Data-driven government

Cross-case comparison of data stewardship in data ecosystems

van Donge, W.; Bharosa, N.; Janssen, M. F.W.H.A.

DOI
10.1016/j.giq.2021.101642

Publication date
2022

Document Version
Final published version

Published in
Government Information Quarterly

Government Information Quarterly 39 (2022) 101642

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Data-driven government: Cross-case comparison of data stewardship in data ecosystems

W. van Donge *, N. Bharosa, M.F.W.H.A. Janssen

Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Keywords:
Data stewardship
Data exchange
System quality
Information quality
Governance
Government-business data ecosystems
Digital transformation
Interdependence

ABSTRACT

Government agencies are becoming more data-driven and need high-quality data to fulfill their roles in society. In the past, each agency organized its own data exchange system according to its own needs. Today, data is distributed over many organizations, and government agencies need to adopt an ecosystem approach for data exchange. Fundamental in the ecosystem approach is the dependence on other parties for the execution of stewardship strategies. Data-driven government agencies increasingly depend on other organizations for high-quality data and data stewardship across organizations is becoming more critical. While there is ample research on data stewardship within organizations, little is known about data stewardship in ecosystems. More specifically, it is unclear which data stewardship strategies government agencies can employ in ecosystems. The main goal of this explorative paper is to identify and compare data stewardship strategies used in empirical government-business ecosystems. Following an explorative case study approach, this paper reveals three different configurations of inter-organizational data stewardship: 1) the government-led ecosystem, 2) the government-business-led ecosystem, and 3) the regulation-led ecosystem. The case studies expose a wide array of data stewardship strategies across ecosystems. While the ecosystem approach provides advantages such as cost-sharing and innovation by private parties, government agencies become increasingly dependent on private parties to gain high-quality data and provide distributed infrastructure components. Maximizing the benefits and minimizing the risks of the ecosystem approach requires government agencies to be cautious when selecting a specific ecosystem configuration.

1. Introduction

Public agencies fulfill a wide array of public tasks, ranging from tax collection and social benefits allocation, to the procurement of services (Lindgren & Jansson, 2013). By its very nature, these tasks are data-intensive. For the planning, implementation, execution, and enforcement of policies, agencies need high-quality data. The latter refers to correct, timely, complete, and suitable data for automated data processing. Data exchange among organizations can improve coordination of process execution and service delivery (Zheng, Yang, Pardo, & Jiang, 2009). Additionally, we know that poor data quality can lead to inefficiency and economic losses (Haug, Zachariassen, & van Liempd, 2011) and incorrect decisions with a negative impact on the lives of citizens (Jaeger & Bertot, 2010). Data is often collected by citizens, companies, and governmental agencies. These parties might be required by law to submit high-quality data. Still, they can also be asked to voluntarily contribute (Klievink, Bharosa, & Tan, 2016) or provide data for other reasons, like gaining advantages. The digital age provides new opportunities for governments to improve their content and services (Chen, 2002).

Given the distributed nature of collecting data, governments turn to a data ecosystem approach. In data ecosystems, multiple participants interact to produce, exchange and eventually exploit data (Oliveira & Lóscio, 2018). Ecosystems grow organically and have no common goal. Instead, actors in an ecosystem have organizational goals (that may overlap or compete). In ecosystems specialized parties provide one or more processes for the data exchange. Examples of these processes include data preparation, data standardization (syntax and semantics), data storage, data quality assurance, validation, identification, authentication, authorization, reconciliation, routing, archiving, and delivery (Bharosa, Hietbrink, Mosterd, & Van Oosterhout, 2018).

While most of these data exchange processes were previously facilitated by agency-specific infrastructures, data stewardship is not straightforward in a data ecosystem. Parties might store similar data or

* Corresponding author.
E-mail addresses: wendy@digicampus.tech (W. van Donge), n.bharosa@tudelft.nl (N. Bharosa), m.f.w.h.a.janssen@tudelft.nl (M.F.W.H.A. Janssen).

https://doi.org/10.1016/j.giq.2021.101642
Received 31 January 2021; Received in revised form 28 April 2021; Accepted 21 September 2021
Available online 30 December 2021

have valuable data for each other, which calls for data sharing and collaboration. However, not all data might be shared because of regulations. The lack of formats and standards and challenges like organizational change, legal regulations, security, privacy, outdated information infrastructures, and lack of IT funding makes data sharing more difficult (Chen, 2002). The shift towards data ecosystems requires dealing with the scattered execution of data exchange processes and the accompanying responsibilities.

Data stewardship refers to a set of tasks focused on assuring that the right data gets to the right processes in the proper format, and is compliant with the prevailing laws and regulations (Dawes, 1996, 2010). Data stewardship is about taking responsibility for the data and arranging the necessary activities for data exchange processes (Van Donge, Bharosa, & Janssen, 2020). This type of stewardship links responsibilities to data (Rosenbaum, 2010), which may include the responsibility for acquiring, storing, safeguarding and the use of data, as well as monitoring compliance with law and regulations. For government agencies, such data stewardship responsibilities are usually bounded by laws and regulations (Dawes, 1996).

Nonetheless, embracing a data ecosystems approach requires rethinking the distribution of those responsibilities. Both public and private organizations can fulfill data stewardship responsibilities, since both types of organizations might collect and process part of the data. Operating as a data-driven government in data ecosystems demands new stewardship strategies. A strategy is a set of choices and decisions that together chart a high-level course of action to achieve high-level goals (Mosley, Brackett, Earley, & Henderson, 2009). There is a lack of research about data stewardship strategies in ecosystems, in which agencies do not control the entire spectrum of components for data exchange.

The explorative research presented in this paper aims to identify inter-organizational data stewardship strategies in data ecosystems. Moreover, we search for distinct configurations of strategies found in empirical data ecosystems. This paper proceeds as follows. Section 2 provides a more in-depth review of the concept of data stewardship. Section 3 outlines the research approach. Section 4 presents the case studies on government-business data ecosystems. Section 5 presents the comparison framework for studying data stewardship configurations and strategies. The framework is substantiated by conducting a cross-case analysis. Subsequently, Section 6 discusses the main findings. Finally, Section 7 presents the conclusions, limitations, and avenues for further research.

2. The need for data stewardship in organizations and ecosystems

In 1996, a study by Sharon Dawes proposed the policy principle of data stewardship for government agencies striving to increase openness and transparency through information access and dissemination (Dawes, 1996). This policy regards governments as stewards instead of owners of data. "*Stewardship focuses on the accuracy, integrity, and preservation of information holdings.*" (Dawes, 1996, p. 393). Instead of having a fixed responsibility for the data, stewardship implies that an organizational entity is responsible for the accuracy, validity, security, use, description, and preservation of the organization's data, regardless of the source (Dawes, 1996). Stewardship demands that government information be acquired, used, and managed as a resource in accordance with organizational, jurisdictional, or societal value across purposes and over time (Dawes, 1996). It thus promotes two essential requirements for information-based transparency: it protects government information from damage, loss, or misuse, and it makes information "fit for use."

There is no centralized responsibility in stewardship (Dawes, 1996), but rather a decentralized and scattered set of responsibilities, as activities needed can be performed by different entities. This makes the concept also suitable for data ecosystems in which public and private organizations (voluntarily) share information and can have different goals and responsibilities. Wende (2007) showed how data stewards are held responsible for the overall data quality. Governance structures are an important component of data stewardship for ensuring clear responsibilities, allocation of decision-making, and ensuring accountability for data quality management.

The original, intra-organizational focus of data stewardship was in line with the conceptualization of information systems at the time. Many organizations, both public and private, had their internal information system and had sole responsibility for data stewardship. Over the years, the boundaries of information systems have expanded beyond the border of a single organization. "*Information systems have been migrating from a hierarchical/monolithic to a network-based structure, where the set of potential data sources that organizations can use has dramatically increased in size and scope.*" (Batini, Cappiello, Francalanci, & Maurino, 2009, p. 16:2).

The same is true for data-driven government agencies that must continuously collect data. For instance, Dutch government agencies such as the Tax Office, Chamber of Commerce, Office of Statistics, Office of Education, and the Social Housing Office rely on the same inter-organizational data exchange components (e.g., eID, data preparation software, data taxonomies, data reporting software, API, validation services) that are provided, managed, and updated beyond their organizational boundaries (Bharosa et al., 2013). The new aspect is that these government agencies have moved away from building and managing these components by themselves and collaborate with others who also want to use (part of this) data. In this light, data stewardship should ensure that data can be used across organizational/sectoral boundaries and system components that assure reliable and secure operations are universally applicable (i.e., not designed for a specific purpose only).

Oliveira and Lóscio (2018) showed that data ecosystems have two essential properties. First, the networked character, where actors are loosely coupled and the systems have multiple levels and dimensions. Such networks need clear governance arrangements for ensuring the use of similar components enabling data exchange. Secondly, the self-organizing environment, as every data ecosystem has feedback loops. There is no hierarchical control but a need for horizontal governance. Where Oliveira and Lóscio (2018) investigated data ecosystems from a general perspective, Ryazanova provides insights into public-private ecosystems. Ryazanova, Pétercsák, Heaphy, Connolly, and Donnellan (2016) show how, in these ecosystems, value is created by the collaboration among multiple participants. Hence, data ecosystems have characteristics of having various public and private participants creating value in a loosely coupled, multi-leveled, and self-organizing environment.

What is unique about these collaborations is the horizontal relationship and governance. There is no hierarchical or centralized governance, like within a single organization. With numerous parties being involved, various interests need to be addressed and balanced (Klievink et al., 2016). Especially between the private and public sector, the values pursued are different. The interdependence in (government) critical processes calls for arranging data stewardship to ensure data usage and keep an overview of the data exchange process. It is therefore important to understand which data stewardship strategies governments can execute in government-business ecosystems. We can understand how data stewardship is arranged by studying the configuration. Configuration in this research refers to the set of design variables present in the government-business ecosystem. It provides an answer to what design choices were made in the context of the data stewardship strategies.

Based on the configuration approach and Dawes (1996), we define data stewardship in ecosystems as "*the distribution of authorities over the network organizations to ensure data quality, system quality, and governance quality* ". Stewardship configurations should secure and protect information, and enable its use in the operation and for innovation. This requires defining three aspects. First, which organization takes care of certain data. Second, which organization manages the data quality, and

third, how data can be used. Data stewards should ensure compliance with laws, including that sensitive data will not be shared or that data will not be misused. Nevertheless, data stewards should enable the use of information by employing privacy enhancement technologies.

3. Research approach

3.1. Explorative research

The research subject of the present study is *'the configuration of data stewardship strategies in empirical government-business data ecosystems'*. As this is a new phenomenon and only a few empirical cases are available, an explorative research approach is taken. Three cases are investigated and compared with each other. By continuously moving back and forth from the cases to the literature, a comparison framework emerged. This framework is partly based on previous work (Van Donge et al., 2020), showing the need for understanding data governance, data quality and system quality. These three categories contain multi-dimensional concepts (e.g., data quality can include over 30 dimensions), and it is not clear up-front which dimensions should be selected and included in the comparison framework. Therefore, the development of the comparison framework also depended on the insights gained from the explorative case studies. The result of the framework is an overview of the important dimensions in governance, system quality, and data quality for the public-private ecosystem. For each case, we will elaborate on the configuration of the data stewardship strategy and analyze the choices made for each of these dimensions of governance, system quality, and data quality. Several iterations were needed to select the appropriate dimensions for identifying data stewardship strategies in practice.

3.2. Case study selection

We explored three empirical cases in this study in search of data stewardship strategies. Table 1 outlines the case study selection criteria.

Based on the criteria in Table 1, we found that—in the Netherlands—only three cases matched these criteria. These three cases are (1) Single Window, (2) Standard Business Reporting, and (3) *E*-procurement (Peppol). The most relevant aspects of the cases are described in Section 4.

3.3. Data collection for the explorative case studies

3.3.1. Document analysis

The explorative case studies were investigated by analyzing documents and conducting expert interviews. First, we searched for the publicly available documentation on each of the cases. For Standard Business Reporting and E-procurement we found most documents online, whereas most of the documentation for the Single Window case was provided by the interview respondents. The document corpus included descriptions of the horizontal governance, data exchange, standards and service level agreements (SLAs). The analysis confirmed our premise that data stewardship strategies can be found in three categories: data governance, data quality and system quality. Inspired by the reference to the various data governance, data quality and system quality dimensions found in the documents, we made an initial selection of dimensions in each category that is already defined in the literature. Section 5 provides an overview of the dimensions selected for the comparison framework based on the reference to these dimensions in the case study documents.

Table 1
Case selection criteria.

#	Criteria	Definition
1	B2G and G2B	Bi-directional data exchange: business-to-government and government-to-business
2	Data-driven government	The data demanded by government agencies is imperative for handling its primary processes and fulfilling public tasks.
3	Multiple actor groups	At least three groups of actors are involved; government agencies, data providers (businesses), intermediaries, and/or software providers.
4	Volume	More than a million messages are exchanged each year.
5	Openness	Any organization, public or private, can join a part of the ecosystem against a minimal set of entry conditions (i.e., adopt one or more standards).
6	Variety	There are a variety of channels (e.g., system to system or portal based) and components that can be used for data exchange.

3.3.2. Semi-structured expert interviews

The many dimensions for each category were used to develop a semi-structured interview protocol. This interview protocol (available upon request) was used to conduct semi-structured interviews with case experts. We conducted a total of nine semi-structured interviews (three for each case) to discuss the data stewardship strategies. Semi-structured interviews are well suited for explorative studies since they allow to have a dialogue (Whiting, 2008) with the respondents during the interviews. This approach enables respondents to reflect on their own experiences allowing new insights to emerge (Dearnley, 2005; Krauss et al., 2009). Additionally, during the interviews the findings of document research were validated. Interview respondents were selected based on three criteria. First and foremost, the respondents must have over five years of experience in the ecosystem, ensuring they have deep knowledge of the subject matter. A second criterion was that respondents had a tactical level role (e.g., program managers or project managers), ensuring that they had expertise on the ecosystem's governance and the technology levels. The third and final criterion was that all respondents represented the government in the ecosystem. This allowed us to capture the government's perspective on data stewardship strategies. Based on these criteria, we reached out to nine respondents. All nine respondents were open to an interview. An overview of the respondents is available on request.

The interview questions included the data stewardship strategies employed in the areas of governance, data quality and system quality elements. We asked respondents to relate to their own experiences during the interview of sixty (60) to ninety (90) minutes. The interviews were audiotaped, transcribed, and sent to the respondent to be validated. Four out of the nine respondents have given more details as a response to the validation email. The interview transcripts are compiled using Microsoft Word and are available on request.

4. Case studies descriptions

Section 4 presents the case descriptions of Single Window, Standard Business Reporting and *E*-Procurement (Peppol).

4.1. Case 1: Single window

To fulfill the European directive 2010/65, several European countries participated in "the Advanced National Network for Administrations (AnNa) project" in the period of 2012–2015. Here the foundations were built for the Maritime Single Window (now Single Window because air cargo messages are also in scope). The Single Window ecosystem is developed to exchange maritime and air notifications to Customs and the Royal Netherlands Military Police.

The objective of the Single Window (SW) is to have a single point of contact. This means data providers report data once. The SW will duplicate messages for multiple information requestors. One message enters the SW, which converts to 22 message types, each for one or more receivers. This data contains information on the arrival of ships or aircraft, the cargo, the people on board, customs declarations, etc.

The data suppliers need to provide data through an intermediary such as a Port Community System, a regional hub, or a web portal directly linked to SW. This data is often time-critical. If data is not

Table 2
Facts and numbers of Single Window.

Type of data	Information on the arrival of ships or aircraft, the cargo, the people on board, customs declarations, etc.
Amount of messages	47 million (2018)
Message types	22
Stakeholders	• the suppliers of data; captains and commanders, shipping companies, shipping agents, airlines, air cargo handlers, ship suppliers, provisioners and bunkerers. • Port community systems, intermediary, hubs; providing IT-solutions for communicating with the government. • Data receiving Government organization, Customs Tax authorities (Belastingdienst Douane), Rijkswaterstaat, Royal Netherlands Mare Chaussee. • Service providing government organizations, Customs Tax authorities (Belastingdienst Douane) heretofore Logius. • Policy, Ministry of Infrastructure & Water Management • Port authorities
Gateway to government	Centralized: Single Window, run by government organizations
Data governance	Participated by receiving parties

provided, an aircraft or ship cannot enter. Therefore, the availability of the system needs to be high. In this process, there are several stakeholders, as shown in Table 2. Not all of the stakeholders are part of the governance. The governance of the data ecosystem, managed by the ministry of Infrastructure and Water Management (I&W), exists out of the receiving parties and is first and foremost created to align the requirements to ensure data and system quality.

4.2. Case 2: Standard business reporting

Standard Business Reporting is the Dutch national standard for business-to-government financial reporting (Bharosa et al., 2018). Table 3 presents the facts and numbers of SBR. Together with accountants, software providers and banks, the Dutch government developed a framework of standards and agreements for data exchange. By standardizing the labels and language of business administration, data should be unambiguous and reusable for different types of business reports (e.g., tax, annual reports, statistics reports). Standard Business Reporting has three foundational components: the Dutch Taxonomy, XBRL, and the shared digital gateway infrastructure for public agencies.

The first version of the digital gateway infrastructure was launched in 2005 called the Government Transaction Gateway. This was a new way of communication with the government using the XBRL language. In 2006, the first financial statements were filed this way. In 2010, the joint governments invested in the governance of SBR and widespread use. Since 2017, the digital signature based on a national public-key infrastructure was developed for accountants to sign the financial statements and to get rid of the paper signature.

Table 3
Facts and numbers of SBR.

Type of data	Business reports
Amount of messages	44,4 million (2019)
Message types	59
Stakeholders	• Governmental organizations as receivers: Tax authority (Belastingdienst), Chamber of Commerce (KvK), executive education service (DUO) of the Ministry of Education, Culture and Science, The Housing Corporation Authority, The Social Housing Guarantee Fund and The Ministry of the Interior and Kingdom Relations (together in the project "SBR-wonen") • Data suppliers: Banks, taxpayers, schooling systems, housing corporation, etc. • Intermediaries • Software developers • Accountancy organizations • Governmental Shared service provider: Logius • Policy: The Ministry of the Interior and Kingdom Relations, Ministry of Finance, Ministry of Economic Affairs. • Other organizations: XBRL-Netherlands, NLdigital, Council for Annual Reporting.
Gateway to government	Centralized: Digipoort, run by government organizations
Data Governance	Participated by stakeholders and stakeholders representatives

As a government-business data ecosystem, SBR has a sound legal foundation. Information providers must provide data to government agencies and the latter must provide the appropriate (electronic) channels for doing so. Data quality and system quality are of the utmost importance since there are legal consequences when failing to provide data to government agencies within pre-defined timeframes. Therefore, SBR employs a plethora of data checks and controls, including non-reputable electronic signatures, data validators, time stamps, and an audit trail system storing data for up to seven years.

4.3. Case 3: E-procurement (Peppol)

In 2008, the Pan-European Public Procurement On-Line (PEPPOL) project was initiated to simplify electronic procurement across borders. The objective was to develop a procurement standard for the governments across Europe to communicate electronically. Procurement messages are all messages on supplier exchange, including the purchase and sale of supplies, equipment, works, and services. This includes order messages, catalogs, and invoice messages. The goal of e-procurement through Peppol is efficiency and cost reduction. In September 2020, the OpenPeppol Association was established. The PEPPOL association is based on three pillars; a network, the document specifications, and the legal framework that defines network governance.

Peppol follows a four (4) corner model, with (end) users in corners 1 and 4, and service providers (or access points) at corners 2 and 3. Through the Peppol network, data that complies with the document specification is transferred. OpenPeppol and the Peppol Authorities are responsible for creating and monitoring the legal framework.

A country can decide to follow the regulations in detail by OpenPeppol or create a local Peppol Authority. Peppol authorities have delegated tasks and responsibilities from OpenPeppol. An authority can develop country-specific rules and standards, contract access points, and supervise. Since October 2020, the Dutch government has become a Peppol authority and started the Netherland Peppol Authority (NPa). Unlike Standard Business Reporting and Single Window, Peppol has no legal foundation in the Netherlands. The government requires either the direct use of the Digipoort, ROAP (the Access point of the Dutch

Table 4
Facts and numbers of Peppol.

Type of data	e-procurement messages
Amount of messages	2.1 million (2019) 6.1 million (2020)
Message types[a]	11
Stakeholders	• OpenPeppol foundation • Access points • Peppol Authorities (Netherlands: NPa fulfilled by Ministry of the Interior and Kingdom Relations, Logius and Netherlands Enterprise Agency) • Independent supervision in the Netherlands: Radiocommunications Agency Netherlands (Agency of Ministry of Economic Affairs and Climate Policy) • End-users (government and Businesses) • Software developers • Intermediaries
Gateway	Decentralized: four corner model, interaction provided by access point/service providers. Centralized: Digipoort, managed by government organizations
Data governance	Participated by stakeholders and stakeholders representatives

[a] These numbers were collected through self-registration by the access points and did not take the global activity by foreign access points into account.

government), or a direct connection with the governmental purchasing system. Using Peppol is based on creating business value for both governmental and business organizations. Table 4 presents an overview of the facts and numbers of Peppol.

5. Cross-case analysis

Data ecosystems need data governance to ensure that the data provided by systems, has a certain quality (Van Donge et al., 2020). We will analyze configurations of data stewardships from the governance, data quality, and system quality perspective. This comparison framework emerged from both literature and the cases. Using the evolving framework, we were able to describe the case studies as three different configurations. In Tables 5, 6, and 7 the different strategies of each case are presented and described for each dimension. In the following subsections we will first explain the dimensions which appeared during the research by going back and forth between cases and the literature.

5.1. Governance

Data stewards in an ecosystem have the task of ensuring adequate governance of data exchange processes. "*Data governance refers to who holds the decision rights and is held accountable for an organization's decision-making about its data assets*."(Khatri & Brown, 2010, p. 149). Where traditional single organizations have less fragmented responsibilities and accountabilities and easier control over data (Khatri & Brown, 2010), government-business collaborations experience more delegated authority which can result into unwanted situations (Brinkerhoff & Brinkerhoff, 2011). **Accountability** refers to actors being called to account to some authority for their actions (Bovens, 2007). This means organizations in the ecosystem have to justify their decisions and actions. The question arises to whom they have an obligation to explain their behavior and who can pose consequences. It is the governance that includes the definition, assignment, and implementation of responsibilities (Rosenbaum, 2010). This is related to the **decision making rights**, which determine how much influence the actors are allowed on the final decision (Bruijn & Heuvelhof, 2008). Governance needs to be transparent and accounted for (AlShamsi, Salloum, Alshurideh, & Abdallah, 2021; Devaney, 2016; Lockwood, 2009).

In the governance context, **transparency** refers to increased visibility of the decision-making processes for stakeholders. Additionally, it requires the availability of achievements and failures in the performance of the authority in the governance (Lockwood, 2009).

Adaptiveness/resilience. Resilience is the change a system can absorb until it sets new processes and structures (Lockwood, 2009). In data ecosystems, governance can either be designed to be resilient or adaptive to change.

Inclusiveness/openness can be defined as enabling all stakeholders to be able to participate in the decision-making process (Lockwood, 2009). Inclusiveness entails fairness – those that are affected by a system should have a voice in the governance of a system. Systems without inclusive participation are prone to transformation resistance and blocking power from those excluded in the governance.

Effectiveness and **efficiency** are concepts are related to output and outcome. "Both effectiveness and efficiency are valued intrinsically in public governance; they are values that constitute the core of public governance's legitimacy" (de Graaf & Paanakker, 2015, p. 2.). Efficiency is about is getting results (supported decisions, changes, etc.) with minimum resource spending (e.g., time and money) (Ernst, 2019).

Funding and **resolving conflicts.** Several authors acknowledge the importance of the allocation of resources and funding (OECD, 2002; Provan & Kenis, 2008) and clear conflict resolution (Provan & Kenis, 2008). Inter-organizational governance must consist of structures to resolve conflicts (Provan & Kenis, 2008).

In the following parts we will take a closer look at the similarities and differences in governance between the three cases. In Table 5 the data strategy is described for each dimension based on the interviews and analyzing documentation.

Table 5
Cross case analysis on data stewardship strategies for governance.

	Single window	Standard business reporting	Peppol
Strategic decision making (decision rights on a strategic level)	Government decides, no private respresentation on strategic level	Partial: some aspects government only, other aspects public-private decision making	Decisions based on consensus between public and private parties
Transparency	Comprehensive description for internal use only	Comprehensive documentation publicly available	Comprehensive documentation publicly available
Accountability	Shared (governmental) accountability. Every organization is responsible for its own processes	Shared responsibilities, but actors are rarely held accountable	Contractual requirements and accountability. Compliance is not strictly supervised
Efficiency of the governance structure	Top down decision making and standard setting	Focus on public-private consensus and commitment; are deemed more important than efficiency	Focus on consensus and commitment
Effectiveness of the governance structure	Align bilateral relationships with own objectives and upfront of communication with market parties	Co-creation with all stakeholders resulting in low perceived effectiveness	No explicit strategy is found
Funding	Only governmental party pay	Several governmental parties pay	Authorities and Access Points pay
Resolving conflicts	The policy-making agency solves conflicts	Fixed escalation process (flow from working groups to tactical level and strategic level)	Escalation process with overruling powers of (public-private) Management Committee on international level
Adaptiveness of the governance	Low adaptiveness. Exceptionally a project group is created	There is the possibility of creating new working groups on specific topics which can result in adaptation	There is the possibility of creating new working groups on specific topics which can result in adaptation
Inclusiveness	Governance only for governments with a legal basis	Open to all, but focus to representatives of sectors, to ensure adoption	Participating in the governance is open to all, entry to governance costs 1500 euro annually

Single Window and Standard Business Reporting benefit from a strong legal foundation for data exchange, allocating most decision-making rights to public agencies. In contrast, e-procurement using Peppol is rooted in private law.

In Peppol, everyone can join the governance by paying a fee, and the governance is clearly defined. From an OpenPeppol perspective, there are several layers, with one management committee (MC). This committee of seven members is chosen by all actors in the governance including the end-users, access points, and Peppol Authorities. However, these represent a large group with all different voices. Respondent 1 clarifies that they are too small to feel heard individually. Additionally, respondent 1 noticed: "*there are 400 service providers and only 15 Peppol authorities. If the service providers would unite, they could potentially decide what happens in OpenPeppol*". Still, the adaptivity of the governance is

high and decisions can be made on multiple levels. The large and final decisions will reach the MC. Multiple parties in different countries can create general formats, rules, requirements, and obligations in different working groups. Working group decisions are based on consensus, but these are often not visited by many stakeholders. As Respondent 6 noticed: "*there are few meetings in which everyone is present, so a decision can be made with not everyone represented. The question is how supported such a decision will be then*". Additionally, respondent 1 states: "*consensus means everyone has to agree, at Peppol this means we will get back to it later*". In the OpenPeppol Foundation, the choice has been made for distributed authority to some level. OpenPeppol's governance sets the framework for countries with their own Peppol Authority to specify further. This distributed authority structure provided countries to have their own governance over their country-specific rules, within the boundaries provided by Peppol. National Service providers need to be contracted with a countries' Peppol authority and meet their requirements, on top of the OpenPeppol requirements.

For SW, the choice was made only to include government parties in the governance. There are multiple levels within the government, each with its own decision-making rights. Here we see a high level of efficiency and effectiveness, governmental funding, and one party as a harmonizing organization. For SW, the Ministry of Infrastructure and Water Management is in the lead to resolve conflicts and harmonize. As respondent 4 states: "*there is a tight governance, in the dossier agreement & procedures everything is written down, including roles, responsibilities and who participates. This has been working to our satisfaction for four years*". Respondent 9 noticed how in the design of SW, the choice was made to keep the market side 'as-is'. Not changing anything in the format or way of interaction, and solely focusing on the government's Single Window function. It is up to the government organizations themselves to align with the market. Respondent 4 states: "*A change almost only occurs when there is a legal basis*". Therefore, voting by private parties is neither needed nor wanted. This means private organizations have to adjust to the government's data exchange system and build their own software systems in accordance with governmental specifications.

Standard Business Reporting has a legal foundation. This means that public agencies such as the Tax Authority and Chamber of Commerce have a legal mandate to demand financial business data using the channels, standards and components they see fit. Nonetheless, much of the governance is based on consensus and the governance structure allows both public and private parties to participate in decision-making on multiple levels. Even though there are multiple levels with their own decision-making rights, respondent 3 states: "*there is much vertical replacement going on, where people of lower levels are asked to replace colleagues of higher levels.* (..) *Discussions will then intertwine, and decision-making rights of each level will get tangled.*" The strategic level of the governance starts to discuss operational data stewardship questions such as the use of formulas in the national taxonomy. With this many parties, we see the complexity of the governance structure increases. Sometimes, the decision-making process becomes slow and bureaucratic. As respondent 8 states: "*there are always parties not happy with decisions. If a decision would lead to a release of administrative burden, companies facilitating services to deal with these burdens are not happy. (..) There are always opposite interests*".

Compared to Single Window, the governance structure in Standard Business Reporting provides far more decision rights to market parties. Since decisions on Standard Business Reporting standards affect software parties and intermediates, they have a vote in the governance. However, the data-requesting party has the responsibility for the formats. The governance has to take all voices into account, but the final decision remains with the government. Respondent 2: "*this is the power of the government, (..) it is the only place you can fill out your tax forms*". So even though many parties are present, intermediates are still not happy with the yearly changes in the data formats, which cost them time and money.

Even though there are multiple levels of governance, two of the three cases show problems with the implementation of the governance. In Peppol, decisions are made on different levels of governance. Yet, they are all made through consensus between the present stakeholders. Without consensus, decisions are escalated to a higher level in the governance. If there still is no consensus, the final decision is made by the elected management committee. Respondent 7 stated: " *Sometimes decisions are made top-down. When this happens, there is lots of critique from lower levels*". On the other hand, participation in the governance is not mandatory, and sometimes attendance is low. Consequently, decisions made earlier are often debated again once actors attend governance meetings, although it looks like there was already agreement before. "*This is not efficient governance*" according to one respondent.

5.2. Data quality

Data quality refers to the quality of the information systems output (Delone & McLean, 1992). There is no consensus on the dimensions of data quality or their definitions. Fisher and Kingma (2001) found that the following dimensions are most frequently used to study data quality: accuracy, completeness, consistency, relevancy, timeliness, and format. **Accuracy** is defined as the extent to which the recorded value corresponds to the real-world value (Fisher & Kingma, 2001). The integrity of the data set should be high, and there should be a lack of errors in the data set, or errors should be easily identified (Wang & Strong, 1996). **Completeness** refers to "*the degree to which values are present in a data collection*" (Fisher & Kingma, 2001, p. 110). An incomplete data set lacks values. As more variables are unknown in a data set, the completeness decreases and the quality of the data set decreases. **Consistency** refers to whether or not semantic rules are defined for the exchanging of data (Batini et al., 2009), which means that the meaning of the data element aligns for all parties (Fisher & Kingma, 2001). **Relevant** data is data that can provide value for the user of the data (i.e., to what extent the receiver of the data can actually use the data) (Fisher & Kingma, 2001). **Timeliness** is a concept on which there is a lot of debate. In this paper, timeliness implies that the data set is not outdated (Fisher & Kingma, 2001). In a timely data set, the modifications in the real world are (directly) represented in the data set. **Fitness for use** refers to the **format** in which data is presented and if it suits the purpose of the data user (Fisher & Kingma, 2001). We consider format and fitness for use as related concepts. "*Format is related to the presentation layout of information outputs.*" (Gorla, Somers, & Wong, 2010, p. 213).

In the following parts we will take a closer look at the similarities and differences in arranging data quality between the three cases. In Table 6 the data strategy is described for each dimension, in this way showing the differences among cases.

Configuring data quality is difficult due to the fact that the data steward is not the owner of the data. Both for Standard Business Reporting and Single Window, the government legally enforces data quality checks at the data provider. Since garbage in is garbage out, it remains the responsibility of the data providing parties (mix of reporting entity, intermediary and software provider) to guarantee the quality of the data.

There are three main data stewardship strategies presented by the respondents for data quality assurance. First, establishing a uniform data format with strict specifications. All three ecosystems follow a certain type of format. Single Window uses the Message Implementation Guideline (MIG) which is created by the governmental organizations, based on the datamodel of the World Customs Organization. Respondent 9: "*The data quality was not so much an issue as definitions were set and knew no variation.(..) The market already knew the World Customs Organization Datamodel.*". Standard Business Reporting allows for more involvement of the market in establishing the data taxonomies used, based on XBRL. Respondent 8: "*In Standard Business Reporting we chose to rely on XBRL taxonomies for data standardization. This implies a working process to harmonize definitions between different sectors and message types, where possible. All aiming to increase data quality*". Yet, in the end, it still

Table 6
Cross case analysis on data stewardship strategies for data quality assurance.

	Single Window	Standard Business Reporting	Peppol
Data Format	Government specific implementation of the Message Implementation Guideline based on the World Customs Organization Datamodel	Taxonomy (XBRL), created by both government and private organizations	Peppol BIS (XML), created by the ecosystem
Accuracy	Two levels of accuracy verification, first on the community system level, and second by the receiving parties	Accountant checks correctness and reliability of data and provides assurance	Automated checks by receiving parties
Completeness	Collaboration by data receiving parties to create a reporting format. Government system checks the completeness of the report	Both providing software as an intermediate gateway (Digipoort) checks the completeness of the report	No check on completeness between data sender and data receiver
Consistency	Enforced by format	Enforced by format	Enforced by format
Relevance	non-relevant messages are rejected by the receiver	Accountant validates relevance. Intermediate gateway (Digipoort) rejects non-relevant data	Non-relevant data rejected by the receiver
Timeliness	Agreements on providing pre-arrival till arrival data	SLAs on reporting period. An accountant evaluated timeliness	SLAs on business process throughput times

remains to be compliant with the laws and regulations of the government. On the international level Peppol is completely created by an international collaboration of multiple parties, both public and private. This means the format is also open for negotiation. Respondent 7 stated that "*only last year we already had 200 change requests, on which there is a voting soon*". For national specifications, there is an alignment between the Peppol authority of the government and the market, where the government still has the final say.

Secondly, each ecosystem has data validation services on syntax and to some extent on the content of the messages. However, not the entire message can be validated, since in all ecosystems the transferring party is not permitted to check the actual content of the message. Digipoort uses XBRL message validators based on business rules to automatically check message completeness and consistency. As respondent 2 states: "*different modules of the message have different numbers, in the end all these numbers are presented in an overview, each of the components have to match*". The Single Window message format contains internal consistency rules. For some message types a conversion is made, in which a check on completeness of fields is performed. However, each data receiving organization validates the data with its own standards. In practice, as respondent 5 noticed, the implication is that data for one governmental organization is fine, whereas another organization would reject the data.

Additionally to the system checks on data quality, in Standard Business Reporting the explicit choice has been made to have an accountant verify the data in advance and give a sign of approval. Respondent 2 noticed: "*it are these accountants which provide valuable input for the taxonomy development*". Respondent 8 noticed how responsibilities for data governance are moving towards the market.

The Peppol network has no validation built in the systems by default. Access providers can choose themselves to provide validation services. It is a push of data, which could mean that organizations transmit whatever they like. Respondent 6 stated: "*It is like a mailbox, which is open for spam as well. You don't know what is thrown into the box*". Here the peppol authority is the point of contact for non-compliance issues. It is the task of the peppol authority to ensure all parties use the formats agreed upon. An access point can implement a validation of the message before sending it to the end-user. The Dutch governmental access point, for example, has to some extent built data format validations into their systems. This enables checking on data quality before delivering data to the public agencies for processing.

The third strategy to ensure data quality is to include various mechanisms for the identification and authentication of data providers in the data ecosystem. For Standard Business Reporting, a compressive public-key infrastructure-based identification scheme managed by the government, ensures that company A cannot deliver fake tax reports or annual accounts on behalf of company B. In addition, registered accountants can only sign for data assurance statements and Digipoort checks if the accountant is still registered. As respondent 7 noticed, in Peppol it is in the interest of the providing party to supply data of high quality. No business will hinder their own business activities since this results in a delay of payment, orders, and other business transactions. Still, the data steward, the Dutch Peppol authority, requires all joint access points to have a know-your-customer process, to ensure parties exchanging data are known within the ecosystem. Especially when the ecosystem transforms from upload portals and email to secure (encrypted) system-to-system (or application-to-application) data exchange, it is increasingly important to know the sender.

5.3. System quality

System quality can be seen as the quality of the information processing system itself, with more engineering-oriented performance characteristics of the system (Delone & McLean, 1992). In the literature, system quality is less coherent and extensive than the information system literature. Delone and McLean (1992) have identified multiple variables for system quality, including flexibility of the system, response time, accessibility, ease of use, system reliability, sophistication, and integration of systems (Delone & McLean, 1992). We added the concepts of auditability and robustness. **Auditability** is important, since government organizations need to be able to rely on the systems as they are part of governmental decision-making. Robustness was added because we believe a system not only needs to be flexible to change, but also have the ability to cope with external pressures and be resilient to environmental changes. Fricke and Schulz (2005) defined **robustness** as a "system's ability to be insensitive towards changing environments". Having low robustness means that a system is not resilient and will fail during pressure from outside.

Ease of use. One of the most important system quality dimensions is ease of use. This dimension refers to the perceived user-friendliness of graphical user interfaces by the end-users entering and sharing data. Data ecosystems entail various user groups across data exchanges, intermediates, third-parties, or software providers creating 'ease of use for end-users, but also intermediates creating easily usable technology for other intermediaries.

System flexibility refers to a system's ability to deal with the variation of the requirements of the business process (Gebauer & Schober, 2006). This includes the extensibility of a system. Having not enough flexibility in the system can limit the success of the system, which prevents the use of the system in certain situations. In the end, this can result in a decrease in the lifetime of a system (Furukawa & Minami, 2013).

Response time of a system refers to "*The degree to which a system offers quick (or timely) responses to requests for information or action*" (Nelson, Todd, & Wixom, 2005, p. 201).

System reliability is in this research defined into two concepts,

system availability and **system integrity**. System availability refers to the uptime of the system (Nelson et al., 2005). In contrast, system integrity can be defined as trusting the data exchanged through the system is free of error and not manipulated (Närman, Schönherr, Johnson, Ekstedt, & Chenine, 2008).

Sophistication of a system relates to whether a system uses state-of-the-art technology. A sophisticated system uses modern technology and has user-friendly interfaces (Gorla et al., 2010).

Accessibility refers to whether data is accessible when needed by the user of the system. Nelson et al. (2005) define accessibility as "*The degree to which a system and the information it contains can be accessed with relatively low effort*" (p. 205).

Integrations/interoperability. In data ecosystems having interoperability is especially important. Interoperability is a multi-dimensional concept and can refer to technical, procedural, semantical, organizational, and legal interoperability. Systems created by different organizations should be able to interact with one another without limitations. Each limitation is a barrier to creating public value through a data ecosystem and will keep organizations from joining a data ecosystem.

Auditability is the ability to track whether someone or something has accessed and/or modified data (Khatri & Brown, 2010). Auditability is an indicator of the integrity of the information system as a whole. Auditability often translates to having accessible audit trails and log files of data exchanges.

In the following paragraphs we will take a closer look at the similarities and differences in arranging system quality between the three cases. Table 7 contains the data strategy for each dimension based on the interviews and document analysis of the cases.

In all ecosystems, the main governmental infrastructure component, the Digipoort (Standard Business Reporting), the Single Window, and ROAP (Peppol), are only one part of the entire ecosystem. Yet, the scope of our analyses focuses on these components since they reflect the majority of data strategies.

All ecosystems are based on some type of covenant. Either it is through an agreement document between (governmental) organizations, a legal contract, or national laws and regulations. These contracts or agreements bind parties in the system to some extent to deliver system quality. It is the task of the data stewards to keep all parties to these agreements or contracts. Note that there is a difference in how stringent the data steward can be depending on the legal instruments at hand.

Peppol has the highest flexibility of the systems since it is still developing its standards. Moreover, prescribing standards in this early stage may pose impediments for parties to adopt Peppol. As the ecosystem matures, it is likely that requirements for data exchange standards will be added over time. As respondent 6 stated "*the electronic invoice is not that time-critical, you have 30 days to pay, so uptime and response time are less important. Once the e-order part of the ecosystem is working the uptime should be high, since you do not want to wait too long for orders to be delivered*". The ease-of use is higher than in the other data ecosystems, as it is not only an open standard, but it is also managed by the working groups in the governance. There is a high level of knowledge sharing, which results in a supported format, applicable for all participating organizations.

At the time of building the MSW in the Netherlands, the choice was made to stay as close as possible to the existing standards and techniques. Respondent 9 noticed: "*The goal of the initiative of SW was to unburdening the market, and implementing the European guidelines within the deadlines*". This means technical innovations and possible investments on both the government as well as business-side were not possible. At this point, changing formats or standards would mean changing the entire data ecosystem. Part of the business side is rooting for change, however, change is will be a large investment from both government and market. In the SW system the uptime and response time are of great importance. As one respondent stated the KPIs are agreed upon by the senior policy officials of the participating public organizations. Therefore, the performance of the system is constantly monitored. As respondent 9 stated there is a mix of message types, and not all message types of the single point of contact require high up-time. Still, the decision was made to run all messages over the same infrastructure. For increasing the quality of the system, there are two government data centers, if one fails, the other takes over, just to always ensure uptime. Additional penetration tests and risk analyses are performed to ensure the quality of the system.

Standard Business Reporting provides data exchange for a range of message types, including dozens of tax report types and annual reports. One of the respondents noticed that the government system processing all messages should be able to deal with diverse message types and peak load. However, as respondent 8 mentioned, there is always a physical limit to the number of messages we can process. Having more government data centers brings more security, but would also increase costs.

Table 7
Cross case analysis of data stewardship strategies for system quality.

	Single Window	Standard Business Reporting	Peppol
Data accessibility	Accessibility is determined by data provider	Data is automatically distributed to relevant public organizations, others can be given access by the the data provider	Accessibility is determined by data provider
Response time	Minimum response time of SW defined in SLAs	Minimum response time of Digipoort defined in SLAs	KPI in contracts
Ease of use by data providers and recievers	Let the market to provide easy-to-use data sharing services service and software	Let the market to provide easy-to-use data sharing services service and software	Let the market to provide easy-to-use data sharing services service and software
Ease of use by intermediates	Use of open standards, however market needs to adjust to governmental requirements	Use of open standards, tuned between the organizations in the data ecosystem. Provide technical workgroups for knowledge exchange	Use of open standards, managed by governance in data ecosystem. Provide technical workgroups for knowledge exchange
Flexibility	Limited through a tightly coupled architecture, including a shared gateway	Limited through a tightly coupled architecture, including a shared gateway	High flexibility created through flexible architecture with multiple access points
Robustness	Mirrored datacenter, Agreements on 24/7 availability	Scalability of the data centers	Four-corner model
System Availability	Uptime requirements defined in SLAs	Uptime requirements defined in SLA	No explicit strategy found
System integrity	SMTP to encrypt data exchange	Two-sided encryption based on public key infrastructure certificates	Open standards for secure exchange of documents, digital signature used by intermediates
Level of sophistication	Actors still use proven (and old) data exchange protocols	Message exchange protocols are outdated, new message exchange protocols are currently tested	Constant improvements through workgroups
Integration / interoperability	Pre-defined message exchange protocol (SMTP), single interface	Multiple pre-defined message exchange protocols based on web service standards	Pre-defined message exchange protocols
Auditability	No explicit strategy was found	Logging of messages enabling audit trail	No explicit strategy was found

Additionally, the system requires high levels of auditability and integrity, since the government makes decisions and can sanction information providers based on the data received. Respondent 3 stated "*If there would be a court case, the judge will accept a digital timestamp of Digipoort on the message as evidence that the information was not provided on time.*". Data stewardship strategies employed to guarantee a high level of integrity and auditability include the use of non-refutable digital signatures, time stamps and a comprehensive audit trail in Digipoort.

6. Discussion

We conclude this paper with three discussion points. First, the transformation of data stewardship in government-business data ecosystems. Second, data strategies and configurations. Third, the changing role of the government in the ecosystem.

6.1. Transformation of data stewardship

The research contributions of Dawes (1996, 2010) laid the groundwork for this study. Based on additional literature (see Section 2) we extend Dawes' conceptualization of data stewardship by focusing on three foundations: (1) setting and running an inter-organizational governance structure, (2) assuring data quality, and (3) assuring system quality.

When looking at government-business data exchange, the role of data stewardship has transformed from intra-organizational data management to inter-organizational data governance. The Standard Business Reporting case is an illustrative example. In the past, organizations like the Dutch Tax Authority focused on their own organizational infrastructure for data exchange. Today, the Dutch Tax Authority relies on a shared government information infrastructure as well as an ecosystem of software providers and financial intermediaries. As a result, any change (for instance, on the data standard) must be planned with the ecosystem. If not, data quality might be lowered. Additional to contracts, there are multiple agreements between the public and private organizations in the data ecosystem. It is up to the data steward to ensure the creation and compliance with these agreements. In this data ecosystem setting, new functions like gatekeeping (monitoring data quality and system quality) are becoming increasingly important, as an ecosystem is not static. Organizations, especially intermediaries, have the possibility to enter and leave the data ecosystem. Data stewards have to maintain a close watch to ensure the interoperability and integrity of the ecosystem.

6.2. Configurations of data stewardship

The case studies reveal three types of configurations: (1) Government-led data ecosystems, (2) Government-business-led data ecosystems, and (3) regulatory-led ecosystems. The three configurations of data stewardship in these ecosystems exhibit significant differences in giving substance to the role of data stewardship. Table 8 presents the configurations in the different ecosystems.

The *government-led* configuration shows a dominant focus on the governmental part of the data exchange process. This is reflected by the limited involvement of private parties in the governance and the centralized financial model. Because of the time-critical nature of data, system quality is especially important in this configuration. It requires robust systems, but as a consequence there is little room for innovations. The data stewardship strategy employed here is to strive for robustness in the SLAs on the governmental part of the ecosystem. The other parts of the ecosystem are the responsibility of the software providers and intermediates. Government organizations can determine what data is needed, and adjust this if necessary. Therefore the government has a strong stewardship position. The disadvantage of this strategy are high costs involved in the maintenance and innovation of the system.

The *regulatory-led* configuration is more focused on data quality from an end-to-end perspective. This configuration also has SLAs for the shared governmental infrastructure to push the robustness of the system, but additionally strives to have an overall high data quality from data provider to data receiver. There is an importance on system quality of having high levels of integrity and auditability, since the data is necessary for decision-making and sanctioning the data provider based on the exchanged information. The formats of the information are fixed in laws and regulations, which means data receiving organizations do not have a direct say in the format. This has to be changed by changing the laws and regulations. This is a large difference to the government-led strategy, where governmental organizations have more freedom to change the requested data. Both public and private parties participate in the governance, though it is only open to thoroughly vetted parties.

The *government-business-led* configuration is highly flexible in comparison to the other configurations. In this configuration, we see how the government functions as data steward on a national level based on an international framework, in which service providers have a high level of freedom. All stakeholders can participate in the further development of the ecosystem on a national level, and on a national level there is a governance structure for stakeholders. In the data exchange ecosystem there are agreements and formats to ensure data quality, however, there is no strong push on data quality in specifications. This creates a highly flexible and interoperable system, but has the disadvantage that government is highly dependent on private organizations for realizing the data-driven government.

6.3. What is data stewardship in data ecosystems?

We define a stewardship configurations as "*the distribution of authorities over the network organizations to ensure data quality, system quality and governance quality*". This research is explorative and has only exposed the tip of the iceberg. We found that the ecosystem approach can be implemented in various configurations. Each of these configurations shows us the different choices in organizing data sharing and provides us some first insights into the trade-offs inherent to interorganizational data stewardship. Having more flexibility and interoperability might lead to less robustness. Also, a high level of market

Table 8
Configuration of stewardship in government-business data ecosystems.

Type of ecosystem	Initiator	Governance	Participation	Focus	System quality	Data quality	Data Usability	Corresponding Case
Government-led ecosystem	Government	Exclusive governance based on incentives to participate	Obliged (all)	Focus on government benefits	Robust and governmental monitoring	Government negotiates format and validate.	Primarily for governments	Single Window
Regulation-led ecosystem	Government	Inclusive governance based on enforcement using regulations	Voluntary	Focus on government and business benefits	High level of integrity and auditability	Set by laws and regulations	Primarily for governments	Standard Business Reporting
Government-business-led ecosystem	European commission	Open governance based on incentives	Voluntary	Focus on business benefits	Flexible and interoperable	Consensus-based between business and governments	Governments and companies	Peppol

involvement in decision-making may lead to indecisiveness, whereas low involvement might lead to having no support base for innovation and changes in the ecosystem. These configurations demonstrate that data-driven government agencies have more choices and options in how they can organize data exchange and effectuate data stewardship responsibilities than previously presented in the literature. Knowing these configurations will help organizations to have more insight into their own configurations.

7. Conclusions, limitations and further research

A data-driven government requires data stewardship going beyond the boundaries of the government. Data is collected and maintained by many public and private organizations and is needed by the government for their decision- and policy-making. In government-business data ecosystems, we found that the original scope of data stewardship that focuses on data quality management must be extended to include governance (allocation of responsibilities and decision rights) and system quality assurance from an end-to-end data exchange perspective. This is needed because an ecosystem itself lacks a common goal, while individual actors have specific goals.

The empirical cases explored reveal three different configurations for data stewardship: 1) the government-led data ecosystem, 2) the government-business-led data ecosystem, and 3) the regulation-led data ecosystem. Each of the configurations has different advantages and disadvantages. Since this is explorative research with a limited number of case studies, further research can reveal other possible configurations.

This research is limited by its explorative nature and has three main limitations. First, the empirical cases are from the Dutch context. Public agencies in the Netherlands are increasingly making use of public-private collaborations for data exchange. Not all countries experience this development. The second limitation lies in the selection of dimensions for the comparison framework. While the literature on governance, data quality and system quality has identified a long list of dimensions, we filtered the most relevant dimensions based on previous empirical research on data stewardship in practice. While the interviews did not yield other relevant dimensions, studying additional cases may surface the importance of other dimensions. Third, the number of cases and interviews is limited. Additional data ecosystems may reveal unique data stewardship configurations. As this is a explorative research, there is a limitation towards the robustness of the framework. In order to increase the validity and reliability of the framework and configurations we used both documents as interviews, and constantly compared this to the literature. More research is needed towards the generalizability of the found data strategies.

Government-business data ecosystems provide a relatively uncharted research area inviting the development of new theories and methods. In these ecosystems, multiple parties, both public and private, share data and develop and maintain pioneering software technologies. We must ask critical questions regarding what actors are gaining influence and how public values can be preserved in government-business data ecosystems. Further research should focus on expanding the knowledge on these ecosystems by answering questions like; who are the data stewards? And how do they assure data quality as well as the quality of the data delivery channels? How do you manage problems/incidents in data exchange processes, acknowledging that a failure may lie in an infrastructure component beyond the organizational boundaries (e.g., in commercial reporting software)? And how do they make sure that any future modifications/updates meet the required quality levels?

Authors statement

All authors agree on final version of the “Towards data-driven government: empirical configurations of data stewardship in government-business data ecosystems”- paper and were fully involved in the research and writing process.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.giq.2021.101642.

References

AlShamsi, M., Salloum, S. A., Alshurideh, M., & Abdallah, S. (2021). Artificial intelligence and blockchain for transparency in governance. In *, Vol. 912. Studies in computational intelligence*. https://doi.org/10.1007/978-3-030-51920-9_11

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys, 41*(3). https://doi.org/10.1145/1541880.1541883

Bharosa, N., Hietbrink, F., Mosterd, L., & Van Oosterhout, R. (2018). Steering the adoption of standard business reporting for cross domain information exchange. In *Proceedings of the 19th annual international conference on digital government research: Governance in the data age* (pp. 1–10). https://doi.org/10.1145/3209281.3209325

Bharosa, N., Janssen, M., van Wijk, R., de Winne, N., van der Voort, H., Hulstijn, J., Tan, Y., & hua.. (2013). Tapping into existing information flows: The transformation to compliance by design in business-to-government information exchange. *Government Information Quarterly, 30*, S9–S18. https://doi.org/10.1016/j.giq.2012.08.006

Bovens, M. (2007). Analyzing and Assessing Accountability: A Conceptual Framework. *European Law Journal, 13*, 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Brinkerhoff, D. W., & Brinkerhoff, J. M. (2011). Public-private partnerships: Perspectives on purposes, publicness, and good governance. *Public Administration and Development, 31*(1), 2–14. https://doi.org/10.1002/pad.584

Chen, H. (2002). Digital government: Technologies and practices. *Decision Support Systems, 34*, 223–227.

Dawes, S. S. (1996). Interagency information sharing: Expected benefits, abstract. *Journal of Policy Analysis and Management, 15*(3), 377–394.

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly, 27*(4), 377–383.

de Bruijn, J. A., & ten Heuvelhof, E. F. (2008). Management in Networks. *On multi-actor decision making. Routledge - Taylor & Francis Group*.

Dearnley, C. (2005). A reflection on the use of semi-structured interviews. *Nurse Researcher, 13*(1), 19–28. https://doi.org/10.7748/nr2005.07.13.1.19.c5997

Delone, W. H., & McLean, E. R. (1992). The quest for the dependent variable. Information Systems Research. *Information Systems Research, 3*(1), 60–95.

Devaney, L. (2016). Good governance? Perceptions of accountability, transparency and effectiveness in Irish food risk governance. *Food Policy, 62*, 1–10. https://doi.org/10.1016/j.foodpol.2016.04.003

Ernst, A. (2019). How participation influences the perception of fairness, efficiency and effectiveness in environmental governance: An empirical analysis. *Journal of Environmental Management, 238*, 368–381. https://doi.org/10.1016/j.jenvman.2019.03.020

Fisher, C. W., & Kingma, B. R. (2001). Criticality of data quality as exemplified in two disasters. *Information and Management, 39*(2), 109–116. https://doi.org/10.1016/S0378-7206(01)00083-0

Fricke, E., & Schulz, A. P. (2005). Design for changeability (DfC): Principles to enable changes in systems throughout their entire lifecycle. *Systems Engineering, 8*(4), no. https://doi.org/10.1002/sys.20039

Furukawa, M., & Minami, A. (2013). A study on the ‘flexibility’ of information systems (part 1): Why do they need to be flexible? *International Journal of Business and Management, 8*(20), 48–61. https://doi.org/10.5539/ijbm.v8n20p48

Gebauer, J., & Schober, F. (2006). Information system flexibility and the cost efficiency of business processes 1. *Journal of the Association for Information Systems* (Vol. 7).

Gorla, N., Somers, T. M., & Wong, B. (2010). Organizational impact of system quality, information quality, and service quality. *Journal of Strategic Information Systems, 19*(3), 207–228. https://doi.org/10.1016/j.jsis.2010.05.001

de Graaf, G., & Paanakker, H. (2015). Good governance: Performance values and procedural values in conflict. *The American Review of Public Administration, 45*(6), 635–652. https://doi.org/10.1177/0275074014529361

Haug, A., Zachariassen, F., & van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management, 4*(2), 168–193. https://doi.org/10.3926/jiem.2011.v4n2.p168-193

Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly, 27*(4), 371–376. https://doi.org/10.1016/j.giq.2010.05.003

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148–152. https://doi.org/10.1145/1629175.1629210

Klievink, B., Bharosa, N., & Tan, Y. H. (2016). The collaborative realization of public values and business goals: Governance and infrastructure of public-private information platforms. *Government Information Quarterly, 33*(1), 67–79. https://doi.org/10.1016/j.giq.2015.12.002

Krauss, S., Hamzah, A., Omar, Z., Suandi, T., Ismail, I. A., & Zahari, M. Z. (2009). *Preliminary investigation and interview guide development for studying how malaysian farmers form their mental models of farming* (Vol. 14). Retrieved from https://nsuworks.nova.edu/tqr/vol14/iss2/3.

Lindgren, I., & Jansson, G. (2013). Electronic services in the public sector: A conceptual framework. *Government Information Quarterly, 30*(2), 163–172. https://doi.org/10.1016/j.giq.2012.10.005

Lockwood, M. (2009). Good governance for terrestrial protected areas: A framework, principles and performance outcomes. *Journal of Environmental Management, 91*, 754–766. https://doi.org/10.1016/j.jenvman.2009.10.005

Mosley, M., Brackett, M., Earley, S., & Henderson, D. (2009). The DAMA guide to the data management body of knowledge. In *Technics Publications*. LLC Post. https://doi.org/10.1161/CIRCULATIONAHA.108.834176.

Närman, P., Schönherr, M., Johnson, P., Ekstedt, M., & Chenine, M. (2008). Using enterprise architecture models for system quality analysis. In *Proceedings - 12th IEEE International Enterprise Distributed Object Computing Conference, EDOC 2008* (pp. 14–23). https://doi.org/10.1109/EDOC.2008.26

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems, 21*(4), 199–235. https://doi.org/10.1080/07421222.2005.11045823

OECD. (2002). Models of public budgeting and accounting reform. Retrieved from www.SourceOECD.org.

Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a data ecosystem?. In *ACM international conference proceeding series*. https://doi.org/10.1145/3209281.3209335

Provan, K. G., & Kenis, P. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of Public Administration Research and Theory, 18*(2), 229–252. https://doi.org/10.1093/jopart/mum015

Rosenbaum, S. (2010). Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research, 45*(5 PART 2), 1442–1455. https://doi.org/10.1111/j.1475-6773.2010.01140.x

Ryazanova, O., Pétercsák, R., Heaphy, L., Connolly, N., & Donnellan, B. (2016). *Perception of value in public-private ecosystems: Transforming the Dublin docklands through smart technologies* (pp. 1–10).

Van Donge, W., Bharosa, N., & Janssen, M. F. W. H. A. (2020). Future government data strategies: Data-driven enterprise or data steward?: Exploring definitions and challenges for the government as data enterprise. *ACM International Conference Proceeding Series, 196–204*. https://doi.org/10.1145/3396956.3396975

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33.

Wende, K. (2007). A model for data governance - organising accountabilities for data quality management. In *ACIS 2007 Proceedings - 18th Australasian conference on information systems* (pp. 416–425).

Whiting, L. S. (2008). Semi-structured interviews: Guidance for novice researchers. *Nursing Standard, 22*(23), 35–40.

Zheng, L., Yang, T. M., Pardo, T., & Jiang, Y. (2009). Understanding the "boundary" in information sharing and integration. In *Proceedings of the 42nd Annual Hawaii international conference on system sciences* (pp. 1–10). HICSS. https://doi.org/10.1109/HICSS.2009.481.

W. van Donge MSc is a PhD candidate at the Faculty of Technology, Policy, and Management of Delft University of Technology. Her research interest include the design and governance of public-private data ecosystems, in particular how agreement frameworks guide decision making, data quality and system quality assurance. She also works at Logius, a shared service center for public agencies in the Netherlands.

N. Bharosa is senior researcher in Govtech at the Faculty of Technology, Policy, and Management of Delft University of Technology. He is also head of research at Digicampus, a quadruple helix innovation partnership of governments, companies, citizens, and academia. He has successfully led research projects in multiple public-private data ecosystems.

M. F. W. H. A. Janssen is full professor in ICT & Governance at the Faculty of Technology, Policy, and Management of Delft University of Technology. He is also honorary visiting professor at Bradford university, UK and KU-Leuven, Belgium. He is listed as one of the world's 100 most influential people in digital government in 2018 and 2019. He has published over 500 publications, an h-index of over 60 and more than 15 K citations.