



Deliverable 3.3

Design artefact and evaluation criteria for testing

31-03-2022

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

Funding Programme: BRAIN-be 2.0 Call 2019

Contract Number: B2/191/P3/DIGI4FED

Work Package: 3

Author(s): Dr. Evrim Tan, Dr. Mathias Sabbe, Prof.dr. Catherine Fallon

This communication reflects only the author's view and BELSPO is not responsible for any use that may be made of the information it contains.

Copyright © 2020 by the DIGI4FED consortium

The DIGI4FED consortium consists of the following partners:

KU Leuven (KUL)

UNamur (UN)

UAntwerpen (UA)

ULiège (UL)

Executive Summary

This deliverable presents the design artefact - a governance model for the use of new digital technologies in the fight against fraud in the social security and taxation domain- and the evaluation method and criteria to be used on assessing the design artefact. The deliverable is divided into two parts. The first part highlights the key findings from previous research gathered through interviews, experiments, and living labs, and based on those findings presents an initial governance design for Belgian federal government to be tested. The second part elaborates on the evaluation criteria to assess the model and the methodology (Delphi survey) to be employed on testing the model according to the evaluation criteria.

TABLE OF CONTENTS

<i>List of Figures & Tables</i>	<i>4</i>
<i>List of Acronyms</i>	<i>5</i>
<i>1 Introduction.....</i>	<i>6</i>
<i>2. Input for the design artefact.....</i>	<i>7</i>
<i>2.1. Objectives of the design artefact</i>	<i>7</i>
<i>2.2. Findings from interview data, experiments, and living lab</i>	<i>8</i>
2.2.1. Findings from interview data	8
2.2.2. Findings from living lab	12
2.2.3. Findings from experiments	15
<i>3. Development of the design artefact</i>	<i>17</i>
<i>3.1. Contingencies</i>	<i>17</i>
<i>3.2. Data prosumers</i>	<i>18</i>
<i>3.3. Data governance goals</i>	<i>19</i>
<i>3.4. Design values</i>	<i>20</i>
<i>3.5. Governance of</i>	<i>21</i>
<i>3.6. Governance by.....</i>	<i>22</i>
<i>3.7. the key features of the design artefact.....</i>	<i>23</i>
<i>4. Selection of Criteria for model evaluation.....</i>	<i>29</i>
<i>4.1. Evaluation criteria in the design science research literature.....</i>	<i>29</i>
<i>4.2. Evaluation criteria in other research fields</i>	<i>36</i>
<i>4.3. Evaluation criteria for our design artefact.....</i>	<i>43</i>
<i>4.4. Summary – Evaluation criteria for testing</i>	<i>44</i>
<i>5. Model testing - Designing the Delphi survey.....</i>	<i>45</i>
<i>5.1. The Delphi method.....</i>	<i>45</i>
<i>5.2. A few cautionary points</i>	<i>47</i>
<i>5.3. Designing the Delphi survey</i>	<i>48</i>
<i>5.4. The Survey participants</i>	<i>52</i>
<i>5.5. The survey questionnaire</i>	<i>54</i>
<i>5.6. Survey analysis & model testing</i>	<i>62</i>

Conclusion64

References.....65

LIST OF FIGURES & TABLES

Figure 1. The 6-step model for OGD governance design7

Figure 2. ISM Model11

Figure 3. MICMAC diagram12

Figure 4. Hevner’s three cycle view of design science research29

Figure 5. Categories of evaluation criteria (Pratt et al., 2014)35

Figure 6. Hierarchy of evaluation criteria (Hevner et al., 2018)36

Table 1. Key drivers in the technology adoption8

Table 2. Challenges from Living Lab13

Table 3. Solutions from Living Lab15

Table 4. Evaluation criteria in the design science research literature30

Table 5. Grouping the most common criteria according to their similarities34

Table 6. Evaluation criteria in other research fields37

Table 7. Grouping criteria according to their similarities41

Table 8. Evaluation criteria for our design artefact43

Table 9. Anticipated tension points48

Table 10. Correspondence between the 6-step model, the survey dimensions, and the evaluation criteria50

Table 11. Survey planning52

Table 12. Survey dimensions, questions and corresponding evaluation criteria62

Table 13. Evaluation criteria, corresponding questions, and testing method63

LIST OF ACRONYMS

AI: Artificial intelligence

BCT: Blockchain technology

BOSA: Federal Public Service Policy and Support

CSI: Comité de Sécurité de l'information

DIGI4FED: Digital (R)evolution in Belgian Federal Government: An Open Governance Ecosystem for Big Data, Artificial Intelligence, and Blockchain.

DPA: Data protection authority

DLT: Distributed ledger technology

DPO: Data protection officer

EBSI: European Blockchain Services Infrastructure

EC: European Commission

ESSIF: European Self Sovereign Identity Framework

EU HLEG: European Union high level expert group

FPS: Federal Public Service

FRM: Final reachability matrix

GDPR: General Data Protection Regulation

IoT: Internet of Things

IRM: Initial reachability matrix

ISM: Interpretive structural modelling

IT: Information Technologies

MICMAC: Matrices d'Impacts Croises Multiplication Appliqué a un Classement

PKI: Public key infrastructure

OGD: Open government data

SSIM: Structural Self-Interaction Matrix

WP: Work package

1 INTRODUCTION

The present deliverable is part of the work package three (WP3). Its objectives are threefold. First, through this deliverable, we aim to present a design artefact that our team of researchers developed based on the results of the previous research steps of the DIGI4FED project. Taking stock of the insights gained through interview data, living lab workshops, and experiments, this design artefact embodies a first version of our governance model for the integration of new technologies (such as AI and BCT) in Belgian federal public policies.

The second objective of this deliverable is to justify the evaluation criteria that will be used to test our design artefact. For that purpose, a short literature review was conducted to identify the most common criteria used in the design science research literature as well as in the public administration and public policy evaluation literatures. Based on the review results and the specificities of our design artefact, this document presents a short list of criteria that will be used to test the model.

Finally, the third objective of this deliverable is to explain the methods that will be used to test the design artefact. More specifically, this document provides details regarding the survey method that will be used to collect the data with the purpose of testing our proposed model. We specify how this survey will be implemented, with which participants, and how the collected data will be analyzed in testing the proposed model.

The design of the deliverable is as follows. In section 2 we present the input that served as the groundwork on which we developed our design artefact. Such input includes interview data, results from the living-lab workshops, and vignette experiments. In section 3, we introduce the design artefact for our OGD governance model. The proposed model is presented in accordance with the 6-step model for OGD governance design. Section 4 presents the literature review and the selection process for the evaluation criteria that will be used to test our proposed governance model. The last section of this deliverable (section 5) presents our methods for testing the design artefact. More specifically this section presents the Delphi method, the construction of our survey questionnaire, as well as our methods for data analysis and testing the proposed model.

2. INPUT FOR THE DESIGN ARTEFACT

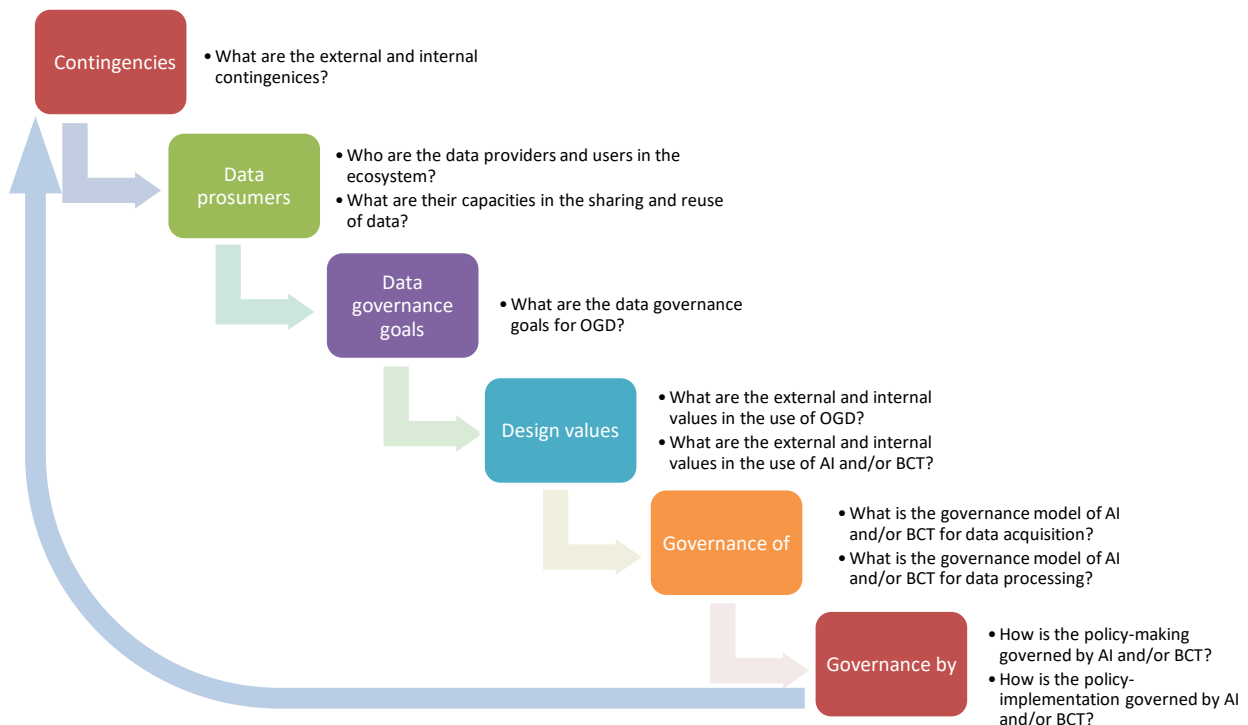
2.1. OBJECTIVES OF THE DESIGN ARTEFACT

The overarching objective of DIGI4FED project is to develop and test an OGD governance model for Belgian federal government in fighting against tax and social security fraud. This objective along with findings from previous deliverables from WP1, WP2, and WP3, will set the basis on the development of the design artefact. Based on the previous findings, the following characteristics are identified for the design artefact:

- It will be compatible with big data, AI, and blockchain solutions
- It will platform-based
- It should address the identified challenges in Living Lab (WP1), experiments (WP2), individual interviews with key stakeholders (WP1)
- It should comply with solutions proposed in Living Lab (WP1)
- It should comply with the findings from interview data and experiments on technology adoption in administration and society (see Section 2.2.).
- It should comply with the institutional framework in taxation and social security domains

In the development of the design artefact, we will follow the heuristics suggested by the 6-step model in OGD governance design (see D.3.2.1). Figure 1 shows the steps to follow on the development of the design artefact. The abovementioned input will be used to address the questions for each step of the model.

Figure 1. The 6-step model for OGD governance design



2.2. FINDINGS FROM INTERVIEW DATA, EXPERIMENTS, AND LIVING LAB

In this section, we will share the findings from the analysis of the interviews with key experts (WP1), the Living Lab (WP1), and experiments with citizen panels (WP2). The details of each of these research inputs are already published in previous deliverables, or in the process of being published as a journal article. Therefore, in this section we will just highlight the main findings from each research.

2.2.1. FINDINGS FROM INTERVIEW DATA

The 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 Belgium¹. 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 interpretive structural modelling to explore the underlying relationships among identified drivers. The adoption models have provided the conceptual framework to categorise the identified themes and subthemes into key drivers. Table 1 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 Appendix.

Table 1. Key drivers in the technology adoption

DRIVERS	CODES	CONSTRUCTS	ELEMENTS	DEFINITIONS
PERFORMANCE EXPECTANCY	1	Technological maturity	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Legal challenge Control of data Democratic challenge Administrative challenges Societal challenges 	This variable captures the perception of stakeholders about the risk of using new digital technologies in the fight against fraud.

¹ The list of interviewees and their organisations can be found in the Appendix.

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

A thematic analysis across interview data has revealed 13 variables 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 in the interrelationship among these drivers. Understanding how changes in certain variables influence changes in other variables is 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 matrix (SSIM). In developing SSIM, there are four possible ways of

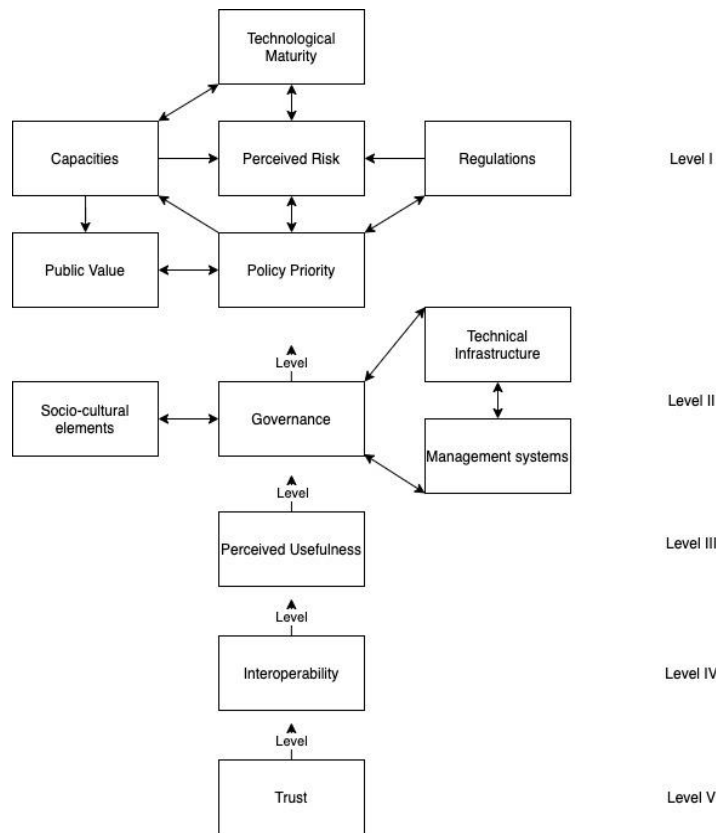
relating variables to each other, that are represented by 'V', 'A', 'X', and 'O' symbols (Hughes et al. 2020). The symbols are to be interpreted as: V = Variable i influences variable j; A = Variable j influences variable i; X = Both variables i and j are influenced by each other; O = Variables i and 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, $R(P_i)$, 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, $A(P_i)$ consists of the variable itself and other elements that might help in achieving it. The variables for which the interaction of these sets, $R(P_i) \cap A(P_i) = R(P_i)$, 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 diagram.

- Level 1 is the highest level in the hierarchy of ISM diagram and contains the following variables: technological maturity, capacities, skills and competencies, perceived risk, public value, policy priorities, regulation. This means that ...
- Level 2 is the second level in the hierarchy of ISM diagram 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 diagram and contains only perceived usefulness.
- Level 4 is the fourth level in the hierarchy of ISM diagram and contains only interoperability.
- Level 5 is the lowest level in the hierarchy of ISM diagram and contains only trust.

Figure 2. ISM Model

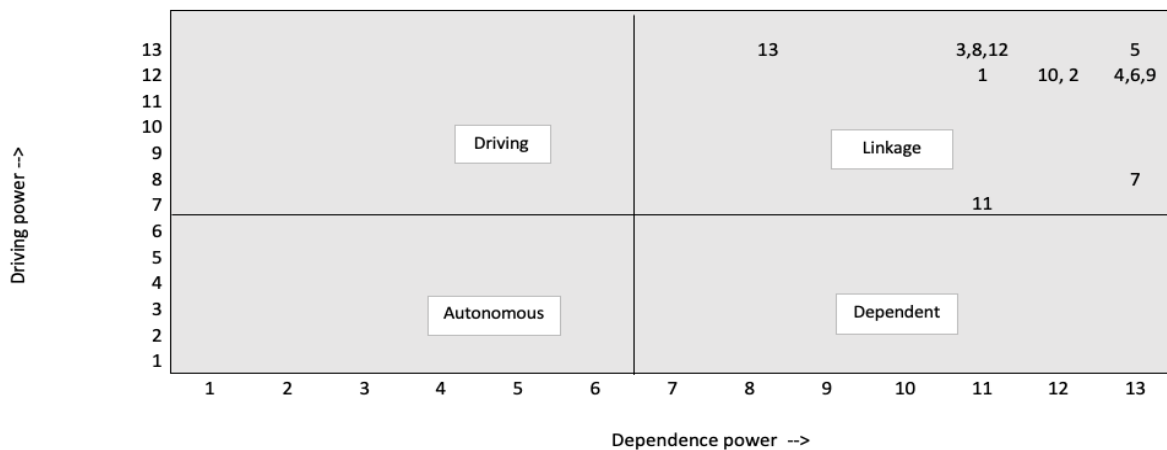


Based on the level partitioning, and the direction of relationships identified in SSIM, we have developed the ISM model for the adoption of new digital technologies in the fight against fraud in taxation and social security systems. The ISM model in Figure 2 shows the direct and indirect relationships among variables. Through the analysis of the relationships, policymakers can develop policy strategies to support digital transformation of administration. 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.

Before completing the ISM analysis, we visually categorized the driving and dependency powers among the variables through a MICMAC diagram. The MICMAC diagram 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 3. MICMAC diagram



The MICMAC diagram in Figure 3 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.

2.2.2. FINDINGS FROM LIVING LAB

During the co-creation phase of the living lab, 18 participants from the key stakeholder organizations in the taxation and social security domains in Belgium discussed the main challenges and proposed solutions concerning the integration of big data, AI and BCT in federal public organizations. The details of these activities

can be found in deliverable D.1.3 Activity report. Below we share the main highlights from the challenges and proposed solutions attained from the living lab co-creation phase.

Table 2. Challenges from Living Lab

Topics	Challenges	Illustrative cues
Topic 1: Citizens' acceptability of data exploitation	Challenge 1: How to generate an acceptance of the new anti-fraud tools?	<ul style="list-style-type: none"> Control tools, but what about the added value for citizens and companies when, for example, the "only once" principle is not respected? There are many fears and negative perceptions among the population that are sometimes reflected in the media
	Challenge 2: How to preserve citizens' trust regarding data use and analysis?	<ul style="list-style-type: none"> Some citizens have a lack of trust in public authorities A lack of transparency is often associated with mistrust Trust depends on citizens' ability to control their data (which is not always the case with issues related to the digital divide)
	Challenge 3: How to balance effectiveness in the fight against fraud with the need for transparency in the analysis process?	<ul style="list-style-type: none"> The adaptability of fraudsters conflicts with the need for transparency
	Challenge 4: How to ensure the explainability of data analysis processes?	<ul style="list-style-type: none"> One of the central issues is to ensure explicability and avoid "black boxes" Many actors support probabilistic applications of AI (thus going beyond "rule-based" algorithms)
Topic 2: Articulation of data use with regulations	Challenge 5: How to address the interpretation and implementation issues of the GDPR?	<ul style="list-style-type: none"> There are divergent interpretations by DPOs: difficulties of coordination and exchange between administrations There is no precise legal framework for the secondary use of data The role of the DPA is questioned by different actors

	<p>Challenge 6: How to articulate the complexity of legislations with the adoption of new IT tools in the fight against fraud?</p>	<ul style="list-style-type: none"> • A thorough revision of specific legislation (tax and social security) would be necessary beforehand to facilitate the integration of digital tools • Legislations sometimes lack flexibility in relation to the digital framework (use of data, public procurement, etc.)
<p>Topic 3: Resources and data expertise within administrations</p>	<p>Challenge 7: How to address recruitment issues and the lack of internal expertise (data & AI)?</p>	<ul style="list-style-type: none"> • Many public organizations have a limited ability to develop new systems "in house" • There are some difficulties in interpreting data (lack of a classification culture) and using AI (limited maturity of technology in public organizations)
	<p>Challenge 8: How to limit dependence on external (private) IT systems providers?</p>	<ul style="list-style-type: none"> • Little public sector control over the development and use of these systems by private actors • Risks related to the access to sensitive data by external providers
	<p>Challenge 9: How to organize a digital governance in a context where financial resources are lacking?</p>	<ul style="list-style-type: none"> • Choosing private software that are less expensive but problematic in terms of confidentiality (e.g. Google Analytics) • Delegation of data collection to intermediary actors (health insurers, unions)
	<p>Challenge 10: How to address the lack of coordination between administrations?</p>	<ul style="list-style-type: none"> • Loss of energy and resources to develop similar ICT projects separately • Difficulties in ensuring the interoperability of the databases of different administrations

Table 3. Solutions from Living Lab

Solutions	Solution to challenge n°
SOLUTION 1 - Pooling resources, expertise, and capabilities	7; 9; 10
SOLUTION 2 - Re-thinking laws and regulations by actively integrating IT tools and data use in reflections	2; 3; 4; 6
SOLUTION 3 - Privileging and supporting EU (or national) external IT contractors	5; 8; 9
SOLUTION 4 - Developing new technical solutions for a fairer use of data	2; 3; 4; 8; 9
SOLUTION 5 - Making communicational and pedagogical efforts to justify policies and societal choices	1; 2; 4
SOLUTION 6 - Humanizing interactions with citizens	1; 2
SOLUTION 7 - Re-thinking public procurement processes and requirements	5; 6; 8
SOLUTION 8 - Adopting common evaluation criteria and quality standards (e.g., ISO certification) in administrations	2; 10
SOLUTION 9 - Formalizing and streamlining DPOs' decision-making processes	5; 10
SOLUTION 10 - Promoting training and maintaining IT fitness in the public sector	4; 7; 8
SOLUTION 11 - Redefining and clarifying the role of certain FPSs (e.g., BOSA)	10

2.2.3. FINDINGS FROM EXPERIMENTS

The third set of findings for the design artefact stems from the survey experiments conducted with citizen panels. The survey experiments test whether measures taken to enhance the trustworthiness in governmental AI projects will increase perceived trustworthiness, policy support and behavioral trust, or whether such attitudes are better explained by pre-existing perceptions and characteristics of citizens. Moreover, the experiments tested these factors among the public by fielding two survey experiments with information on hypothetical AI projects in the Belgian federal government, with intervention groups receiving additional sections of information on ways in which the federal government could make such projects more trustworthy. The details of the survey experiments will be shared in an upcoming deliverable by WP2. In this subsection, we will share the main findings from the experiments.

Interventions focused on different measures to enhance trustworthiness as used in practice by several interviewed governments and as proposed in the data science, computer science and/or legal literatures. More specifically, interventions tested were information on 1) emphasizing legal compliance, 2) applying the no-harm principle and ensuring the explainability of AI predictions, 3) the use of anonymized or internal data-gathering, 4) retaining a human in the loop, 5) ensuring fairness and non-discrimination and 6) enhancing technical robustness of AI-projects. Preliminary results for all (combinations of) interventions provide little evidence that presenting information on trustworthiness to lay, non-user citizens enhances perceived trustworthiness as

compared to the control group. Instead, perceived trustworthiness is predicted by prior attitudes and personal characteristics.

Seven hypotheses are tested. These were:

H1: Measures taken to enhance the trustworthiness of an AI project will lead to higher perceived trustworthiness, trust and policy support

H2: Measures taken to enhance the trustworthiness of an AI project will not affect perceived trustworthiness, trust and policy support

H3: Perceived trustworthiness of government positively affects trust in and support for specific governmental AI projects.

H4: Perceived trustworthiness of AI positively affects trust in and support for specific governmental AI projects.

H5: Privacy concern negatively affects trust in and support for specific governmental AI projects.

H6: Perceptions that one is a member of a group that is discriminated against negatively affect trust in and support for specific governmental AI projects.

H7: Familiarity with AI positively affects trust in and support for specific governmental AI projects.

The results suggest that influencing perceived trustworthiness of the AI projects, policy support and behavioural trust (in the form of citizens being willing to provide their data to governmental AI projects) by providing information on trustworthiness is exceedingly difficult. Hence the efficacy of communication strategies on governmental AI-projects may be limited. Even where single models showed minor support for hypothesis 1 (expecting that such measures could improve citizen attitudes towards governmental AI), R2 values remained very low, F-tests for the entire model remained non-significant and results were not robust when incorporating control variables. What is more, there was some support for the opposite hypothesis, namely that some interventions (in particular, legal information) would have no effect on the public's attitudes. Instead, perceptions of specific governmental AI projects seem to be formed by pre-existing attitudes, perceptions, and characteristics. Consistent support was found across both experiments for hypotheses 3, 4 and 5, expecting positive effects from general perceived trustworthiness of government and general perceived trustworthiness of AI, as well as negative effects of privacy concern.

The implications of these findings are important. First, they suggest that the difficulty of managing pre-existing perceptions is an important oversight in the current public administration, data science and AI ethics literature, as well as in recently introduced policy frameworks on trustworthy or ethical AI (e.g., the EU HLEG's trustworthy AI guidelines). Contributions and policy reports in these areas continue to point to AI and project design as the most crucial factor in determining trustworthiness, implicitly assuming that actual trustworthiness will translate to perceived trustworthiness, trust and – in turn – policy support (e.g., Gunning & Aha, 2019; Floridi et al., 2018). The results for hypotheses 1-5 together not only show that such assumptions may be overly optimistic, but also that stable negative attitudes towards governmental AI projects are prevalent among relatively large portions of our respondents. This implies that policymakers and data scientists who develop governmental AI projects may not be able to 'win' a major portion of the public's trust through project design and should instead consider negative citizen attitudes to be a strategic risk to their projects. Such an approach could be more focused on the

potential trust breaches that applying an AI for a certain public task may have and the proportionality of relying on a controversial application of AI (e.g. risk-profiling citizens or facial recognition in public spaces).

Support for hypotheses 6 and 7 was less consistent, but still points to the potential relevance of perceived discrimination and professional interactions with AI. Discrimination may of particular value to incorporate in future research, as groups that feel systematically mistreated by society (and its governments) may be less inclined to accept certain applications of AI in government (Kääriäinen & Niemi, 2014). If this is indeed the case, this would produce an additional risk for specific groups in society, as any suspicion of or confirmed AI bias towards these groups may produce disproportionate legitimacy issues.

3. DEVELOPMENT OF THE DESIGN ARTEFACT

In this section, we will highlight the key components of the design artefact. In the development of the design artefact, we will follow the steps described in Figure 1. The design choices in each step of the 6-step model will determine the choices in subsequent steps. In what follows, we introduce the key drivers in the selection of the design elements, and then we translate them into tangible policy objectives and solutions. At the end of the section, we summarize the key characteristics of the proposed design artefact for the OGD governance model in the fight against fraud in social security and taxation.

3.1. CONTINGENCIES

The first set of decisions in the development of the design artefact are related to the contingency conditions. There are two types of contingencies: external and internal contingencies. External contingencies refer to the environmental context of a public sector organization that impedes the available design choices. Internal contingencies are the more immediate and direct causes of the characteristics of the platform. Hence, the first step in OGD governance design is the identification of external contingencies that may impede the available design choices in platform governance, and accordingly assessing how external contingencies may affect the existing strategies, internal and external rules, and structures for the platform governance.

The research input detailed in previous section suggest the following contingency conditions:

Contingencies	Research input	Source
External contingency	Policy intervention should start from trust dimensions	ISM
External contingency	Protection of personal data is crucial for trust	Experiments
External contingency	Some citizens do not trust administrations	Living Lab
Internal contingency	Influencing perceived trustworthiness of the AI projects, policy support and behavioral trust (in the form of citizens being willing to provide their data to governmental AI projects) by providing information on trustworthiness is exceedingly difficult.	Experiments

Internal contingency	Technical robustness in the form of blockchain technology can create trust among citizens towards data analytics and AI projects.	Experiments
Internal contingency	A lack of transparency is associated with mistrust	Living Lab
Internal contingency	Trust depends on citizen's abilities to control their data	Living Lab

Based on these contingency conditions, we can have the following design objectives:

- The platform system needs to prioritize transparency in the usage of data, citizens need to be able to control their data, the system needs to ensure the protection of personal data, and preferably citizens can self-govern their data.

Contingencies	Research input	Source
External contingency	Interoperability is the second element to assess in policy interventions.	ISM
External contingency	Loss of energy in developing similar ICT projects separately	Living Lab
External contingency	Difficulties in ensuring the interoperability of the databases of different administrations	Living Lab
Internal contingency	Adopting common evaluation criteria quality standards such as ISO certifications in public administration are needed for OGD governance	Living Lab
Internal contingency	The roles of FPSs need to be redefined. Specifically, the role of BOSA needs to be redefined and clarified in the OGD governance	Living Lab

Based on these contingency conditions, we can have the following design objectives:

- The platform system needs to adopt common technical standards that comply with international standards,
- The platform needs to be compatible for all administrations in the OGD ecosystem.
- The platform is to be led likely by BOSA

3.2. DATA PROSUMERS

The second set of decisions in the development of the design artefact are related to the identification of data prosumers. Data prosumers are the actors that supply the platform with raw and/or processed data and reuse the data through the platform for public, commercial, social, or academic purposes. In a platform ecosystem, the data prosumers can be public or private, and they could be communities or individual users that can access

and reuse the data available in the platform. This step is not only about identifying data providers (and beneficiaries of OGD) but also about assessing the existing capacities among prosumers, and about cultivating the data for the growth of the platform ecosystem.

As such, the following research input can be used for the selection of the data prosumers:

Research input	Source
'Data prosumers' is related to the Level II factors in ISM model. Level II corresponds to the relations between socio-cultural dimensions, governance, technical infrastructure, and management systems.	ISM
There are risks related to the access to sensitive data by external providers	Living Lab
Privileging and supporting EU/national external contractors could be a solution to limited 'in-house' abilities to develop new IT systems	Living Lab
There are information asymmetries between government and citizens about data analytics and AI projects	Interviews

Based on these inputs, we can have the following design objectives on identifying data prosumers.

- The system infrastructure needs to be developed by the public sector or controlled by it.
- The system needs to allow public sector organizations to control the access of external providers to sensitive data.
- The system needs to allow external users (e.g., European level) and commercial users to use platform data without undermining privacy concerns.

3.3. DATA GOVERNANCE GOALS

Data governance goals aim to maximize the value of OGD for the users in the platform ecosystem. The goals therefore need to align the public policy objectives and goals on the (re)use of OGD with the governance design principles. Data governance goals include decisions about the purpose of data use, desirable behaviors, appropriate governance mechanisms, aligning business goals, and the overall interactions between data subjects, owners, and managers.

The following research input is related to the selection of the data governance goals.

Research input	Source
Data governance goals are related to the Level II factors in ISM model. Level II corresponds to the relations between socio-cultural dimensions, governance, technical infrastructure, and management systems.	ISM
Perceived usefulness affects the choices at Level II (ISM).	ISM
The adaptability of fraudsters conflicts with the need for transparency – there is a need to balance effectiveness in the fight against fraud with transparency	Living Lab

Little public sector control over the development and use of the systems (e.g. software) provided by private actors	Living Lab
Many actors support probabilistic applications of AI (thus going beyond "rule-based" algorithms)	Living Lab
There is a need for rethinking public procurement processes and requirements	Living Lab
Fishbowl transparency on fairness has its limits due to bounded rationality of citizens	Experiments
Many citizens are wary of how, where, when and for what purpose their data is used	Experiments

Based on these inputs, we can have the following design objectives on setting data governance goals:

- The system needs to be able to use predictive technologies (based on machine-learning) yet be able to provide transparency and traceability about how technology is used for decisions.
- The system needs to provide incentives to private contractors to support and develop the system infrastructure without holding the data infrastructure or keeping monopoly on the software/applications.
- The system needs to be compatible with the digital divide, digital cultural differences, and willingness to share data of public.

3.4. DESIGN VALUES

Design values stem from the internal and external value propositions associated with the OGD governance. Internal values represent the value of OGD to a particular organization based on organizational, technical and managerial investment requirements and expectations. External values stem from the expectations from the platform users and beneficiaries on the (re)use of OGD in public policy processes and service provisions. An effective governance design process needs to identify the value propositions associated with the way OGD is used and align them with data governance goals, in an order of importance.

The following research input is rendered significant for the identification of design values:

Research input	Source
Design values are public values related to AI/blockchain solutions and the use of open data. Design values are related to Level I considerations.	ISM
Design values in the fight against fraud concern with the appropriateness of technology, ensuring tax fairness, and respecting privacy	Interviews
Design values are directly influenced by capacity conditions and policy priorities, and they directly influence policy priorities	ISM
The citizens should control tools in data governance, but the system should also respect 'only-once' principle	Living Lab
There are many fears and negative perceptions among the population about new anti-fraud technologies	Living Lab
There are some difficulties in interpreting data (e.g., lack of a classification culture) and using AI (e.g., limited maturity of technology in public organizations)	Living Lab

Choosing private software is less expensive but problematic in terms of confidentiality (e.g., Google Analytics)	Living Lab
The lack of financial resources pushes administrations to delegate data collection to intermediary actors (e.g. ,health insurers, unions, banks, etc.)- they have less direct control over which data is collected, how is it collected and processed.	Living Lab
Pooling resources, expertise, and capabilities could be a solution to capacity problems	Living Lab
Making communicational and pedagogical efforts to justify choices, promoting training and maintaining IT fitness in the public sector could be a solution to the reservations on the use of anti-fraud technologies	Living Lab

These inputs can be interpreted into the following design objectives on identifying the design values:

- The platform needs to allow easy access (only-once) to users to get information about fraud decisions, while preserving the privacy of individual data.
- The platform needs to provide better tools for the public sector organizations to ensure tax fairness and privacy protection without creating additional financial and administrative strains.
- The platform needs to allow outside solutions to be developed and integrated without undermining public sector control over data collection and analysis processes.
- How the system operates, what type of data is used, and the access/limitations of stakeholder organizations about data should be communicated unequivocally to citizens and other organizations.

3.5. GOVERNANCE OF

Governance of decisions concerns with the use of digital technologies (AI, BCT) in data collection and processing. AI and BCT have different governance attributions in the system architecture choices, with different value propositions and trade-offs influencing the architecture design of platform ecosystem. Additionally, considerations pertaining to technology readiness and maturity, associated organizational and human resource capacities, and design principles (e.g. explainability, transparency, etc.) are also important for the governance of the technology and its inclusion in the system architecture design.

The following research input is related to the ‘governance of’ decisions.

Research input	Source
The interrelationships between technological maturity, capacity conditions, and perceived risk expected to influence governance of decisions	ISM
One of the central issues is to ensure explainability and avoid "black boxes"	Living Lab
There is absence of AI use cases they are as effective as current data matching operations	Interviews
Blockchain did not reach the maturity, and it is difficult to find good use cases	Interviews
Some technological concerns include risks of bias with AI algorithm, misinterpretation, possible errors in probabilistic AI, systemic biases due to algorithm, bias in training data, using too many algorithms	Interviews

There is no robust evidence for measures to enhance the trustworthiness of AI projects (e.g. XAI) in government will positively affect perceived trustworthiness, policy support and behavioral trust	Experiments
Developing new technical solutions for a fairer use of data could address some perceived risks with fraud detection technologies	Living Lab

These inputs can be interpreted into the following design objectives on ‘governance of’ decisions:

- The platform needs to support probabilistic applications of AI but should be careful with black box systems.
- The platform needs to provide a secure environment to test the stability of AI/BCT solutions in data collection and analysis.
- Advanced fraud detection technologies need to gradually be introduced in the platform to test the effectiveness of new technological solutions.

3.6. GOVERNANCE BY

Governance by decisions concerns about the impact of digital technologies (AI, BCT) in policy making and implementation. The techno-social power dynamics among platform prosumers, regulatory processes in the implementation of technology in public policy processes, and the control of the platform functions by the automated and human agents must be considered in this last step.

The following research input is related to the ‘governance by’ decisions.

Research input	Source
The interrelationships between regulations, policy priorities, and perceived risks expected to influence governance by decisions	ISM
There are divergent interpretations by DPOs: difficulties of coordination and exchange between administrations	Living Lab
There is no precise legal framework for the secondary use of data	Living Lab
The role and ability of the Belgian DPA is questioned by different actors due to conflict of interest	Living Lab
A thorough revision of specific legislation (tax and social security) would be necessary beforehand to facilitate the integration of digital tools	Living Lab
Legislations sometimes lack flexibility in relation to the digital framework (use of data, public procurement, etc.)	Living Lab
Re-thinking laws and regulations by actively integrating IT tools and data use in reflections could be a solution for the shortcomings with existing regulations	Living Lab
Formalizing and streamlining DPOs' decision-making processes could be a solution for the interpretation and implementation issues of the GDPR	Living Lab

These inputs can be interpreted into the following design objectives on ‘governance by’ decisions:

- There needs to be clear regulative standards and rules in interpreting the GDPR issues in platform governance.
- The role of DPOs/DPA needs to be clearly stated and formalized.
- The national legislations on taxation/social security should be compatible with the use of digital technologies (including secondary use of data).
- There need to be flexibility in relation to the digital framework.

3.7. THE KEY FEATURES OF THE DESIGN ARTEFACT

The previous section highlighted the design objectives of an OGD platform based on the research input. In this section, we will highlight the key features of the design artefact in compliance with the design objectives. Below we list the key features of the model and discuss how the proposed feature of the governance model complies with the design objectives. Under each point, we will argue the expected benefits and potential issues that may arise that may potentially hinder the ability of the model to yield the expected benefits.

1. The OGD governance model should help fostering trust among prosumers with a citizen-centric approach by:

- Improving citizens' ability to control and manage their personal data. One solution could be the widespread implementation of a digital wallet² system for citizens. These digital wallets would give individuals the ability to control and manage public and private actors' access to their personal data. Such systems, which are often based on the principle of self-sovereign identity, are attracting some interest from public authorities. However, there is no strategic consultation between the different levels of government (federal, regional, and local) at this stage.
- Favoring a decentralized approach to data management that promotes transparency and protection of personal data. Wallet solutions can be implemented in centralized, hybrid, or decentralized data verification infrastructures. Findings from experiments suggest that large portions of our respondents have low level of trust towards governmental AI projects and citizen-centric approaches in data management can engender more trust among users. Therefore, we expect the digital wallet solutions to be better suitable in either hybrid models that introduce SSI features (e.g., verifiable credentials) on the current PKI or in decentralized models that rely on blockchain solutions.

Although this will not be directly tested in the current model, it appears that the BCT and SSI solutions that are being developed at the European level (EBSI and ESSIF) can constitute interesting options³ to support hybrid and decentralized models from a technical, operational, regulative, and interoperability perspectives. As such, EBSI and ESSIF already offer more developed use cases in comparison to other SSI initiatives and, since these initiatives stem from the EU, they are being developed in compliance with the EU ethical and regulative standards regarding the use of personal data. Furthermore, these solutions are devised and are specifically developed as a public sector oriented European infrastructure. In the end, the proposed platform should enable citizens to control what data is shared and with which public administration. It

² This is further supported by the fact that Belgium has set the strategic objective to issue digital wallets to every Belgian citizen by 2023. The development of an EBSI compliant digital wallet and platform would likely be supported by the federal government and the EU.

³ Other SSI alternatives exist, such as the SOLID initiative developed by Tim Berners-Lee and with which the Flemish government recently started a collaboration.

should enhance the overall transparency regarding data use and provide citizens with the ability to better track and govern their data in a secure way.

Some uncertainties should be taken into consideration at this point. Nowadays, some of these proposed solutions are at the cutting edge of innovation and are still under active development. Although developments are swift (policy developments, technical progress, new use cases, etc.), this still raises up several questions regarding the socio-political and technical maturity that are required for this approach to yield its anticipated benefits in terms of trust and endorsement by system users. Besides technical maturity, one can also interrogate our socio-political readiness to the adoption of such a novel platform. For instance, to what extent can we expect some resistance to the proposed solutions within the administration? This resistance can arise for various reasons such as disagreements regarding technical aspects of the proposed model (are decentralized technology solutions the way to go?) or administrative culture resistance. On another level, we might also wonder if governments – including regional partners – will be able to reach a political consensus regarding the adoption of such a platform and how much time would policy development take? Overall, the uncertainties raised by technical maturity and the potential lack of socio-political readiness are susceptible to undermine prosumers' trust and general endorsement of the proposed OGD platform.

2. The OGD governance model would probably improve the GDPR compliance of new digital technologies in the fight against fraud and facilitate the compatibility of national legislations in relation to the digital framework by:

- Promoting the development of a legislation that is more compatible with the digital framework. In addition to legal experts, the involvement of other experts (such as computer scientists) in the reflection process could help ensure that legal instruments are compatible with the new digital framework. The use of a regulatory sandbox mechanism could also help to develop and test different regulatory frameworks for the use of new technologies in the fight against fraud. However, there are still some unknowns as to the viability of this experimental approach to law.
- Simplifying the task of privacy actors by adopting solutions and tools developed at the European level. To strengthen the compatibility of the system with national legislation and the GDPR, it would be possible to rely on certain tools developed in compliance with EU ethical and regulatory standards (such as ESSIF and EBSI) for the construction of a data exchange platform with decentralized features. Such an approach could facilitate the interpretation of rules and decision-making by DPOs and other privacy actors (DPA, CSI). This could also contribute to the development of a national legislation on taxation and social security that is more compatible with the use of new technologies.

This approach should also help in providing clear regulative standards and rules in interpreting the GDPR issues. In particular, the fact that the proposed OGD governance platform relies on GDPR compliant solutions should help in clarifying and formalizing the role of data controllers (i.e., DPOs, DPA) according to the GDPR rules. Similarly, the GDPR compliancy of the platform should likely improve legislative flexibility in relation to the digital framework, thus facilitating the development of national legislations on taxation and social security that are compatible with the use of digital technologies. In that regard, the use of a regulatory sandbox approach might further help to develop and test different regulatory frameworks for the use of new digital technologies and automated tools in the fight against fraud.

There are still some uncertainties pertaining to the technological and technical features offered by the proposed solutions. Although these solutions offer encouraging perspectives in addressing concrete GDPR

issues, the legal assessment regarding the GDPR compliance of some solutions developed at the EU level is still pending. Furthermore, the proposed design artefact does not guarantee an improved flexibility and compatibility of national legislations in relation to the digital framework. Although a regulatory sandbox approach might help in developing an adapted regulatory framework compatible with use of new digital technologies, there are still many uncertainties regarding the design and methodology of experimental law-making.

3. The proposed OGD governance platform should also address the concerns regarding public sector's current dependence on external IT providers while preserving its resources by:

- Promoting the internalization of digital skills through the development and technical management of a data exchange platform. One possibility could be to internalize technical skills by assigning a federal service organization (e.g., FPS BOSA or a new entity along the lines of Smals) with the task of developing and managing a common data exchange platform across administrations. This would strengthen internal skills while providing a solution that meets high standards in terms of transparency and data protection.
- Promoting the adoption of national or European solutions that meet strict data protection requirements. Favoring data analysis or storage solutions developed by national (development of a Belgian cloud, for example) or European (e.g., EBSI, Single digital gateway, etc.) actors could help maintaining administrations' fraud-fighting capabilities while reducing their dependence on private solutions that are problematic in terms of personal data protection.
- Allowing public authorities to take on the role of key regulators in data governance. Public managers could be granted the ability to control the access to personal data stored on the common data exchange platform mentioned earlier. They could then establish specific rules for actors who wish to access the data for commercial or public purposes. They could also choose to share some specific data with European counterparts to strengthen transnational anti-fraud capabilities for example.
- Stimulating private actors to provide administrations with efficient tools without creating additional financial constraints. The public managers of the data exchange platform could decide to authorize the use of some carefully anonymized personal data to private actors who wish to develop new predictive tools using this data as training data. The newly developed tools could then be integrated by administrations, thus reinforcing their fraud detection capabilities without requiring additional resources the immediate future.

The design artefact gives a central role to the public sector in leading the adoption and management of the new system infrastructure. More specifically, an organization such as BOSA or another entity inspired by the Smals could be expected to play a leading role in the development of the platform and its subsequent management. This organization could also be expected to develop and test the platform as use case in specific domains such as the fight against tax fraud. To assist the governance of the OGD platform, this organization would benefit from the active collaboration with other actors from the public and the private sector. Overall, the proposed system should allow public sector organizations to control the access of external IT providers to sensitive data. In fact, external users⁴ (i.e., administrations from other member states) and commercial users should be able to access the platform without undermining data privacy concerns. When accessing the platform, private actors, such as app developers, and European counterparts would only be able to use the transaction data kept by the Belgian federal government. As such, the

⁴ Regional and local government are not considered as external users.

platform data should contain the data which is currently stored in crossroad bank databases (or in other public databases) as well as transaction data retrieved from digital wallets.

Before being shared through the OGD platform, the transaction data retrieved from digital wallets would have to be curated in a way that makes impossible the identification of data issuers by outside actors (such as tech developers, or European counterparts). Since the accessible data does not contain sensitive information about the identity of individual users (the data is pseudonymized), this data can be shared with other non-governmental sources. In doing so, this should provide incentives to private contractors to support and further develop the system infrastructure, notably by using the data provided by the platform as training data. The new solutions developed by external actors could then be integrated by public organizations, thus mitigating the costs and difficulties associated with the development of new in-house solutions. Similarly, due to the open data characteristics of the platform, and because the shared data would be pseudonymized, external actors should also be able to run fraud detection technologies in the platform without undermining GDPR principles. This would in turn strengthen fraud detection capabilities without requiring additional resources, expertise, and capabilities from the public sector.

Although the design artefact proposes a compelling case in addressing the concerns regarding the public sector's limited resources and dependence on external IT providers, some uncertainties can be pinpointed at that point. First, it is unclear whether if the federal administration possesses the needed internal expertise and resources to carry such a central role in the development and the governance of the OGD platform. For instance, does BOSA hold the adequate technical skills that would be required for the large-scale distribution of digital wallets and deployment of this decentralized platform? More broadly, what additional resources should be allocated to spearhead the development and management of this new system? Would creating a new dedicated organization in charge of such developments (e.g., a new Smals inspired organization) be more adapted for this purpose? A second concern arises regarding the data that should be allowed for sharing with technology developers (transaction data and crossroad banks data). We would have to ensure that these actors have access to enough information to develop better predictive technologies all the while being unable identify data issuers. For instance, which rules or solutions should be followed so that the data shared through the OGD platform ensures the anonymity of data issuers? Who should be in charge of curating the data to be used by the private actors for tech development? How to assess whether the shared data contains enough information to develop better predictive solutions? This is why sandboxes, benchmarking (to compare the existing solutions), and testing will be needed.

4. The proposed design artefact should allow for the use of predictive probabilistic applications of AI without compromising on transparency regarding how this technology is used for decisions by:

- Promoting the adoption of new predictive tools that allow for a better traceability of analyses. It would be possible to develop and adopt explainable AI (XAI) solutions which allow to explain the relative weight of certain data in the results obtained by the analysis. In the same way, some progress is being made in the decoding of black box algorithms, thus indicating that it would be possible to improve the explicability of decisions resulting from such algorithms.
- Improving the reliability of these new tools by conducting pilot experiments in a secure environment. Such experiments could be conducted using anonymized training data provided by the public managers of the data exchange platform. Conducted on a large scale, such an approach could significantly enhance the reliability and technological maturity of the proposed solutions. These experiments could also be integrated into public procurement processes to test new solutions before they are introduced in administrations. However, the development and adoption of such tools should be consistent with

the principle of proportionality defended by the GDPR and should minimize the risks of socio-economic inequalities reproduction.

Our proposed system addresses black boxes by advocating for the use of XAI systems, which allow to explain the weight of a certain input data in reaching an outcome. However, XAI solutions will need to be tested in the platform to ensure that they provide sufficient explainability if a fraud decision is contested. As such, the reliability and the effectiveness of AI solutions in fraud detection using data from an EBSI compliant platform remains to be tested in a secure way, for instance by allowing IT providers to use platform data in sandboxes. As such, supporting predictive applications of AI while avoiding black box remains a challenge for developers. The proposed system should allow for more transparency in fraud decisions while maintaining the ability to use predictive algorithms. However, further testing will be needed to ensure the technical reliability and effectiveness of (X)AI solutions. Furthermore, it is unclear if the proposed technical and processual solutions for justifying fraud decisions would generate additional trust among the public. This will also have to be tested. The suitability of the proposed system regarding the existing organizational processes for fraud decisions that can also be questioned. Is the proposed system appropriate with regard to existing operational processes in the fight against fraud? Would it interfere with existing legislative rules in the current framework?

5. The OGD governance platform should contribute to the creation of common socio-technical standards and a greater interoperability among the platform stakeholders at the national and at the EU level with:

- The adoption of a data exchange platform that would operate based on common criteria and technical standards. The implementation of such a platform would stimulate data interoperability and collaboration between all actors in charge of fighting tax fraud and social security infringements. Under certain conditions, this solution could also facilitate cooperation with European counterparts who have set up an equivalent system.
- The creation of an entity in charge of coordinating, developing, and monitoring digital projects within the administrations. One solution could be to empower some existing Federal Public services (FPS), such as the FPS BOSA, or to create a new entity along the lines of the Smals organization, but which would offer its expertise to all FPSs. This new entity would ensure the harmonization of projects and the development of a common strategic vision. It would also be responsible for the development and management of the data exchange platform or the deployment of digital wallets, for example.

Here a first concern arises regarding the actors that will assist in the development of the platform. These idea providers should be carefully selected since they will help to establish common socio-technical standards among the OGD platform stakeholders. However, it is unclear if these idea providers should be limited to a few specialized actors or if these should involve all the potentially interested parties: not only all Belgian governments (e.g., the Flemish Region, the Walloon region, communities, etc.), but also stakeholder groups and particularly RGD/privacy defenders. To improve the coherence and agreement on common standards among platform stakeholders, a debate should perhaps be held regarding which actors will be associated in the development of the common platform. Second, another potential challenge would be to achieve organizational interoperability between federal and regional levels if they end up choosing different technical solutions. For instance, the Flemish government's choice to work with SOLID suggests that federal and regional levels are simultaneously pursuing different digital wallet solutions. This again stresses the need to create a synergy, or a dialogue, between all the interested parties at the federal and regional level. Last, but not least, we can again interrogate our socio-political readiness to the adoption of such a novel data exchange platform. For instance, how likely are we to encounter resistance to these socio-

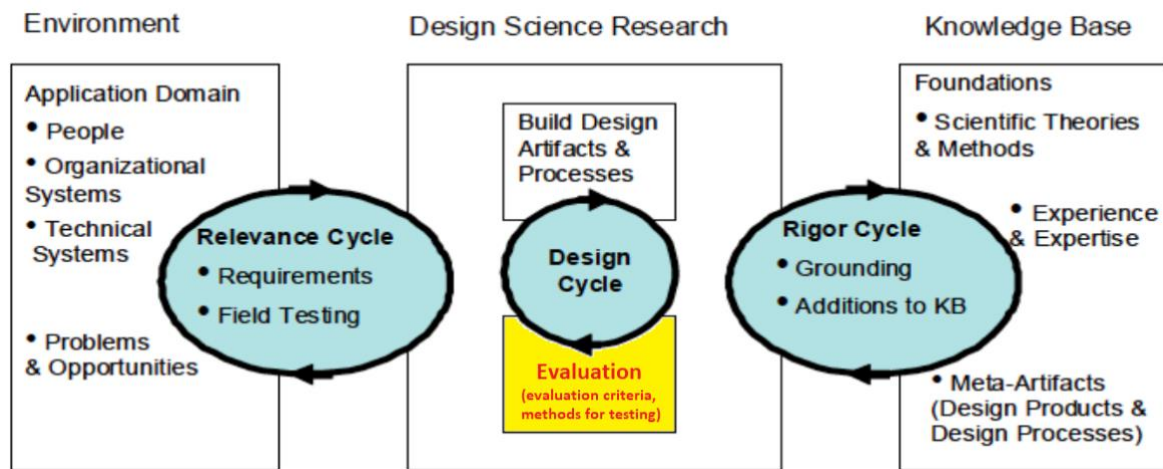
technical standards due to factors such as digital culture differences and digital divide? Actively involving citizens into the development and testing process of the platform would perhaps generate wider public endorsement in creating new common socio-technical standards.

In sum, the proposed design artefact is a governance model for an OGD platform, that support predictive analytics and engender trusts among users through citizen-centric solutions in the share and (re)use of data. This platform may rely on a decentralized approach and support digital wallet solutions provided by the Belgian government (e.g., BOSA or another entity) and possibly by some EU actors. This governance model should also integrate the databases from intermediary actors (i.e., banks, health insurers, unions, etc.) and crossroad banks. In the end, the proposed platform consists in an integrated IT architecture that offers solutions and opportunities that reach far beyond the specific areas of public intervention that are the fight against tax fraud and social security infringements. However, this design artefact also supposes some radical changes to the existing system since it suggests a transition from a centralized system architecture (current model) to a decentralized or at least to a hybrid model. Furthermore, it assumes a seamless and coherent integration of the OGD platform among the various components of the Belgian federal landscape.

4. SELECTION OF CRITERIA FOR MODEL EVALUATION

Following Hevner’s three cycle view of design science research, the “design cycle” comprises both the construction of design artefacts (see section 3) and the evaluation of their applicability. The present section – as well as the next one – specifically addresses the evaluation part of the design cycle. The evaluation of the design artefact includes both the identification and selection of relevant evaluation criteria for testing the design artefact (section 4) as well as the methods used for collecting and analyzing the relevant data to conduct (section 5). The following figure presents Hevner’s (2007) three cycle view and highlights the present stage of the research process.

Figure 4. Hevner’s three cycle view of design science research



The design science research literature recommends the selection of evaluation criteria to test the design artefact. The selection of appropriate evaluation criteria was performed in three steps. First, we conducted a short literature review on the evaluation criteria that are generally used in design science research literature. Indeed, the design science research approach attaches great importance to the testing and assessment of design artefacts with the use of criteria (Sonnenberg & Vom Brocke, 2011). This literature thus provides great insight on which specific criteria are generally used and for what purposes. Second, we also conducted a short literature review on the evaluation criteria that are used in other literatures, such as public policy evaluation. This was done so that we could broaden our perspective on which criteria should be adopted to evaluate a design artifact that may have strong implications for various public policy sectors. Finally, and taking stock of the results of both literature reviews, we selected a list of criteria that appear well suited to the evaluation of our design artefact. This was done so the selected criteria address the several potential showstoppers that were highlighted in the concluding piece of the previous section (section 3.7).

4.1. EVALUATION CRITERIA IN THE DESIGN SCIENCE RESEARCH LITERATURE

The following table presents the results of our literature review regarding the evaluation criteria that are generally used in design science research literature. For each reference, we listed the evaluation criteria that are used or recommended by the authors. We also provide some short comments regarding how authors categorize or envision the use of the listed criteria.

Table 4. Evaluation criteria in the design science research literature

Design science research literature		
References	Evaluation criteria	Comments
<p>Sonnenberg, C., & Vom Brocke, J. (2011, October). Evaluation patterns for design science research artefacts. In European design science symposium (pp. 71-83). Springer, Berlin, Heidelberg.</p>	<ul style="list-style-type: none"> • Accessibility • Applicability • Ease of use • Effectiveness • Efficiency • Elegance • Completeness • External consistency • Feasibility • Fidelity with real world phenomenon • generality • Impact on artefact environment and user • Internal consistency • Level of detail • Operationality • Robustness • Simplicity • Suitability • Understandability 	<p>“When choosing evaluation criteria, a design researcher should pay attention to balance the interests of practitioners and researchers [cf. 8] which is a central aim of design science research. E.g., practitioners are interested in the applicability and usefulness of an artefact (relevance) whereas researchers are interested in the validity of the artefact and thus aim at structuring their evaluations appropriately in order to ensure rigour in the process.” (Sonnenberg & Vom Brocke, 2011, p. 4)</p>
<p>Aier, S., & Fischer, C. (2011). Criteria of progress for information systems design theories. <i>Information Systems and E-Business Management</i>, 9(1), 133-172.</p>	<ul style="list-style-type: none"> • Utility • Internal consistency • external consistency • broad purpose and scope • simplicity • fruitfulness of further research 	<p>Those criteria are independent of an artefact type. They particularly apply for evaluating design theories.</p>
<p>Rosemann, M., & Vessey, I. (2008). Toward improving the relevance of information systems research to practice: the role of applicability checks. <i>Mis Quarterly</i>, 1-22.</p>	<ul style="list-style-type: none"> • Importance • Suitability • Accessibility 	<p>“Using the dimensions of importance, accessibility, and suitability, we thereby view research relevance as the degree to which practitioners can readily comprehend research as promising a solution potentially applicable to a problem existing in their organization.” (Roseman & Vessey, 2008, p. 3)</p>
<p>March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. <i>Decision</i></p>	<p>For construct evaluation:</p> <ul style="list-style-type: none"> ○ Completeness ○ Simplicity ○ Elegance ○ Understandability ○ Ease of use <p>For model evaluation:</p>	<ul style="list-style-type: none"> • Construct evaluation • Model evaluation • Here, “methods” should be understood as the type of technical solution that are designed to solve a problem. • “Evaluation of instantiations considers the efficiency and effectiveness of the artifact and its

<p><i>support systems</i>, 15(4), 251-266.</p>	<ul style="list-style-type: none"> ○ Fidelity with real world phenomena ○ Completeness ○ Level of detail ○ Robustness ○ Internal consistency <p>For methods evaluation:</p> <ul style="list-style-type: none"> ○ Operationality ○ Efficiency ○ Generality ○ Ease of use <p>For Instantiation evaluation:</p> <ul style="list-style-type: none"> ○ Effectiveness ○ Efficiency ○ Impact on the environment and on the artefact's users 	<p>impacts on the environment and its users" (March & Smith, 1995, p. 261)</p>
<p>Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. <i>MIS quarterly</i>, 75-105.</p>	<ul style="list-style-type: none"> ● Utility ● Quality ● Efficacy ● Functionality ● Completeness ● Consistency ● Accuracy ● Performance ● Reliability ● Usability ● Fit with the organization 	<p>"The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well executed evaluation methods. Evaluation is a crucial component of the research process. The business environment establishes the requirements upon which the evaluation of the artifact is based. This environment includes the technical infrastructure which itself is incrementally built by the implementation of new IT artifacts. Thus, evaluation includes the integration of the artifact within the technical infrastructure of the business environment." (Hevner et al., 2004, p. 85)</p> <p>"Design, in all of its realizations (e.g., architecture, landscaping, art, music), has style. Given the problem and solution requirements, sufficient degrees of freedom remain to express a variety of forms and functions in the artifact that are aesthetically pleasing to both the designer and the user. Good designers bring an element of style to their work (Norman 1988). Thus, we posit that design evaluation should include an assessment of the artifact's style." (Hevner et al., 2004, p. 88)</p>
<p>Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. <i>MIS quarterly</i>, 337-355.</p>	<ul style="list-style-type: none"> ● Validity ● Utility ● Quality ● Efficacy 	<p>"The artifact is evaluated in terms of criteria that can include validity, utility, quality, and efficacy. Validity means that the artifact works and does what it is meant to do; that it is dependable in operational terms in achieving its goals. The utility criteria assesses whether the achievement of goals has value outside the development environment. A rigorous design evaluation may draw from many potential techniques, such as analytics, case studies, experiments, or simulations (see Hevner et al. 2004) and naturalistic evaluations (Carlsson 2010). Further sources for methods of evaluation include Pries-Heje et al. (2008) and Sein et al. (2011)." (Gregor & Hevner, 2013, p. 351)</p>

<p>Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). Artifact Evaluation in Information Systems Design-Science Research- a Holistic View. <i>PACIS</i>, 23, 1-16.</p>	<p>Goal:</p> <ul style="list-style-type: none"> ○ Efficacy ○ Validity ○ Generality <p>Environment:</p> <ul style="list-style-type: none"> ○ Consistency with people ○ Consistency with organization ○ Consistency with technology <p>Structure of artefact:</p> <ul style="list-style-type: none"> ○ Completeness ○ Simplicity ○ Clarity ○ Style ○ Homomorphism ○ Level of detail ○ Consistency <p>Activity:</p> <ul style="list-style-type: none"> ○ Completeness ○ Consistency ○ Accuracy ○ Performance ○ Efficiency <p>Evolution:</p> <ul style="list-style-type: none"> ○ Robustness ○ Learning capability 	<p>Following an extensive literature review, they propose a hierarchy of evaluation criteria. They categorize these criteria and add sub-criteria for each of them: “We built this hierarchy as a result of a consensus between all of us. This consensus was attained as prescribed by the hierarchical theory of justification (Laudan, 1984) based on three interrelated levels. At the factual level, we collected all criteria in the literature with their descriptions and explanations. At the methodological level, by applying linguistic rules (mainly synonym, antonym, hyponym relationships), we produced the evaluation criteria and subcriteria. Finally, the axiological level refers, in our research, to general systems theory leading to the main system dimensions at the highest level of our hierarchy” (Prat et al., 2014, p. 7)</p>
<p>Hevner, A., Prat, N., Comyn-Wattiau, I., & Akoka, J. (2018, December). A pragmatic approach for identifying and managing design science research goals and evaluation criteria. In <i>AIS SIGPrag Pre-ICIS workshop on "Practice-based Design and Innovation of Digital Artifacts"</i>.</p>	<p>Evolution Goals</p> <ul style="list-style-type: none"> ○ Learning capability ○ Robustness ○ Applicability <p>Innovation goals</p> <ul style="list-style-type: none"> ○ Innovativeness ○ Inventiveness ○ Research contribution <p>Cognitive and aesthetic goals</p> <ul style="list-style-type: none"> ○ Understandability ○ Openness ○ Style <p>Interaction and communication goals</p> <ul style="list-style-type: none"> ○ Ethicality ○ Alignment with business ○ Embeddedness in a design system <p>Safety Goals</p> <ul style="list-style-type: none"> ○ Security ○ Integrity ○ Privacy ○ Confidentiality <p>Utilitarian goals</p> <ul style="list-style-type: none"> ○ Functionality 	<p>Authors propose a hierarchy of Design science research goals and link them with specific evaluation criteria. They propose a pyramidal approach whereby “many of the evaluation criteria appear at the bottom level of the hierarchy (utilitarian criteria). These criteria are inherently associated with the artifact: they qualify its behavior, e.g. functionality, and/or its structure, e.g. decomposability. As we move upwards in the hierarchy, subjectivity plays a greater role, the criteria being more related to the perception of individuals (e.g. style). At the upper level, innovation criteria measure the new knowledge added to the knowledge bases of the application domain and the degree to which the artifact can be generalized to new situations. At the top level, evolution criteria measure the extent to which the artifact is capable of accommodating changes in its context (e.g., malleability) and/or its goals (e.g. learning capability).” (Hevner et al., 2018, pp. 8-9)</p>

	<ul style="list-style-type: none"> ○ Usefulness ○ Ease of use ○ Accuracy ○ Performance ○ Completeness ○ Simplicity ○ Decomposability ○ Diversity 	
--	--	--

Some of the most commonly cited criteria are: Accessibility (Sonnenberg & Vom Brocke, 2011; Roseman & Vessey, 2008); Accuracy (Hevner et al., 2018; Pratt et al., 2014; Hevner et al., 2004); Completeness (Hevner et al., 2018; Pratt et al., 2014; Sonnenberg & Vom Brocke, 2011; Hevner et al., 2004; March & Smith, 1995); Internal and external consistency (Pratt et al., 2014; Sonnenberg & Vom Brocke, 2011; Aier & Fischer, 2011; Hevner et al., 2004; March & Smith, 1995); Ease of use (Hevner et al., 2018; Sonnenberg & Vom Brocke, 2011; March & Smith, 1995); Effectiveness (Sonnenberg & Vom Brocke, 2011; March & Smith, 1995); Efficacy (Pratt et al., 2014; Gregor & Hevner, 2013; Hevner et al., 2004); Efficiency (Pratt et al., 2014; Sonnenberg & Vom Brocke, 2011; March & Smith, 1995); Elegance (Sonnenberg & Vom Brocke, 2011; March & Smith, 1995); Functionality (Hevner et al., 2018; Hevner et al., 2004); Generality (Pratt et al., 2014; Sonnenberg & Vom Brocke, 2011; March & Smith, 1995); Learning capability (Hevner et al., 2018; Pratt et al., 2014); Level of detail (Pratt et al., 2014; Sonnenberg & Vom Brocke, 2011; March & Smith, 1995); Operationality (Sonnenberg & Vom Brocke, 2011; March & Smith, 1995); Performance (Hevner et al., 2018; Pratt et al., 2014; Hevner et al., 2004); Quality (Gregor & Hevner, 2013; Hevner et al., 2004); Robustness (Hevner et al., 2018; Pratt et al., 2014; Sonnenberg & Vom Brocke, 2011; March & Smith, 1995); Simplicity (Hevner et al., 2018; Pratt et al., 2014; Sonnenberg & Vom Brocke, 2011; Aier & Fischer, 2011; March & Smith, 1995); Style (Hevner et al., 2018; Pratt et al., 2014); Suitability (Sonnenberg & Vom Brocke, 2011; Rosemann & Vessey, 2008); Understandability (Hevner et al., 2018; Sonnenberg & Vom Brocke, 2011; March & Smith, 1995); Utility (Gregor & Hevner, 2013; Aier & Fischer, 2011; Hevner et al., 2004); Validity (Pratt et al., 2014; Gregor & Hevner, 2013).

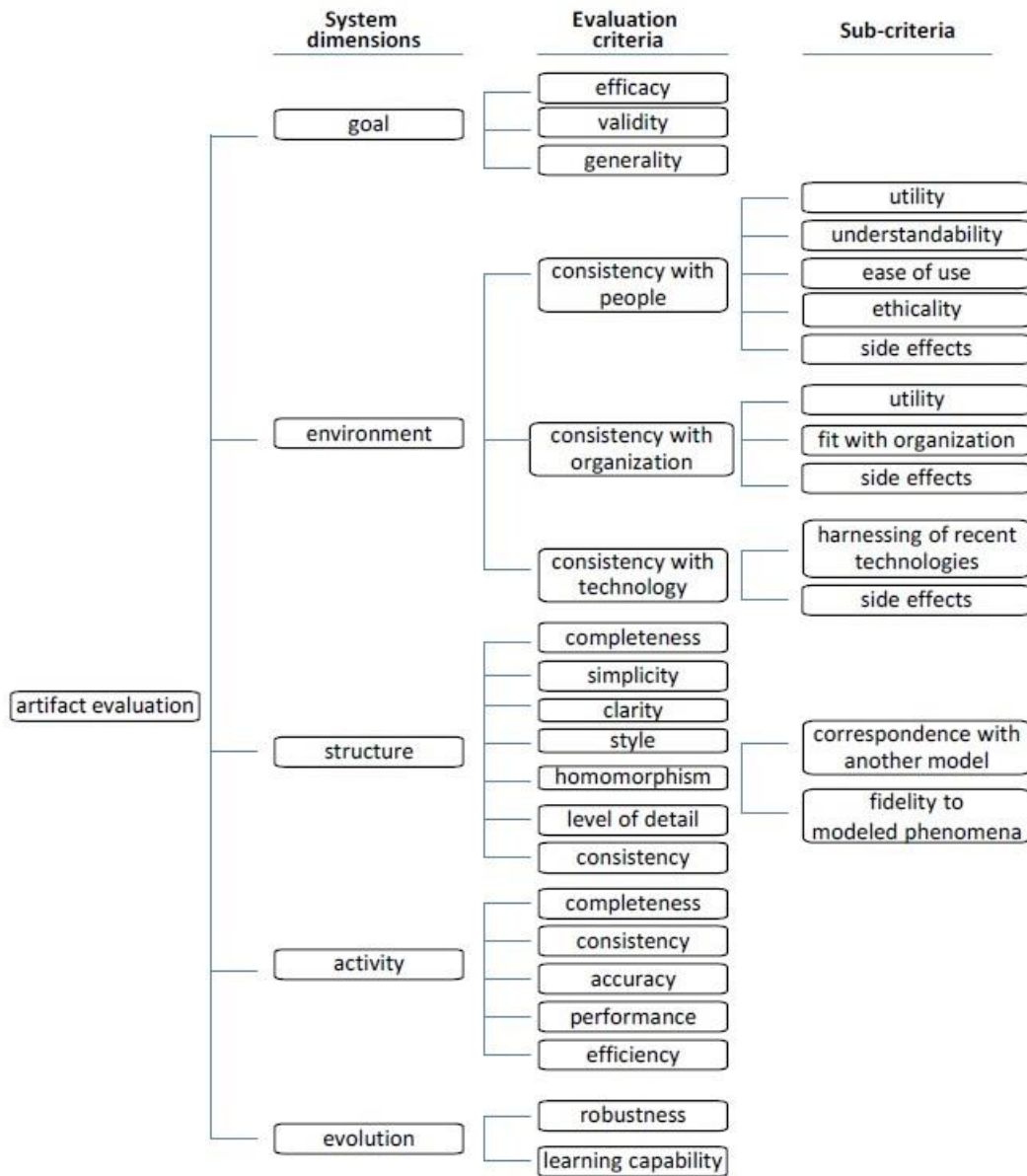
However, many of these evaluation criteria are synonyms that convey a similar meaning and substance. The following table groups the aforementioned evaluation criteria according to their similarities. This allows us to propose a shorter list of key criteria that should also account for most of the common evaluation criteria used in the design science research literature. These criteria are efficacy, simplicity, consistency, operationality, robustness, and suitability.

Table 5. Grouping the most common criteria according to their similarities

Grouping criterion	Evaluation criteria
Efficacy	Effectiveness; Efficacy; Efficiency; Performance
Simplicity	Understandability; Simplicity; Elegance; Style; Ease of use; Accessibility
Consistency	Completeness; Accuracy; Internal and external consistency; Level of detail; Quality
Operationality	Operationality
Robustness	Robustness; Learning capability
Suitability	Suitability; Utility; Validity; Generality; Functionality

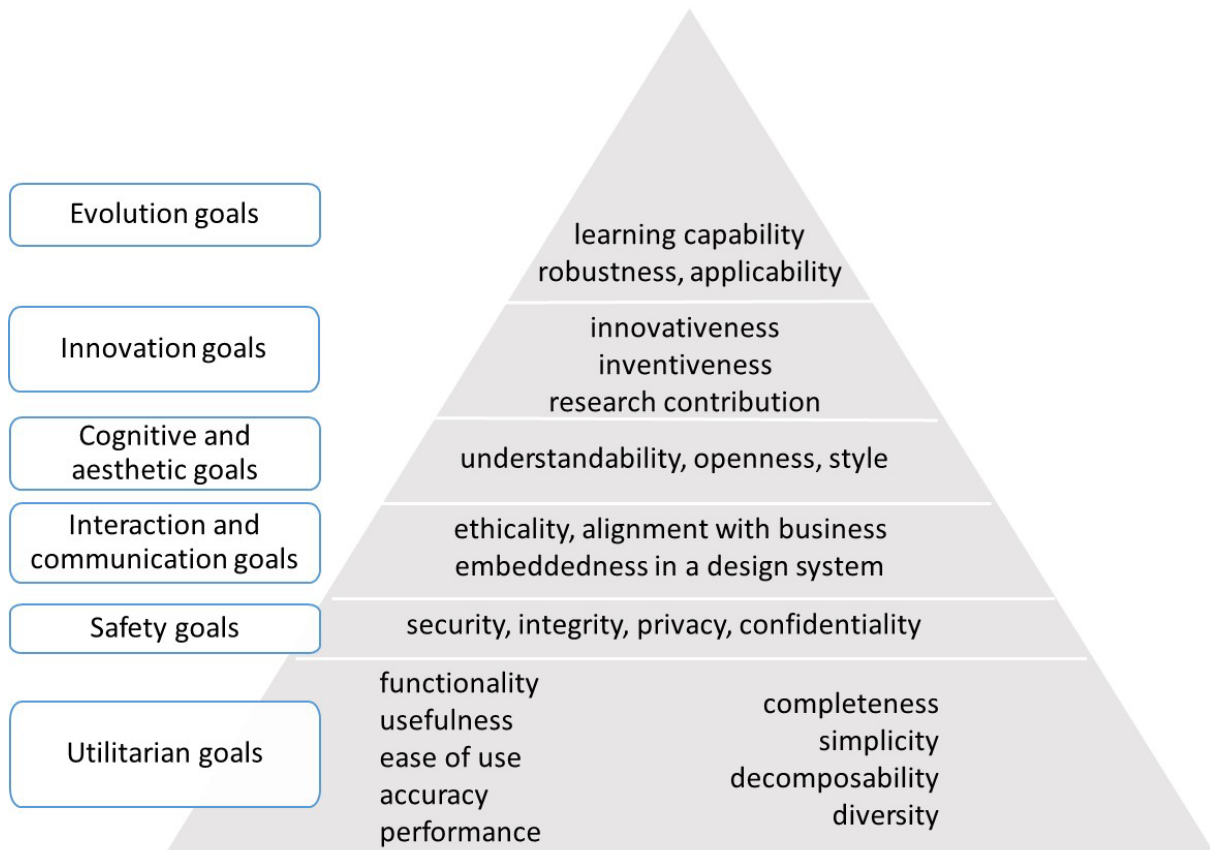
Some authors also develop a hierarchy, or a categorization, of the proposed evaluation criteria. For instance, after conducting an extensive literature review on the evaluation criteria used in design science research, Pratt et al. (2014) suggest a categorization of evaluation criteria for assessing design artefacts. The criteria are categorized according to the various aspects of the design artefact that are expected to be evaluated. The authors distinguish between five system dimensions comprising the criteria that evaluate the design artefact's ability to reach its goals (1); the criteria that evaluate the design artefact's effect on the environment in which it is embedded (2); the criteria that evaluate the structure and overall coherence of the design artefact (3); the criteria that evaluate the activities envisioned by the design artefact (4); the criteria that evaluate the design artefact's ability to evolve and adapt over time (5). For some evaluation criteria, the authors also highlight a set of sub-criteria that can be adopted to further evaluate the design artefact. The figure 6 describes the various system dimensions, evaluation criteria, and sub-criteria suggested by Pratt et al. (2014).

Figure 5. Categories of evaluation criteria (Pratt et al., 2014)



Similarly, Hevner et al. (2018) propose a hierarchy of design science research goals and link them with specific evaluation criteria. According to the authors, most criteria are at the bottom of the hierarchy since they are dedicated to the evaluation of the utilitarian aspects of the design artefact such as its functionality, its structure, or its completeness. However, when moving upwards in the hierarchy, it is the perception of the concerned individuals that tends to gain in importance (Hevner et al., 2018). These criteria are for instance those that aim at evaluating the interaction and communication goals of the design artefact or its cognitive and aesthetic goals. Moving upwards, the authors place the innovation criteria, which assess the extent to which “new knowledge added to the knowledge bases of the application domain and the degree to which the artifact can be generalized to new situations” (Hevner et al., 2018, p. 9). At the top of the hierarchy, Hevner et al. (2018) place the evolution criteria which evaluate the design artefact’s ability to evolve and adapt over time. This pyramidal hierarchy of design science research goals and associated evaluation criteria is illustrated by the figure 7.

Figure 6. Hierarchy of evaluation criteria (Hevner et al., 2018)



4.2. EVALUATION CRITERIA IN OTHER RESEARCH FIELDS

The following table presents the results of a short literature review on the evaluation criteria that are used in other research fields such as public administration, public policy evaluation, and planning. As for the previous literature review, we listed the evaluation criteria that are used or recommended in each of the references. Again, we provide some short comments regarding how the authors categorize or envision the use of the listed criteria.

Table 6. Evaluation criteria in other research fields

Public administration, public policy evaluation, planning, etc.		
References	Evaluation criteria	Comments
<p>Albarelo, L., Van Haeperen, B., Aubin, D., & Fallon, C. (2016). Penser l'évaluation des politiques publiques. De Boeck Supérieur.</p>	<ul style="list-style-type: none"> • Relevance • Quality of the design • Quality of the implementation process • Effectivity • Effectiveness • Allocative efficiency • Productive efficiency 	<p>The relevance (1) assesses the relationship that exists, or should exist, between the objectives defined in the public policy and the problem to be solved. The quality of the design (2) questions the internal coherence of the program and the rationality of its logic of action, particularly with regard to the identification of target groups and the choice of instruments. The quality of the implementation process (3) questions the link between the program as it was conceived and its actual realization in the field, i.e. the outputs achieved during the implementation of the policy.</p> <p>The effectivity (4) evaluates the effects of a public policy, along with effectiveness. It measures the degree to which the (normative) objectives of a policy are in adequation with the actual behavior of the target groups. The effectiveness (5) questions the relationship between the expected (final) effects of a policy and those that actually appear in social reality. The allocative efficiency (6) compares the resources invested in public policy with the results obtained in a cost-benefit analysis. The productive efficiency (7) relates the outputs produced to the resources invested. It only examines the managerial performance. (Abarelo et al., 2016, pp. 20-21)</p>
<p>Haarich, S. N. (2018). Building a new tool to evaluate networks and multi-stakeholder governance systems. Evaluation, 24(2), 202-219.</p>	<p>GOCAPASS – Dimensions and Factors of Governance Performance:</p> <p>D1 Policy Capacity</p> <ol style="list-style-type: none"> 1. Analysis for decision-making 2. Translate needs into action 3. Responsibilities 4. Vision- and goal-setting 5. Priority-setting 6. Reflection and feedback <p>D2 Managerial Capacity</p> <ol style="list-style-type: none"> 1. Capacity to budget 2. Adequate implementation support 3. Implementation resources 4. Monitoring and control 5. Stakeholder engagement 	<p>The authors present the GOCAPASS tool which “enables evaluators to assess governance systems and networks involving multiple stakeholders” (Haarich, 2018, p. 202)</p>

	<p>D3 Networking and Cooperation Capacity</p> <ol style="list-style-type: none"> 1. Trust as precondition 2. Exchange of information 3. Communication flows 4. Knowledge management 5. Active cooperation in policy-making and implementation <p>D4 Enabling Environment and Governance Levers</p> <ol style="list-style-type: none"> 1. Leadership 2. Mutual support (within the system) 3. Learning facilities in the system 4. Openness and adaptability (learning from others) 5. Commitment from the environment 	
<p>Lemire, S., Peck, L. R., & Porowski, A. (2020). The growth of the evaluation tree in the policy analysis forest: Recent developments in evaluation. <i>Policy studies journal</i>, 48, S47-S70.</p>	<p>No specific criteria were mentioned</p>	<p>“Cultural awareness is fundamental, especially in many evaluation contexts” “American Evaluation Association’s <i>Guiding Principles for Evaluators</i>. Revised in 2018, the principles now emphasize cultural competence as a core element of professional practice. According to the standards, “culturally competent” evaluators are defined as those “who draw upon a wide range of evaluation theories and methods to design and carry out an evaluation that is optimally matched to the context. In constructing a model or theory of how the evaluation operates, the evaluator reflects the diverse values and perspectives of key stakeholder groups, for good practice” (American Evaluation Association, 2018, p. 3).” (Lemire et al., 2020, p. 61)</p>
<p>Patton, M. Q. (2021). Evaluation Criteria for Evaluating Transformation: Implications for the Coronavirus Pandemic and the Global Climate Emergency. <i>American Journal of Evaluation</i>, 42(1), 53-89.</p>	<p>See Patton (2021, p. 78):</p> <ul style="list-style-type: none"> • Transformation fidelity involves examining the connection between transformation rhetoric and reality—evaluating the scale, scope, and pace of actual transformational engagement which means reality-testing the vision against what is being done and accomplished. • Complex systems framing connects system thinking and complexity concepts to define the processes, nature, trajectories, and results of transformational engagements. 	<p>The author presents and discusses six criteria for evaluating transformations stemming from interventions. They are illustrated using the case of the pandemic and the Global Alliance for the Future of Food. Criteria are inspired from the OECD DAC criteria that are being developed since 1984.</p> <p>“The formal logic of evaluation involves four steps:</p> <ol style="list-style-type: none"> 1) define the criteria that will be used to evaluate something; 2) set standards of performance on those criteria; 3) measure the actual performance; and 4) synthesize the results to reach an evaluative judgment” (Patton, 2021, p. 57)

	<ul style="list-style-type: none"> • Eco-efficient full cost accounting makes transparent the interconnections between an intervention’s direct costs and benefits in relation to broader environmental and human/societal systems costs and benefits (economic externalities). • Adaptive sustainability invites evaluation of adaptive capacity interconnections between environmental ecosystems and human systems sustainability over time, the interrelationships between people and nature, and, in the triple bottom line economic framing, the interdependence of people, planet, and profits. • Diversity/equity/inclusion focuses attention on the interconnections between who is engaged in and affected by processes of inclusion and diversity toward transformational aspirations of greater equity. • Interconnectedness momentum calls attention to the transformational implications of aligning and integrating across divisions, silos, differing perspectives, historical divisions, and competing interests toward a vision of a more sustainable and equitable future. 	<p>The author stresses the necessity to develop evaluation criteria that are contextually meaningful: “I want to reiterate that the criteria I have offered for evaluating transformation are meant to be illustrative of what is possible. The applications of the criteria to the coronavirus pandemic and the Global Alliance for the Future of Food provide illustrations of contextual meaningfulness rather than suggesting these specific criteria be standardized and adopted as universal. Schwandt (2018) has articulated definitively the case against treating the DAC criteria, or any set of criteria, as universal.” (Patton, 2021, p. 78)</p>
<p>Van Bueren, E., & ten Heuvelhof, E. (2005). Improving governance arrangements in support of sustainable cities. <i>Environment and planning B: Planning and Design</i>, 32(1), 47-66.</p>	<ul style="list-style-type: none"> • Satisfactory policy outcomes for actors • Cognitive learning about the nature of policy • Changed patterns of interaction • Social learning about implementation of policy 	<p>According to the authors:</p> <ul style="list-style-type: none"> • Governance should be considered and understood as a process • The satisfaction of stakeholders is a key element to consider when evaluating a governance plan • It is important to assess the changes in the patterns of interactions between actors (indicative of institutional change) <p>“Theories on network management and process management teach us that governance arrangements are successful when the different stakeholders are satisfied with the course and outcomes of the policy process (Teisman, 1995). This satisfaction is an indication that the arrangement has offered stakeholders, with their different and sometimes competing and conflicting perceptions, enough room to participate and to articulate and exchange their points of view, knowledge, and information (Klijn and Koppenjan, 2000). A second criterion for policy success can be found in the patterns of interaction. Changed interaction patterns are an indicator of institutional change: the interdependencies between actors have changed, and this</p>

		may also lead to changes in the problem-solving space (Klijn and Koppenjan, 2000; Teisman, 1995).” (Van Bueren & ten Heuvelhof, p. 49)
Olejniczak, K., Borkowska-Waszak, S., Domaradzka-Widła, A., & Park, Y. (2020). Policy labs: the next frontier of policy design and evaluation?. <i>Policy & Politics</i> , 48(1), 89-110.	<ul style="list-style-type: none"> • Utility • Legality • Transparency • Effectiveness • Efficiency • Sustainability 	“The key criteria used in the assessment of public policy programmes and interventions include utility, legality, transparency, effectiveness, efficiency and sustainability (Olejniczak and Mazur, 2014). Depending on the criteria used, different pictures of policy outcomes emerge. Evaluators are aware of this issue, and they are cautious about clarifying what criteria can be applied” (Olejniczak et al., 2020, p. 93)
Alexander, E. R., & Faludi, A. (1989). Planning and plan implementation: notes on evaluation criteria. <i>Environment and planning B: Planning and Design</i> , 16(2), 127-140.	<ul style="list-style-type: none"> • Conformity • Rational process • Optimality ex ante • Optimality ex post • Utilisation 	The goal of the authors is to develop a rigorous approach for assessing the quality of plans. They provide five criteria for comprehensive evaluation (Alexander & Faludi, 1989, pp. 135-136): <ul style="list-style-type: none"> • Conformity: "To what degree do operational decisions, implementation decisions, and actual outputs, outcomes, and impacts conform to the goals, objectives, intentions, and instructions expressed in the policy, plan, or programme being evaluated?" • Rational process: "A rational process here means conforming to certain normative requirements in process and method." • Optimality ex ante: "Could the strategy or the courses of action prescribed in the policy or plan under assessment be considered optimal?" • Optimality ex post: "Was the strategy or were the courses of action prescribed in the policy or plan under assessment in fact optimal?" • Utilization: " whether the policy or plan was used as a frame of reference for operational decisions."
Leach, W. D., Pelkey, N. W., & Sabatier, P. A. (2002). Stakeholder partnerships as collaborative policymaking: Evaluation criteria applied to watershed management in California and Washington. <i>Journal of Policy Analysis and</i>	<ul style="list-style-type: none"> • Perceived effects of the partnership on specific problems in the watershed • Perceived effects of the partnership on human and social capital • The extent of agreement reached among the stakeholders • Implementation of restoration projects • Monitoring projects • Education and outreach projects 	“Rather than an exhaustive list of criteria, a manageable set of six disjoint criteria have been selected that together can adequately measure partnership success. Following Connell and Kubisch (1998), measures of outcomes, actions, and agreements have been included, specifically, measures of: perceived effects of the partnership on specific problems in the watershed; perceived effects of the partnership on human and social capital; the extent of agreement reached among the stakeholders; implementation of restoration projects; monitoring projects; and education and outreach projects.” (Leach et al., 2002, p. 652)

<i>Management: The Journal of the Association for Public Policy Analysis and Management, 21(4), 645-670.</i>		
--	--	--

In comparison with the design research literature, there is less consistency in the evaluation criteria mentioned by authors. The most common shared criteria are efficiency and effectiveness (Albareello et al., 2016; Olejniczak et al., 2020), which are also commonly used in the design research literature. Although a quick glance suggests that the listed criteria generally differ from those that are highlighted in the design science research literature, some general commonalities can be drawn once these criteria are grouped into general categories. Table 7 clusters the various criteria cited by the authors according to seven main categories or grouping criteria. This allows us to identify a limited number of more general, yet relevant, criteria that can be listed as follows: 1. efficacy; 2. sustainability / robustness; 3. feasibility / operationality; 4. stakeholder endorsement; 5. cognitive aspects; 6. suitability; 7. consistency.

This allows us to identify broad similarities with the criteria used in design science research. For instance, the sustainability criteria – which is either referred to as “openness and adaptability” (Haarich, 2018), “sustainability” (Olejniczak et al., 2020), or “Adaptive sustainability” (Patton, 2021) – usually aims at interrogating the models’ ability to adapt and evolve. This idea of sustainability can also be found in the design science research literature under the form of the “robustness” criteria, the “evolution” system dimension (Pratt et al. 2014) or the “evolution goals” of a given model (Hevner et al., 2018). Such similarities with design science research criteria can also be observed for the following categories: Efficacy; Consistency; Suitability; and Feasibility / operationality.

Table 7. Grouping criteria according to their similarities

Grouping criterion	Evaluation criteria
Efficacy	Effectiveness; Effectivity; Efficiency; Allocative efficiency; Productive efficiency; Utilization; Changed patterns of interaction; Optimality ex post; Perceived effects on specific problems; Perceived effects on human and social capital; Translate needs into action
Sustainability / Robustness	Adaptive sustainability; Sustainability; Interconnectedness momentum; Openness and adaptability; Commitment from the environment
Feasibility / operationality	Capacity to budget; Implementation resources; Eco-efficient full cost accounting; Complex systems framing; Legality; Adequate implementation support; Leadership; Monitoring and control; Monitoring projects; Responsibilities; Analysis for decision-making

Stakeholder endorsement	Stakeholder engagement; Active cooperation in policy-making and implementation; Diversity/equity/inclusion; Trust as precondition; Satisfactory policy outcomes for actors; Extent of agreement reached among stakeholders
Cognitive aspects	Cognitive learning about the nature of policy; Communication flows; Education and outreach projects; Exchange of information; Reflection and feedback; Social learning about implementation of policy; Knowledge management; Learning facilities in the system
Suitability	Utility; Relevance; Vision- and goal-setting; Priority-setting; Optimality ex ante
Consistency	Conformity; Quality of the design; Quality of the implementation process; Rational process; Transformation fidelity

The first contrast with the criteria listed in the design science literature lies in the importance given to cognitive aspects of the model under assessment. Scholars emphasize the importance of learning and knowledge exchange among stakeholders. For instance, Van Bueren and ten Heuvelhof (2005) highlight the necessity to assess stakeholders’ “cognitive learning about the nature of policy” and the degree of “social learning about the implementation of policy”. Similarly, the third dimension of the GOCAPASS tool – which “enables evaluators to assess governance systems and networks involving multiple stakeholders” (Haarich, 2018, p. 202) – highlights the need to assess information exchanges, communication flows, and knowledge management.

A second difference is scholars’ emphasis on the embeddedness of evaluation criteria in a meaningful context. For instance, Lemire, Peck, and Porowski (2020) highlight the importance of cultural awareness when developing evaluation criteria. These authors note that “culturally competent” evaluators are defined as those “who draw upon a wide range of evaluation theories and methods to design and carry out an evaluation that is optimally matched to the context” (Lemire et al., 2020, p. 61). Similarly, Patton (2021) stresses that rather than adopting criteria that are “standardized and adopted as universal” (p. 78) he would rather develop evaluation criteria that are contextually meaningful. Haarich (2018) insists on evaluating the political, managerial, and technical capacities that are required for the proper operation of governance models. This alludes to the necessity of evaluating the feasibility of a model given the specific context in which it is embedded.

Finally, the third contrast lies in the idea of assessing stakeholders’ endorsement of the model under scrutiny. In other words, authors recommend testing the models’ acceptability. For instance, stakeholders’ satisfaction and agreement with the design solutions of the model are key elements to consider when evaluating a governance model. As Van Bueren and ten Heuvelhof (2005) note: “governance arrangements are successful when the different stakeholders are satisfied with the course and outcomes of the policy process” (Van Bueren & ten Heuvelhof, 2005, p. 49). Similarly, “the extent of agreement reached among the stakeholders” is put to the fore by Leach et al. (2002). To some extent, stakeholders’ endorsement and satisfaction is also linked with the idea of inclusivity of the model (Patton, 2021). In other words, the proposed model should make sure that all the stakeholders can be included in a fair and equitable way.

4.3. EVALUATION CRITERIA FOR OUR DESIGN ARTEFACT

Based on the results of our literature review, the following table presents the evaluation criteria that will be used to assess our design artefact. More precisely, a Delphi survey will be constructed to test the proposed model. The survey will evaluate several key dimensions of the design artefact and, for each of these dimensions, specific questions will assess the model with the help of these criteria. More details about the survey method and construction will be given in the following section of this deliverable.

Table 8. Evaluation criteria for our design artefact

System dimension	Evaluation criteria	Short description
Goal	Overall efficacy	Is the proposed governance model perceived as an effective and efficient solution for the integration of new technologies in the fight against tax fraud and social security infringements?
	Overall suitability	Is the proposed governance model perceived as a useful and relevant solution for the integration of new technologies in the fight against tax fraud and social security infringements?
Environment	Feasibility	To what extent are the design solutions of the proposed governance model considered feasible on a political / financial / technological / legal level?
	Stakeholder endorsement	To what extent stakeholders agree with and trust the design solutions of the proposed governance model? Are they satisfied with the outcomes?
	Cognitive aspects	Does the proposed governance model promote knowledge and information exchange among stakeholders? Does it foster cognitive learning about the policy and its implications?
Structure	Overall consistency	Is the proposed governance model perceived as coherent and consistent? Is it viewed as sufficiently detailed and comprehensive?
	Overall simplicity	Is the proposed governance model considered simple enough to be understandable and accessible by stakeholders?
Evolution	Sustainability	Is the proposed governance model resilient to changes and unforeseen events? To what extent is it able to adapt and evolve in response to future challenges?

To test our design artefact, we selected eight evaluation criteria that can be categorized according to the key system dimensions proposed by Pratt et al. (2014). First, the *Goal* system dimension addresses how the design artefact envisions to reach its objectives and with what effectiveness. Two evaluation criteria were defined: the overall efficacy and the overall suitability of the governance model. The overall efficacy criterion assesses if the proposed governance model is perceived as an effective and efficient solution for the integration of new technologies in the fight against tax fraud and social security infringements. The overall suitability criterion assesses if the proposed governance model is perceived as a useful and relevant solution for the integration of new technologies in the fight against tax fraud and social security infringements. Second, the *Environment* system dimension addresses the proposed governance model's fitness, or adequacy, with the cultural, social, and technological environment in which it is supposed to be implemented. Three evaluation criteria were selected: feasibility, stakeholder endorsement, and the cognitive aspects of the proposed model. The feasibility criterion assesses if the design solutions of the proposed governance model can be considered feasible on a political / financial / technological / legal level. The stakeholder criterion assesses if stakeholders agree with and trust the design solutions of the proposed governance model and if they are satisfied with the outcomes. The cognitive criterion assesses if the proposed governance model promotes knowledge and information exchange among stakeholders and if it fosters cognitive learning about the policy and its implications. Third, the *Structure* system dimension addresses how the proposed model is designed and how its various dimensions are articulated to one another. Two evaluation criteria were selected: the overall consistency and the overall simplicity of the proposed governance model. The overall consistency criterion assesses if the proposed governance model is perceived as coherent, consistent, and sufficiently detailed. The overall simplicity criterion assesses if the proposed governance model is considered simple enough to be understandable and accessible by stakeholders. Fourth, the *Evolution* system dimension addresses the proposed governance model's ability to evolve and adapt to changes or unforeseen events. Sustainability was selected as an evaluation criterion for this dimension. This criterion assesses if the proposed governance model is resilient to changes and unforeseen events and if it can adapt and evolve in response to future challenges.

We chose not to include Pratt et al.'s (2014) *Activity* system dimension and its associated evaluation criteria. Indeed, this dimension addresses how specific changes in stakeholders' activities brought by the model (via new rules and regulations) may affect the day-to-day practices of practitioners. However, our governance model mostly proposes general policy orientations and principles regarding the integration of new technologies in Belgian federal public policies. As such, the design solutions that are proposed in our design artefact do not provide much detail regarding the specific changes that might affect the daily practices and activities of social and social fraud fighting practitioners. Furthermore, assessing the models' effect on day-to-day practice supposes that changes have already been implemented et experienced by practitioners, which is not the case.

4.4. SUMMARY – EVALUATION CRITERIA FOR TESTING

In a nutshell, eight evaluation criteria were selected for the purpose of testing our design artefact. These criteria are the following: Overall efficacy; Overall suitability; Feasibility; Stakeholder endorsement; Cognitive aspects; Overall consistency; Overall simplicity; Sustainability.

The next section presents how the data needed to evaluate our design artefact will be collected (with the help of a Delphi Survey) and analyzed. These evaluation criteria will play a key role in the design of the survey questions and during the analysis of participants' answers.

5. MODEL TESTING - DESIGNING THE DELPHI SURVEY

The present section builds on the previous one to address the evaluation part of Hevner's design cycle (see figure 4). Here, we describe the steps and the methods that we will use in conducting an evaluation of our design artefact. In other words, we present how the data needed to evaluate our design artefact will be collected and analyzed.

More specifically, we provide some details about the survey method that was selected for testing purposes (the Delphi method), the construction and the design of the survey, the survey participants, the survey questionnaire, and how the data will be analyzed for the purpose of model evaluation.

5.1. THE DELPHI METHOD

The Delphi method is "a social research technique whose aim is to obtain a reliable group opinion using a group of experts. It is a method of structuring communication between a group of people who can provide valuable contributions in order to resolve a complex problem" (Landeta, 2006, p. 468). Initially developed by researchers at the RAND corporation in the 1950s and used as a forecasting tool, the Delphi method is the oldest and most widely known of the so-called expert methods (Brady, 2015; Landeta, 2006; Linstone & Turoff, 1975). This technique relies on a multi-round survey process to obtain positions regarding a topic of interest from a panel of experts in a given field. By eliciting such positions, the researchers try to determine the level of consensus or dissensus regarding the investigated topic among the surveyed experts.

Although there are no strict guidelines, Delphi surveys are generally designed with two survey rounds (Shariff, 2015; Sharkey & Sharples, 2001) and can go up to six rounds (Skulmoski et al., 2007). The number of rounds is usually determined in an iterative manner, depending on the objectives of the study. If the goal is to establish consensus among experts, then three or more rounds may be needed. However, "if the goal is to understand nuances (a goal in qualitative research)", then two rounds might be sufficient to uncover sufficient information or reach theoretical saturation (Skulmoski et al., 2007, p. 11). One round can also prove sufficient if the collected data is sufficiently rich. Usually, the first round may use an open-ended format to collect qualitative comments from respondents regarding the issue at stake. Data collected via interviews or focus groups can be used to inform the first round of the survey (Hasson et al., 2000). In between each round, researchers analyze and summarize participants' answers. The obtained results are then fed back to participants through a process of controlled feedback. Summarized responses serve as the groundwork on which the questions of the following rounds of the survey are based. Such a process can be "repeated until consensus is reached or until the number of returns for each round decreases" (Hasson et al., 2000, p. 1010). Participants are not physically brought together during the process. This facilitates the feeding back of opinions in a non-adversarial manner and reduces the risk of hierarchical bias among panel of participants. In turn, this also promotes participants' mutual understanding and learning about the groups' collective opinion.

The Delphi method is a flexible and pragmatic approach that can be used to answer many research questions with both qualitative and quantitative data sources. Although it is often used in conjunction with some quantitative techniques, many scholars rely on qualitative methods to interpret and analyze the collected data (Brady, 2015; Skulmoski et al., 2007). Indeed, the Delphi method lends itself well to interpretivist approaches by which researchers are "interested in how the social world is interpreted, understood and experienced" (Skulmoski et al., 2007, p. 9). It provides a structured process to rigorously capture rich and detailed qualitative data that allow qualitative researchers to interpret an object of study in terms of the meaning that respondents

place onto them. The thematic analysis (Paillé & Muchielli, 2021; Strauss & Corbin, 1998; Miles & Huberman, 1994) is generally recommended for qualitative Delphi data analysis (Brady, 2015; Linstone & Turoff, 1975).

In the past decades, the Delphi method – that initially centered on the collection of scientific and technical expertise – progressively evolved to integrate a wider pool of participants such as practitioners and individuals with lay expertise on the investigated topic. Nowadays, the panel of participants is no longer necessarily exclusively composed of experts. Depending on the investigated topic, the panel can be made up of all individuals that are deemed relevant either due to their specific competences, their involvement, or their concrete experience (Brady, 2015; Skulmoski et al., 2007; Alder & Ziglio, 1996; Linstone & Turoff, 1975). The aim of the method thus becomes the mobilization of a field expertise via the collection of the multiple perceptions and representations that participants have of a situation. This method thus provides a very complete source of information on the issue at hand while allowing and a first approach to contradictory debate thanks to its multi-round approach. Furthermore, by appealing to respondents' concrete experiences, this approach tends to generate a greater investment and involvement among the associated actors. So far, the Delphi method has been widely used in the field of health sciences, but it has also been successfully implemented in social sciences and information technology (Brady, 2015; Landeta, 2006; Muckherjee et al., 2005; Hasson et al., 2000). The Delphi method is also regularly used for public policy analysis and evaluation. Indeed, the Delphi method allows to apprehend the studied policy sector in all its complexity, and thus provides policy makers with a better understanding of policy design and implementation (Brady, 2015; Alder & Ziglio, 1996; Linstone & Turoff, 1975). Furthermore, by minimizing hierarchical relationships in the communication process, the Delphi method “promotes intersubjective understanding between the decentralized units and with the centralized management” (Fallon, 2018). This can encourage innovation with the emergence of new forms of collaborative governance and knowledge sharing.

The Delphi method provides many advantages that tie particularly well with the objectives of the testing phase of the living lab process that is being implemented as part of the DIGI4FED project:

- First, this method allows to collect a wide variety of opinions by facilitating the expression of a diversity of points of view and experiences. This makes it possible to account for a given problem in all its complexity, by bringing out its multiple constitutive facets. This aspect will be critical to ensure the most comprehensive assessment of our design artifact. It is on that basis that we aim to provide a POC governance model and policy recommendations that account for the diversity of stakeholders' perspectives.
- Second, the Delphi method provides a comprehensive source of information that can serve to initiate debates on a solid and reliable base. As such, Delphi studies are often geared towards informing policy, practice, or decision-making (Brady, 2015; Alder & Ziglio, 1996) because this method can lead to the creation of a consensus on the recommendations, opinions, or modes of action that may emerge from the multiple rounds of the survey. This aspect is particularly relevant in the case of the DIGI4FED project. Indeed, survey participants' reactions to our design artefact will be critical in the development of a POC governance model that reflects a certain degree of consensus among stakeholders. In addition, the conclusions brought by the survey should contribute to the production of policy recommendations to the Belgian federal authorities that may ideally carry a high level of social acceptability.
- Third, Delphi surveys allow participants to express their views on future scenarios that are, in their eyes, both possible and desirable. Such a process can result in high levels of personal investment, with participants taking ownership of the debates and their conclusions. Additionally, Delphi surveys can prompt a learning effect among participants about the issue in question (e.g., Van Dijk, 1990; Schneider, 1971). These elements authorize and encourage a wider dissemination of the debate to the public sphere, thus

facilitating the transition from collective reflection to joint action. In the DIGI4FED project, the testing phase of the living lab process aims to elicit stakeholders' reactions to a proposed model for the future of digital governance regarding the fight against tax fraud and social security infringements. In that perspective, it is key to reach high levels of investment and involvement among stakeholders during that critical step. Although the goal of DIGI4FED is not necessarily to generate joint action among the public, the project strives to contribute to the public debate about the possible and desirable futures of digital governance.

5.2. A FEW CAUTIONARY POINTS

A few cautionary elements should be mentioned before getting started with the construction of the survey. First, when presenting our design artefact to survey participants, we should primarily put the emphasis on its potential in resolving issues pertaining to the integration of new technologies in the fight against tax fraud and social security infringements. Indeed, the proposed model suggests some radical changes (i.e., transition from a centralized system architecture to a decentralized one) and offers solutions and opportunities that might affect many policy sectors beyond the mere fight against tax fraud and social security infringements. As a consequence, a possible issue is that we might involuntarily elicit reactions and comments about the impact of our design solutions in other policy sectors during the testing phase. For instance, if widely endorsed by public authorities, the proposed governance model would probably also have deep implications regarding the management and use of citizen's health data in the health care sector. However, we did not interview stakeholders who are involved with the management and use of data in this policy sector. Hence, it would be difficult for us to provide definite answers regarding potential implications of such a governance model for health care data management.

Second, it is critical to reaffirm that our main motivation is to construct a governance model and not to provide a turnkey IT system solution to implement as is by public authorities. Ultimately, the federal and the regional governments should, and will, be the ones to make the choices that will define the future of digital governance in Belgium. This project merely elicits possible design solutions and orientations that appear well suited to address a series of identified issues. As such, it is the systematic identification of key drivers, challenges, and solutions via the analysis of empirical material that allowed the researchers to formulate a series of informed design solutions. It is important to highlight this aspect to avoid any misunderstandings that could arise due to misplaced expectations about the outcomes and objectives of the project.

Third, given the innovative nature of the proposed model, a few precautionary steps should be taken to ensure participants' best comprehension of the issues at stake in the survey. Participants might be unfamiliar with some of the concepts and technologies that are evoked in our design solutions. To address this issue, survey participants should be provided with enough technical information and context before asking our questions. As such, all participants should be introduced to the key assumptions of the proposed OGD governance model. Ideally, participants should also be presented with concrete cases and examples so that they can embed our design solutions in their actual day to day practices. For instance, in addition to a description of the current administrative processes used in the fight against tax fraud and social security infringements, we could provide a short explanation of how those processes would look like using the design solutions inspired by the proposed model. It is thus imperative to ensure that our survey questionnaire addresses all the relevant issues and that it is understandable by all participants before its submission. To do so, we aim to pre-test the questionnaire via a pilot application (Landeta, 2006) with external actors such as other researchers, EBSI experts and foreign tax and social fraud officials.

5.3. DESIGNING THE DELPHI SURVEY

With the Delphi survey, our ambition is to conduct an analysis that will examine the various *tension points* (see: Plesch et al., 2013; Sharma & Yang, 2015) of the proposed governance model. The idea is to identify key dimensions that of the proposed model that are susceptible to produce *tensions*, or *frictions*, caused by the design characteristics of the model and its possible socio-political implications. For that purpose, the summary of the design artefact section (section 3.7) already provides some key insights regarding the aspects of the model that will likely be tested. Five main expected benefits of the proposed model were put to the fore:

1. The OGD platform helps fostering trust among prosumers, and thus improves the overall endorsement of the system by its users
2. The proposed design artefact would probably improve the GDPR compliance of new digital technologies in the fight against fraud and facilitate the compatibility of national legislations in relation to the digital framework
3. The proposed OGD platform also addresses the concerns regarding public sector's current dependence on external IT providers while preserving its resources
4. The proposed design artefact should allow for the use of predictive probabilistic applications of AI without compromising on transparency regarding how this technology is used for decisions
5. The OGD platform contributes to the creation of common socio-technical standards and a greater interoperability among the platform stakeholders at the national and at the EU level

However, and as previously developed, each of these dimensions also comes with possible shortcomings that might hinder the ability of the model to reach is expected benefits. The challenge will be to assess these friction points with participants while taking stock of the evaluation criteria that were selected in the previous section. The following table presents a simple framework that will be used to guide the construction of the questionnaire for the Delphi survey.

Table 9. Anticipated tension points

Dimensions	Anticipated tension points (and examples)
1. Trust among prosumers and endorsement of the system	<u>Uncertainties regarding the technical maturity of the proposed design solutions.</u> For instance, how close are we from an eventual testing or a pilot stage for this platform?
	<u>Uncertainties regarding our technical readiness to the proposed design solutions.</u> For instance, does BOSA actually hold the adequate technical skills that would be required for the large-scale distribution of digital wallets and deployment of this decentralized platform?
	<u>Uncertainties regarding our socio-political readiness to the proposed design solutions.</u> For instance, to what extent can we expect some resistance to the proposed solutions within the administration? Will governments be able to reach a political consensus regarding the adoption of such a platform and how much time would policy development take?
2. GDPR compliance and compatibility of	<u>Uncertainties regarding the actual compliance of the proposed design solutions with GDPR requirements.</u> For instance, the assessment regarding the GDPR compliance of EBSI solutions is still pending at the EU level.

national legislations	<u>Uncertainties regarding the compatibility of national legislations in relation to the digital framework.</u> For instance, and, although a regulatory sandbox approach might help in developing an adapted regulatory framework, there are still many unknowns regarding the design and methodology of experimental law-making.
3. Public sector's resources and dependence on external IT providers	<u>Uncertainties regarding the predominant role allocated to some specific administrations.</u> For instance, does BOSA possess the needed expertise and resources to carry such a central role in the development of the platform? Would creating a new dedicated organization in charge of such developments be less adapted for this purpose?
	<u>Uncertainties regarding our ability to allocate sufficient resources to implement the proposed design solutions.</u> For instance, which resources will be allocated to BOSA to spearhead the development and management of this new system?
	<u>Uncertainties regarding the curation and sharing of data with non-public actors?</u> For instance, what data should be allowed for sharing with private actors so that they can develop better predictive technologies without identifying data issuers? Who should oversee the curation of the data used by these actors?
4. Predictive AI and transparency	<u>Uncertainties regarding the technical maturity of XAI solutions.</u> Further testing is still needed to ensure the technical reliability and effectiveness of (X)AI solutions using BCT.
	<u>Uncertainties regarding the appropriateness of the proposed design solutions.</u> For instance, is the proposed system appropriate with regard to existing operational processes in the fight against fraud? Would it interfere with existing legislative rules in the current framework?
5. Socio-technical standards and interoperability among the platform stakeholders	<u>Uncertainties regarding the involvement of various actors in the development of the proposed design solutions.</u> For instance, which actors will assist BOSA in the development of the EBSI compliant platform? Shouldn't we involve all the potentially interested parties to agree on common standards?
	<u>Uncertainties regarding the acceptance of new socio-technical standards by potential users.</u> For instance, how likely are we to encounter resistance to these new standards due to digital culture differences and digital divide?

Taking into consideration these tension points as well as the structure of our proposed model, the (first) round of the Delphi survey questionnaire will be constructed around six key dimensions:

1. Fostering citizen trust
2. Promoting coordination and interoperability between administrations
3. Limiting public actors' dependence towards private actors
4. Values and data governance
5. Improving the explainability of the analysis process
6. Ensuring the legal compliance of the system

Each of these survey dimensions can be linked with a specific aspect of our proposed governance model. For instance, the *Contingencies* of our model are addressed in the first (*Dimension A: Fostering citizen trust*) and second (*Dimension B: Promoting coordination and interoperability between administrations*) dimensions of the

survey. The following table presents the correspondences between the various steps of our proposed governance model and the survey dimensions.

For each dimension, several possible tension points can be highlighted. For instance, one of the key dimensions that will be assessed in the questionnaire is the ability of the proposed model to foster trust among citizens (dimension A). On that regard, we identified several potential tension points that might hinder the proposed model’s ability to yield its expected benefits in terms of trust. Such anticipated tensions points are, for example, the uncertainties regarding the efficacy of the proposed design solutions and the uncertainties regarding our socio-political readiness to the proposed design solutions. These tension points can be linked with the specific evaluation criteria that were selected in the previous section (in that case, these criteria are efficacy and suitability). These evaluation criteria will guide the construction of our survey questions for each of the survey dimensions. Three to four questions will be constructed for each survey dimension.

Table 10. Correspondence between the 6-step model, the survey dimensions, and the evaluation criteria

Steps of the proposed model	Survey dimensions	Evaluation criteria
Contingencies	A. Fostering citizen trust	Overall efficacy; Overall suitability
	B. Promoting coordination and interoperability between administrations	Overall suitability; Overall efficacy; Feasibility
Data prosumers	C. Limiting public actors’ dependance towards private actors	Overall efficacy; Overall suitability; Feasibility
Data governance goals	D. Values and data governance	Overall suitability; Sustainability; Overall simplicity; Cognitive aspects
Design values		
Governance of	E. Improving the explainability of the analysis process	Overall efficacy; Feasibility; Sustainability
Governance by	F. Ensuring the legal compliance of the system	Feasibility; Overall simplicity; Sustainability

The questions will be framed so that respondents can position themselves regarding several propositions pertaining to each of the survey dimensions. Some questions will be formulated in a closed-ended way, thus asking participants to what extent they agree, partially agree, or disagree regarding specific statements. Participants will then be invited to specify why they selected a specific answer. Some other questions will be formulated in a deliberately broad and open-ended fashion (Skulmoski et al., 2007; Adler & Ziglio, 1996; Linstone & Turloff, 1975) to encourage participants to develop their arguments.

Questions will not be formulated in ways that directly and explicitly allude to each of the anticipated tension points identified here. Although, we expect to see similar frictions and tensions emerging from participants’

answers, we will avoid framing our questions in ways that risk orienting participants' answers. Instead, we will use the evaluation criteria that were selected in the previous section to loosely guide the construction of our questions for each dimension. Indeed, these evaluation criteria link with the anticipated tension points because they take stock of both the existing literature for model evaluation and link them with the potential shortcomings that our team or researchers identified regarding the proposed design artefact (section 3.7). Furthermore, relying on these evaluation criteria for constructing the questionnaire is critical to ensure that we can collect the required data to adequately assess the proposed governance model. When issuing our survey questionnaires, we should make sure to specify that there are no right or wrong answers. Participants' responses should be grounded on their concrete experience and knowledge on the matter at hand.

In the end, the Delphi survey should contribute to the identification of specific areas of tension in the model and help us to further explore and consolidate our understanding of those tension points. This will provide us the needed material to construct our POC governance model and formulate the policy recommendations destined to the federal authorities.

No second survey round was planned in advance. Panelists' responses to the survey will be analyzed and summarized by the team of researchers. For that purpose, simple descriptive statistics and a thematic analysis will be implemented to synthesize and structure the collected data. The obtained results will then be fed back to participants a few weeks after the end of the survey. However, depending on the results of the survey (unexpected results, emergence of unforeseen topics, etc.) a second survey round can be envisioned. In that case, the second survey round should be based on participants' answers to the first round. These would serve as the groundwork on which the second-round questions will be based. For the second round of the survey, we could expect to formulate the following type of questions: 1. Questions that were widely discussed in the first round and for which we will propose a synthesis; 2. Additional questions to clarify certain information; 3. New questions written at the request of the respondents themselves because, according to them, they were omitted in the first questionnaire

On a practical level, the Delphi survey will be organized online with the help of the SPIDEL software. SPIDEL⁵ is a new online qualitative survey tool developed in house by the SPIRAL research center at the ULiège. This tool was specifically developed for use in Delphi surveys. After gathering all the e-mail addresses, names and occupation of our panel of experts, we will send an invitation e-mail containing a link to the online survey. After activation, the questionnaire will be available for three weeks. During that period, we will send two reminder emails to the experts who did not react to our inquiry. The collected data will then be analyzed and synthesized to formulate results and new questions that will be fed back to participants a few weeks after the end of the first round of the survey. The following table synthesizes the key steps that will be undertaken for implementing the survey.

⁵ <http://spidel.be/>

Table 11. Survey planning

Approximate date	Tasks
24/02/2022	Activating the online survey questionnaire and sending the invitation emails for the first round
08/03/2022	Sending reminder emails 1.1
14/03/2022	Sending reminder emails 1.2
20/03/2022	Deactivating the online survey questionnaire
From 20/03/2022 onwards	Analyzing participant's answers
Approx. 15/04/2022	Sending feedback of the survey results to participants

The survey questionnaires will be issued in French, Dutch, and English versions. The Dutch version will be translated from French by a team of linguists while the English version will be translated from French by the researchers.

5.4. THE SURVEY PARTICIPANTS

The sample of individuals that will be invited to participate to the survey will be constructed with a purposive sampling approach. In other words, the idea is not to obtain a generalizable sample, but to gain input from a “sample of individuals with specific expertise on a topic” (Brady, 2015, p. 2). This sample will be relatively heterogeneous as it will comprise various categories of stakeholders who are invested in the integration of new IT tools in the fight against tax fraud and social security infringements in Belgium. Given the relative heterogeneity of our sample, a larger sample size may be required for our survey⁶ (Skulmoski et al., 2007; Delbecq et al., 1975). The advantage of larger samples is that they provide better external verification. In comparison, the interpretation of results might require some caution when the number of surveyed experts is limited (Wynekoop & Walz, 2000). Therefore, we expect to send invitations for our survey to a total of 110 individuals. The projected survey participants are as follows:

- Participants to the interviews that were conducted during the exploration phase of the living lab
- Participants to the scenario workshops and solution-oriented workshops
- The follow-up committee members
- People that were recommended to us by the follow-up committee members

We choose to include in the survey all the stakeholders who already participated to the previous steps of the research. This choice was made for several reasons. First, most these participants are already somewhat familiar

⁶ Smaller samples sizes of ten to fifteen people can be sufficient when there is a high degree of homogeneity among panel members (Skulmoski et al., 2007).

with the research and its objectives. This will facilitate contacts and probably improve the response rate of the survey. Furthermore, we have already mentioned to some stakeholders (in particular with the participants to the living-lab workshops) that they would likely be contacted for further steps of the research. Most of them appeared willing to contribute to the next steps of the research. Second, this is also a pragmatic way of mobilizing the pool of expertise that we have already identified in the previous steps of the research. As mentioned in previous deliverables, these stakeholders were already selected because they are either: 1) directly involved in the use of data (for instance the CTIF); 2) impacted or have an impact on the use of data in the public sector (e.g., Crossroad Bank for social Security); 3) affected in one way or another by the policies (Banks, Startups, Tech companies, etc.); 4) academics chosen specifically based on their research area.

Since the Belgian federal administration uses both French and Dutch as working languages, we should expect language differences among panel of participants. We should also take into consideration that some of the experts integrated in our projected panel of participants are English speakers. Therefore, to account for language differences, we developed French, Dutch, and English versions of our survey questionnaire.

5.5. THE SURVEY QUESTIONNAIRE

This subsection presents the English version of our survey questionnaire as well as the contact emails that will be used to invite participants.

Invitation to participate in a survey

Object: Invitation to participate in the BELSPO Brain DIGI4FED survey

Dear {Last name}

Over the past few months, you have been invited to participate in the DIGI4FED project, either through an interview with our researchers or through a participation in one or more participatory workshops. Your contribution was decisive, and we would like to thank you for it!

As a reminder, DIGI4FED is a research project financed by BELSPO whose objective is to determine how Big Data can contribute to improving public services through the integration of new technologies, such as artificial intelligence (AI) and blockchain, within the federal administration. We wanted this research to be as participatory as possible, by involving stakeholders in all stages of the project.

The current step consists in defining a governance model for the integration of new technologies in the fight against tax and social fraud. To this end, we have established the first version of a model proposing possible solutions to a series of issues identified through the interviews and workshops.

We would therefore be very grateful to receive your informed opinion on the central aspects of this model to adapt it as best as possible based on your feedback. Your participation in this survey will take about 40 minutes.

{Survey Link}

The survey will close on March 16, 2022, at 23:00. We will make sure to inform you of the results of the survey during the month of April. If you wish to receive more information, please contact us at mathias.sabbe@uliege.be (or 0479/01.18.19).

We thank you in advance for your participation.

Kind regards,

The DIGI4FED team

Welcome!

This survey is divided into six sections, each corresponding to an important dimension of the proposed model.

Each of these dimensions presents several possible solutions that can be considered cumulatively or complementarily.

- There are no right or wrong answers to the questions. The objective is to gather your impressions based on your knowledge and experience.
- Although the survey can be completed in 40-45 minutes, we encourage you to take as much time as necessary to develop your arguments.
- It is possible to pause the survey and complete it later by using your invitation link.
- Your responses will be treated anonymously and used for research purposes only.

A. Fostering citizen trust

According to interviews and workshops, **the lack of citizen trust regarding the use of their personal data by public authorities** is a central issue in a context where administrations are using new technologies, such as AI. Some citizens are distrustful of the public authorities that use their data. This distrust is further accentuated by a lack of transparency regarding the (re)use of their personal data (in case of non-compliance with the 'only once' principle, for example). In this context, the ability of citizens to retain a certain degree of control over their personal data is often put forward as a factor that might help in fostering citizen trust.

The governance model could therefore ensure that citizens' trust is enhanced by:

- **Improving citizens' ability to control and manage their personal data.** One solution could be the widespread implementation of a digital wallet system for citizens. These digital wallets would give individuals the ability to control and manage public and private actors' access to their personal data. Such systems, which are often based on the principle of self-sovereign identity, are attracting some interest from public authorities. However, there is no strategic consultation between the different levels of government (federal, regional, and local) at this stage.
- **Favoring a decentralized approach to data management that promotes transparency and protection of personal data.** One possibility could be the implementation of a decentralized platform for data exchange between administrations that is inspired by the model proposed by the Crossroads Bank for Social Security (BCSS). Indeed, this system already respects a certain number of good practices that make it possible to meet the requirements in terms of transparency and protection of personal data. This approach could also allow to avoid an excessive centralization of data.

1.1 - Do you think that such solutions can provide an effective response to the lack of trust among citizens?

Yes Partly No No opinion

Could you specify why?

1.2 - Do you think that the generalization of a digital wallet system could be an adequate solution to a lack of citizen trust?

Yes Partly No No opinion

Could you specify why?

1.3 - Do you think that a decentralized data management could be an appropriate response to a lack of citizen trust?

Yes Partly No No opinion

Could you specify why?

B. Promoting coordination and interoperability between administrations

Results from the workshops and the interviews suggest that **the lack of interoperability and coordination between administrations** is a central issue faced by administrations using new technologies in the fight against fraud. This dynamic tends to result in a considerable waste of energy and resources when administrations develop similar digital projects separately. In addition, this lack of coordination also makes it difficult for administrations to share data and make their respective databases interoperable.

The governance model could therefore ensure to improve the coordination and interoperability between administrations with:

- **The adoption of a data exchange platform that would operate based on common criteria and technical standards.** The implementation of such a platform would stimulate data interoperability and collaboration between all actors in charge of fighting tax fraud and social security infringements. Under certain conditions, this solution could also facilitate cooperation with European counterparts who have set up an equivalent system.
- **The creation of an entity in charge of coordinating, developing, and monitoring digital projects within the administrations.** One solution could be to empower some existing Federal Public services (FPS), such as the FPS BOSA, or to create a new entity along the lines of the Smals organization, but which would offer its expertise to all FPSs. This new entity would ensure the harmonization of projects and the development of a common strategic vision. It would also be responsible for the development and management of the data exchange platform or the deployment of digital wallets, for example.

2.1 - In your experience, what is the most important barrier to coordination and interoperability between administrations?

2.2 - Do you think that the proposed solutions could help overcome this obstacle?

Yes Partly No No opinion

Could you specify why?

2.3 - Do you think that such a data exchange platform between administrations is feasible (legally, organizationally, etc.)?

Yes Partly No No opinion

Could you specify why?

2.4 - What do you think would be the most appropriate organizational choice to ensure the coordination, development and monitoring of future digital projects? Could this role be played by the FPS BOSA or by a new entity on the model of Smals (not-for-profit organization) as indicated by some actors? Do you see other possibilities?

c. Limiting public actors' dependance towards private actors

According to workshops and interviews, a central issue in the governance of the ecosystem of actors involved in the fight against fraud is **public actors' dependance towards private IT solution providers**. Public actors tend to lack the internal capabilities to develop their own systems and analysis tools. They rely on solutions provided by private actors and therefore have relatively little control over the development and use of these systems. This can cause issues with respect to the protection of personal data since private actors are likely to have access to this data (whose storage can also be problematic).

The governance model could therefore ensure that public actors are strengthened by:

- **Promoting the internalization of digital skills through the development and technical management of a data exchange platform.** One possibility could be to internalize technical skills by assigning a federal service (e.g., FPS BOSA or a new entity along the lines of Smals) with the task of developing and managing a common data exchange platform across administrations. This would strengthen internal skills while providing a solution that meets high standards in terms of transparency and data protection.
- **Promoting the adoption of national or European solutions that meet strict data protection requirements.** Favoring data analysis or storage solutions developed by national (development of a Belgian cloud, for example) or European (AI4europe, ESSIF, EBSI, etc.) actors could help maintaining administrations' fraud-fighting capabilities while reducing their dependence on private solutions that are problematic in terms of personal data protection.

3.1 – Which of the following statements best illustrates your view regarding the proposed solutions to limit the dependence of public actors towards private actors?

- They limit the dependence of public actors
 They partially limit the dependence of public actors
 They do not limit the dependence of public actors
 There is no dependence of public actors
 No opinion

Could you specify why?

3.2 – Should the solutions proposed by national or European actors be preferred? Why or why not?

3.3 - Do you think that a federal service would be able to assume the development and the technical management of a new data exchange platform?

- Yes Partly No No opinion

Could you specify why?

D. Values and data governance

Several issues pertaining to the expectations of administrations and end users have been identified. The workshops and interviews have underlined **the limited acceptability of new fraud fighting tools among companies and citizens**. These new tools (AI in particular) tend to be negatively perceived and their acceptance is difficult as long as their added value is not directly perceived. On their part, public actors are experiencing **difficulties in organizing the digital governance in a context where financial resources are lacking**. Due to budgetary constraints, administrations tend to choose private IT solutions, which are less costly but sometimes problematic in terms of confidentiality (Google Analytics until recently, for example).

The governance model could therefore ensure that these expectations are met by:

- **Improving the communication about how the system works.** To improve the acceptability of the new tools, one solution could be to deploy more communication and pedagogical efforts in the justification of the choices that govern the large-scale use of personal data in the fight against fraud.
- **Allowing public authorities to take on the role of key regulators in data governance.** Public managers could be granted the ability to control the access to personal data stored on the common data exchange platform mentioned earlier. They could then establish specific rules for actors who wish to access the data for commercial or public purposes. They could also choose to share some specific data with European counterparts to strengthen transnational anti-fraud capabilities for example.
- **Stimulating private actors to provide administrations with efficient tools without creating additional financial constraints.** The public managers of the data exchange platform could decide to authorize the use of some carefully anonymized personal data to private actors who wish to develop new predictive tools using this data as training data. The newly developed tools could then be integrated by administrations, thus reinforcing their fraud detection capabilities without requiring additional resources the immediate future.

4.1 - Do you think it is relevant that the public managers should endorse the role of key regulators in data governance?

Yes Partly No No opinion

Could you specify why?

4.2 - Do you think that a (supervised) contribution from private actors could help administrations to adapt and maintain their fraud detection capacities?

Yes Partly No No opinion

Could you specify why?

4.3 - Do you think that the proposed solutions allow for a governance of data that can be understood by all actors concerned by the protection of personal data?

Yes Partly No No opinion

Could you specify why?

4.4 - Do you think that better communication about the use of new technologies in the fight against fraud could lead to a better understanding and acceptance of these tools by citizens and companies?

Yes Partly No No opinion

Could you specify why?

E. Improving the explainability of the analysis process

The workshops and interviews suggest that the **difficulty to ensure the explicability of decisions based on predictive algorithms** is a central issue presiding over the technological choices for data analysis in the fight against fraud. AI solutions can be effective when used in the fight against fraud. However, they can also result in blackboxes by increasing the opacity of the data analysis processes. As a result, the explanation of fraud related decisions to citizens and companies can be undermined.

The governance model could therefore ensure to improve the explicability of analysis processes by:

- **Promoting the adoption of new predictive tools that allow for a better traceability of analyses.** It would be possible to develop and adopt explainable AI (XAI) solutions which allow to explain the relative weight of certain data in the results obtained by the analysis. In the same way, some progresses are being made in the decoding of blackbox algorithms, thus indicating that it would be possible to improve the explicability of decisions resulting from such algorithms.
- **Improving the reliability of these new tools by conducting pilot experiments in a secure environment.** Such experiments could be conducted using anonymized training data provided by the public managers of the data exchange platform. Conducted on a large scale, such an approach could significantly enhance the reliability and technological maturity of the proposed solutions. These experiments could also be integrated into public procurement processes to test new solutions before they are introduced in administrations. However, the development and adoption of such tools should be consistent with the principle of proportionality defended by the GDPR and should minimize the risks of socio-economic inequalities reproduction.

5.1 - Do you think that the proposed solutions can improve the explainability of decisions based on AI?

Yes Partly No No opinion

Could you specify why?

5.2 - Do you think that the development and implementation of these new, explainable, tools is within the reach of the different actors involved?

Yes Partly No No opinion

Could you specify why?

5.3 - To what extent would the increased explicability obtained by these new tools be compatible with the effectiveness in the fight against future fraud?

F. Ensuring the legal compliance of the system

According to the workshops and the interviews, two issues raise questions about the **compatibility between the current legislation and the technological choices in the fight against fraud**. The first issue is the interpretation and implementation difficulties associated with the General Data Protection Regulation (GDPR). The lack of perceived clarity of the GDPR complicates the task of data privacy actors and leads to divergent interpretations between the data protection officers (DPOs) of the different administrations. Another issue is the **difficulty of articulating complex legislations with the adoption of new IT tools in the fight against fraud**. Existing legislation (on social and fiscal matters, public procurement rules, etc.) is sometimes unsuited to the integration of new digital tools.

The governance model could therefore ensure that the overall compatibility of the system with legislation is improved by:

- **Promoting the development of a legislation that is more compatible with the digital framework.** In addition to legal experts, the involvement of other experts (such as computer scientists) in the reflection process could help ensure that legal instruments are compatible with the new digital framework. The use of a regulatory sandbox mechanism could also help to develop and test different regulatory frameworks for the use of new technologies in the fight against fraud. However, there are still some unknowns as to the viability of this experimental approach to law.
- **Simplifying the task of privacy actors by adopting solutions and tools developed at the European level.** To strengthen the compatibility of the system with national legislation and the GDPR, it would be possible to rely on certain tools developed in compliance with EU ethical and regulatory standards (such as ESSIF and EBSI) for the construction of a data exchange platform. Such an approach could facilitate the interpretation of rules and decision-making by DPOs and other privacy actors (DPA, CSI). This could also contribute to the development of a national legislation on taxation and social security that is more compatible with the use of new technologies.

6.1 - Do you think that the proposed solutions can provide a clear and understandable response to the challenges faced by DPOs?

Yes Partly No No opinion

Could you specify why?

6.2 - Do you think that the proposed solutions can provide a sustainable response to the challenges faced by privacy actors?

Yes Partly No No opinion

Could you specify why?

6.3 - Do you think that these new approaches to law development (such as regulatory sandboxes) can be implemented?

Yes Partly No No opinion

Could you specify why?

Thank you for your input!

Are there any other issues you would like to see mentioned?

End of the survey

With this message, we confirm your participation in the DIGI4FED survey. We thank you warmly for your participation.

We will keep you informed of the results of this survey very soon. If you have any questions, please do not hesitate to contact me at: mathias.sabbe@uliege.be.

Your answers have been saved.

The browser above allows you to re-visit the questionnaire.

If you wish, you can come back to your answers using the link in your invitation email until the closing date on **Mercredi 16 mars 2022**

5.6. SURVEY ANALYSIS & MODEL TESTING

Each of the survey dimensions contains three to four questions that are linked to specific evaluation criteria. The following table synthesizes the correspondence between the survey dimensions, the survey questions, and the associated evaluation criteria.

Table 12. Survey dimensions, questions and corresponding evaluation criteria

Survey dimension	Question and evaluation criteria
A. Fostering citizen trust	Question 1.1: Overall efficacy Question 1.2: Overall suitability Question 1.3: Overall suitability
B. Promoting coordination and interoperability between administrations	Question 2.1: Overall suitability Question 2.2: Overall efficacy Question 2.3: Feasibility Question 2.4: Overall suitability
C. Limiting public actors' dependence towards private actors	Question 3.1: Overall efficacy Question 3.2: Overall suitability Question 3.3: Feasibility
D. Values and data governance	Question 4.1: Overall suitability Question 4.2: Sustainability Question 4.3: Overall simplicity Question 4.4: Cognitive aspects
E. Improving the explainability of the analysis process	Question 5.1: Overall efficacy Question 5.2: Feasibility Question 5.3: Sustainability
F. Ensuring the legal compliance of the system	Question 6.1: Overall simplicity Question 6.2: Sustainability Question 6.3: Feasibility

Participants' answers to the survey questions will be analyzed in several ways. First, simple descriptive statistics will be performed based on the answers to the scale survey questions where participants are asked to express their level of agreement regarding statements about the feasibility, efficacy, simplicity, etc. of our proposed design solutions (possible answers are: Yes; Partly; No; No opinion). Any additional comments left by participants to explain their position will be analyzed with the help of a qualitative, thematic, analysis (Paillé & Mucchielli, 2021). This analysis will be conducted in order to extract any additional information that might help us improve, or reframe our design artefact. The following questions will be analyzed using this approach: Question 1.1; Question 1.2; Question 1.3; Question 2.2; Question 2.3; Question 3.1; Question 3.3; Question 4.1; Question 4.2; Question 4.3; Question 4.4; Question 5.1; Question 5.2; Question 6.1; Question 6.2; Question 6.3.

Second, the answers to the open-ended questions in the survey questionnaire will be qualitatively analyzed using a thematic analysis. More specifically, our team of researchers will rely on open coding when assessing panelists' views on the validity of our design solutions. Inductive coding will also be used to extract any additional information that might contribute to revise our design artefact. The following questions will be analyzed using this approach: Question 2.1; Question 2.4; Question 3.2; Question 5.3.

The following table presents each evaluation criteria along with the corresponding question and method of analysis.

Table 13. Evaluation criteria, corresponding questions, and testing method

Evaluation criteria	Questions	Analysis / testing method
Overall efficacy	Question 1.1; Question 2.2; Question 3.1; Question 5.1	Descriptive statistics (supplemented by a thematic analysis of respondents' answers)
Overall suitability	Question 1.2; Question 1.3; Question 4.1	Descriptive statistics (supplemented by a thematic analysis of respondents' answers)
	Question 2.1; Question 2.4; Question 3.2	Qualitative analysis with the help of a thematic analysis
Feasibility	Question 2.3; Question 3.3; Question 5.2; Question 6.3	Descriptive statistics (supplemented by a thematic analysis of respondents' answers)
Stakeholder endorsement	/	Qualitative analysis across all of respondents' answers with the help of a thematic analysis (open coding)
Cognitive aspects	Question 4.4	Descriptive statistics (supplemented by a thematic analysis of respondents' answers)
Overall consistency	/	Qualitative analysis across all of respondents' answers with the help of a thematic analysis (open coding)
Overall simplicity	Question 4.3; Question 6.1	Descriptive statistics (supplemented by a thematic analysis of respondents' answers)
Sustainability	Question 4.2; Question 6.2	Descriptive statistics (supplemented by a thematic analysis of respondents' answers)
	Question 5.3	Qualitative analysis with the help of a thematic analysis

Two evaluation criteria were not addressed by dedicated questions (stakeholder endorsement and overall consistency). Instead, these criteria will be assessed with the help of a thematic analysis of participants' full set of answers across the survey questions. Open coding will be used to assess participants' endorsement of the model and their perception of the model's consistency.

In order to test the model, some benchmarks will be applied. We will apply a 65% benchmark for each criterion. In other words, we will consider that our design artefact respects a given evaluation criterion if 65% of

participants either agree, and partially agree in their answers to all the questions pertaining to a given evaluation criterion. For instance, if less than 65% of participants agree, or partially agree with the idea that the proposed model provides a clear and understandable response to the challenges faced by DPOs (question 6.1), this would mean that the simplicity criterion is not respected for this specific aspect of the model. If less than 65% participants either agree or partially agree with the propositions that are assessed in the questions 6.1 and 4.3, this would indicate that our design artefact does not respect the simplicity criterion.

CONCLUSION

This deliverable focused on the presentation of our design artefact. In other words, based on the inputs from the previous steps of the research (interviews, living lab workshops, vignette studies, and ISM), we proposed a first version of our governance model for the integration of new technologies in Belgian federal public policies regarding the fight against tax fraud and social security infringements. This design artefact is divided into six main steps presenting key aspects of our model: the model contingencies; the data prosumers; the data governance goals; the design values of the model; the governance of (new technologies); and the governance by (new technologies).

To test this design artefact, we propose to rely on a list of pre-selected evaluation criteria. Based on a literature review on the evaluation criteria that are most commonly used for testing in design science research as well as in the planning and public policy evaluation literatures, we selected the following criteria: Overall efficacy; Overall suitability; Feasibility; Stakeholder endorsement; Cognitive aspects; Overall consistency; Overall simplicity; Sustainability. These criteria will be used to evaluate our design artefact with the help of a Delphi survey. This qualitative survey method will aim at collecting experts' and practitioners' positions regarding central aspects of our governance model.

The analysis of survey participants' answers will be conducted using a combination of qualitative (thematic analysis) and quantitative methods (descriptive statistics). Survey results will help us to test the viability of our model based on the evaluation criteria. The qualitative analysis of participant's answers will also provide us some additional input that might help us revise our design artefact and contribute to the construction of a proof of concept governance model.

REFERENCES

- Adler, M., & Ziglio, E. (1996). *Gazing into the oracle: The Delphi method and its application to social policy and public health*. Jessica Kingsley Publishers.
- Ahmad, M., X.-W. Tang, J.-N. Qiu, & 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>
- Aier, S., & Fischer, C. (2011). Criteria of progress for information systems design theories. *Information Systems and E-Business Management*, 9(1), 133-172.
- Albarelo, L., Van Haepere, B., Aubin, D., & Fallon, C. (2016). *Penser l'évaluation des politiques publiques*. De Boeck Supérieur.
- Alexander, E. R., & Faludi, A. (1989). Planning and plan implementation: notes on evaluation criteria. *Environment and planning B: Planning and Design*, 16(2), 127-140.
- Brady, S. R. (2015). Utilizing and adapting the Delphi method for use in qualitative research. *International Journal of Qualitative Methods*, 14(5), 1609406915621381.
- Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). Group techniques for program planning: A guide to nominal group and Delphi processes. Scott, Foresman,.
- Fallon, C. (2018). Interpretative analysis of decentralized policy with the use of an online Delphi, ECPR General Conference, Hambourg, 25/8/2018.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689-707.
- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, 337-355.
- 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>
- Haarich, S. N. (2018). Building a new tool to evaluate networks and multi-stakeholder governance systems. *Evaluation*, 24(2), 202-219.
- Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of advanced nursing*, 32(4), 1008-1015.
- Helms, C., Gardner, A., & McInnes, E. (2017). The use of advanced web-based survey design in Delphi research. *Journal of advanced nursing*, 73(12), 3168-3177.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 75-105.

- Hevner, A., Prat, N., Comyn-Wattiau, I., & Akoka, J. (2018, December). A pragmatic approach for identifying and managing design science research goals and evaluation criteria. In *AIS SIGPrag Pre-ICIS workshop on "Practice-based Design and Innovation of Digital Artifacts"*.
- Hughes, D.L., N.P. Rana, & Y.K. Dwivedi. (2020). "Elucidation of IS project success factors: an interpretive structural modelling approach". *Annals of Operations Research* 285: 35–66. DOI: <https://doi.org/10.1007/s10479-019-03146-w>
- Janssen, M., S. Luthra, S. Mangla, N. P. Rana, & 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, & Y. K. Dwivedi. (2018). Trustworthiness of digital government services: deriving a comprehensive theory through interpretive structural modelling, *Public Management Review* 20(5): 647-671. DOI: [10.1080/14719037.2017.1305689](https://doi.org/10.1080/14719037.2017.1305689)
- Kääriäinen, J., & Niemi, J. (2014). Distrust of the police in a Nordic welfare state: victimization, discrimination, and trust in the police by Russian and Somali minorities in Helsinki. *Journal of ethnicity in criminal justice*, 12(1), 4-24.
- 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.
- Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological forecasting and social change*, 73(5), 467-482.
- Leach, W. D., Pelkey, N. W., & Sabatier, P. A. (2002). Stakeholder partnerships as collaborative policymaking: Evaluation criteria applied to watershed management in California and Washington. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 21(4), 645-670.
- Lemire, S., Peck, L. R., & Porowski, A. (2020). The growth of the evaluation tree in the policy analysis forest: Recent developments in evaluation. *Policy studies journal*, 48, S47-S70.
- Linstone H.A. & Turoff M. (1975) *The Delphi Method. Techniques and Applications*. Addison-Wesley, Reading, Massachusetts.
- March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems*, 15(4), 251-266.
- Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Mukherjee, N., Hugel, J., Sutherland, W. J., McNeill, J., Van Opstal, M., Dahdouh-Guebas, F., & Koedam, N. (2015). The Delphi technique in ecology and biological conservation: applications and guidelines. *Methods in Ecology and Evolution*, 6(9), 1097-1109.
- Olejniczak, K., Borkowska-Waszak, S., Domaradzka-Widła, A., & Park, Y. (2020). Policy labs: the next frontier of policy design and evaluation?. *Policy & Politics*, 48(1), 89-110.
- Paillé, P., & Mucchielli, A. (2021). *L'analyse qualitative en sciences humaines et sociales-5e éd.* Armand Colin.

- Patton, M. Q. (2021). Evaluation Criteria for Evaluating Transformation: Implications for the Coronavirus Pandemic and the Global Climate Emergency. *American Journal of Evaluation*, 42(1), 53-89.
- Plesch, C., Kaendler, C., Rummel, N., Wiedmann, M., & Spada, H. (2013). Identifying areas of tension in the field of technology-enhanced learning: results of an international delphi study. *Computers & Education*, 65, 92-105.
- Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). Artifact Evaluation in Information Systems Design-Science Research-a Holistic View. *PACIS*, 23, 1-16.
- Rosemann, M., & Vessey, I. (2008). Toward improving the relevance of information systems research to practice: the role of applicability checks. *Mis Quarterly*, 1-22.
- Schneider, J. B. (1971). The policy Delphi: A regional planning application. *Technological Forecasting and Social Change*, 3, 481-497.
- Shariff, N. (2015). Utilizing the Delphi survey approach: A review. *J Nurs Care*, 4(3), 246.
- Sharkey, S. B., & Sharples, A. Y. (2001). An approach to consensus building using the Delphi technique: developing a learning resource in mental health. *Nurse education today*, 21(5), 398-408.
- Sharma, R. S., & Yang, Y. (2015). A hybrid scenario planning methodology for interactive digital media. *Long Range Planning*, 48(6), 412-429.
- Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research*, 6(1), 1-21.
- Sonnenberg, C., & Vom Brocke, J. (2011, October). Evaluation patterns for design science research artefacts. In *European design science symposium* (pp. 71-83). Springer, Berlin, Heidelberg.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques* (pp. 1-312). Thousand oaks, CA: Sage publications.
- Van Bueren, E., & ten Heuvelhof, E. (2005). Improving governance arrangements in support of sustainable cities. *Environment and planning B: Planning and Design*, 32(1), 47-66.
- Van Dijk, J. A. (1990). Delphi method as a learning instrument: bank employees discussing an automation project. *Technological Forecasting and Social Change*, 37(4), 399-407.
- Warfield, J. N. (1974). Developing interconnection matrices in structural modelling. *IEEE Transactions on Systems, Man, & Cybernetics* (1): 81-87.
- Wynekoop, J. L., & Walz, D. B. (2000). Investigating traits of top performing software developers. *Information Technology & People*.