Government Information Quarterly xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Government Information Quarterly

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

Data governance: Organizing data for trustworthy Artificial Intelligence

Marijn Janssen[a,*], Paul Brous[a], Elsa Estevez[b,c], Luis S. Barbosa[d,e], Tomasz Janowski[f,g]

[a] Faculty of Technology, Policy and Management, Delft University of Technology, the Netherlands
[b] Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (UNS), Argentina
[c] Instituto de Ciencias e Ingeniería de la Computación UNS-CONICET, Argentina
[d] Department of Computer Science, University of Minho, Portugal
[e] United Nations University Operating Unit on Policy-driven Electronic Governance (UNU-EGOV), Portugal
[f] Department of Informatics in Management, Faculty of Management and Economics, Gdańsk University of Technology, Poland
[g] Department for e-Governance and Administration, Faculty of Business and Globalization, Danube University Krems, Austria

ARTICLE INFO

Keywords:
Big data
Data governance
AI
Algorithmic governance
Information sharing
Artificial Intelligence
Trusted frameworks

ABSTRACT

The rise of Big, Open and Linked Data (BOLD) enables Big Data Algorithmic Systems (BDAS) which are often based on machine learning, neural networks and other forms of Artificial Intelligence (AI). As such systems are increasingly requested to make decisions that are consequential to individuals, communities and society at large, their failures cannot be tolerated, and they are subject to stringent regulatory and ethical requirements. However, they all rely on data which is not only big, open and linked but varied, dynamic and streamed at high speeds in real-time. Managing such data is challenging. To overcome such challenges and utilize opportunities for BDAS, organizations are increasingly developing advanced data governance capabilities. This paper reviews challenges and approaches to data governance for such systems, and proposes a framework for data governance for trustworthy BDAS. The framework promotes the stewardship of data, processes and algorithms, the controlled opening of data and algorithms to enable external scrutiny, trusted information sharing within and between organizations, risk-based governance, system-level controls, and data control through shared ownership and self-sovereign identities. The framework is based on 13 design principles and is proposed incrementally, for a single organization and multiple networked organizations.

1. Introduction

Organizations in general, and public sector organizations in particular, increasingly collect and use *Big and Open Linked Data* (BOLD) (Janssen, Matheus, & Zuiderwijk, 2015). The rise of BOLD, combined with machine learning and other forms of Artificial Intelligence (AI) results in the increasing use of *Big Data Algorithmic Systems* (BDAS). Such systems are used to make decisions about: access to affordable loans amid the shortage of credit files; matching of skills and jobs to promote access to employment; implementing admission to schools while helping individuals choose the right school; and mitigating risks of disparities in the treatment of individuals by law enforcement while helping build trust between the public and law enforcement (Executive Office of the President, 2016).

The use of BDAS for improving and opening government is met with a lot of enthusiasm. However, BDAS rely heavily on the use of data combined from various sources, some controlled by the organization itself, others controlled by partner organizations, yet others controlled by unknown entities. Without control over such data to ensure quality and compliance, BDAS would be too risky to be entrusted with consequential decisions. Therefore, many organizations are turning to data governance as a means to exercise control over the quality of their data and over compliance with relevant legal and ethical requirements in order to guarantee the delivery of trustworthy decisions. The concept of trustworthiness, which can be directly controlled or indirectly influenced (Yang & Anguelov, 2013), refers to properties through which a trusted entity is serving the interests of the trustor (Levi & Stoker, 2000). In the situation under study, the trustor (an organization) entrusts its system (BDAS, which itself uses BOLD and AI) in making sound decisions.

Data governance is about allocating authority and control over data (Brackett & Earley, 2009) and the exercise of such authority through decision-making in data-related matters (Plotkin, 2013). To fulfil its goals, data governance should focus not just on data, but on the systems

* Corresponding author.
E-mail addresses: m.f.w.h.a.janssen@tudelft.nl (M. Janssen), paul.a.brous01@gmail.com (P. Brous), ece@cs.uns.edu.ar (E. Estevez), lsb@di.uminho.pt (L.S. Barbosa), tomasz.janowski@pg.edu.pl (T. Janowski).

https://doi.org/10.1016/j.giq.2020.101493

through which data is collected, managed and used. Specifically, people are essential in these systems (Benfeldt, Persson, & Madsen, 2020); thus data governance should provide incentives and sanctions to stimulate desirable behaviour of the persons involved in collecting, managing and using data. Beyond a single organization, data governance depends on collaboration between organizations and persons that make up the system. This multi-organizational context requires trusted frameworks to ensure reliable data-sharing between all organizations involved, that the right data is securely and reliably shared between participating organizations, while complying with General Data Protecting Regulation (GDPR) (European Parliament and European Council, 2016) and other relevant laws and regulations.

Consistent with this context, we define data governance as:

> *Organizations and their personnel defining, applying and monitoring the patterns of rules and authorities for directing the proper functioning of, and ensuring the accountability for, the entire life-cycle of data and algorithms within and across organizations.*

This definition takes into account both data and data processing by AI and other algorithms, considers that both data and algorithms change during their respective life-cycles, accounts for the personnel responsible for creating and use of data and algorithms, and adopts a systems (multi-organizational) view.

Data governance is a success factor for BDAS (Brous, Janssen, & Krans, 2020) and has an overall positive effect on the performance of organizations that apply BDAS (Zhang, Zhao, & Kumar, 2016). The purpose is to increase the value of data and minimize data-related costs and risks (Abraham, Schneider, & vom Brocke, J., 2019). Given the consequential and repetitive nature of the BDAS decision-making, mistakes in data governance that affect the working of such systems can have profound legal, financial and social implications on the organizations involved, citizens and businesses, and society at large. Such mistakes can result in systemic bias, unlawful decisions, large financial exposures, political crises, lives lost or any combination thereof. In the interconnected world, where data is collected by (and about) governments, businesses and citizens, and is processed by different entities using various algorithms, dependencies grow, mistakes accumulate, and accountability is gradually lost in the process.

The rationale outlined above directly leads to the goal of this article. The goal is threefold. First, to define and conceptualize data governance for AI-based BDAS. Second, to review the challenges and approaches to such governance. Third, to propose the concept of trusted AI-based BDAS and a framework for data governance for such systems.

The rest of the article is structured as follows. Section 2 introduces the concept of data governance, followed by data governance for AI-based BDAS. Different forms of data governance for AI-based BDAS are outlined in Section 3. Section 4 formulates the main proposal: trusted AI-based BDAS and a data governance framework for such systems. The proposal consists of: system-level governance model of BDAS in Section 4.1, data stewardship and base registries as the foundation for data governance in Section 4.2, and the trusted framework and self-sovereign identities for data sharing in Section 4.3. Finally, essential data governance principles are outlined in Section 5.

2. Data governance

Data governance has been given scant attention and is often overlooked by organizations in their efforts to realize BDAS and create Fair, Accountable and Transparent (FAT) algorithms. Often the focus is on experimenting with AI, but acquiring and preparing data for AI, which often consumes most of the time, is given less consideration. However, the ubiquitous nature of data, when using large volumes and varieties of data from multiple sources, the uncertain impact of data flows on data quality, and lack of awareness about the importance of data quality, all complicate governance. Data quality consists of many dimensions including accuracy, timeliness, completeness, consistency, objectivity, believability and relevance, which all determine whether data is fit for use (Strong, Lee, & Wang, 1997).

Data collection and sharing has become easier over the last decade, partly due to interoperability solutions. However, interoperability also facilitates inaccurate data to flow smoothly across systems and contaminate them in an exponential manner (Dasu, 2013). The same technological advancements, with data collected from heterogeneous resources, stored in various ways and having different qualities, make data types and structures increasingly complex (Dasu, 2013). Besides, vast volumes of data complicate *entity resolution* – identifying the same real-world entity like e.g. a person, within a single database (de-duplication) or in multiple databases (record linkage). Furthermore, even if accurate data is stored, *data glitches* may produce inaccuracies over time. These glitches originate from changes in the environment, which produce discrepancies between the reality and how data capture this reality, e.g. after the change of the residential address, the old address remains on record. Hence, high levels of information quality are hard to achieve.

While most government organizations recognize today that data is crucial, the creation of a culture which treats data as an asset and which helps public servants make data-driven decisions is challenging (Benfeldt et al., 2020). While data governance should help lower the cost of data management and create value from data, data is often fragmented over many organizations which implement different data policies. This can result in unclear responsibility, diffused accountability, and unknown data quality, which, in turn, might undermine the fitness-to-use of such data within BDAS.

BDAS, and the AI algorithms embedded in them, are increasingly used to make consequential decisions. However, such decisions may be incorrect, and responsibility for this may be challenging to determine. Low data quality and unclear dependencies between data and algorithms can easily bias or skew the outcomes of the AI algorithms. Shared roles and joint operations performed among departments and organizations may dilute responsibilities. For instance, who is responsible when an algorithm provides wrong outputs due to anomalies in data collected through multiple sensors? The causal relationship between an event and a system failure might be difficult to establish without proper data governance.

While applying BDAS in consequential decision-making is conditional on data quality, perfect data quality does not exist. On one side, data providers can argue that data quality never reaches 100% and even include such arguments in data usage agreements. On another, the BDAS operators might blame poor data quality for arriving at wrong decisions. This raises the challenge of defining and sharing responsibilities between data providers, algorithms provides and BDAS operators as part of data governance, particularly when multiple organizations engage in such governance jointly. Yet there is limited understanding of what constitutes data governance for BDAS in general and AI-based BDAS in particular, even less how to carry out system-level data governance, between different organizations.

Terminologically, the terms governing, governance and management are all different. Governance is the organizing logic through which the management of data – collection, storage, processing, using, sharing and destroying – takes place. Governing comprises activities conducted to create and execute this logic. These include not only the management of data but also decisions made over data: *who* can make such decisions and therefore *influence* how data is accessed, controlled, used and benefited from (Khatri & Brown, 2010), and *how* data can be used (or potentially used) and by whom. The scope of usage decisions is essential as they determine what is expected from data governance. In any case, designing data governance requires stepping back from the daily routine (Khatri & Brown, 2010).

Conceptually, it transpires that data governance is the exercise of authority and control over the management of data (Brackett & Earley, 2009, p. 19). It can also be viewed as "the exercise of decision making and authority for data-related matters. It is a system of decision rights

and accountabilities for information-related processes, executed according to the agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods" (Plotkin, 2013, pp. 1–2). Thus the goals of data governance are ensuring the quality and proper use of data, meeting compliance requirements, and helping utilize data to create public value. Fulfilling these goals requires mechanisms for personal data protection, security, non-discrimination and equal treatment; covering the entire life-cycle from creating, to processing and sharing, to destroying data; and addressing technical, institutional and social implications of data sharing. Hence, we define data governance as *organizations and their personnel defining, applying and monitoring the patterns of rules and authorities for directing the proper functioning of, and ensuring the accountability for, the entire life-cycle of data and algorithms within and across organizations.*

Structurally, data governance is exercised through policies, incentives and sanctions, as needed to create an organizational culture where data is treated as an asset, and behaviour that supports or violates this treatment is rewarded or sanctioned respectively. Data governance includes: standardizing data – creating metadata to be able to integrate datasets and ensure the same interpretation of data; allocating relevant procedures and decision authorities to ensure data stewardship and data quality; monitoring data usage, e.g. ensuring risk assessment before using data to carry out consequential decisions; and monitoring such systems as part of the data life-cycle.

3. Data governance approaches

A common challenge with data governance is that the data flow and logic may not follow the structure of an organization. The mismatch between organizational structure and data usage can easily result in data silos, duplications, unclear responsibilities, and missing control of data over its entire life-cycle. This is particularly the case for BDAS, which are typically crossing departmental boundaries, not bound to any single function or process, and have to deal with data originating in multiple departmental silos. Numerous challenges are arising in this scenario. There might be a lack of established mechanisms for data governance for dealing with BDAS while involving various organizations. Another is ad-hoc handling of data without procedures and processes and secure data infrastructure, which might easily result in individual data items being accessible to non-authorized persons.

Given such challenges, the choice of the data governance approach is critical, albeit this choice is not always clear or even explicit (Koltay, 2016). Three approaches to data governance are planning and control, organizational, and risk-based. These approaches are not mutually exclusive and can be used to complement each other. The three approaches are depicted in Fig. 1 and described below. While they are typically applied within organizations, they can be also used between organizations. Risk-assement, defining responsibilities, and joint planning and control can be done between organizations, but require the establishment of trusted frameworks.

The *planning and control* approach, often used by IT-governance frameworks (De Haes, Van Grembergen, & Debreceny, 2013), is based on the annual cycle of planning and control. In each cycle, objectives are set, budgets are allocated, and projects are defined, implemented, monitored and evaluated. Budgets and other resources are allocated to projects and departments for executing activities, subject to set priorities. In turn, projects and departments must compete. Projects are evaluated on their performance and close alignment between business and technology goals. Planning can initiate infrastructure projects aimed at, for example, improving data quality or exploring the potential and risks of AI-based BDAS in various application areas. In this approach, data governance is carried out through policies and procedures that are repeatable, verifiable and auditable. The approach is often criticized for not easily adapting to change (Janssen & van der Voort, 2016). However, continuous monitoring can help adjust project plans and resource allocations on an ongoing basis.

The *organizational* approach to data governance emphasizes structure, responsibility, accountability and reporting (Mullon & Ngoepe, 2019). This approach, using the principle of top-level design, sets up organizational structures for data governance, and treats data governance as a defining authority (Mullon & Ngoepe, 2019). Consistent with this principle, the approach recommends setting up decision-making structures in the areas of data, AI, privacy or ethics, such as Chief Data Officers (CDA), Chief AI Officers (CAIO), Chief Privacy Officers (CPO) or Chief Ethics Officers (CEO). Within such structures, it also includes the responsibility for data stewardship (Rosenbaum, 2010).

With the advent of GDPR and AI, the *risk-based* approach has gained attention as a way to identify risks of BDAS and introduce appropriate governance mechanisms to address them (Ladley, 2019). This approach is often advocated as a foundation for data governance (Rothstein, Borraz, & Huber, 2013) and an effective solution to AI-specific risks such as data or algorithmic errors, data or algorithmic bias, or even data-embedded discrimination (Janssen & Kuk, 2016). These problems arise due to, in large part, sensitive attributes embedded in data sets which are used by machine learning algorithms to search for patterns (Beretta et al., 2018). Example risks include missing, stolen, outdated, inaccurate or biased data. Regular assessment is needed to establish such risks and appropriate action undertaken to manage them. Actions can be taken incidentally, preventively, or both, depending on the governance mechanism. For example, each AI project could be subject to a risk audit to anticipate and address the possible undesirable effects of the AI algorithms.

These approaches can be all in place and complement each other. Nevertheless, different governance mechanisms should be introduced with care, as too much governance can result in excessive overhead and

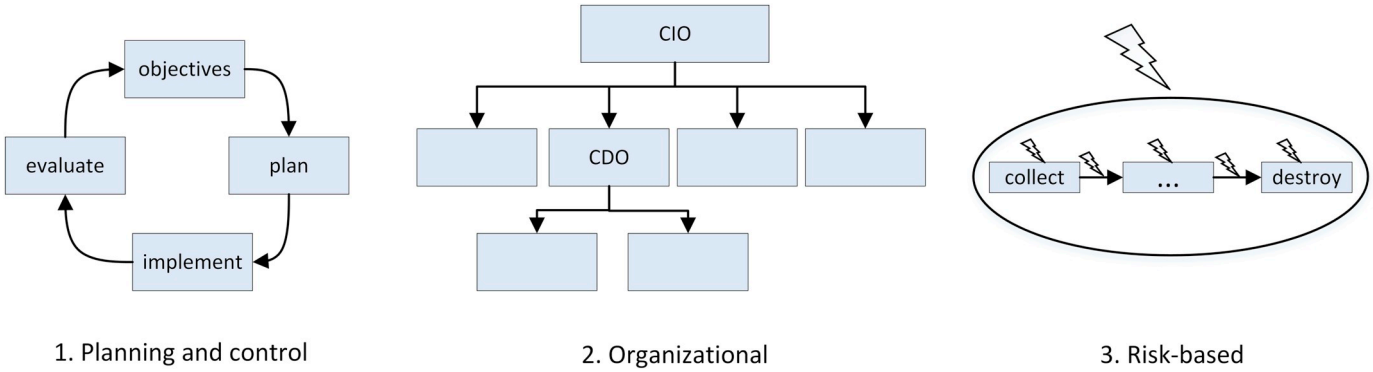

Fig. 1. Tripartite approaches to data governance.

lower performance. Too little governance, however, might result in unclear responsibilities, uncontrolled risks and not taking the right precautions and actions.

4. Data governance for trusted BDAS

This section aims to formulate the main proposal of this article: the concept of trusted AI-based BDAS and a framework for data governance for such systems. The proposal consists of three elements: system-level governance model for BDAS (Section 4.1), data stewardship and based registries (Section 4.2), and the trusted data-sharing framework based on self-sovereign identities and data-sharing agreements (Section 4.3).

4.1. System-level controls for BDAS

AI comes with immense opportunities but also with risks. It may violate privacy, discriminate, avoid accountability, manipulate and misinform public opinion, and be used for surveillance. For instance, it can recognize faces in photos and video streams to help determine people's whereabouts and their behaviour patterns. AI capabilities generate immense powers, which are dangerous if systems are allowed to expand without proper oversight, accountability and governance. Data is the basis for BDAS, but the outcomes of BDAS should be monitored as well, as they are part of the data life-cycle. Therefore, the entire BDAS should be subject to effective data governance.

Data and algorithms can be opened for inspection, although primarily intended for machine processing. Even if presented in formats accessible to people, few will be able to understand them, let alone scrutinize. Lacking opportunities for proper public scrutiny, risk assessment, regular audits, sampling protocols, validity and quality checking, clear responsibilities, and other inspection mechanisms are needed.

Fig. 2 depicts a system-level governance model that enables thinking about data governance within BDAS. As data and AI are intrinsically connected, the model considers them together. The input data is shown on the left, the output on the right, and algorithmic processing in the middle. The model takes into account how BDAS operate with data. Without this, effective governance is not possible.

BDAS enable automatic decision-making within public institutions. However, the decision-making authority is hidden from the user directly affected by the outcomes, public officers become merely mediators rather than decision-makers, and automated public services become "*hidden bureaucrat*" (Wihlborg, Larsson, & Hedström, 2016). Hence, accountability should be designed at the system level. This system-level accountability design should cover not only the internal working of algorithms, but how their usage is organized, how they are fed with data, how the data is controlled, how the outcomes are checked, and how the entire system is audited.

As shown at the top of Fig. 2, systems are guided by regulations, for instance, regulations concerning data protection or Freedom of Information (FOI). Within the regulatory context, policies, principles and procedures are formulated. Policies prescribe how users should or should not behave concerning data and algorithms, and specifies the means of rewarding or sanctioning such behaviours. Principles are normative and directive; they determine the organizing logic of data governance. The main principles for data governance in BDAS are summarized in Section 5. Data governance is based on the expectations and values of the society, which is ultimately affected by the outcomes of BDAS. The basis for taking these into account is creating a culture where data is treated as an asset, while regulations and commonly accepted public values are guiding data-based decisions. Furthermore, the professionals involved in gathering and processing data must adhere to professional norms in their respective areas.

These expectations and values are translated in the planning and control, organizational, or risk-based governace approaches, resulting in the assignment of decision-making authority and the establishment of processes and procedures. Procedures include the annual planning and control cycles, the introduction of independent ethics commitees, the regular audits, taking of data samples, etc. Controls and audits should be applied to inputs, processes and outputs in the data usage processes, which all must be monitored using qualitative and quantitative measures.

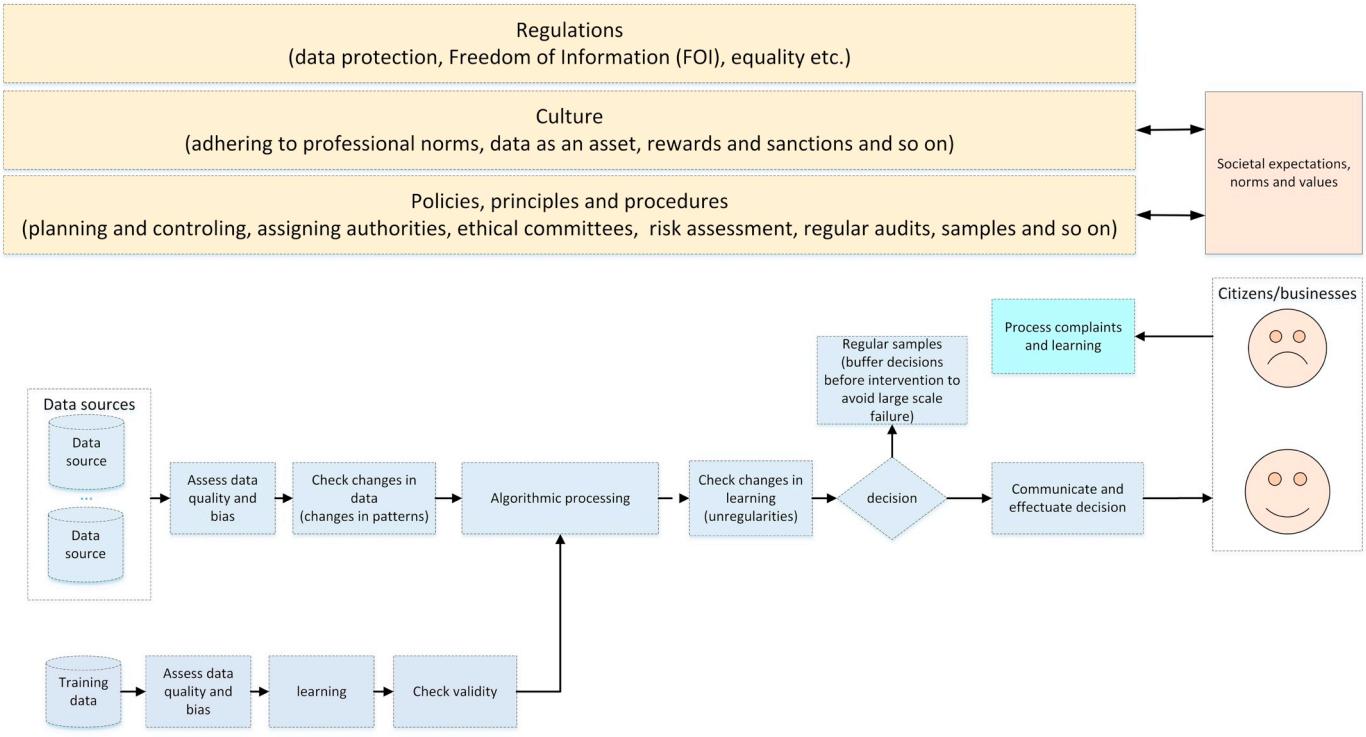

Fig. 2. System-level governance model of BDAS.

The bottom left part of Fig. 2 depicts the learning process through which the AI algorithm is fed with training data to learn to make decisions. Most BDAS are based on a type of machine learning algorithms that identify patterns in data, which they use for descriptive, predictive or prescriptive goals. The risks of training algorithms using historical data are that mistakes, inconsistencies and bias embedded in such data will be reflected in the working of the algorithms. Another risk area is poor generalization of the algorithms to situations outside the data upon which they were trained. Hence, oversight should cover both data quality issues, including the presence of bias, and the fitness of algorithms and the data fed into them to the problems they are asked to address. Part of the learning process is also checking the validity of the results, including the determination of false positives and false negatives. Although these controls are needed, AI algorithms are often opaque, controls alone are insufficient, and additional governance mechanisms may be required.

AI algorithms and their implementations can range from black-box to white-box approaches. For the sake of accountability and transparency, the causality between inputs and decisions should be explained to guarantee the fairness of the results. While people should decide on the rules to be able to explain the causality between data, rules and decisions, AI algorithms can be used to improve upon these rules and derive new ones. These requirements suggest that algorithmic processing should take a white-box rather than a black-box approach.

The middle part of Fig. 2 depicts the main elements of the decision-making process. In the ideal scenario, this process should adhere to compliance-by-design, i.e. BDAS should be designed to comply with all relevant norms and regulations. However, this scenario would have to guarantee unambiguous data ownership, monitoring of data sources and data quality, controlled adherence to standards, and compliance with other specific requirements. Various data sources used as inputs should be checked for quality, bias and other properties. As changes to data may easily result in the wrong outcomes produced by the algorithms, the closeness of new data to training data should be evaluated. Therefore, the changing pattern principle suggests investigating the reasons for introducing changes, and any time a change in pattern is discovered, the outcomes should be validated again. This can be done using samples or comparing the outcomes with previous results or manual decisions. Once the results are communicated, the persons involved should have an opportunity to file appeals. Such appeals can be used to scrutinize and further improve the BDAS.

BDAS are often opaque, and their working is difficult to understand. Similarly to opening data to the public for the sake of transparency and accountability, the AI algorithms can be opened as well. While most people will be unable to make much sense of the algorithms, this would allow auditors, experts, scientists, citizen scientists and other professionals, equipped with proper tools, to check the working of the algorithms. However, the opening of algorithms should be done with care as it might reveal bugs in the working systems, which could be exploited by hackers. Hence, the model assumes that data, algorithms and processes will be opened to controlled groups for scrutiny.

Data governance contains mechanisms to incentivize correct behaviours and sanction incorrect ones. Whereas misbehaviour and mistakes can be sanctioned, incentives including monetary rewards could be offered for uncovering errors, discrimination, bias or other undesirable features in data, algorithms acting on such data, and processes governing the use of data and algorithms. Such incentives are used in the bug bounty programmes that encourage people to spot and report back issues with the software. However, incentives cannot replace proper policing by audit organizations. The more controls are embedded in governance processes, the higher auditability and problem detection rates, and the more opportunities for learning and improvement.

Creating sound data governance requires a balance between complete control, which is unreachable, and lack of control, which is dangerous, into the mode of governance that is necessary and feasible. Risk-based approaches to data governance are suited to maximize the value created from data while mitigating risks and reducing costs. Controls can play similar preventive and detective roles, guided by socio-technical arrangements within the organizations concerned. To ensure sound data governance, some organizations appoint Chief Data Officers (CDOs) or Chief Algorithmic Officers (CAOs). In contrast, others keep it within the remit of Chief Information Officers (CIOs), Chief Technology Officers (CTOs) or even Chief Privacy Officers (CPOs). The planning and control approach also engages Ethics Committees and related bodies to decide about the use of BDAS. Indeed, all public servants should be sensitized and trained to take responsibility for ensuring data quality and proper data sharing.

4.2. Data stewardship and base registries

The foundation of data governance is *responsible data collection*. If sensitive data, like gender, race, residential address, health status or political preference is collected, then it can be misused or abused. As data that is not collected cannot be misused or abused, the collection of data should be minimized. Nevertheless, there are numerous situations where sensitive data must be collected and shared for the sake of transparency, fraud detection, service improvements, or better decisions. Such data, once collected, must be secured to guard against misuse or abuse.

The ownership of data is often challenging to establish, and multiple persons might claim the rights to it. An analogy is the ownership of the electricity used in a household, which could be assigned to the house owner, the person currently renting the house, the utility company that provides the electricity, or the government that taxes electricity usage. Furthermore, in law, tangible goods are more natural to own than intangible ones, like data which can be shared without limits. In the case of data, the notion of stewardship is preferred over ownership. It draws attention to the provision of trusted and authentic data, to responsible data use and sharing, and to the presence of multiple stewards taking care of similar data. Data stewardship is a team effort where responsibilities and expertise are divided among members, who can manage data on behalf of others.

Data stewards should ensure responsible information sharing. Plotkin (2013) view data stewardship as a way to formalize accountabilities for managing information resources on behalf of and in the best interest of others. Dawes (2010) formulated the stewardship and usefulness principles for information sharing: stewardship should assure data quality, validity and security; manage risks; manage and preserve data; and make public officials and organizations responsible for handling information with care and integrity. Others emphasize that stewards are also responsible for information security (Cuganesan, Hart, & Steele, 2017). The usefulness principle should ensure that data can be used for innovation; in particular for BDAS. Dawes (2010) further argued that stewardship and usefulness are complementary principles and should guide information-based transparency. In contrast, achieving both information usefulness and effective stewardship over information can be competing (Cuganesan et al., 2017). The stewardship principle might deal with information sharing risks by withholding that information and thus reducing transparency and accountability. The other way around, the usefulness principes could result in data protection being violated by data stewards.

From an operational perspective, data should be managed by one organization or department and used by other organizations or departments. When collecting data, data stewards should follow the principle of collecting data at the source (Hammer, 1990) and separating sensitive and non-sensitive data (Janssen, Matheus, Longo, & Weerakkody, 2017). The data can be stored in a *base registry,* which according to the European Commission (2017, p. 37), is “a trusted and authoritative source of information which can and should be digitally re-used by others, where one organization is responsible and accountable for the collection, use, updating and preservation of information”. When other organizations spot incorrect data, then they should report

this to the data steward who is in charge of the base registry, and wait for a response of the steward before using this data. The steward will investigate the issue with the correctness of data and, if needed, update the data and inform the reporting organization, that subsequently can start using this data.

Even if data is pooled and linked at the conceptual level, physical storage and responsibility for data should be distributed to reduce vulnerability. Pooling and linking should ensure that data is interrelated and can easily be combined, if needed. However, governance mechanisms should ensure that data can only be shared if the right conditions, like authorization by multiple persons or approval by the data protection officer, are met. When information cannot be shared with another department to avoid violating regulations or due to the conflict of interest, we use the intra-organizational information barrier called the *Chinese wall*. A leading principle is to minimize the number of persons having access to data – if somebody does not need to access the data, such access should not be granted. To avoid the situation of a single entity exercising control over data without others' consent, distributed responsibilities and separation of concerns make systems less vulnerable.

Data stewardship and base registries provide a foundation for data and information sharing. However, mechanisms are needed to ensure responsible data sharing, allowing data sharing when required, but also blocking data sharing when necessary. For this, trusted data sharing frameworks are required.

4.3. Trusted data sharing framework

BDAS often depend on data sources external to the organization, which can be easily manipulated or misused. Hence, data and algorithms should not be indiscriminately shared with every organization, and organizations should not have access to information which they do not need. For a single organization, these types of decisions are made within the framework of data governance. However, as BDAS often depend on data sources that are external to the organization, there is a need to extend such data governance to cover multiple organizations. This extension relies on trusted data sharing frameworks, or trusted frameworks for short, that guide data exchange within and between participating organizations while ensuring compliance with regulations and the realization of public value.

Trusted frameworks should ensure that the right data is shared, that such sharing is carried out securely and reliably, and that the sharing complies with the regulations like GDPR (European Parliament and European Council, 2016). Trusted frameworks limit access to data to those who are authorized, and ensure non-repudiation of the data origin. A non-repudiation service provides a recipient with the *proof of origin*, which constitutes a legally-valid evidence that a particular person or organization provided the data.

Trusted frameworks are a standard mechanism for governing relationships and are used in payment systems, in domain registration systems, or in mobile networks to provide roaming services. In information sharing, traditional *trusted third parties* (TTPs) that use certificates to provide non-repudiation services, are being replaced with distributed ledger or blockchain technology (DLT) (Dunphy & Petitcolas, 2018; Ølnes, Ubacht, & Janssen, 2017). As DLT ensures decentralized execution, avoiding the creation of a single point of failure or misuse, they are less vulnerable. Nevertheless, such a decentralized information sharing system needs to be guided by proper governance mechanisms.

Trusted frameworks need to provide identification, authentication and authorization services. Identification refers to a person or organization claiming to have a particular identity. Authentication makes sure that the identified person or organization are what they claim to be; it verifies the claim made by the identifying party. Authorization happens when a person or an organization has been identified and granted access to data, specifying what they can do with this data. This also results in non-repudiation, that the data origin cannot be disputed.

Increasingly, trusted frameworks which are based on agreements among participating parties are used for data sharing and service provision. The agreements refer to a collaboration of various public and private parties to share data or provide services on topics, such as identification and authentication, part of the efforts to ensure interoperability and compliance, for instance with GDPR. A trusted framework contains several elements to regulate data sharing or other types of services, which may include:

- a list of requirements for trusted data sharing;
- a set of standards for realizing trusted data sharing;
- a collection of contracts and agreements for trusted data sharing;
- an authorization scheme who should have access to which data under what circumstances;
- a certification mechanism to record adherence of different parties to the rules;
- an auditing mechanism to verify compliance with requirements, agreements, contracts; and
- a mechanism to enforce compliance with the rules and agreements.

These elements should cover the entire data life cycle and should be adaptable to changing circumstances. For instance, a trusted framework should ensure that personal information is only shared when given consent to do so, that only defined parties can have access to specified data, that a variety of public and private organizations can legitimately gain access to such data, etc.

Data sharing should be based on the *'need to know' principle*. A minimum amount of data should be shared and only for specified purposes. For example, when a service requests information if a person is eligible for voting, her age should not be shared, but only whether or not she is 18. The same applies to other types of information, such as birth certificates, diplomas, social security, tax services and others. Instead of providing detailed personal information, an answer should be submitted to a specific question. This reduces the chances of using the data for unrelated purposes.

Critical aspects of data governance are identity and trusted data sharing, which are not part of the standard Internet infrastructure. Therefore organizations collect and store identity information such as name, gender, age, profession and others to be able to identify persons. This results in fragmented data landscape where such information is stored in many places. This raises security risks for identity information to be stolen, privacy risks with handling personal information, and data integrity risks as data might not be consistent with each other and the states of the real world. Therefore, this data should be treated as an asset that can be re-used by other organizations using secure and reliable storage and sharing services. Organizations or users can control the data. The first situation entails organizations defining base registries using data stewardship. The second involves users exercising Self-Sovereign Identity (SSI). SSI provides control over and ownership of data to citizens, who can give consent to share such data with others. Interoperability is created to be able to share data based on commonly agreed standards specified by the trusted framework.

The data-sharing governed by trusted frameworks should adhere to the *informing principle*. When the government shares data about a particular person or organization, this person or organization should be made aware of this to avoid misuse and verify the correctness of such data. Then, when the data is hacked or shared for illegitimate purposes, this is immediately revelated. Furthermore, access to their own data not only results in transparency but also empowers citizens and organizations to be in control of their data, including data about themselves. However, this also requires *agreements* to regulate access, openness and inclusion.

5. Essential data governance principles

Although the foundation of trustworthy BDAS is sound data governance, this area is often overlooked. Data governance for BDAS is a

Table 1
Overview of data governance principles.

Name	Description
1. Evaluate data quality and bias	When data is used by BDAS, its quality, and possible embedded bias should be evaluated.
2. Detect changing patterns	When the outcomes of the algorithms change, their validity should be checked, and the reasons for such changes investigated.
3. Need to know	Minimize the amount of data that is shared by only sharing what is necessary, e.g. answers to questions instead of complete datasets.
4. Bug bounty	Rewards could be used to encourage people to spot errors and issues and report them back.
5. Inform when sharing	When governments share data about a person or an organization, these entities should be informed to ensure transparency and avoid misuse.
6. Data separation	Separate personal from non-personal data, and sensitive from non-sensitive data (Janssen et al., 2017).
7. Citizens control of data	Empower citizens and organizations to be in control and check the accuracy of their data.
8. Collecting data at the source	Collect data at the source to ensure its correctness and to know how such data is collected (Hammer, 1990).
9. Minimize authorization to access data	If a party does not need data, access should not be granted.
10. Distributed storage of data	Distributed systems are less vulnerable and avoid easily combining data without permission.
11. Data stewards	Assign data stewards to formalize accountability for managing information resources while adhering to the principle of the separation of concerns (Dawes, 2010).
12. Separations of concerns	Responsibilities for data should be distributed in such a way that no single person can misuse or abuse data.
13. Usefulness	Data should be recognized as a valuable asset that can be used by BDAS (Dawes, 2010).

complex field, and the development of BDAS without due attention to data governance is a significant risk. Data governance can be viewed as organizations and their personnel defining, applying and monitoring the patterns of rules and authorities for directing the proper functioning of, and ensuring the accountability for, the entire life-cycle of data and algorithms within and across organizations. Data governance can help mitigate the issues of transparency, accountability, fairness, discrimination, and trust. Available approaches to data governance can be based on clear organizational structures, responsibilities and accountabilities, planning and control cycles, and risks. The latter is particularly relevant. BDAS face the risks of violating the privacy, using data for undesired purposes, allowing bias or discrimination in data to inform algorithmic decisions, making wrong decisions, and so on. Although greater transparency into the inner workings of algorithms is necessary, it is insufficient for effective oversight. For that, system-level governance is needed. Organizations and their personnel need to work in concert to carry out effective data governance. Apart from socio-technical measures focused on controls and safeguards, an organizational culture that promotes awareness and ethical value of data and algorithms is part of data governance.

Table 1 provides an overview of the 12 main data governance principles that are discussed in this article. Although these principles might at first appear simple, they are challenging to realize. There are hardly any good practices for successful adoption and application of data governance for BDAS. Scarce research exists about trusted frameworks and SSIs, and there is no consensus yet about how they should be realized in BDAS. Adhering to these principles will help improve data governance and contribute to trustworthy BDAS. However, there is a need for technologies like base registries and self-sovereign identities to make trustworthy BDAS work. The realization of data governance in multi-organizational settings will also require the use of trusted data-sharing frameworks to guide inter-organizational data exchange, and ensure compliance with regulations as well as the creation of public value. While the foundation of BDAS data governance is responsible data collection, citizens' control of data, and data stewardship, this new research field has to continue advancing before it creates a solid research foundation for trusted BDAS.

References

Executive Office of the President (2016). *Big data: A report on algorithmic systems, opportunity, and civil rights.* Executive Office of the President.

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management, 49*, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data governance as a collective action problem. *Information Systems Frontiers, 22*(2), 299–313. https://doi.org/10.1007/s10796-019-09923-z.

Beretta, E., Vetrò, A., Lepri, B., & De Martin, J. C. (2018). *Ethical and socially-aware data labels. Paper presented at the Annual International Symposium on Information Management and Big Data.* .

Brackett, M., & Earley, P. S. (2009). *The DAMA guide to the data management body of knowledge (DAMA-DMBOK guide)*.

Brous, P., Janssen, M., & Krans, R. (2020). *Data governance as success factor for data science. Responsible Design, Implementation and Use of Information and Communication Technology: 19th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2020, Skukuza, South Africa, April 6–8, 2020, Proceedings, Part I, 12066*. 431–442. https://doi.org/10.1007/978-3-030-44999-5_36.

Cuganesan, S., Hart, A., & Steele, C. (2017). Managing information sharing and stewardship for public-sector collaboration: A management control approach. *Public Management Review, 19*(6), 862–879.

Dasu, T. (2013). Data glitches: Monsters in your data. In S. Sadiq (Ed.). *Handbook of data quality: Research and practice* (pp. 163–178). Berlin, Heidelberg: Springer Berlin Heidelberg.

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

De Haes, S., Van Grembergen, W., & Debreceny, R. S. (2013). COBIT 5 and enterprise governance of information technology: Building blocks and research opportunities. *Journal of Information Systems, 27*(1), 307–324.

Dunphy, P., & Petitcolas, F. A. (2018). A first look at identity management schemes on the blockchain. *IEEE Security & Privacy, 16*(4), 20–29.

European Commission (2017). New European Interoperability Framework - Promoting seamless services and data flows for European public administrations. Retrieved from https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf.

Hammer, M. (1990). Reengineering work: don't automate, obliterate. *Harvard Business Review, 68*(4), 104–112.

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly, 33*(3), 371–377. https://doi.org/10.1016/j.giq.2016.08.011.

Janssen, M., Matheus, R., Longo, J., & Weerakkody, V. (2017). *Transparency-by-design as a foundation for open government.* Transforming Government: People, Process and Policy.

Janssen, M., Matheus, R., & Zuiderwijk, A. (2015). Big and open linked data (BOLD) to create smart cities and citizens: Insights from smart energy and mobility cases. In E. Tambouris, M. Janssen, H. J. Scholl, M. A. Wimmer, K. Tarabanis, M. Gascó, ... P. Parycek (Vol. Eds.), *Electronic Government. 9248. Electronic Government* (pp. 79–90). Springer International Publishing.

Janssen, M., & van der Voort, H. (2016). Adaptive governance: Towards a stable, accountable and responsive government. *Government Information Quarterly, 33*(1), 1–5. https://doi.org/10.1016/j.giq.2016.02.003.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 63*(1), 148–152.

Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA Journal, 42*(4), 303–312.

Ladley, J. (2019). *Data governance: How to design, deploy, and sustain an effective data governance program.* Academic Press.

Levi, M., & Stoker, L. (2000). Political trust and trustworthiness. *Annual Review of Political Science, 3*(1), 475–507.

Mullon, P. A., & Ngoepe, M. (2019). An integrated framework to elevate information governance to a national level in South Africa. *Records Management Journal, 29*(1/2), 103–116.

Ølnes, S., Ubacht, J., & Janssen, M. (2017). Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. *Government Information Quarterly, 34*(3), 355–364. https://doi.org/10.1016/j.giq.2017.09.007.

European Parliament and European Council (2016). Regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46. *Official Journal of the European Union, 59*(1–88), 294.

Plotkin, D. (2013). *Data stewardship: An actionable guide to effective data management and data governance: Newnes*.

Rothstein, H., Borraz, O., & Huber, M. (2013). Risk and the limits of governance: Exploring varied patterns of risk-based governance across Europe. *Regulation & Governance, 7*(2), 215–235. https://doi.org/10.1111/j.1748-5991.2012.01153.x.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM, 40*(5), 103–110.

Wihlborg, E., Larsson, H., & Hedström, K. (2016). "The Computer Says No!"–A Case Study on Automated Decision-Making in Public Authorities. *Paper presented at the 2016 49th Hawaii International Conference on System Sciences (HICSS)*.

Yang, K., & Anguelov, L. G. (2013). Trustworthiness of public service. *Public Administration Reformation* (pp. 73–89). Routledge.

Marijn Janssen is a full Professor in ICT & Governance and head of the Information and Communication Technology (ICT) research group of the Technology, Policy and Management (TPM) Faculty of Delft University of Technology.

Paul Brous is researcher at the Information and Communication Technology (ICT) research group of the Technology, Policy and Management (TPM) Faculty of Delft University of Technology.

Elsa Estevez is the Chair holder of the UNESCO Chair on Knowledge Societies and Digital Governance at Universidad Nacional del Sur, Independent Researcher at the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), and full professor at Universidad Nacional de La Plata, all in Argentina.

Luís Soares Barbosa is the deputy head of UNU-EGOV and full professor at the Department of Informatics at the University of Minho.

Tomasz Janowski is head of the Department of Informatics in Management at the Faculty of Economics and Management, Gdańsk University of Technology, Poland and invited professor at the Department for *E*-Governance and Administration, Faculty of Business and Globalization, Danube University Krems, Austria.