From Data Governance by design to Data Governance as a Service

A transformative human-centric Data Governance framework

Armend Duzha*
Dept. of Digital Systems, University of Piraeus, Greece

Emmanouil Alexakis
Dept. of Digital Systems, University of Piraeus, Greece

Dimosthenis Kyriazis
Dept. of Digital Systems, University of Piraeus, Greece

Louis Sahi
IRIT, University of Toulouse III, France

Ali Mohamed Kandi
IRIT, University of Toulouse III, France

ABSTRACT

Nowadays, companies and governments collect data along every phase of a product/service lifecycle. The data is acquired and stored in various and continuous ways, contributing to a large and unique data fingerprint for every product/service in use. Thus, establishing policies, processes and procedures (P3) around data and subsequently enacting those to compile and use such data for effective management and decision-making is extremely important. Data governance (DG) plays an essential role in a dynamic environment with multiple entities and actors, complex IT infrastructures, and heterogeneous administrative domains. Indeed, not only is it beneficial for existing products/services, but it can also support appropriate adjustments during the design of new ones. This research provides an overview of the existing literature and current state-of-the-art in the domain of data processing and governance aiming to evaluate existing approaches and to investigate their limitations. To this extent, this study introduces a novel approach for data governance as a service (DGaaS), which provides a framework for (private or public) organizations that facilitate alignment with their vision, goals and legal requirements. It is flexible and adaptable to the needs of the different sectors, and aims at increasing the value of data and minimize data processing related costs and risks.

CCS CONCEPTS

• **Collaborative and social computing systems and tools**; • **Data management systems**; • **Privacy policies**;

KEYWORDS

Data Governance, Data Processing, Data Quality, Decentralization, Trust

*__CRediT Author Statement__: Armend Duzha: Conceptualization (lead), Investigation, Visualization, Writing – Original Draft (lead), Writing – Review and Editing (lead). Emmanouil Alexakis: Conceptualization, Writing – Review and Editing. Dimosthenis Kyriazis: Validation, Writing – Review and Editing. Louis Sahi: Writing – Original Draft. Ali Mohamed Kandi: Writing – Review and Editing.

ICCBDC 2023, August 17–19, 2023, Manchester, United Kingdom

ACM ISBN 979-8-4007-0733-9/23/08.
https://doi.org/10.1145/3616131.3616145

ACM Reference Format:
Armend Duzha, Emmanouil Alexakis, Dimosthenis Kyriazis, Louis Sahi, and Ali Mohamed Kandi. 2023. From Data Governance by design to Data Governance as a Service: A transformative human-centric Data Governance framework. In *2023 7th International Conference on Cloud and Big Data Computing (ICCBDC 2023), August 17–19, 2023, Manchester, United Kingdom.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3616131.3616145

1 INTRODUCTION

Many organizations consider data as one of the most important assets [1] [2] [3]. By leveraging data processing (DP), they transform raw data into valuable information and tangible business/economic benefit. Data governance (DG) plays a crucial role in defining, implementing, and monitoring the context, responsibilities, tools and stakeholders involved in DP. It refers to the system of decisions and accountabilities that regulate and guide information-related processes, ensuring adherence to pre-defined policies [4]. DG outlines permissible activities, assign responsibilities to different entities and stakeholders, and determine the data to be used. It also enforces rules to enable proper data management (DM) throughout its entire lifecycle [5].

In the context of AI-based systems, DG takes a broader role as the system of processes and infrastructures that enable organizations to align AI tools and technologies with their strategies, objectives, and values while maximizing the value of data [6]. To this effect, [7] [8] propose data governance by design (DGbD), which provides matrices that facilitate the design of effective data governance frameworks (DGFs) for organizations. DGbD offers a comprehensive and proactive approach to managing and upholding organizations' applications and processes. It ensures through oversight of the associated data throughout its entire lifecycle, ranging from requirements elicitation and engineering to execution management.

Data is sensitive and ubiquitous in the organization's activities. The emergence of distributed systems has led to collaborative data processing (CoDP), involving multiple entities and actors [9] [10]. In this context, DG encompasses a set of principles that govern data flows processed by different entities and actors. It establishes rights and responsibilities that guide DP according to agreed-upon models, and enforcing actions that can be taken, by whom, with which data, and for what purposes [4]. However, traditional approaches such as DGbD are no longer sufficient in addressing the evolving landscape. Indeed, the field of DG warrants further research and development [11]. To address the challenges posed by decentralization and distribution, there is a need for a flexible and scalable DGF.

This research proposes a novel framework that guides organizations in designing, developing, and implementing DG systems aligned with their vision, business objectives, and legal constraints [12], while at the same time increasing awareness among individuals and teams involved in the value chain. It introduces the concept of data governance as a service (DGaaS), which provides structured and comprehensive support within the organizational context and business environment [13]. DGaaS is a robust framework that goes beyond traditional DG approaches. It provides a range of services and support to address the complexities of DG in distributed environments. By adopting DGaaS, organizations can leverage a structured and scalable approach to ensure effective DG aligned with their specific needs. The DGaaS framework serves as a valuable resource for organizations, facilitating the management, control, and utilization of data assets while promoting compliance, collaboration, and innovation in an increasingly interconnected landscape.

1.1 Background and motivation

The increasing popularity of DG is closely tied with the growing recognition of value inherited in data as observed by Zygmuntowski et al. [14]. Before the emergence of AI and Big Data, DG primarily focused on the data control and management and it constituted a task that was performed mostly by private organizations. Today, there is a growing concern among policymakers to facilitate data flows across companies and public administrations, aiming to enhance data re-use, sharing, and publication to maximize its value. Yet, data often involves various, sometimes conflicting interests, such as privacy or intellectual property. Therefore, a DGF should enable data flows to be 'as open as possible and as closed as necessary', ensuring that the interests of different stakeholders are respected when re-using data from other entities.

Scholars have recently emphasized that DG involves exercising authority and control over the entire lifecycle with the objective of increasing the data value while minimizing associated costs and risks [15]. Others advocate for holistic approaches to DG, differentiating between various layers: the regulatory layer, the organizational layer, and the technological layer [16]. Another perspective, put forth by Micheli et al., adopts a social-science lens to understand DG. They explore power relations among entities and actors affected by or influencing DP and how the generated data value is distributed among them [17].

These diverse perspectives highlight the multifaceted nature of DG, encompassing both technical and social dimensions. By incorporating these perspectives, organizations can develop comprehensive DG strategies that align with their specific goals, uphold stakeholder interests, and optimize the value derived from data. Overall, this research can contribute to advancing knowledge, providing practical insights, and addressing critical issues related to DG, and organizational effectiveness in the data-driven era. In the next section, we outline the main challenges of current DG systems.

1.2 Challenges posed by current systems

The establishment of an optimal DG system lacks consensus, both within a country and across borders [18] (*Challenge 1*), mainly due to the absence of harmonization among diverse national laws and regulations. Recognizing the need for additional measures, beyond the General Data Protection Regulation (GDPR) [19], the European Commission has emphasized the importance of fostering private-public data sharing frameworks and requiring data sharing among companies. This raises questions about data control, storage, and sharing conditions (*Challenge 2*), and the operation of effective DG systems (*Challenge 3*). In response to these challenges, the EU Strategy for Data [20] has been introduced, highlighting the development of common European Data Spaces (DS) as a crucial enabler for the Digital Economy. However, the current state of data usage is fragmented, with data being stored and processed in isolated silos, following either legacy or proprietary formats [21], leading to the need for extensive re-processing when shared with other stakeholders (*Challenge 4*). This involves addressing semantic, syntactic, and technical disparities, resulting in redundant resource consumption. Moreover, the creation of cross-sectional DS presents significant difficulties, necessitating the establishment of sector-specific DS based on domain characteristics and standards (*Challenge 5*), such as the European Health Data Space (EHDS) [22].

The challenges posed by AI and Big Data, including the substantially larger volumes and complex DP, call for a comprehensive DGF. The European Union is currently considering additional legislation, such as the Data Act (DA) [23] to enhance data access and utilization, the Data Governance Act (DGA) [24] to establish data exchange platforms, the Digital Markets Act (DMA) [25] to ensure that data guardians do not impede innovation, and the AI Act (AIA) [26], the regulatory and legal framework for AI. These initiatives aim to unlock data currently confined within silos and effectively enforce the broad data protection framework outlined in the GDPR.

Within this context, this paper presents the conceptual architectural of an integrated environment designed to provide trustworthy mechanisms for DP using AI and Big Data analytics. The proposed framework, known as DGaaS, aims to contribute to the research, technological, and societal landscapes by addressing ethical, legal and regulatory concerns and trust issues. The remainder of the paper is organized as follows. Section 2 provides an overview of the current state-of-the-art and emerging technologies on data processing and governance, with an emphasis on reliability and trust. Section 3 introduces the methodology used to develop the proposed DGaaS solution. Section 4 describes the DGaaS framework, outlining the high-level architecture and the various integrated components to tackle the challenges associated with the use of AI and Big Data. Finally, Section 5 concludes the paper, summarizing the key findings and highlighting the significance of the proposed framework.

2 STATE OF THE ART AND EMERGING SOLUTIONS

With the advent of cutting-edge technologies such as quantum computing, edge computing, and the Internet of Things (IoT), organizations today are confronted with the need to handle vast and diverse amount of data to gain valuable insights and knowledge. By engaging in DP, organizations can effectively comprehend customers' needs and accomplish their business goals. This capability enables them to maintain their competitiveness, swiftly adapt to evolving circumstances, and capitalize on emerging opportunities.

These technologies are reshaping DP in different ways, offering new possibilities for enhanced efficiency, scalability, and connectivity. Simultaneously, they provide viable solutions that can be implemented in decentralized and collaborative environments, ensuring greater reliability and trust.

Quantum computing leverages the principles of quantum mechanics to perform computations that surpass the capabilities of classical computers. It utilizes quantum bits (qubits) that can represent multiple states simultaneously, enabling parallel processing and solving complex problems at an unprecedented speed [27]. Quantum computing has the potential to revolutionize DP by accelerating tasks such as optimization, security, and simulation [28]. It holds promise for solving computationally intensive problems that are currently impractical or infeasible with standard computing, while at the same time, provides stronger data security measures, protecting sensitive information from potential cyber threats, and optimized resource allocation, evading to more efficient utilization of resources, improved task scheduling, and better load balancing.

Edge computing refers to DP at or near the node of the network, closer to the data source or end-user devices. By decentralizing computational power, edge computing reduces latency, enhances real-time decision-making, and alleviates network congestion [29]. It enables DP to occur closer to where the data is generated, resulting in faster response times and improved efficiency [30]. This enhances data privacy and security, reducing the risk of data breaches or unauthorized access during data transmission. Edge computing is particularly valuable for applications and industries in decentralized settings such as autonomous vehicles, industrial IoT, and smart cities that require low latency, high bandwidth, and real-time processing.

IoT refers to the process of transforming vast amount of data from heterogeneous internet-connected devices (e.g., sensors, appliances, cyber-physical systems) to business intelligence facilitating automation, monitoring, and control of various processes [31] [32]. It has transformed industries such as healthcare, manufacturing, agriculture, and transportation by enabling efficient resource management, predictive maintenance, and enhanced decision-making based on real-time data [33].

As these technologies continue to advance, their integration and synergy with traditional and decentralized DP systems will further transform the way data is processed, enabling new insights, applications, and possibilities across various industries and domains.

2.1 Data processing

Data processing (DP) is performed in a sequential series of steps aimed at extracting valuable information and insights to enhance organizational governance. As emphasizes by Shah et al., the main conceptual components of DP are *data management* and *data governance* [34].

2.1.1 Data Management layer. The data management (DM) encompasses a range of processes:

- **Planning**: determines the activities to be performed during the entire data lifecycle. The outcome is a comprehensive data management plan that outlines policies, processes and procedures (P3) governing data handling. It defines roles, responsibilities, and access rights, serving as the foundation for data flows in projects involving multiple entities [35].
- **Collection**: involves gathering relevant data from various sources [36], such as social media, IoT devices, surveys, medical records, satellite observations, and statistics. Data may exist in structured, unstructured and semi-structured formats.
- **Pre-processing**: encompasses the integration, filtering and enrichment of data to obtain datasets that reduce complexity in DP. Integration consolidates all data into a unified structure, ensuring coherence and homogeneity [37]. Filtering identifies incomplete, incorrect, inaccurate, or irrelevant data, which are then replaced, modified, or removed [38]. Enrichment involves enhancing collected raw data with relevant contextual information obtained from additional sources.
- **Analysis**: encompasses inspecting, transforming, and modelling data to extract knowledge and valuable insights from raw or pre-processed data. Many data analysis techniques are employed to understand what is happening (descriptive analytics), why something is happening (diagnostic reasoning), and what is most likely to happen (predictive analytics). These techniques include data mining [39], clustering [40] [41], classification [38] [42], and regression [43].
- **Visualization:** focuses on representing analysis results through charts, infographics, and animations. Visual displays effectively communicate complex data relationships and explain the meaning of the retrieved information in a user-friendly manner. There are three main categories of data visualizations: exploratory, explicatory, and explanatory visualization [44], catering to different purposes and data volumes. Examples include dashboards, reports, and oral presentations.
- **Storage:** aims to securely save the data and ensure its availability throughout the entire data lifecycle. Data availability emphasizes that data can be accessed or downloaded by data consumers at any time [45]. Data Lakes, utilizing tools like NoSQL, NewSQL, Big Data Query platforms, Hadoop Distributed File System (HDFS), and cloud storage technologies, serve as repository for ingesting and querying raw data from different sources and formats [46] [47].
- **Access:** provides a communication system and implements access control policies between data providers and data consumers. It regulates, monitors, and documents the actors involved, mechanisms used, and duration of access [48].
- **Share/publish:** involves distributing and transferring data through information systems, data catalogues, websites, social networks, and other platforms. Sharing and publication of data can be beneficial for researchers, companies, governments, and citizens.
- **Re-use:** encompasses two activities: data and information re-use by data consumers, and collecting feedback from data consumers to improve published data.
- **Archive:** involves cataloguing, indexing, and related actions based on specific parameters or business rules. It maintains records of obsolete data that are eventually destroyed, removed from the system, or no longer required.

2.1.2 Data Governance layer. The data governance (DG) includes nine tasks:

- **Data architecture:** captures the technological capabilities around DP, including design documents at different levels of abstraction, encompassing principles governing data collection, storage, analysis, and storage within the organization's systems [49]. The outcome is the enterprise data model, which includes data assets, metadata definitions, entities and their relations, and business rules.
- **Data modelling:** includes identifying, analyzing, and scoping data requirements, and then representing and communicating them in a standardized form known as data model [50]. Each model contains of elements such as entities, relationships, facts, keys, and attributes. Different schemes including relational, dimensional, object-oriented, fact-based, and NoSQL, are employed.
- **Data operations:** encompasses the design, implementation, and support of stored data to maximize its value throughout the entire lifecycle. It includes operational support, focusing on monitoring and fine-tuning data storage activities from initial implementation to data acquisition, backup, and retrieval. Technical support focuses on specifying technical requirements for defining the technical architecture, deploying and administering relevant technology, and resolving any potential issues.
- **Data security:** involves defining, developing, and implementing security and data protection policies and procedures to ensure proper authentication, authorization, access control, privacy, and confidentiality. Data confidentiality safeguards against unauthorized access, while privacy ensures that personal data, such as identity, location, and other sensitive information remain confidential and protected from misuse and malicious purposes [51].
- **Data integration and interoperability:** deals with the movement and consolidation of data within and between data stores, applications, and organizations. Integration ensures data is consolidated into consistent forms, while interoperability enables multiple systems to communicate with each other.
- **Document and content management:** involves controlling and monitoring the acquisition, storage, access, and use of data and documents in any form or medium, particularly unstructured and semi-unstructured information.
- **Data warehousing and business intelligence:** encompasses planning, implementation, and control activities to provide evidence-based decision support and a knowledge base for the reporting, querying, and analysis by relevant stakeholders.
- **Metadata:** aids organizations in better understand their data, systems, and workflows. It contributes to data processing, maintenance, integration, security, auditing, and governance. Metadata includes information about technical and business processes, rules, constraints, and data assets (e.g., databases, data elements, data models), concepts (e.g., application systems, software code, technology infrastructure), and connections (relationships) between the data and concepts.
- **Data quality:** involves managing data through its lifecycle by defining standards and measuring data against those standards to ensure its fitness for use by data consumers. Data quality assessment includes accuracy, accessibility, completeness for the intended purpose, and interpretability [52] [53].

2.2 Collaborative vs decentralized

When multiple infrastructures and services are interconnected to perform various stages of DM, it forms a distributed data management (DDM) ecosystem [54]. Similarly, when multiple entities, situated in different teams or organizations, participate in the decision-making and DG, it is referred to as decentralized data governance (DDG) [55]. The combination of DDG with DDM leads to collaborative data processing (CoDP). CoDP represents a community-based approach to managing, leveraging, and sharing data, in contrast to centralized data processing (CDP), where a single entity governs the decision-making throughout the data lifecycle. The adoption of CoDP and CDP is determined by the specific DP context, enabling organizations to establish a flexible DGF [56].

A centralized approach to DG offers benefits in terms of streamlined coordination between entities, efficient communication, and the ability to monitor information integrity and project progress over time [57]. However, a centralized architecture may face limitations in terms of scalability, such as constrained storage, bandwidth, computing power, energy consumption, and security capabilities, making it inadequate for handling large volumes of data [58]. Moreover, a centralized architecture is susceptible to reliability issues since it represents a single point of failure. In the event of failure, the entire infrastructure becomes compromised [59]. Consequently, organizations are increasingly considering CoDP as it enables unification, cooperation, and healthy competition among entities, addressing the shortcoming of the centralized approach.

In response to the limitations of CDP, CoDP has emerged as a viable solution [60]. It involves the active participation and interaction of both humans and computer systems [9]. This approach leverages the combination and filtering of data from various systems to enrich knowledge [61], reduce storage costs, and enhance security [62]. By utilizing multiple systems working together, CoDP addresses the capacity, reliability, and trust issues associated with a single system [63]. Achieving this requires a combination of collaborative tools and distributed systems, enabling a decentralized approach to DP where a diverse range of individuals or groups within and beyond an organization can participate. Two key concepts arise from CoDP: decentralization and distribution. Distribution pertains to the physical infrastructure of a system, consisting of autonomous processing entities interconnected through a computer network, collaborating to achieve shared objectives [62]. It also involves the distribution of datasets across multiple sites in a partial or complete way [63][54]. In the context of DP, distribution focuses on DM, while decentralization pertains to DG and decision making, involving multiple entities overseeing the overall ecosystem.

An example of CoDP can be found in the work of Bin Liu et al., who proposed distributed DP systems for e-business [64]. Their solution utilizes a distributed DGF and multiple systems for data collection and mining. By employing a multi-agent model in diverse data environments, the model facilitates distributed DP and consolidates the outputs. This approach reduces network traffic, enhance security, and addresses data privacy concerns [9]. Additionally, Abreha et al. introduced the concept of federated learning (FL) in an edge computing environment. FL encourages privacy-preservation collaboration among data producers, eliminating the need to combine data that may contain personal or confidential information [65]. While distributed systems share some similarities with federated systems, the key distinction lies in their objectives. Federated systems aim to provide a unified, integrated view of distributed data (a virtual database), whereas distributed systems focus on facilitating common access to distributed data (a distributed file system with built-in DM capabilities) [66]. FL mitigates data communication costs and resolves issues related to bandwidth limitations, privacy, and legal considerations. However, employing a DDG approach raises concerns regarding trust, as the self-interest of an entity processing the data may conflict with the overall benefit of other entities [67]. Furthermore, the reliability of the assets used for DP and the quality of the results must be continuously monitored.

2.3 Reliability and trust

Within the field of DP, two critical challenges stand out: reliability and trust. Although these concepts are closely interconnected, they possess distinct characteristics. Reliability is an objective measure that pertain the physical infrastructure of a system, encompassing hardware, firmware, and software. Trust, on the other hand, is subjective and resolves around resource sharing, decision-making, and a specific purpose [68]. Moreover, reliability is a context-based property of a system that correlates its operation and goals or expected behaviors. Generally, it is assessed through a set of criteria and dimensions, such as probability, intended functions, time dependence, specific conditions, performance scalability, fault prevention, fault [69]. Trust, in contrast, is a psychological condition characterized by the willingness to embrace vulnerability based upon positive anticipations regarding the intentions or actions of another entity [70]. Furthermore, in the realm of DP, Anhalt-Depies et al assert that trust plays a pivotal role in cultivating the social cohesion necessary for collaboration as well as for conflict resolution and negotiation [71].

The emergence of CoDP has brought about a shift in how these concepts are approached. Previously, centralized third parties were relied upon for any form of exchange between entities. Now, there have been a paradigm shift towards distributed systems with DDG, leading to discussions on distributed reliability and decentralized trust. Distributed reliability is defined by Akimova et al as a property enabling the preservation over time of the defined parameters required to fulfil the main function under the influence of malfunctions (failures and breakdowns) of hardware, software and data errors, human mistakes, and working environment when maintenance and servicing parameters are specified [72]. In the context of CoDP, distributed reliability concerns every stage of the DM, the transitions between stages, and the ultimate objective of the DP. It ensures that each phase of the DM process is carried out by multiple systems with fault tolerance, resulting in high-quality outputs. Decentralized trust represents a challenge as data typically accept vulnerability arising from human behaviors (i.e., the trustee), rather than machines [73]. However, establishing the overall reliability of distributed systems during DP is crucial to instill trust among data consumers in the results obtained through DP. Hence, reliability is regarded as an indicator of trust.

2.4 The role of blockchain

Blockchain has recently emerged as a disruptive innovation with a wide range of applications, able to redesign interactions in business activities and society at large [74]. It consists of a permanent, distributed, digital ledger, recording transactions across a peer-to-peer network [75]. Unlike other peer-to-peer networks, a blockchain does not duplicate the value being transferred. Instead, it functions as a register, noting that a value has been successfully transferred from one entity to another. It enables an open network where participants can interact without prior knowledge or trust in each other. Transactions are automatically verified and recorded by network nodes using cryptographic algorithms, eliminating the need for human intervention, central authority, or third-party intermediaries such as financial institutions or other organizations. Even in the presence of unreliable, dishonest, or malicious nodes, the network can accurately verify transactions and safeguard the integrity of the ledger. This is achieved through a mathematical mechanism known as proof-of-work, rendering human intervention or centralized control unnecessary. The underlying principle behind this protocol is *decentralized trust* or trust-by-computation.

When it comes to governance, the most significant innovation of the blockchain is that it provides the ability to experiment with new organizational structures and distributed governance models, which are more transparent and less hierarchical than traditional ones. Indeed, blockchain allows the creation of so-called decentralized organizations, enabling entities and actors to coordinate themselves in a peer-to-peer manner, according to a set of protocols and rules incorporated into self-executing smart contract code [76]. Nevertheless, a comprehensive assessment considering multiple factors is currently lacking to determine the viability of a blockchain-based governance model in meeting the necessary legal, ethical and technical requirements [77].

The above-described aspects serve as fundamental pillars for establishing an effective and scalable DG system, and thus shall be considered when designing and implementing DG programs within an organization.

3 METHODOLOGY

This study adopted an **evolutionary approach** (common in product development activities) by iterating a series of activities (as illustrated in Figure 1), aiming to successfully identifying the appropriate features that DGF needs to possess in order to achieve high technology acceptance and wider impact. The methodology involved a comprehensive analysis of relevant scholars in data processing and governance, evaluating any remaining challenges and risks, and creating a standardized schema for enhanced understanding and automation.

The starting point was the Data Lifecycle Framework (DaLiF), as proposed by Sha et la [34], which was then enhanced and further developed using an **iterative co-creation approach** to ensures adherence to the principles outlined in the DMBoK [3]. The result of the first prototype is depicted in Figure 2.

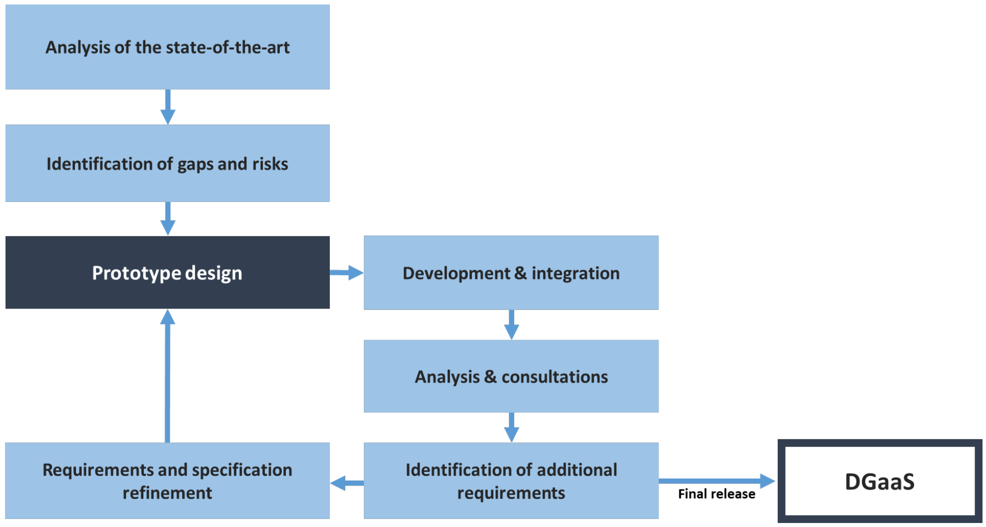

Figure 1: DGaaS Evolutionary Process

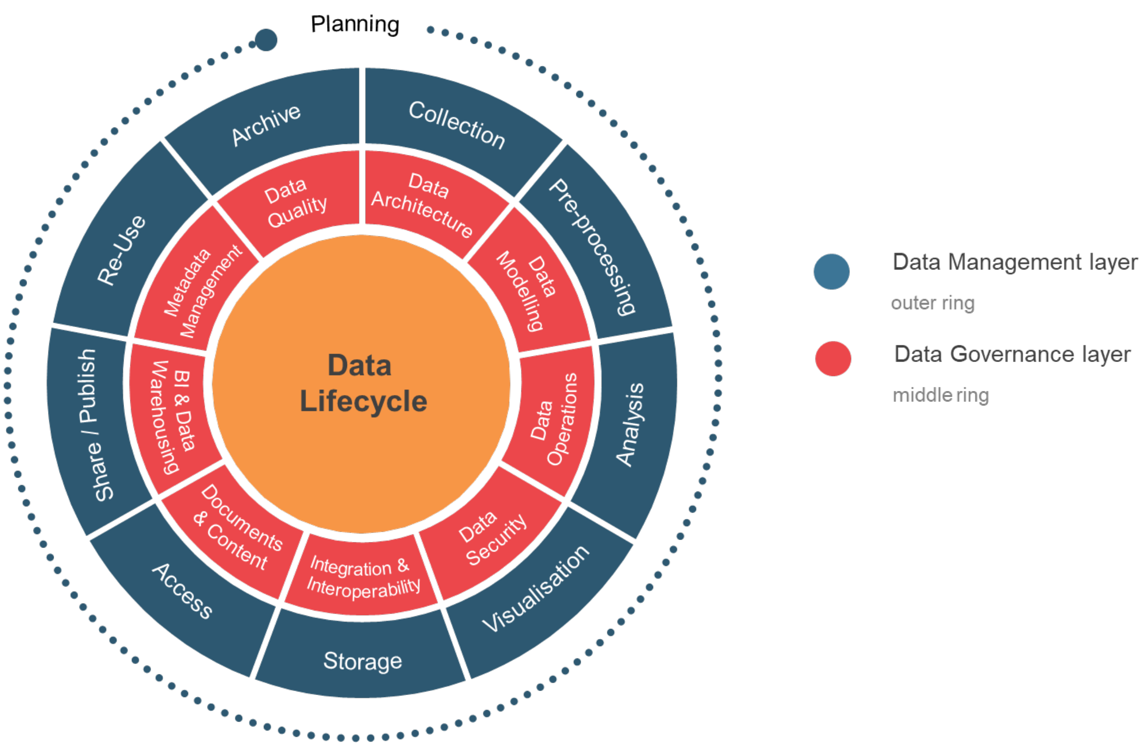

Figure 2: The Data Lifecycle (re-worked from Shah et al [34] and DMBoK [3])

4 THE DATA GOVERNANCE AS A SERVICE

We present a DGaaS framework, depicted in Figure 3, which adopts a concentric circles structure, with people, process, and technology at its core. Moving outward, we emphasize the importance of security and privacy on one side, and ethical, legal and regulatory compliance on the other. These are combined with a blockchain-enabled transactions tracking mechanism, ensuring transparency throughout the system. The subsequent circle encompasses ***collaborative*** actions, such as data architecture and data modelling. The outermost ring contains areas where DG plays a critical role, such as metadata management and business intelligence (as detailed in sub-section 2.1.2).

The proposed DGaaS framework follows a bottom-up approach, empowering and providing support to data consumers (individuals

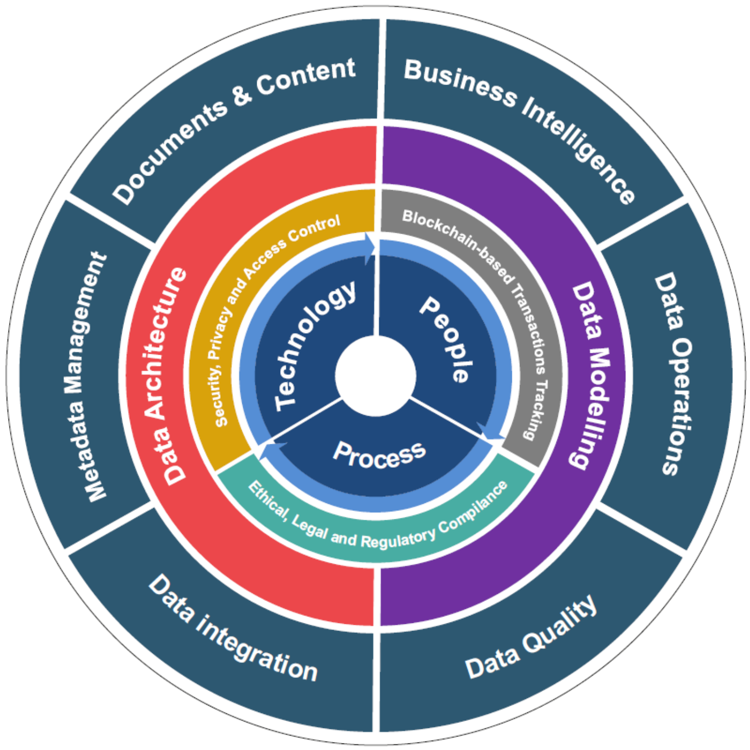

Figure 3: Data Governance as a Service (DGaaS) Framework

or teams). It is flexible and adaptable to the needs of the different sectors and stakeholders, and aims at increasing the value of data and minimize DP-related costs and risks. While security and privacy remain paramount as in a centralized approach, top management remains responsible for resources allocation and communicating the overall strategy. Management bodies still maintain their relevance, but P3 are implemented as close to the point of the data usage as possible, ensuring maximum value extraction.

Overall, the framework introduces three main novel components (as illustrated in the second circle) in order to address the organization's technical, business and operational requirements:

Security, Privacy, and Access Control module

We propose the adoption of a privacy-preserving ***human-centric*** federated identity management approach based on a trusted execution environment (TEE). The TEE is located between the identity providers and service providers. An identity provider can store some of the user's attributes in the TEE. Information that is endorsed by identity provider can then be accessed by service providers. The latter use the information to provide fine-grained access control and offer personalized services. Prior to user authentication, the service provider has to authenticate to the TEE and prove that it is authorized to access certain personal attributes. The TEE verifies the acceptability of the service provider's information request. This verification ensures that only information from identity providers is queried, for which the identity providers (or their representative) gave their consent. The authorization information is included in the certificate (or credential) of the service provider. Additionally, the user may further restrict access to personal information through a policy or an explicit consent. If the query is acceptable, the TEE forwards this request to the identity provider(s) that can provide the information.

Blockchain-based Transactions Tracking module

A novel blockchain-based information exchange system with built-in smart contracts for decentralized, scalable, and interoperable data processing and governance. This will ensure the tamper-proof, decentralized logging, auditing, and tracking of actions leading to traceability and non-repudiation. Moreover, the blockchain mechanisms will allow for logging, auditing and tracking of the transactional data activities (e.g., acquisition, sharing, access) between the involved entities and stakeholders in a holistic manner that preserves confidentiality and data security. The impact of such solution will be tangible via specific privacy metrics, including i) attack surface exposed to untrusted software, ii) amount of the trusted computing base used by security tools in terms of source lines of code (SLOC), and iii) number of access points to the TET.

Ethical, Legal and Regulatory Compliance module

The DGaaS is designed following the principles, protocols and recommendations outlined in the GDPR, the Data Act (DA), the Data Governance Act (DGA), the AI Act (AIA), but also ethics guidelines for trustworthy AI developed by the High-Level Expert Group. It seeks to analyses, predict, quantify, and monitor transparency, accountability and trustworthiness of DP activities using AI and Big Data, considering individual rights to ensure meaningful redress for people affected by these technologies.

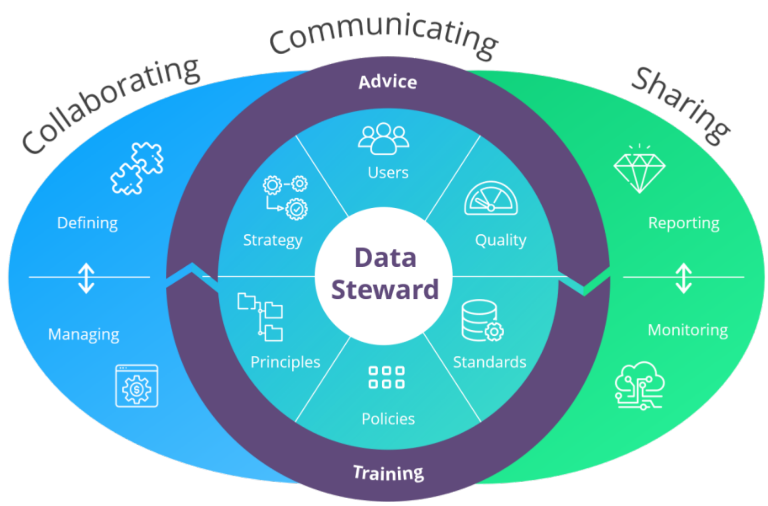

Figure 4: Data Stewardship

4.1 Data stewardship

To foster trust in the system and promote widespread adoption, it is crucial to implement specialized data stewardship (see Figure 4) practices. The data stewardship is defined as accountability and responsibility for data and processes that ensure effective control and use of data assets to help an organization get value from its data [3]. These practices ensure that data collection, usage, and sharing adhere to fundamental principles outlined in the GDPR such as notice and consent, purpose limitation, data minimization, retention restrictions and usage limitations. By mandating these measures, organizations can demonstrate their commitment to responsible data handling, which in turn enhances trust among stakeholders and encourages broader acceptance of the system.

DGaaS is a ***transformative*** process that relies heavily on the collaborative efforts from different stakeholders. To effectively implement it, organizations must establish an *agile governance model*, encompassing Strategy, Organization, Directives, Measurement, and Change Management [78]. Here is an enhanced breakdown of each component:

- **Strategy:** The Data Strategy establishes a clear alignment between the organization's goals, objectives, and desired values to be accomplished. It provides a roadmap for leveraging data effectively.
- **Organization:** entails defining the roles, responsibilities, and accountabilities necessary for executing the Data Strategy. It clarifies the involvement of different stakeholders, articulates their motivations, and sets clear expectations upon them.
- **Directives:** encompasses the P3 that support and align with the organization's Data Strategy. They ensure these P3 are effectively implemented and that the work conducted adds tangible value.
- **Measurement:** involves assessing the impact and value generated by the DG program. It incorporates both qualitative and quantitative metrics to gain a better understanding of the economic, social and environmental returns. These metrics allow organizations to gauge the effectiveness of their DG practices and make data-driven decisions for continuous improvement.
- **Change management:** focuses on the effective communication, comprehensive training, and seamless integration to embed DG practices within the organization's culture and workflows. It aims to foster awareness, understanding, and adoption of DG practices to support the organization in advancing its mission and driving meaningful change.

By embracing the DGaaS framework and effectively implementing these steps, organizations can navigate the journey of DG, fostering collaboration, maximizing value creation, and ultimately unlocking the full potential of their data assets.

4.2 Implementing DGaaS

When initiating a DG program, organizations face several challenges that need to be addressed for its successful implementation. Here are some key challenges associated with DGaaS:

- **Change management and adoption:** persuading business stakeholders to accept implementing a DG program can be a significant obstacle. It requires organizational change management efforts that usually involves training, education, and promoting a cultural shift towards DG practices. Stakeholder engagement and effective communication strategies are crucial for driving adoption and achieving the desired outcomes.
- **Financing:** securing necessary funds can be problematic as it may require determining funding levels for tools, addressing resource and compensation constraints, and understanding how to deliver tangible value. While traditional costs associated with data, such as storage expenses, are relatively quantifiable, assessing its value to the organization proves more complex. It is essential to recognize that data itself holds no intrinsic value, and individuals or organizations treating it as an abstract object may overlook its true

worth. The DG systems often come with substantial costs and present complex integration and operational activities [79]. Moreover, its financing could often be susceptible to annual variations, making it more challenging to adequately plan compared to a continuous funding stream.
- **Resource and time constraints:** Data governance and stewardship activities, such as data collection and definition of data assets, are resources and time-consuming, despite the ability of tools to learn data structure and format directly from database structures. Moreover, many organizations nowadays operate in hybrid environments, incorporating both on-premises and cloud-based infrastructures. DGaaS must offer solutions that effectively govern data across multiple platforms and infrastructures, providing consistent governance practices, data policies, and compliance measures.
- **Learning curve, skills and commitment:** Acquiring the necessary skills and the learning curve associated with DG tools pose time investments and may hinder business adoption, presenting serious commitment challenges.
- **Establishing business priorities:** Defining a DGF, setting priorities, determining ROI, establishing Key Performance Indicators (KPIs), and establishing metadata loading templates and processes can be daunting as a first-time exercise.

The concept of DGaaS offers hope for organizations lacing the momentum, size or buy-in required to initiate an organic DG organization. By overcoming these challenges, organizations can unlock the full potential of DGaaS and realize its benefits in terms of improved data management, enhanced decision-making, and compliance with regulatory requirements.

4.3 Expected impacts of DGaaS

The DGaaS can be offered to several stakeholders and more specifically to: i) *private organizations* for planning, designing and implementing data-driven solutions to align their vision and business objectives with their customers' needs; ii) *public administrations* for organizing, planning, and monitoring the timely provision of appropriate and enhanced policies leading to efficient decision-making; and (iii) *researchers* involved in DP activities, to facilitate data discovery, interoperability and use.

Moreover, DGaaS can boost the ability of organizations to exploit and monetize their data assets [80], while at the same time developing novel services or enhancing the existing ones. It will open opportunities stemming from access to and consumption of data in an open, decentralized, blockchain-based environment instead of the current centralized and siloed models and solutions.

Furthermore, by adopting DGaaS, organizations can distribute the implementation and DG program progress risk through the expertise provided across the whole ecosystem, ranging from strategy development to ongoing data stewardship. Unlike in-house staffing, DGaaS offers a flexible support structure that can be adopted to project's needs, timelines, and constraints. During the initial set-up and metadata capture phase, projects typically require higher data stewardship resources. However, the ongoing project demands decrease significantly. DGaaS facilitates staffing flexibility by combining onshore and offshore resources and services. This flexible staffing approach not only enhances cost-effectiveness, but also enables the provision of timely solutions by leveraging round-the-clock stewardship capabilities.

DGaaS delivers core stewardship services at an optimal cost, allowing organizations to allocate their internal staff to higher-value tasks. DGaaS can enhance the effectiveness of DG programs while only investing a fraction of a total program cost.

5 CONCLUSIONS

This paper investigates whether effective data processing can be achieved without formal data governance. First, it emphasized the challenges posed by current systems both from technological and legal perspective. Second, it provided an overview of the state-of-the-art solutions and emerging technologies aiming to evaluate existing approaches on data processing and governance and to investigate their limitations. We elaborated on these technologies with a special focus on decentralization, collaboration, reliability and trust. Next, we presented an innovative framework, called Data Governance as a Service (DGaaS), providing a clear and structured description of the defined elements, processed and relationships. This paper underscores the importance of DGaaS together with principles, processes and procedures (P3) for managing data effectively throughout the entire lifecycle. The DGaaS also enables collaboration from different levels of the organizations and it also provides the ability to align various data related programs with business objectives.

Given the ever-growing value of data in the modern digital age era, continuous improvement of the performance and efficiency of technologies and methodologies targeted at data management, analysis and treatment is required. Moreover, adequate training for human resources towards implementing an agile governance structure and collaborative environment must be considered. This is very important when adopting DGaaS framework in order to maintain the competitive edge in the global market. We plan to further investigate the applicability of DGaaS in specific sectors, such as smart city and eHealth, and better understand the potential weaknesses that may arise in order to design additional measures.

ACKNOWLEDGMENTS

The authors would like to thank Dr Arianna Rossi from University of Luxembourg and Dr Salvatore Rinzivillo from National Research Council of Italy for their valuable feedback. This research has received funding from the European Union's Horizon 2020 Innovative Training Networks program under Grant Agreement No. 956562.

REFERENCES

[1] R. Kitchin, "The Data Revolution: A Critical Analysis of Big Data, Open Data and Data Infrastructures", SAGE Publications, 2021.
[2] G. Bellinger, D. Castro, and A. Mills, "Data, Information, Knowledge, and Wisdom", 2011.
[3] D. Henderson, S. Earley, and L. Sebastian-Coleman, "The DAMA International's Guide to the Data Management Body of Knowledge (DMBoK2)", 2nd Edition, Technics Publications, Basking Ridge, New Jersey, USA, 2017.
[4] M. Janssen *et al.* "Data governance: Organizing data for trustworthy Artificial Intelligence", Government Information Quarterly, Vol. 37, No. 3, pp. 101493, July 2020. DOI: 10.1016/j.giq.2020.101493.
[5] O. B. Nielsen, "A Comprehensive Review of Data Governance Literature", IRIS Selected Papers, Vol. 3, No. 8, 2017. Available at: https://aisel.aisnet.org/iris2017/3.
[6] M. Mantymaki *et al.* "Defining organizational AI governance", AI and Ethics, 2022. DOI: 10.1007/s43681-022-00143-x.

[7] V. Khatri and C. V. Brown, “Designing Data Governance”, Communications ACM, Vol. 53, No. 1, pp. 148–152, January 2010. DOI: 10.1145/1629175.1629210

[8] J. Ladley, “Data Governance: How to Design, Deploy, and Sustain and effective Data Governance Program”, 2nd Edition, Academic Publisher, November 2019. DOI: 10.1016/C2017-0-03353-0.

[9] M. van Steen, G. Pierre, and S. Voulgaris, “Challenges in Very Large Distributed Systems”, Journal of Internet Services and Applications, Vol. 3, No. 1, pp. 59–66, May 2012.

[10] J. Qiu *et al.*, “A Survey of Machine Learning for Big Data Processing”, EURASIP Journal on Advances in Signal Processing, No. 1, pp. 67, May 2016.

[11] M. Al-Ruithe, E. Benkhelifa, and K. Hameed, “A Systematic Literature Review of Data Governance and Cloud Data Governance”, Personal and Ubiquitous Computing, Vol. 23, No. 5, pp. 839–859, November 2019.

[12] J. Yebenes and M. Zorrilla, “Towards a Data Governance Framework for Third Generation Platforms”, Computer Science 151, The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019), November 2022.

[13] M. Al-Ruithe, E. Benkhelifa, and K. Hameed, “A Conceptual Framework for Designing Data Governance for Cloud Computing,” Procedia Computer Science, Vol. 94, pp. 160-167, 2016. DOI: 10.1016/j.procs.2016.08.025.

[14] J. J. Zygmuntowski, L. Zoboli, and P. F. Nemitz, “Embedding European values in data governance: a case for public data commons”, Internet Policy Review, Vol. 10, No. 3, September 2021. DOI: 10.14763/2021.3.1572.

[15] R. Abraham, J. Schneider, and J. V. Brocke, “Data governance: A conceptual framework, structured review, and research agenda”, International Journal of Information Management, Vol. 49, pp. 424-438, January 2019.

[16] M. Grafenstein, “Reconciling Conflicting Interests in Data through Data Governance. An Analytical Framework (and a brief discussion of the Data Governance Act Draft, the Data Act Draft, the AI Regulation Draft, as well as the GDPR), HIIG Discussion Paper Series No. 2, May 2022.

[17] M. Micheli *et al.* “Emerging models of data governance in the age of datafication”, Big Data & Society, Vol. 7, No. 2, 2020.

[18] S. Tiwari, S. Sharma, S. Shetty, and F. Packer, “The design of a Data Governance system”, BIS Paper, No. 124, 2022.

[19] 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 (General Data Protection Regulation).

[20] Communication from the European Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions – A European strategy for data. COM/2022/66

[21] J. Patel. “Bridging Data Silos Using Big Data Integration”, International Journal of Database Management Systems, Vol. 11, No. 3, pp. 1-6, June 2019. DOI: 10.5121/ijdms.2019.11301

[22] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197

[23] Proposal for a Regulation of the European Parliament and of the Council on harmonized rules on fair access to and use of data (Data Act). COM/2022/68 final.

[24] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). COM/2020/767 final.

[25] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

[26] Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). COM/2021/206.

[27] R. Rietsche *et al.*, “Quantum computing”, Electronic Markets, Vol. 32, pp. 2525–2536 (2022). DOI: 10.1007/s12525-022-00570-y.

[28] A. Bayerstadler *et al.*, “Industry quantum computing applications”, EPJ Quantum Technology, Vol. 8, No. 25, 2021. DOI: 10.1140/epjqt/s40507-021-00114-x.

[29] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, "Edge Computing: Vision and Challenges", IEEE Internet of Things Journal, Vol. 3, No. 5, pp. 637-646, October 2016, DOI: 10.1109/JIOT.2016.2579198.

[30] H. Liu, S. Li, and W. Sun; “Resource Allocation for Edge Computing without Using Cloud Center in Smart Home Environment: A Pricing Approach”, Sensors, Vol. 20, No. 22, 2022. DOI: 10.3390/s20226545.

[31] P. Barnaghi, M. Bauer, *et al.*, “IoT Analytics: Collect, Process, Analyse, and Present Massive Amounts of Operational Data – Research and Innovation Challenges”, in O. Vermesan and P. Fries (eds) Building the Hyperconnected Society- Internet of Things Research and Innovation Value Chains, Ecosystems and Markets, River Publishers, New York, NY, 2022. DOI: 10.1201/9781003337454.

[32] S. Kumar, P. Tiwari, and M. Zymbler, “Internet of Things is a revolutionary approach for future technology enhancement: a review”, Journal of Big Data, Vol. 6, July 2019. DOI: 10.1186/s40537-019-0268-2.

[33] S. Nižetić, P. Šolić, D. González-de-Artaza, and L. Patrono, “Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future”, Journal of cleaner production, November 2020. DOI: 10.1016/j.jclepro.2020.122877.

[34] S. I. H. Shah, V. Peristeras, and I. Magnisalis, “DaLiF: A Data Lifecycle Framework for Data-Driven Governments”, Journal of Big Data, Vol. 8, No. 1, June 2021. DOI: 10.1186/s40537-021-00481-3.

[35] B. Denkena, L. E. Lorenzen, and J. Schmidt, “Adaptive Process Planning”, Production Engineering, Vol. 6, No. 1, pp. 55-67, February 2012. DOI: 10.1007/s11740-011-0353-7.

[36] Y. Rui, G. Heo, and S. E. Whang, “A Survey on Data Collection for Machine Learning: A Big Data – AI Integration perspective”, 2019. Available online at: https://arxiv.org/pdf/1811.03402.pdf.

[37] M. Lenzerini, “Data integration: a theoretical perspective”, Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS ’02). ACM, New York, NY, USA, 2002. DOI: 10.1145/543613.543644.

[38] R. Nisbet, G. Miner, and K. Yale, “Handbook of Statistical Analysis and Data Mining Applications”, 2nd Edition, Academic Press, 2018. DOI: 10.1016/C2012-0-06451-4.

[39] T. Calders and B. Custers, “What Is Data Mining and How Does It Work?”, In: B. Custers, T. Calders, B. Schermer, T. Zarsky (eds) “Discrimination and Privacy in the Information Society”, Studies in Applied Philosophy, Epistemology and Rational Ethics, Vol 3. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-30487-3_2

[40] P. Berkhin, “A Survey of Clustering Data Mining Techniques”, in J. Kogan, C. Nicholas, and M. Teboull (eds) Grouping Multidimensional Data: Recent Advances in Clustering, Berlin, Heidelberg: Springer, 2006. DOI: 10.1007/3-540-28349-8_2.

[41] G. J. Oyewole, and G. A. Thopil, “Data clustering: application and trends”, Artificial Intelligence Review, 2022. DOI: 10.1007/s10462-022-10325-y

[42] I. G. Maglogiannis, “Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth”, Information Retrieval and Pervasive Technologies, IOS Press, 2007.

[43] P. B. Palmer, and D. G. O’Connell, “Regression analysis for prediction: understanding the process”, Cardiopulmonary Physical Therapy Journal, Vol. 20, No. 3, pp. 23-26, September 2009.

[44] H. Hinterberger, “Data Visualization”, In: L. Liu, and M. T. Ozsu (eds) Encyclopedia of Database Systems. Springer, Boston, MA. 2009. DOI: 10.1007/978-0-387-39940-9_1370.

[45] L. Tedersoo, R. Küngas, E. Oras, E. *et al.*, “Data sharing practices and data availability upon request differ across scientific disciplines”, Scientific Data, Vol. 8, No. 192, 2021. DOI: 10.1038/s41597-021-00981-0.

[46] S. Mazumdar, D. Seybold, K. Kritikos, and Y. Verginadis, “A survey on data storage and placement methodologies for Cloud-Big Data ecosystem, Journal of Big Data, Vol. 6, No. 15, 2019. DOI: 10.1186/s40537-019-0178-3.

[47] A. Nambiar, and M. Divyansh, "An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management", Big Data and Cognitive Computing, Vol. 6, No. 4, 2022.DOI: 10.3390/bdcc6040132.

[48] W. Lee, S. J. Stolfo, and K. W. Mok, “Adaptive Intrusion Detection: A Data Mining Approach”, Artificial Intelligence Review, Vol. 14, No. 6, pp. 533–567, December 2000. DOI: 10.1023/A:1006624031083.

[49] R. Cao and M. Iansiti, “Digital transformation, data architecture, and legacy systems”, Journal of Digital economy, Vol. 1, No. 1, pp. 1-19, 2022. DOI: 10.1016/j.dec.2022.07.001.

[50] J. Parsons, “Data Modeling”, In: Błażewicz, J., Kubiak, W., Morzy, T., Rusinkiewicz, M. (eds) Handbook on Data Management in Information Systems. International Handbooks on Information Systems. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-540-24742-5_3.

[51] P. Yang, N. Xiong, and J. Ren, “Data Security and Privacy Protection for Cloud Storage: A Survey”, IEEE Access, Vol. 8, pp. 131723–131740, 2020.

[52] N. Côrte-Real, P. Ruivo, and T. Oliveira, “Leveraging Internet of Things and Big Data Analytics Initiatives in European and American Firms: Is Data Quality a Way to Extract Business Value?”, Information & Management, Vol. 57, No. 1.

[53] S. Choenni, M. Bargh, T. Busker and N. Netten, “Data governance in smart cities: Challenges and solution directions”, Journal of Smart Cities and Society, Vol. 1, No.1, pp. 35-51, 2022. DOI: 10.3233/SCS-210119.

[54] M. T. Ozsu and P. Valduriez, “Data Replication”, In Principles of Distributed Database Systems, 3rd Edition, M. T. Ozsu and P. Valduriez (eds), New York, NY: Springer, 2011, DOI: 10.1007/978-1-4419-8834-8_13.

[55] J. Domingue, A. Third, and M. Ramachandran, “The FARI-TRADE Framework for Assessing Decentralized Data Solutions,” in Companion Proceedings of the 2019 World Wide Web Conference (WWW ’19), ACM, New York, NY, USA 2019, DOI: 10.1145/3308560.3317076.

[56] S. L. Greer *et al.*, “Centralizing and Decentralizing Governance in the COVID-19 Pandemic: The Politics of Credit and Blame”, Health Policy, Vol. 126, No. 5, May 2022.

[57] S. E. Schapke, J. Beetz, M. Konig, and A. Borrmann, “Collaborative Data Management”, In: Borrmann, A., König, M., Koch, C., Beetz, J. (eds) Building Information Modelling, Springer, Cham. 2018. DOI: 10.1007/978-3-319-92862-3_14

[58] L. Geng and H. J. Hamilton, “Interestingness Measures for Data Mining: A Survey”, ACM Computing Surveys, Vol. 38, No. 3, September 2006. DOI:

10.1145/1132960.1132963.

[59] M. Cummings, "Operator Interaction with Centralized Versus Decentralized UAV Architectures", Handbook of Unmanned Aerial Vehicles, pp. 977–992, January 2015. DOI: 10.1007/978-90-481-9707-1_117

[60] D.A. Cieslak, D. Thain, and N.V. Chawla, "Short Paper: Troubleshooting Distributed Systems via Data Mining," in 15th IEEE International Conference on High Performance Distributed Computing (Presented at the 2006 15th IEEE International Conference on High Performance Distributed Computing, Paris, France: IEEE, 2006).

[61] A. L. Prodromidis, P. K. Chan, and S. J. Stolfo, "Meta-Learning in Distributed Data Mining Systems: Issues and Approaches," Advances of Distributed Data Mining, AAAI Press, 2000.

[62] H. A. Abbass and C. Lokan, "DXCS: An XCS System for Distributed Data Mining", Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2005), Washington DC, USA, June 2005.

[63] S. Noel and D. Lemire, "On the Challenges of Collaborative Data Processing", in Collaborative Information Behavior: User Engagement and Communication Sharing, edited by Jonathan Foster, IGI Global, 2010, pp. 55-71. DOI: 10.4018/978-1-61520-797-8.ch004.

[64] B. Liu, S. G. Cao, and W. He, "Distributed Data Mining for E-Business", Information Technology and Management, Vol. 12, No. 2, pp. 67–79, June 2011. DOI: 10.1007/s10799-011-0091-8.

[65] H. G. Abreha, M. Hayajneh, and M. A. Serhani, "Federated Learning in Edge Computing: A Systematic Survey", Sensors, Vol. 22, No. 2, January 2022.

[66] C. Jean-Quartier, M. R. Mazón, M. Lovrić, and S. Stryeck, "Collaborative Data Use between Private and Public Stakeholders: A Regional Case Study", Data, Vol. 7, No. 2. DOI: 10.3390/data7020020

[67] M. Foy, D. Martyn, D. Daly, A. Byrne, C. Aguneche, and R. Brennan, "Blockchain-based governance models for COVID-19 digital health certificates: A legal, technical, ethical and security requirements analysis", Procedia Computer Science, pp. 662–669, 2022. (12th International Conference on Emerging Ubiquitous Systems and Pervasive Networks/11th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare).

[68] C. W. Hart and M. D. Johnson, "Growing the Trust relationship", Cornell University Library, April 1999. Available online at: https://ecommons.cornell.edu/handle/1813/71431.

[69] M. T. Ozsu and P. Valduriez, "Distributed DBMS Reliability," in Principles of Distributed Database Systems, Third Edition, ed. M. Tamer Özsu and Patrick Valduriez (New York, NY: Springer, 2011), 405–457.

[70] S. Pearson, "Privacy, Security and Trust in Cloud Computing", In Privacy and Security for Cloud Computing (eds) by S. Pearson and G. Yee. Computer Communications and Networks. London: Springer, 2013, pp. 3–42.

[71] C. Anhalt-Depies, J. L. Stenglein, B. Zuckerberg, P. A. Townsend, and A. R. Rissman, "Trade-offs and tools for data quality, privacy, transparency, and trust in citizen science", Biological Conservation, October 2019.

[72] G. P. Akimova, A. V. Solovyev, and I. A. Tarkhanov, "Reliability Assessment Method for Geographically Distributed Information Systems", 12th International Conference on Application of Information and Communication Technologies (AICT), 2018.

[73] Y. P. Huang, K. Fan, H. Yang, K. Zhang, and H. Li, "A Collaborative Auditing Blockchain for Trustworthy Data Integrity in Cloud Storage System", IEEE Access, May 2022.

[74] M. Atzori, "Blockchain technology and decentralized governance: Is the state still necessary?", December 2015. Available at SSRN 2709713.

[75] A. Zwitter and J. Hazenberg, "Decentralized Network Governance: Blockchain Technology and the Future of Regulation", Frontiers in Blockchain, Vol. 3, 2020. DOI: 10.3389/fbloc.2020.00012.

[76] A. Wright, and P. De Filippi, "Decentralized Blockchain Technology and the Rise of Lex Cryptographia", March 2015. DOI: 10.2139/ssrn.2580664.

[77] O. Ozer and Y. Zheng, "Trust and Trustworthiness", The Handbook of Behavioral Operations. Ed. by Karen Donohue *et al.* Hoboken, NJ, USA: John Wiley and Sons, Inc., October 2018. DOI: 9781119138341.ch14.

[78] A. L. Dennis, "Establishing Data Governance as a Service", Dataversity, 2022. Available at: https://www.dataversity.net/establishing-data-governance-as-a-service/.

[79] B. Petzold, M. Roggendorf, K. Rowshankish, and C. Sporleder, "Designing data governance that delivers value", McKinsey Digital, June 2020. Available at: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/designing-data-governance-that-delivers-value#/.

[80] J. Ofulue and M. Benyoucef, "Data monetization: insights from a technology-enabled literature review and research agenda, Management Review Quarterly, 2022. DOI: 10.1007/s11301-022-00309-1.