Toward big data and analytics governance: redefining structural governance mechanisms

Martin Fadler
Faculty of Business and Economics
(HEC), University of Lausanne
martin.fadler@unil.ch

Christine Legner
Faculty of Business and Economics
(HEC), University of Lausanne
christine.legner@unil.ch

Abstract

Big Data and Analytics (BDA) enable innovative business models and, simultaneously, increase existing business processes' efficiency and effectiveness. Although BDA's potential is widely recognized, companies face a variety of challenges when adopting BDA and endeavoring to generate business value. Researchers and practitioners emphasize the need for effective governance to delineate data and analytics' roles and responsibilities. Existing studies focus either on data or on analytics governance, even though both approaches are closely interlinked and depend on each other. Our study aims to integrate these two distinct research perspectives into a unified view on structural mechanisms for BDA. Using design science research, we iteratively develop data and analytics roles, clarify their responsibilities and provide guidelines for their organizational assignment. Our study contributes to advancing research on data and analytics governance and supports practitioners managing BDA.

1. Introduction

Companies with a clear strategy to monetize their data can strengthen their competitive advantage through new business models, data-driven insights, and improved business processes [1]. Although the potentials of leveraging Big Data and Analytics (BDA) are well known, enterprises face various challenges with BDA adoption and value generation. Among these challenges are processing large data volumes, ensuring the data quality, protecting privacy-related data, as well as governing data [2]. Researchers point out that "without appropriate organizational structures and governance frameworks in place, it is impossible to collect and analyze data across an enterprise and deliver insights to where they are most needed" [3, p.417]. Despite its practical relevance, data governance is a challenging topic that has received much less attention than IT governance [4]. Existing data governance concepts focus on master data and data

quality [5]–[7], but do not consider new requirements that have emerged with BDA [8]. In the past, analytics governance was associated with business intelligence and data warehousing [9], [10], with research only recently opening up to advanced analytics [11]. However, this research stream has not as yet integrated findings from data governance studies. In the context of BDA, an overarching understanding is required of data and analytics governance. This comprises a better understanding of structural governance mechanisms such as roles and responsibilities [13] which are fundamental to any governance design and considered a catalyst of BDA value creation [6]. The following is therefore our research question: *How do enterprises (re-)define their roles and responsibilities in the Big Data and analytics (BDA) context?*

Our overarching research objective is to develop a reference model for enterprise-wide data and analytics governance. In this paper, we emphasize the structural governance mechanisms, which we will complement with procedural and relational mechanisms in further research. Following the guidelines for design science research [12], we developed a role model through multiple iterations involving data and analytics experts from large corporations. Learning from enterprises with extensive experience in data governance and adopting BDA technologies, we derived unique insights into the inner workings of designing data and analytics governance in large enterprises. We also identified common patterns regarding the allocation of roles and responsibilities.

The remainder of this paper is structured as follows. First, we review literature on IT, data, and analytics governance to clarify the commonalities and differences, as well as the research gap that our study examines. We then motivate and describe our design science research approach and outline the research process. Thereafter we present our framework for data and analytics' roles and responsibilities in detail. Finally, we discuss our results, and provide an outlook on future research.

HICSS

2. Background

Although often considered a practitioner topic, governance has gained increasing interest from IS researchers in recent years. Generally speaking, IT governance is associated with "[…] the distribution of IT decision-making rights and responsibilities among different stakeholders in the enterprise, and defines the procedures and mechanisms for making and monitoring strategic IT decisions" [13, p.7]. An increasing number of studies investigate the phenomenon from different perspectives (see Table 1): (1) the governance of IT artifacts (IT governance), (2) the content of IT artifacts (data/information governance), and (3), more recently, the analysis of IT artifacts' content (analytics governance).

In the following, we describe each form in detail and discuss their commonalities and differences.

Table 1 Prior research on IT, Data, and Analytics Governance

	IT	Data	Analytics
Governance forms	Differentiation of structural, procedural, and relational mechanisms [10], [12] IT decision domains and organizational archetypes [14] Governance in specific contexts: Eco-system/ platform governance [15], [16], application governance [17], and IT consumerization [18]	Structural, procedural, and relational information governance mechanisms [19] Roles, decision areas, and main activities of data quality management [5] Data management's decision domains and locus of accountability [20]	Structural, procedural, and relational analytics governance mechanisms [11] Typical roles/process of analytics competence centers [21] Governance in specific contexts: accountability in algorithmic decision-making [22], AI governance [23], and BDA ownership types [8]
Contingency	Multiple contingencies and their interaction (reinforcing, conflicting, and dominating) in IT governance design [24]	Multiple contingencies in data governance design and data quality management success [5] Uncertainty and business unit similarity regarding a decision to centralize/ decentralize data management [25]	*No studies found*

2.1. IT governance (IT artifact)

Most of IS studies on governance have been conducted with a focus on governing IT artifacts on the firm level [4]. Previous research can be broadly divided into two distinct streams with one focusing on IT governance forms and the other on IT governance contingency influences [26].

Studies on IT governance forms investigate the organizational assignment of the decision-making authority and the organizational structuring of IT activities in order to increase the return on investment [26, p.700]. Typical discussions include the advantages and disadvantages of assigning IT decisions in a central, decentral, or federated organization. While a centralized organization emphasizes control over IT standards and increases the opportunity for an economy of scale, a decentralized allocation allows for customizing IT solutions to meet the business needs and enables flexibility. The federated IT governance form allows companies to gain advantages from the two extremes. Here, a central IT organization seeks economies of scale and provides core IT services to the entire enterprise, while business units retain their flexibility and build their own solutions if the offered services were to not address the business needs sufficiently. In addition, researchers also examined the IT governance architecture [13]. They suggested differentiating between the structural (e.g. roles and responsibilities), procedural (e.g. processes), and relational (e.g. alignment) governance capabilities or mechanisms [13], [27].

Studies on IT governance contingency influences aim to understand which form is most suitable for which company type. They analyze factors that affect individual IT governance framework success [26, p.703], such as the organizational structure, business strategy, industry, and firm size. Researchers investigate either single or multiple contingency factors. The seminal study by Sambamurthy and Zmud applies the theory of multiple contingencies to examine the factors that influence decisions on the IT governance form [24]. It reveals that the contingency factors interact with one another along three scenarios: reinforcing, conflicting, and dominating.

Although IT governance has been extensively researched, innovative technologies and changes in technology lead to new requirements arising. The governance of platforms and eco-systems, for instance, requires stakeholders' involvement beyond the firm's boundaries [15], [16]. Furthermore, application governance becomes more important when companies increasingly rely on Internet-based delivery models, such as software-as-a-service (SaaS), which challenge traditional governance assumptions [17]. Other researchers emphasize that IT consumerization could transform the fundamentals of IT governance [18].

2.2. Data governance (Content)

Compared to IT artifacts' governance, the content perspective has received much less research attention, but BDA and the processing of massive amounts of data have made it more important [4]. While IT governance aims to manage IT assets in the sense of hardware and software components that help support the automation of well-defined tasks, data governance aims to manage data (or information) assets as facts having value or potential value that are documented

[20, p.148]. Consequently, the need for data governance evolved with data's importance for enterprises. [28] describe how data management evolved in three phases: data administration (since the 1980s), quality-oriented data management (since the 1990s) and extensions of strategic data management (since the 2010s). Data governance issues were only raised in the second phase when data quality became a significant success factor in system integration projects (e.g. data warehouse implementations). Similar to IT governance, existing research can be broadly divided into understanding data governance forms and analyzing contingency influences. The first data governance frameworks emerged when companies started managing master data (a company's core data objects, e.g. its business partners or products), which requires data quality management and coordination across business units [29], [30]. [8] use the IT governance framework's general structure suggested by [14] to derive five key decision domains (data principles, data quality, metadata, data access, and data lifecycle) for data governance and map them to the locus of accountability (central to decentral). These authors use an example to illustrate that data principles should, for instance, be centrally defined, while data quality should be managed in a decentralized way.

Other researchers focus on data governance mechanisms. [19] conclude that many of the factors leading organizations to adopt information governance practices are equally relevant when governing physical IT artifacts (p. 170). These authors therefore use the same structure of [13], [27] to define structural, procedural, and relational information governance mechanisms. [5] define typical roles (executive sponsor, data quality board, chief steward, business data steward, and technical data steward), decision areas, and the main activities of data quality management. The business data steward is, for example, a key role in data quality management and responsible for detailing, from a business perspective, corporate-wide data quality standards and policies for his/her area of responsibility [5, p.11]. The authors further analyze the influence of seven contingency factors (e.g. performance strategy or organization structure) on data governance design and data quality management success. While this study has a narrow focus on data quality management, a more recent study examines the influences of similarities between business units and uncertainty on a decision to centralize or decentralize their data management [25]. One of their findings is that when business units are very similar and uncertainty is low, their data management should be centralized, which will, for example, reduce the coordination effort.

Only a few studies have investigated data governance in the BDA context. [31] argues that data governance practices need to balance value creation and risk exposure to gain a competitive advantage and maximize the business value through BDA. This balance requires companies to continuously evaluate the business value of their data assets, while ensuring that the risks are monitored and assessed. [32] argues that information governance is required to facilitate BDA capabilities. [8] endeavor to explicate the fundamentals of a BDA governance model by defining three data ownership types: data owner, data platform owner, and data product owner. The latter study stresses that new roles and responsibilities are required to manage BDA on a firm level. While data owners ensure controlled access and use of data at the data source level, data product owners are accountable to ensure that data products generate business value over their lifecycle. Since data are a strategic asset, [33] introduce the role of a chief data officer (CDO).

2.3. Analytics governance (Analysis of content)

More sophisticated forms of analytics involve artificial intelligence and automated decision-making, requiring new roles and responsibilities, but also leading to new risks. Governance should therefore not be limited to the content, but should also encompass its analysis.

Researchers emphasize that, in addition to IT and data governance, analytics governance mechanisms are required to overcome challenges, such as the alignment between business users and analytics practitioners [11]. Another well-known challenge is that data scientists still spend 80% of their time "[...] finding, cleaning, and organizing data" [34, p.4]. To improve this situation, the alignment with business users and with data management experts is of particular importance. Consequently, researchers have started investigating analytics governance forms. Based on a literature review, [9] suggest a framework of the structural (organization structure, coordination and alignment, and roles and responsibilities), procedural (process model, monitoring and evaluation, and development), and relational mechanisms (shared perceptions, collaboration, and transfer of knowledge). These authors apply their framework to analyze three case studies, with only one case appearing to have reached high analytics governance maturity. Other studies focus on specific areas of analytics governance or particular organizational setups. [21], for instance, examine analytics competence centers to understand how analytics capabilities can be cultivated in enterprises. While the core of these researchers' investigation is strategic by nature and on capability

development, they also identify analytics competence centers' (ACC) typical roles and responsibilities and a general process for ACCs. Their findings can therefore be considered structural and procedural governance mechanisms, although they are not phrased as such.

With the increasing application of artificial intelligence and machine learning, new risks are encountered, for example, discriminatory effects and privacy infringements [35]. These risks have led to increasing public awareness and questions about how these technologies should be governed in the future. [22] investigates accountability in algorithmic decision-making, proposing an algorithmic transparency standard. [23] formulates a research agenda for artificial intelligence (AI) governance to overcome risks such as "[…] labor displacement, inequality, an oligopolistic global market structure, reinforced totalitarianism, shifts and volatility in national power, strategic instability, and an AI race that sacrifices safety and other values" [23, p.1]. While these AI governance considerations could have implications for enterprises, their scope is more political.

2.4. Research gap

BDA has led to new challenges emerging and enterprises struggling to gain a return on their BDA investments. In these situations, governance - specifically defined responsibilities and data accountabilities - is considered a catalyst for BDA value creation [3, p.417]. Existing studies focus either on data or on analytics governance, even though both approaches are closely interlinked and depend on each other [8]. We therefore argue that an integrated view is needed to overcome the challenges and ensure that BDA investments generate value. Nevertheless, most studies have investigated data and analytics governance by uncovering general governance mechanisms. The identified structural, procedural, and relational mechanisms mirror the IT governance literature. While existing studies mainly focus on identifying the generic mechanisms, we lack further details of structural mechanisms in the context of BDA and insights into their design and implementation in enterprises. Research on decision rights is, for instance, part of the structural governance mechanisms stream, but has not investigated the roles and responsibilities in greater detail. While governance models for master data and data quality have elaborated on the roles and responsibilities, the first attempts have been made to define common analytics roles, albeit for educational purposes [36]. Consequently, the definition and assignment of roles and responsibilities for data and analytics are a promising direction, which [4] also is emphasizes.

3. Methodology

In this paper, our objective is to develop structural mechanisms for enterprise-wide data and analytics governance with a particular focus on roles and responsibilities. Using design science research (DSR) [12], we worked closely with five enterprises (see Table 2) for a period of 17 months to understand their current challenges and approaches in term of governing BDA and developed a role model in multiple iterations.

All the enterprises are suitable with regard to supporting our research goals, because they (1) have a high maturity regarding managing data (2) have gained first experiences with adopting BDA (e.g. all the companies have an established data lake), and (3) are in the process of (re-)designing their data and analytics governance models.

Table 2 Companies involved in the research process

Company	Industry	Revenue/ # Employ.	Current situation
A	Consumer goods	50–100 B$ / ~80 000	Central data and analytics management unit to operate a central big data platform and distributed BI infrastructures
B	Public transport-tation	1–50 B$ / ~35 000	Central data management organization, decentralized analytics teams working with multiple analytics infrastructures
C	Industry products	50–100 B$ / ~110 000	Central data management org. and advanced analytics group operating multiple data lakes and data warehouses
D	Consumer goods	1–50 B$ / ~30 000	Central data and analytics management organization with a high business intelligence maturity managing a central enterprise data warehouse and one data lake
E	Manu-facturing	1–50B$ / ~90 000	Central data management organization and central platform team that enable digital innovations and industrialize analytics products

The chosen DSR process model encompasses six steps [12]. Before starting this process, an adequate research entry point had to be selected. All the companies already had an established role model, but with a focus on data management. The major drawbacks were, on the one hand, that the existing roles and responsibilities originated from master data management and could only address the new BDA requirements partly. On the other hand, analytics-related roles were somewhat defined, but not integrated into the overall governance framework. We therefore initiated the research process with an objective-centered solution, asking the question: *what would a*

better role model accomplish? [12]. In the first step (*identify problem & motivate)*, we conducted one expert interview with each company to understand their initial situation. We also held a one-day focus group workshop in February 2019, where we presented the findings and discussed the challenges of managing data lakes compared to traditional business intelligence environments. The group agreed that existing data governance models needed to be extended to cope with BDA requirements and achieve a unified enterprise view. In the second step, we *defined the objectives of the solution*. In March 2019, we allowed the companies to first present their current roles and responsibilities and thereafter discussed the missing roles and responsibilities in the BDA context in a second focus group meeting that was held virtually. During this session, the group reached consensus that (1) the general structuring of data and analytics organizations needed to be reviewed in the context of BDA and (2) that the outcome should be a framework depicting best practices that could be used as reference to design roles and responsibilities for data and analytics governance.

To fulfill these requirements, we developed the model in three *design & development, demonstration,* and *evaluation* iterations (see Figure 1).

In the first iteration (Mar–Aug 2019), we focused on the roles and responsibilities of the emerging analytics platforms. In order to do so, we analyzed each company's role models for their data lakes, and conducted a literature review to attain a list of relevant data and analytics roles. We then defined an area of responsibility for each role, based on an information supply chain depicting the steps needed to provide analytics products. In a focus group, we demonstrated how this first version of the role model could be used to assign roles and allocate their responsibilities, thereafter asking the participants to assess the model's applicability and usefulness. While the companies appreciated the ease with which they understood the framework, they remarked that the focus was too narrowly on analytics platforms and failed to determine the required responsibilities for managing data and analytics enterprise-wide. In addition, we evaluated the list of data and analytics roles by means of a survey, asking companies to assess the completeness of the roles and assign the area of responsibility in the context of their company.

In the second iteration (Sept 2019–Dec 2019), we revised the model, integrating the previous iteration's feedback. We structured the framework by categorizing the roles and responsibilities according to their organizational level with regard to the strategy, governance, and operations. Furthermore, we extended the information supply chain with an additional step to emphasize the creation and maintenance of data on the

source level. We demonstrated the new version by applying it to two analytics products (advanced analytics model and dashboard) during an executive course with a group of 15 professionals (Sept 2019). In addition, we ran a focus group with the involved companies. While the categorization regarding the strategy, governance and operations ensured that all the relevant roles and responsibilities were assigned, it became apparent that the accountable roles had to be allocated first. This led to discussions about the locus of accountability and the decision to identify core data/analytics roles that have to interact with business and IT roles.

In the third iteration (Jan 2019–May 2020), we designed the final version (see Section 4). We demonstrated its general applicability by mapping the companies' role models to the framework and presented the results in a focus group. The group reached consensus that the framework was useful and could be used to allocate roles and responsibilities in the context of BDA, thus fulfilling the solution's defined objectives.
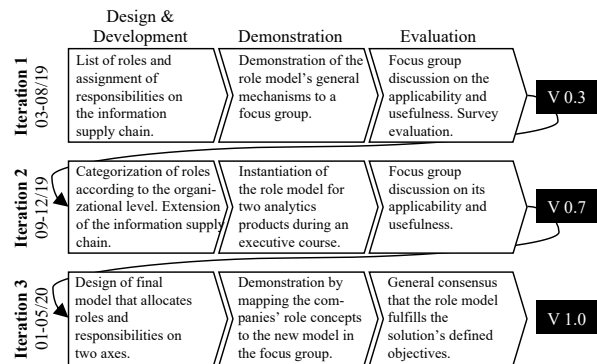

Figure 1 Design science research iterations according to [8]

4. Roles and responsibilities for BDA

In the following, we present the suggested role model for enterprise-wide data and analytics governance. First, we discuss the general principles which guide the design of roles and responsibilities for data and analytics governance in enterprises. Second, we present the roles and responsibilities and illustrate the framework in terms of a typical organization.

4.1. Design principles

At the beginning of our research process, three important design decisions were taken that guided the definition and assignment of roles and responsibilities for data and analytics in enterprises.

Federated model for enterprise-wide data and analytics governance - Any data management and analytics activities require alignment and collaboration with business and IT departments. The complete centralization of data and analytics in an enterprise would potentially increase the economies of scale, but diminish business units' flexibility (no customization) and value generation through data and analytics (no business self-responsibility, because the ownership remains with the data and analytics organization). Conversely, complete decentralization makes business units flexible, but leads to data silos and hinders data sharing and integration across functions [25]. Consequently, a federated approach is needed for enterprise-wide data and analytics governance. This implies that the roles and responsibilities can be assigned to employees who work in different parts of the enterprise:

- A federated data and analytics organization relies on central teams with core data and analytics roles that coordinate data management and analytics delivery activities at the enterprise level. They ensure that business requirements are correctly transformed into data and analytics products.

- Since the ownership of data and analytics products lies with business [8], [37], business roles play an important role. They define business requirements for data and analytics products and own their data and analytics products.

- IT roles support data management and analytics delivery by means of infrastructure and IT services. This includes the operation of analytics products and the development of analytics platforms.

Governance at the intersection between strategy and operations - Generally speaking, governance implements a strategy by means of oversight and control mechanisms [38] and complements strategic as well as operational tasks: *Strategy is doing the right things, operations are doing things right, and governance is ensuring that the right things are done right.*

The suggested model therefore considers governance roles as complementing the roles on the strategic and operational levels.

- The objective and long-term direction for data and analytics are defined at the strategic level. This includes sponsorship, strategic direction, funding, and the coordination of data management and analytics activities at an enterprise-wide level.

- The governance level implements the strategy through oversight and control mechanisms. While

enterprise-wide data and analytics governance is cross-functional, defines the overarching governance framework and controls its implementation, it needs to be detailed for the different business units or departments by defining the standards and the policies of the areas of responsibility.

- The operations level executes the strategy through day-to-day activities, operates the data and analytics product lifecycle based on the defined standards, and takes responsibility for the correctness of the data content and the use of analytics products.
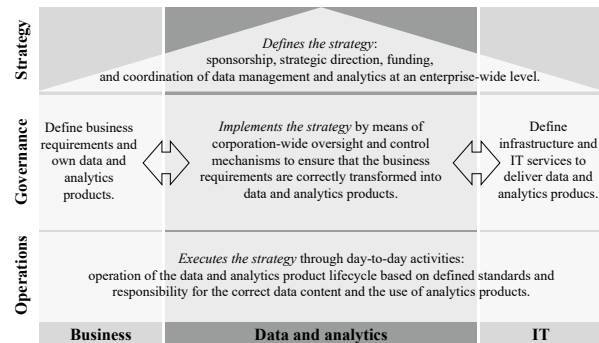

Figure 2 Data and analytics governance at the intersection of strategy and operations

Data and analytics roles facilitate the information supply chain - While the data roles emphasize the provision of data for different business purposes, the analytics roles endeavor to deliver analytics products throughout the enterprise and integrate data across the business units. Both roles clearly depend on each other and facilitate information supply chains. Their responsibilities can be summarized as follows:

- Data roles aim to make data fit for use for business processes, data-driven insights, and digital products and services.

- Analytics roles aim to deliver different types of analytics products, for example, reports, ad-hoc analysis, data science experiments, and production.

According to these design considerations, we allocate roles and responsibilities along two axes (see Figure 2). The horizontal axis defines whether a role is part of the core data and analytics organization, or primarily a business and/or IT role. The vertical axis allocates roles' general responsibility according to the strategy, governance, and operations. Using this framework, roles can be allocated according to the described dimensions. In this regard, it is important to note that a role can be allocated to bordering areas if

they only belong partially to the data and analytics organization. This applies to data owners, for instance, who are part of a business function but are assigned by the data organization.

In the following, we present the governance-related roles and responsibilities in detail. In contrast to existing research, which mostly focuses on the allocation of decision control rights (accountable roles) [39], we define the accountable and responsible roles. This is motivated by our observation that companies are increasingly required to define their roles and responsibilities at this level of detail to enforce governance.

4.2. Roles and responsibilities in detail

Before we delve into the data and analytics governance roles and responsibilities, we briefly define the role of the Chief Data Officer, who plays as an integrating role between the strategy and governance level.

Chief Data Officer as integrating role - The role of the Chief Data Officer (CDO) - also called head of data and analytics or chief data and analytics officer – is becoming of major importance in enterprises, because data are increasingly recognized as a strategic asset [33]. A recent study established that companies with a CDO are twice as likely to have a clear digital strategy [40]. Of the surveyed companies, 67,9% had a CDO assigned in 2018 [40]. A CDO is the head of the central data and analytics organization, is responsible for the overall data management and analytics strategy, and accountable for its implementation. This range of activities requires continuous exchanges with the data and analytics organization's executive sponsor on the business side, as well as with the chief information officer (CIO) on the IT side. In the role model suggested by [5], a CDO fulfills the chief data steward role and extends his or her accountability to the analytics organization.

In addition, companies increasingly establish a dedicated data and analytics board comprised of C-level executives to align the stakeholders on the enterprise level. This board is accountable for defining the data and analytics strategy, controlling its implementation (including compliance requirements), and setting priorities.

Data roles and responsibilities - An effective data governance design (see Table 3) requires data ownership to remain with the business functions. It also requires a rather central organization of data stewards and data architects, who, for instance, set and enforce enterprise-wide standards for data

documentation, or facilitate data unification activities to enable experimentation with and exploration of data lakes.

Two types of ownership need to be distinguished in terms of data ownership: the data definition owner and the data content owner. Both roles are usually assigned to senior executives responsible for a defined business domain (e.g. business process) and who have strategic responsibility (e.g. head of sales). In respect of the data in his/her domain, the data definition owner is accountable for data definitions of business and quality rules, data access policies, data lifecycle, and the conceptual data model. The data definition ensures that data are created and used in controlled ways. The data definition owner collects business requirements for the defined area of responsibility (e.g. a particular data domain like a business partner or product) from other business process owners and from the compliance officer.

Table 3 Data roles and responsibilities

Role name	Decision right and area	Allocation
Data definition owner	Accountable for the data definition in specific areas of responsibility (e.g. a specific data domain).	Business
Data steward	Responsible for the data definition in specific areas of responsibility (e.g. the data field attributes of a data object in a specific data domain).	Data & analytics organization/ Business
Data architect	Responsible for designing, creating, deploying, and managing conceptual and logical data models as well as for the mapping to physical data models. Accountable for the implementation and maintenance of data pipelines.	Data & analytics organization/ IT
Data content owner	Accountable for data creation and maintenance (data lifecycle) according to a specific area of responsibility's data definition.	Business/Shared service center
Data editor	Responsible for data creation and maintenance (data lifecycle) according to a specific area of responsibility's data definition.	Business/Shared service center
Data expert	Responsible for communicating data the definition and for training data editors.	Business/Shared service center

While the data definition owner is accountable, she or he is often at a high hierarchical level, for instance, the head of material or supplier data purchasing. The data steward is responsible for specific areas of responsibility's data definition and is often part of the central data and analytics team. Here, the data steward takes care of certain data fields of a data object in a specific data domain. This includes defining data while enforcing data quality measures and ensuring that data is *fit for use* [41]. The data architect supports the data steward by designing, creating, deploying, and managing conceptual and logical data models, as well as with mapping to physical data models. In the role model defined by [5], the data architect role corresponds to the technical data steward role and complements the business steward. With the

emergence of new data types (e.g. IoT/sensor data) and analytics use cases, the data definition needs to be continuously adapted and serves as a central element to ensure facilitated access and use across the enterprise. The data steward therefore needs to handle data requests from different business functions.

The data content owner's role is usually assigned to executives with operational responsibilities (e.g. the head of sales of a specific country), who are accountable for creating data according to the relevant data definition. This role manages a team of data editors, who are responsible for data creation. The data expert is another typical role on the operations level. This expert has no other major responsibility besides communicating the data definitions to the data editors and training them.

Analytics roles and responsibilities - An effective analytics governance design (see Table 4) requires the requestors and users of analytics products to collaborate with the data and analytics organization and IT.

On the business side, executives in business domains who sponsor and request analytics products usually represent the analytics product requirement owner's role. In this sense, she or he is accountable for the business value and specification of an analytics product's business requirements. Accordingly, the person assuming this role has to stimulate the identification and use of analytics products in her or his area of responsibility in order to increase data-driven decision-making and communicate with important business stakeholders. A business analyst, in the analytics product requirement owner's area of responsibility, is responsible for the specification of the analytics product on the operations level. While the analytics product requirement owner specifies the business requirements, the analytics product lifecycle owner is accountable for implementing these requirements in a specific analytics product, doing so by coordinating its development, deployment, and maintenance. In addition, this analytics product lifecycle owner is responsible for defining analytics product standards and guidelines, assuring quality, and for managing the lifecycle as part of her or his governance responsibility.

On an operations level, the coordinates the data analysts, data scientists, and data engineers responsible for analytics products' development and deployment. In order to do so, she or he involves the business stakeholders to ensure that the business requirements are met. The analytics product lifecycle owner is typically a person with project management experience with technical know-how of analytics product development. The Analytics product architect's role is

meant to ensure applications' reusability and scalability across the enterprise. This architect is responsible for analytics products and analytics product architecture's design, which requires close collaboration with the IT organization. Consequently, this role is allocated to the bordering area of analytics/IT.

Two data governance roles are of particular importance for the analytics organization. The data architect is accountable for data pipelines' implementation and maintenance by providing the data models that data engineers use. The data steward, a key role for data governance, is responsible for managing analytics projects' data requests and for supporting the data onboarding process. This support is of particular importance to increase the analytics practitioners' efficiency and reduce the time spent on finding and preparing data.

Table 4 Analytics roles and responsibilities

Role name	Decision right and area	Allocation
Analytics product requirement owner	Accountable for the business value and the specification of the business requirements of an analytics product.	Business
Analytics product architect	Responsible for the design of analytics products and analytics product architecture.	Data & analytics organization/ IT
Analytics product lifecycle owner	Accountable for the implementation (development and deployment) and maintenance of an analytics product. Responsible for analytics product standards and guidelines, quality assurance, and the lifecycle management.	Data & analytics organization
Business analyst	Responsible for the business value and specification of an analytics product's business requirements.	Business
Data analyst	Responsible for the implementation (development and deployment) and maintenance of reports and ad-hoc analyses.	Data & analytics organization
Data scientist	Responsible for the implementation (development and deployment) and maintenance of advanced analytics models.	Data & analytics organization
Data engineer	Responsible for data pipelines' implementation and maintenance.	Data & analytics organization/ IT
Analytics expert	Responsible for the training of analytics product users.	Business/ Data & analytics organization

A central data and analytics organization ensures that requests for new analytics products (e.g. data science use case) are prioritized and specified within an enterprise-wide demand management process. Although all companies still distinguish between the delivery of BI (e.g. reporting) and advanced analytics products (e.g. predictive modelling), they seek an integrated, unified view on analytics products' demand and delivery in the long term, in order to bundle resources and facilitate BDA capabilities. Business roles' involvement guarantees that the business

requirements are met, and the domain knowledge is transferred to analytics products.

5. Conclusions and outlook

Our study aims to integrate two distinct research perspectives on data and analytics governance into a unified view. Based on a close collaboration with five large corporations with extensive BDA experience, we provide a role model that explicates how structural governance should be pursued in the context of BDA on an enterprise level. More specifically, it defines roles and responsibilities for data and analytics and provides guidelines for their organizational assignment. The enterprise-wide governance of data and analytics requires a federated organization, because business units need to assume ownership of data content and analytics products. On the other hand, a central data and analytics organization ensures that data are "fit for use" and analytics products are delivered efficiently across business functions and in line with the strategic direction. Furthermore, roles and responsibilities for data and analytics need to be assigned on the strategy, governance, and operations levels. This means that data and analytics governance roles have to work hand-in-hand with strategic roles and with those involved in day-to-day operational tasks. The chief data officer (or chief data and analytics officer) plays an important role in linking the strategic and governance levels. Although data and analytics have largely been viewed separately, their interdependence in managing today's information supply chains cannot be questioned. The data steward and data architect each play an integrating role between the data and analytics organizations.

The role model and derived design considerations contribute to IS governance literature in general and to structural governance mechanisms in particular. Practitioner can use the framework to design their own data and analytics organizations.

Our study does have limitations. Since we only collaborated with large, multi-national corporations to build the framework, our results might not be applicable to smaller size organizations. Furthermore, we could not evaluate the actual implementation of the defined roles and responsibilities since most companies have not yet implemented all roles. However, all the companies use the framework to design their own governance models.

As already mentioned, this study contributes to a larger research program in which we develop a reference model for data and analytics governance. The latter not only comprises structural mechanisms, but also procedural and relational ones. The roles presented in this paper help to clarify responsibilities in processes and to define the required interactions. While we have a clear plan for our research activities, we also see interesting avenues for future research: The presented roles can be used as a reference to investigate how companies implement data and analytics governance and compare different organizational setups. Besides comparing centralized to decentralized setups, a more in-depth investigation of the coordination between data and analytics roles with regard to overarching governance goals (for instance, data quality) might be an interesting research opportunity to foster the unified view that we suggest in our study.

6. References

[1] B. Wixom and J. Ross, "How to Monetize Your Data," *MIT Sloan Management Review*, vol. 58, no. 3, 2017.

[2] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *Journal of Business Research*, vol. 70, pp. 263–286, Jan. 2017.

[3] V. Grover, R. H. L. Chiang, T.-P. Liang, and D. Zhang, "Creating Strategic Business Value from Big Data Analytics: A Research Framework," *Journal of Management Information Systems*, vol. 35, no. 2, pp. 388–423, Apr. 2018.

[4] A. Tiwana, B. Konsynski, and N. Venkatraman, "Special Issue: Information Technology and Organizational Governance: The IT Governance Cube," *Journal of Management Information Systems*, vol. 30, no. 3, pp. 7–12, Dec. 2013.

[5] K. Weber, B. Otto, and H. Österle, "One Size Does Not Fit All---A Contingency Approach to Data Governance," *Journal of Data and Information Quality*, vol. 1, no. 1, pp. 1–27, Jun. 2009.

[6] J. J. Korhonen, I. Melleri, K. Hiekkanen, and M. Helenius, "Designing Data Governance Structure: An Organizational Perspective," *Journal on Computing*, vol. 2, no. 4, 2013.

[7] B. Otto, "Data Governance," *Business & Information Systems Engineering*, vol. 3, no. 4, pp. 241–244, Aug. 2011.

[8] M. Fadler and C. Legner, "Who owns data in the enterprise? Rethinking data ownership in times of big data and analytics," Marrakesh, Morocco, 2020.

[9] H. J. Watson, C. Fuller, and T. Ariyachandra, "Data warehouse governance: best practices at Blue Cross and Blue Shield of North Carolina," *Decision Support Systems*, vol. 38, no. 3, pp. 435–450, Dec. 2004.

[10] R. Winter and M. Meyer, "Organization of data warehousing in large service companies - A matrix approach based on data ownership and competence centers," *Journal of Data Warehousing*, vol. 6, no. 4, pp. 23–29, 2001.

[11] J. Baijens, R. W. Helms, and T. Velstra, "Towards a Framework for Data Analytics Governance Mechanisms," 2020.

[12] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, Dec. 2007, doi: 10.2753/MIS0742-1222240302.

[13] R. Peterson, "Crafting Information Technology Governance," *Information Systems Management*, 2004.

[14] P. Weill and J. Ross, "A Matrixed Approach to Designing IT Governance," *MIT Sloan Management Review*, vol. 46, no. 2, pp. 26–34, 2005.

[15] A. Tiwana, B. Konsynski, and A. A. Bush, "Research commentary-Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics," *Information Systems Research*, vol. 21, no. 4, pp. 675–687, 2010.

[16] J. Wareham, P. B. Fox, and J. L. Cano Giner, "Technology Ecosystem Governance," *Organization Science*, vol. 25, no. 4, pp. 1195–1215, Mar. 2014.

[17] T. J. Winkler and C. V. Brown, "Horizontal Allocation of Decision Rights for On-Premise Applications and Software-as-a-Service," *Journal of Management Information Systems*, vol. 30, no. 3, pp. 13–48, Winter 2013.

[18] R. W. Gregory, E. Kaganer, O. Henfridsson, and T. J. Ruch, "IT Consumerization And The Transformation Of It Governance," *MIS Quarterly*, vol. 42, no. 4, pp. 1–9, 2018.

[19] P. P. Tallon, R. V. Ramirez, and J. E. Short, "The Information Artifact in IT Governance: Toward a Theory of Information Governance," *Journal of Management Information Systems*, vol. 30, no. 3, pp. 141–178, Dec. 2013.

[20] V. Khatri and C. V. Brown, "Designing Data Governance," *Communication of the ACM*, vol. 53, no. 1, pp. 148–152, 2010.

[21] R. Schüritz, E. Brand, G. Satzger, and J. Bischhoffshausen, "How to cultivate analytics capabilities within an organization? - Design and types of analytics competency centers," in *Proceedings of the 25th European Conference on Information Systems (ECIS)*, Guimarães, Portugal, Jun. 2017, pp. 389–404.

[22] N. Diakopoulos, "Accountability in algorithmic decision making," *Communications of the ACM*, vol. 59, no. 2, pp. 56–62, Jan. 2016.

[23] A. Dafoe, "AI Governance: A Research Agenda," Centre for the Governance of AI, Future of Humanity Institute, University of Oxford, 2018.

[24] V. Sambamurthy and R. W. Zmud, "Arrangement for Information Technology Governance: A Theory of Multiple Contingencies," *MIS Quarterly*, vol. 23, no. 2, pp. 261–290, Jun. 1999.

[25] C. K. Velu, S. E. Madnick, and M. W. Van Alstyne, "Centralizing Data Management with Considerations of Uncertainty and Information-Based Flexibility," *Journal of Management Information Systems*, vol. 30, no. 3, pp. 179–212, Dec. 2013.

[26] A. E. Brown and G. G. Grant, "Framing the Frameworks: A Review of IT Governance Research," *Communications of the Association for Information Systems*, vol. 15, 2005, Accessed: Apr. 15, 2020.

[Online]. Available: https://aisel.aisnet.org/cais/vol15/iss1/38.

[27] S. De Haes and W. Van Grembergen, "IT Governance and its Mechanisms," *Information Systems Control Journal*, vol. 1, pp. 27–33, 2004.

[28] C. Legner, T. Pentek, and B. Otto, "Accumulating Design Knowledge with Reference Models: Insights from 12 Years of Research on Data Management," *Journal of the Association for Information Systems*, vol. 21, no. 3, 2020.

[29] D. Loshin, Ed., "Master Data Management," in *Master Data Management*, Boston: Morgan Kaufmann, 2009, pp. i–ii.

[30] H. A. Smith and J. D. McKeen, "Developments in Practice XXX: Master Data Management: Salvation Or Snake Oil?," *CAIS*, vol. 23, no. 4, 2008.

[31] P. P. Tallon, "Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost," *Computer*, vol. 46, no. 6, pp. 32–38, Jun. 2013.

[32] P. Mikalef, J. Krogstie, R. van de Wetering, I. O. Pappas, and M. N. Giannakos, "Information Governance in the Big Data Era: Aligning Organizational Capabilities," 2018.

[33] Y. Lee, S. Madnick, R. Wang, F. Wang, and H. Zhang, "A Cubic Framework for the Chief Data Officer (CDO): Succeeding in a World of Big Data," *MIS Quarterly Executive*, vol. 13, no. 1, 2014.

[34] H. Bowne-Anderson, "What Data Scientists Really Do, According to 35 Data Scientists," *Harvard Business Review Digital Articles*, pp. 2–5, Aug. 2018.

[35] B. Custers, "Data Dilemmas in the Information Society: Introduction and Overview," in *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases*, B. Custers, T. Calders, B. Schermer, and T. Zarsky, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 3–26.

[36] J. Saltz, F. Armour, and R. Sharda, "Data Science Roles and the Types of Data Science Programs," *Communications of the Association for Information Systems*, vol. 43, pp. 615–624, 2018.

[37] M. Van Alstyne, E. Brynjolfsson, and S. Madnick, "Why not one big database? Principles for data ownership," *Decision Support Systems*, vol. 15, no. 4, pp. 267–284, Dec. 1995.

[38] A. Tiwana and S. K. Kim, "Discriminating IT Governance," *Information Systems Research*, vol. 26, no. 4, pp. 656–674, Oct. 2015.

[39] T. J. Winkler and M. Wessel, "A Primer on Decision Rights in Information Systems: Review and Recommendations," in *Proceedings of ICIS 2018*, San Francisco, CA, 2018, p. 17.

[40] Forbes, "Forbes Insights: Rethinking The Role of Chief Data Officer," *Forbes*. https://www.forbes.com/sites/insights-intelai/2019/05/22/rethinking-the-role-of-chief-data-officer/ (accessed Jul. 14, 2020).

[41] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–33, 1996.