migrants), training citizens about data protection, information campaigns for citizens and public agents
Policy priorities	EU-level policy priorities	Competion with US, and Asian tech providers (EU wants to lead in producing data standards because it missed out digital transition), Green deal, Digital innovation
	Fight against fraud	Higher priority of figthing tax fraud against social fraud, targeting small business vs big business, too much time and money in fighting social fraud less effectiveness, ineffectiveness of tax policies against large players, finding a balance between fight against fraud and protection of individual liberties, more targeted and proportional measures against tax fraud, use of data and AI to select more problematic cases, government wants total transparency on assets for tax justice, but going other way, the need for a clear vision and consistency in targeting big fraudsters in specific sectors, impression of hunt for money than fraud
	Political support	Need for political support for data sharing, budget allocation, short-termism
	Geo-political aspects	Rare metals and potential conflicts, impact of technology on environment
Perceived usefulness	Automation	to free up some working time for other tasks, improves efficiency, elimination of unnecessary tasks, reduces administrative costs and pushes for simplification, it saves money and brain power that can be used to better look at data and make better decision

	Social security and taxation	for the public sector, for enterprises, for the funding of social security, citizens do not need to go administration, gain speed and efficiency to target and control problematic companies; better target controls, data matching helps to thwart the creativity of fraudsters and allows to detect low-level social fraud that is generally difficult to identify, not only detecting undue payments but ensuring people receive everything they need on time
	Data collection & analysis	the opportunity to automate and reduce the administrative burden while improving services by anticipating problems (e.g. in mutualiteits), solving problems not being able to solve before; anticipating future problems, understanding better the reasons behind succesful practices; better results with fewer and better qualified people, making more effective policies; new IT tools accelerate data analysis while reducing the risk of committing mistakes
	Past experiences	CBSS improved results on fight against social fraud, pre-filling of tax returns , usefulness of relational business data (SQL) rather than big data for fraud detection (ONSS), PoCs of predictive analysis with anonymised data were useful for inspectors, and that there was an enormous gain in efficiency and accuracy in terms of case selection and they wielded larger results.
	Indirect added value of new digital technologies	Better predictive medicine impacts social security, fighting against unfair competition, making public authorities legimate and relevant
Socio-cultural elements	Digital culture	Lack of digital and data culture in administration, lack of digital culture in society, different digital mindset between Wallonia and Flanders (data protection vs innovation with SMEs)
	Digital divide	Generational gap, socio-economic gap, gap between SMEs, digitalization of society, access to Internet,
	Willigness to share data	