Improving Data Value and its Influence on Decision Making through Better Data Frameworks and Management*

Mansoor Ahmed[1,3,*,†], Claudia Roessing[1], Priyanka Singh[1], Gabriel Hogan[2,3,†] and Markus Helfert[1,3,†]

[1]Innovation Value Institute, Maynooth University, Maynooth, Ireland
[3]Adapt Centre for AI-driven Digital Content Technology, Ireland
[2]Dublin City University, Glasnevin, Dublin D9, Ireland

Abstract

Data asset exploitation and management present challenges for organizations in the era of big data. Information Systems have adapted and evolved to meet many of these challenges. However, holistic governance and management of data assets for exploitation is lacking. In this paper, we show the dependencies between data lifecycle, data provenance, consent management, and data value creation, which allow identifying and applying data governance structures appropriate to the needs of individual organizations and their business contexts. We propose a framework using these four perspectives as the main building blocks for implementing holistic data governance. This framework facilitates the development of unique data workflows to meet data governance requirements for differing organizations and underpins the creation of value for organizations to exploit their digital assets.

Keywords

Data Governance; Data Lifecycle, Data Provenance, Data Privacy, Data Value, Consent management,

1. Introduction

In an era of data proliferation, organizations, entities, and even IoT sensors generate vast amounts of data. However, transforming data into a valuable asset that impacts decision-making, innovation, and operational advantage is challenging. This immense data influx necessitates stringent measures to ensure its quality. However, data producers often lack control over the data once it is shared or distributed [63]. This highlights a critical gap in data oversight and management. Data governance emerges here as a fundamental solution, defined broadly as a framework of policies, structures, and processes to manage data assets within organizations.

1.1. Data Governance

Enterprises, entities, and sensors produce huge amounts of data, which is a useful asset. This creates a need to check, ensure, and maintain data quality, including readiness, authenticity, safety, and integrity. Commonly used control mechanisms include policies, structures, and processes called data governance [66]. In this regard, data governance is increasingly becoming an evolving topic in modern information systems literature as organizations necessitate an effective approach to the behavior of their data assets [53]. Data governance is defined by the Data Management Association (DAMA) as: "the exercise of authority, control, and shared decision-making (planning, monitoring, and enforcement) over the

Companion Proceedings of the 17th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling Forum, M4S, FACETE, AEM, Tools and Demos co-located with PoEM 2024, Stockholm, Sweden, December 3-5, 2024

*You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

*Corresponding author.
†These authors contributed equally.
mansoor.ahmed@mu.ie (M. Ahmed); claudia.roessing@mu.ie (C. Roessing); priyanka.singh0074@gmail.com (P. Singh); gabriel.hogan8@mail.dcu.ie (G. Hogan); mmarkus.helfert@mu.ie (M. Helfert)
0000-0003-2034-1403 (M. Ahmed); 0000-0003-3156-8806 (C. Roessing); 0000-0001-6182-6111 (P. Singh); 0000-0002-6913-3739 (G. Hogan); 0000-0001-6546-6408 (M. Helfert)

management of data assets" [25]. Another definition of data governance is "Data governance specifies a cross-functional framework for managing data as a strategic enterprise asset. In doing so, data governance specifies decision rights and accountabilities for an organization's decision-making about its data. Furthermore, data governance formalizes data policies, standards, and procedures and monitors compliance" [2]. The literature reflects the following four pillars, which are essential components of data governance. Of the top 48 data governance publications (citation ranked by Google Scholar), 43 mention consent, data lifecycle, provenance, or value. However, within these, only four publications reference all four pillars of governance, 3 of which were published since 2019; see Figure 1.

2. Background

The definitions outlined above have the commonality of viewing data as an asset, i.e., it can be used to generate value for its owner. While acknowledging these perspectives, we suggest that they do not capture the full breadth of data governance, nor do they identify the foundations that enable the core aspect of an asset, that ability to create value. The authors above identify 'behaviour', 'management', and 'decision making' as some of the key attributes of data governance. We approach data governance from a different perspective and contend that effective data governance relies on four foundational elements of data as an asset: lifecycle, provenance, consent, and value. In proposing this new framework, we examine each of these foundations, validating their inclusion from the existing literature both individually and collectively.

2.1. Data Lifecycle

A data lifecycle is a data management tool containing phases and activities that transform data for a specific purpose while following quality and security requirements [80]. Different lifecycle models are available. General models are for use in any domain, and specialized models are for use only in certain domains, with each model having a different constitution of their phases and activities [20, 82]. During a data lifecycle, the phases and activities are determined depending on the data processing needs and objectives to be reached, and information on how data has been modified is documented.

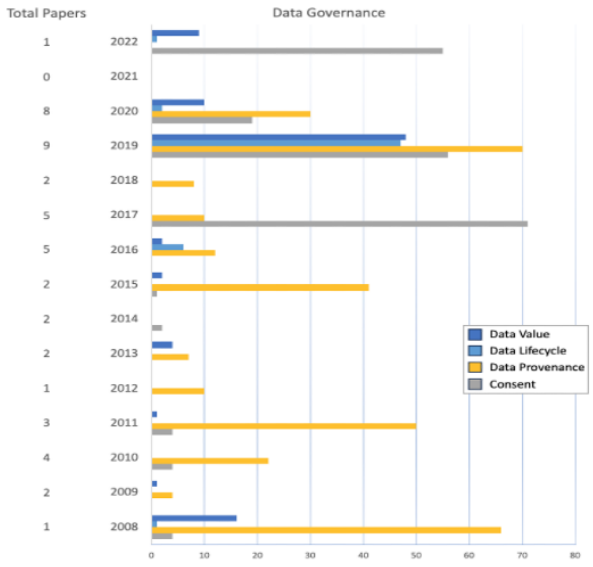

Figure 1: Pillar mentions in most cited papers on Data Governance 2008-2022

2.2. Data Provenance

Data provenance describes the origins and processing of a record. It helps in improving data quality and to increase fairness, accountability, transparency, and explainability [95]. Data provenance is the historical documentation of the data and its sources that traces the influences of the entities and processes on the relevant data. A common record form will specify who has access to it, what is processed, and the purpose of each data item. Data provenance is helpful in data analytics as it requires a tamper-proof data structure [66]. The audit process could be simplified by using a data provenance record to show who collected and processed the data for the dataset in question. Records of data origins and pre-processing can aid in understanding the data's origins and enhance transparency [95].

2.3. Consent Management

The expectation of privacy as a feature of Information Systems has become ubiquitous, and privacy principles [18] are well established. The introduction of the consent provisions of the GDPR [29] is now practiced across all aspects of the data ecosystem, including, for example, provenance [78, 92], data lifecycle [5, 16], value [30, 52]; management [43, 69]; governance [36, 91,99]; and services [41, 55].
In GDPR, informed consent requires that organizations wishing to use data that could be used to identify an individual (personally identifiable information or PII) must have received the active consent of that individual to store and/or process their PII. PII has further categorizations such as sensitive, i.e., financial information, and highly sensitive PII, i.e., medical information, requiring greater protection. Organizations as data controllers and data processors face obligations to protect PII, which requires implementing adequate technology, organizational, and management measures to ensure the protection of the PII. The depiction of the data value creation can be seen in Figure 2.

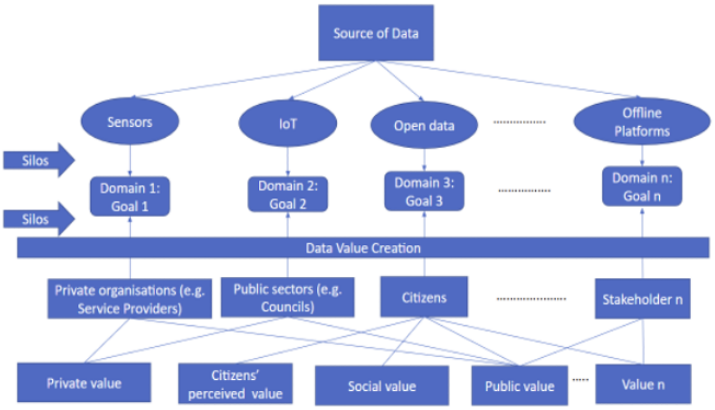

Figure 2: Data Value Creation in Silos

2.4. Data Value

Traditional approaches for evaluating public data value are still in the early stages and often do not adequately reflect its wider impact [61]. Although technocrats have developed strategic models for creating value, end users ultimately need to experience and understand it in their daily lives [54]. The concept of data value is associated with the exploitation of data by the users to accomplish their goals. There is a growing requirement to value data from a data investment point of view and to understand its impact on the economy to improve productivity and value for societal impact [23]. Generated value can be classified as private, social, public, etc.
The data from one domain can also be used in another to increase its value further. For instance, open urban energy data used in the transformation of energy-efficient markets and investments in the real estate sector can also be used to provide value to the general public [32]. Nevertheless, currently, the data value is generated in silos and does not support the data exchange across multiple domains, as

shown in Fig. 2. Municipalities across Europe support open data platforms and deliver services by exploiting it as a part of their digital transformation strategy [71]. Yet there is a lack of understanding about how these data sets can provide value for better outcomes and benefits for their personal use [17].

3. Literature Review

The literature was reviewed in the context of the four proposed pillars of Data Governance: data lifecycle, data provenance, consent management, and data value, as well as the overlapping commonalities between them.

3.1. Data Lifecycle

One of the most valuable resources an organization has is its data. It is essential that data be processed efficiently in order to turn it into knowledge and extract value, which organizations can then use to enhance operations. As described in section 2.1, data lifecycles are used to assist stakeholders in planning and organizing data management. Different data lifecycles are available for use in research and practice. The variation results from the necessity to satisfy various domains' and data processing requirements [20, 82]. These models' representations need to be improved in order to show relevant and necessary elements during data management and to assist stakeholders in analyzing them [15]. Identifying these elements will assist stakeholders in gaining a comprehensive view of data processing, enabling communication among them and assisting them in decision-making. Several studies report the limitations of data lifecycles, stressing the need for models to show more information, which is necessary for data processing nowadays [20, 82]. Furthermore, studies state that there is a need to show transparency in data processing and how important it is to create value [27].

3.2. Provenance

The most valuable asset of an organization is data. An organization's decision is based on this asset, and accurate information is required for making accurate decisions. To ensure data quality, data governance defines policies and standards [65]. For this reason, it is important to know that data comes from trusted and reliable resources. In this regard, data provenance helps identify where the data is coming from, who is creating the data, who is transforming the data, and whether the data has reached the desired location correctly or not. DeStefano, Tao, and Gai [26] outline that one can improve governance by knowing your assets, as you cannot govern well what you do not know. Having less knowledge about data affects the quality, so keeping track of changes made in data can improve data governance in organizations. When making critical decisions, organizations' lack of data governance is a serious concern. Data provenance helps enable tracking by using data life cycle & tracking data stewardship changes [85].

Three stages are very important to optimize data processing and achieve quality data as input, i.e., collection, integration, and filtering of data as input for further analysis. The authors in [27] proposed the smart data lifecycle to manage data. The lifecycle comprises "planning, management, collection, integration, filtering, en-richment, analysis, visualization, access, storage, destruction, archiving, quality, and security". The data enrichment helps create standard repositories and enhances the dimensions in the collected data. The quality of enrichment data depends on continuous and automatic updating of data. Trust is also a fundamental challenge in data governance. In [50], the authors said that data quality assessment must be conducted systematically at all stages and at various points in the life cycle. Traceability in all phases of the data value chain is important to achieve quality. Due to this reason, the value of data heavily relies on data provenance.

3.3. Consent

Privacy and consent management are acknowledged as core features of modern data ecosystems [6]. Consent is a primary concern affecting trust, data quality, transparency, and value [36]. Big data ecosystems such as health informatics [39, 84], big data-intensive research [45, 93], and data sharing between ecosystems [94] have demonstrated sensitivity to consent management. These and other examples [38] illustrate the necessity for business ecosystems and models to transform to consider the management of PII and consent. Privacy considerations are now designer requirements for Information Systems, particularly where data location may be ambiguous, i.e., cloud-based [37].
Organizations differ in their industry, sector, and types with different value propositions and 'business' focus variations. For example, there are different data governance requirements for commercial safety entities and those of a government department whose remit is public access to documents [48], while health organizations may need to weigh the balance between patient privacy and the public good [77]. However, regardless of the different organizational focus, each organization type that is operating under GDPR is required as a data controller or data processor to have traceability of all consents regarding the PII that they control or process so that they can provide assurance that they are compliant with GDPR. The right to consent withdrawal and particularly the right to be forgotten (RTBF) in GDPR have generated differing opinions on their impact on disruptive big data business opportunities. Regardless of discipline, sector, type, or size, each organization faces the challenge of the complexity of consent management within the overall context of information and data governance [72].

3.4. Data Value

Data governance is a crucial aspect of contemporary data management, significantly impacting organizations aiming to leverage their data resources [68] fully. Nonetheless, the understanding of value in the context of governing data seems to be a wicked and persistent challenge [11]. As a part of open data initiatives, many data sets are available for the general public to use, while the resulting societal value from such initiatives is not as prevalent as was predicted [49]. The technological challenge is integrating real-time data from different sources such as mobile apps, web applications, social networks, etc. To address this challenge, Gagliardi et al. [34] developed an ICT-based tool, 'UrbanSense', to assist city authorities in creating new and open services for citizens by integrating their feedback in real time. Nonetheless, the tool will not be useful in cases where data is partially available or unavailable at all. Another challenge is the lack of accessibility and overview of relevant available data and the associated actors [71].
Moreover, Meijer and Potjer [59] found that the complex interactions between different stakeholders contributing to building collective data and influencing its impact are not studied enough in the literature. Their study confirmed that citizen-generated open data could provide guidance for collective governance aimed at generating public value, though this study does not provide any information about the relevance, impact, and distribution of such open data initiatives. Aydin [8] provided a comparative perspective on big data technologies using a big data value chain model but did not include the application in a specific context. Based on the above discussion, our study found that there is a need to understand how value is generated in the context of data governance and the critical roles the different pillars (i.e., provenance, consent management, data life cycle) play in value generation. This has been neglected in the majority of the existing research.

4. Relationship Among Four Pillars

The interconnected relationships between data value, provenance, consent, and the data lifecycle form the foundation for effective data governance and value optimization. Each pillar relationship, data value and provenance, consent and data value, consent and provenance, consent and data lifecycle, and data is collected from various sources and sensors, which is central for creating value-added services for the end users [71]. They create a comprehensive framework where data quality, privacy, and traceability

are continuously reinforced. By understanding and integrating these relationships, organizations can navigate the complexities of data management, maximize its utility, and uphold compliance and ethical standards throughout the data lifecycle.

4.1. Relationship between data value and provenance

Data is collected from various sources and sensors, which is central to creating value-added services for the end users [71]. The key to creating innovative services is the reusing of data to create value globally [1]. Data provenance can assist in re-using data to optimize maximum value, but it is important to ensure that it is of good quality [7] and to know that data comes from trusted and reliable resources. Data provenance describes a record's origins and processing and helps improve data quality [26]. In this regard, data provenance helps identify where the data is coming from, who is creating the data, who is transforming the data, and whether the data has reached the desired location correctly or not. Data knowledge is related to data quality, and tracking changes made to data can improve data governance in an organization. Data quality assessment must be conducted systematically at all stages and various life cycle points [50]. Additionally, traceability in all phases of the data value chain is important to avoid poor quality, which can impact data value. Therefore, there is a need to ensure that data provenance techniques are applied while processing data so that correct decisions and appropriate actions are taken to obtain maximum value out of data.

4.2. Consent and Data Value

Companies collect enormous amounts of data to profile individuals and extract predictive information, resulting in high economic, political, social, and strategic value [72]. Personal data is an asset due to its potential for generating private and commercial value by providing services using personal information [86], which often reduces privacy and social welfare [3]. Under GDPR, personal data must be kept independently and be subject to organizational and technical measures to ensure non-attribution [88]. Data value can be perceived as a function of utility [9], and the distinction between data as capital, data as labor, or data as property (intellectual or otherwise) illustrates the multifaceted nature and relation to value [79]. Property rights in law and the principle of controlling access to and use of the property are also among the foundations for privacy in GDPR. Data value is also contextual to the stakeholder as mentioned above, i.e., private value, social value, public value, etc., and there is potential for value reciprocation [52], though this can be an issue if perceived as transactional under GDPR. The calculation of data value is conditional, i.e., a logical prerequisite for data value is the existence and availability of data [67]. Data value chains are also a well-established value perspective in the literature, and numerous aspects are widely documented [30, 87]. Each of the respective proposing authors acknowledges the effect and impact of consent on their value perspectives and the essential role consent plays in modulating each value proposition.

4.3. Consent and Provenance

Consent management and traceability are facilitated by data provenance. Consents are specific instances of data to be managed for a specific context. The accurate and efficient provenance of consent is equally as important for the assurance, integrity, and quality of the data as the provenance of the data that the consents are linked to. The use of data provenance is a widely accepted approach to the management of consent, and numerous approaches to this problem have been proposed [4] but tend to be siloed by technology [76] and programmatic [48] perspectives. For example, proposed technology approaches incorporating provenance and consent include RDF technology [81]; design patterns [92]; flow auditing in IoT [69]; data privacy traceability [10, 24]; blockchain [35, 57,98]; and smart contracts, [62, 90]. Each of these underlines how the overlapping, though distinctly independent, provenance-consent relationship may be managed as data governance requirements for privacy applied as the efficient provenance of PII consents.

4.4. Consent and Data Lifecylce

This difference in approach (technology vs. programmatic) is also reflected in the relationship between consent and the data lifecycle. Different models and technical solutions are proposed that affirm the close relationship between consent management and the different contexts of the data lifecycle. Consent management is particularly relevant in the data collection, storage (including archiving), security, data processing, and data destruction phases of the data lifecycle. Various approaches are proposed in the literature to provide methods and models confirming the part that consent management now plays as an essential component in the data lifecycle. These range from an abstract personal data lifecycle model [5], a semantic reference architecture for privacy and consent in the data lifecycle [31], relationship mapping of privacy regulation, properties, and lifecycle phases [56], Data Lifecycle phase implementation in blockchain for GDPR compliance, [33], a research data lifecycle for managing and sharing research data with provenance, ethics and informed consent requirements, [22], extended data-lifecycle Data Flow Diagrams to elicit and mitigate privacy threats in IoT [16].

4.5. Provenance and Data Lifecycle

Data lifecycle and data provenance are interconnected, as data provenance can be seen as part of a data lifecycle. Data provenance information is gathered, managed, and updated at all data lifecycle phases [96]. Data provenance is enhanced as data progresses through a data lifecycle by adding details about transformations, processing steps, and usage. In all phases of a data lifecycle, data provenance helps to increase transparency and understanding of the data [50]. Integrating these two pillars improves data traceability and accountability, making identifying error sources easier, understanding data lineage, and making reliable data available to stakeholders for decision-making [28].

5. Existing Frameworks

A number of individual frameworks exist for each of the pillars. Table 1 depicts some of the work in the literature about data governance. These are outlined below for each of the individual pillars. However, they exist in isolation, none of which encompass all four identified pillars or the relationships between them in an effective data governance framework.

5.1. Value

There is a lack of studies that provide guidance to understand how value is created for society, and it requires solutions that can determine how data is produced, published, and promoted. To solve this issue, Abella et al. [1] presented a model for creating value via reusing data that can assist in creating social and economic value for society using data-driven innovation. However, the model has not been evaluated with real data, and the number of considered impacted indicators within the model is limited. Similarly, Pedersen [70] proposes a framework for helping public sector organizations increase value creation and solve society's complex problems using open innovation approaches. Nevertheless, the research findings from this study do not represent the problem in a wider context, and the resulting outcome of the investigated open innovation projects is unknown. Another challenge is the absence of access to an overview of relevant available data and associated actors [71]. To deal with this challenge, Petersen et al. [71] provided an Enterprise Architecture Framework to support existing data spaces for creating value-added services for citizens by composing all accessible data and related information about it. Furthermore, this study defines various data perspectives within the proposed framework, such as interoperability, data security and risk assessment, and data governance. However, this study does not provide any details about how this can be achieved in a real-time application.

Table 1
Data Lifecycle Frameworks

Reference	Framework	Objective	Limitations
64	Master data lifecycle management	Reference model to analyze master data lifecycle	The model was not evaluated using real-life settings. The model doesn't mention data privacy or data security.
7	Value creation assessment	To provide a guideline to show what aspects to consider when evaluating an open data initiative in value creation.	The assessment metric is not accurate
82	Data value chain	To provide a framework with key factors to create value in information-intensive services	It does not consider data security.
51	DaLif	To create a data lifecycle for data-driven governments	The framework needs a practical validation

5.2. Lifecycle

Various frameworks are available in the literature, some of which are listed in Table 1. Lim et al. [51] created a framework for data-based value creation in information-intensive services. The study identified nine essential factors which are considered important for creating data value. The factors are data source, data collection, data analysis, information on the data source, information delivery, customer (information user), value in information use, and provider network. Intending to transform data into knowledge and subsequently create value, Shah et al. [82] created a data lifecycle focused on data-driven governments. The framework has 14 phases and was created from analyzing 76 data lifecycle models. The analysis also allowed authors to define which phases are mandatory or optional in a data life cycle.

In [64], a holistic framework is provided for master data lifecycle management, which contains strategic, tactical, and operational aspects to assist organizations in analyzing a master data lifecycle and identifying areas for improvement. The model is divided into four competencies: data portfolio, data and system design, data supply, and data support. Attard et al. [7] focus on creating data value through open government data. They provide the definition of a data value network, which shows a set of linked activities, data discovery, data curation, data interpretation, data distribution, and data exploitation. The work concludes by adding new aspects to the value creation assessment framework, which guides which aspects should be considered when evaluating an open data initiative in value creation.

5.3. Consent

There are many consent management frameworks (CMF) available in the literature. Bonnici and Coles Kemp [14] propose a principle-based CMF of consent (ethics) theory (policy) norms norm (software) manifestation where the authors argue that both organizational and software processes are essential

to a consent management framework. Tokas and Owe [89] propose a modeling language framework to facilitate data subject management of privacy settings, and various blockchain-based frameworks are proposed [75]. The Interactive Advertising Bureau [40] introduced a Transparency & Consent Framework, including policies and technical specifications with supporting operational and compliance resources for commercial operators. More recently, dynamic consent management frameworks have been proposed using both blockchain [44, 46, 47, 74, 97] and non-blockchain approaches [42, 83]. However, these proposed frameworks do not consider the relationships between these CMFs and other IS paradigms. In providing a singular focus on consent management, the interactions and interfaces within IS and the context of CMFs within data governance are unacknowledged. Without a holistic perspective, the proposed frameworks sit uneasily in data governance and IS. The type of data statement, including lifecycle, province, consent, and value, can be seen in Figure 3.

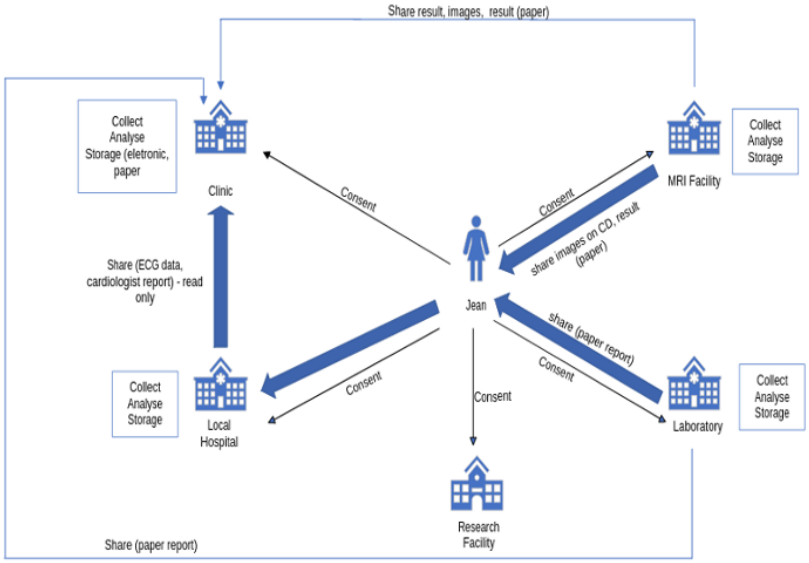

Figure 3: Case Study - Type of data statement - lifecycle, provenance, consent, value

5.4. Case Study

To provide some perspective for the real-world context of our framework, we outline the following use case. This illustrates:

- The complex nature of the data as an asset, with value meaning different things in different contexts and scenarios.
- The different phases of the data lifecycle.
- The opportunities for multiple provenance instances to occur.
- The multiple touch-points for consent in the data journey

Jean is ill and is going for the first time to a medical clinic for treatment, her previous medical records are not available. Jean provides her consent and personally identifiable information (PII) to the clinic so that they can provide treatment to her. She also provides consent to the medical doctors to collect, analyze, and store her sensitive personally identifiable information (SPII) and to share it with other facilities for diagnosis. The clinic collects and stores Jeans PII (name, address, age, date of birth, height, weight, place of birth, family medical history, and any medical conditions), SPII, and consent in a mixture of electronic formats such as database, spreadsheet, and emails, in addition to various physical formats, including paper, dictation tapes, and blood samples. These are initial records establishing the origin of the information (from Jean) and the date of the record, along with the obligations on the clinic

as a data controller. The clinic (data controller) sends her blood samples to a laboratory (data processor) for processing and analysis. Jean has made an MRI with a local radiology facility. She provides her consent and personally identifiable information (PII) to the facility (the data controller) for collecting and analyzing her sensitive medical information. The laboratory (data processor) and the radiology facility send the analysis results to her clinic (data controller) electronically, and a physical copy is provided to Jean. Once this is completed, both facilities dispose of the information. After analysis of her results, she is referred to a hospital for treatment, where her medical history, test measurements, analysis, results, diagnosis, Jean's physician's observations, decisions, and recommendations are forwarded by the clinic (the controller) to the hospital (the processor) and stored there. Jean consents to the treatment and is admitted to the hospital.

The details of her inpatient tests and her treatment are kept on file at the hospital and shared back with her physicians at the clinic. The clinic provides follow-on care for Jean, and she consents to participate in a clinical trial, allowing her data to be used by a research facility to discover new treatments. Her data journey mirrors each of the above steps of Jean's patient journey. When Jean provides her PII and SPII, she provides consent to allow for her data to be managed by each organization as a data controller or data processor. Jean can withdraw or revoke her consent to use her data at all times. Each organization she interacts with follows its own data lifecycle management process, applying it as appropriate to her data, ensuring that Jeans' data is managed appropriately from generation to disposition. In each case, the provenance of her data and the provenance of her consent is established from its origin (Jean) through each activity, operation, use, or transaction on her data and consent, whether collected, received, transferred, or disposed of. Ensuring that Jeans' data is genuine, accurate, verifiable, and trustable as it is used in and applied to her treatment. It is also important to ensure that it is traceable, can be located when required, and verifiably disposed of when Jean requests or when she withdraws consent to the continued use of her data. The combination of consent management, data lifecycle management, and data provenance management enables the creation of value from the data journey, which can be multi-faceted - private (i.e., financial to the organization), social (to citizens with similar challenges), public value (based in knowledge and understanding), or personal to Jean (her health).

6. Proposed Framework

In the context of Jean's healthcare journey, the four pillars, data value, provenance, consent, and data lifecycle, collectively address key gaps and challenges in data governance, creating a reliable, value-driven framework for handling sensitive information. Among these pillars, data value and provenance ensure that Jean's information is traceable, trustworthy, and capable of supporting quality treatment decisions across various medical facilities. Consent and data value illustrate that Jean's permissions provide ethical and legal compliance and enhance her data's value at each step of her care. The interplay between pillars, consent, and provenance provides an additional layer of assurance that Jean's data, along with her consent, is accurately recorded and traceable across all medical and diagnostic stages. Moreover, linking consent to each phase of the data lifecycle, i.e., from collection to secure disposal, creates a clear, structured path for handling Jean's data. Thus empowering her with better control of data and aligning each step with security and privacy standards. These pillars minimize potential loopholes, ensuring Jean's data is ethically managed, secure, and optimized for healthcare outcomes while providing a strong foundation for generating public, social, and private value. The proposed data governance framework is shown in Figure 4.

The implementations of the aforementioned pillars are solution-agnostic (not tied to any specific platform, tool, or specific technology). The choice is left to the practitioners to adopt and apply what best fits their needs. However, we believe that to provide a cohesive data governance framework, each of the four pillars, i.e., data lifecycle, data provenance, consent management, and data value, must be in place. We also believe that excluding any of these pillars limits the effectiveness of a data governance strategy.

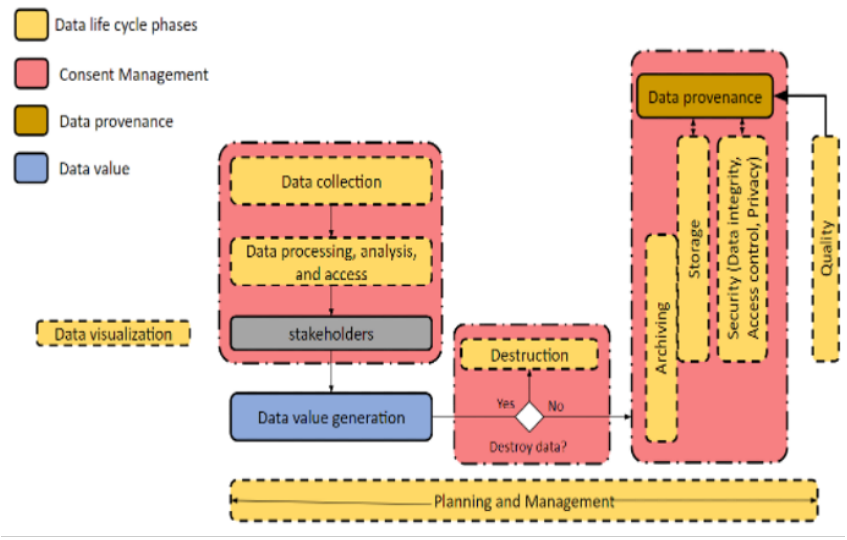

Figure 4: Data Governance Framework

Conclusion

This article highlights the dependencies and relationships between data lifecycle, data provenance, permission management, and data value creation, enabling organizations to design and implement appropriate data governance structures for their needs. To help the creation of data workflows that meet data governance requirements, the authors propose a framework that uses these four pillars as the main building blocks for establishing holistic data governance. Each approach and proposition outlined above addresses data governance in the context of provenance and consent as enablers in the data lifecycle and the subsequent generation of value from the efficient use of quality data assets. Data is only available through the collection, acquisition, or sharing mechanisms, each of which requires consent in specific regulatory environments such as GDPR in Europe, and data assets can only be exploited if they're known and available from technical, legal, and business perspectives.
The tight coupling between the core provenance, consent management, lifecycle, and data value is demonstrated in the literature. These four perspectives together form the main building blocks for implementing holistic data governance, facilitating the development of unique data workflows to meet data governance requirements for differing organizations and their business contexts, and underpinning value creation for organizations exploiting their digital assets.
Future work will focus on advanced data provenance tracking, and we will look deep into adaptive consent management mechanisms. Moreover, we will further explore developing governance models that maximize data value and align well with organizational-specific needs.

Acknowledgments: This publication has emanated from the research conducted with the financial support of Science Foundation Ireland under Grant No. $23/PSF/12107$.

References

[1] Abella, A. et al.: A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems. Cities. 64, 47–53 (2017).
[2] Abraham, R. et al.: Data governance: A conceptual framework, structured review, and research agenda. Int. J. Inf. Manag. 49, 424–438 (2019).
[3] Acquisti, A. et al.: The Economics of Privacy. J. Econ. Lit. 54, 2, 442–492 (2016).
[4] Al-Ruithe, M. et al.: A systematic literature review of data governance and cloud data governance. Pers. Ubiquitous Comput. 23, 5, 839–859 (2019).

[5] Alshammari, M., Simpson, A.: Personal Data Management: An Abstract Personal Data Lifecycle Model. Presented at the Business Process Management Workshops: BPM 2017, Barcelona, Spain, September 10 (2017). https://doi.org/10.1007/978-3-319-74030-0_55.
[6] Anwar, M.J. et al.: Secure big data ecosystem architecture: challenges and solutions. EURASIP J. Wirel. Commun. Netw. 2021, 1, 130 (2021).
[7] Attard, J. et al.: Data-driven governments: Creating value through open government data. Trans. Large-Scale Data- Knowl.-Centered Syst. XXVII Spec. Issue Big Data Complex Urban Syst. 84–110 (2016).
[8] Aydin, A.A.: A Comparative Perspective on Technologies of Big Data Value Chain. IEEE Access. 11, September, 112133–112146 (2023). https://doi.org/10.1109/ACCESS.2023.3323160.
[9] Barker, K.: Privacy Protection or Data Value: Can We Have Both? In: Kumar, N. and Bhatnagar, V. (eds.) Big Data Analytics. pp. 3–20 Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-27057-9_1.
[10] Baum, B. et al.: Opinion paper: Data provenance challenges in biomedical research. Inf. Technol. Munich Ger. 59, 4, 191–196 (2017).
[11] Benfeldt, O. et al.: Data governance as a collective action problem. Inf. Syst. Front. 22, 299–313 (2020).
[12] Bonnici, C.J., Coles-Kemp, L.: Principled Electronic Consent Management: A Preliminary Research Framework. In: 2010 International Conference on Emerging Security Technologies. pp. 119–123 (2010). https://doi.org/10.1109/EST.2010.21.
[13] Bork, D., Roelens, B.: A technique for evaluating and improving the semantic transparency of modeling language notations. Softw. Syst. Model. 20, 4, 939–963 (2021).
[14] Bugeja, J., Jacobsson, A.: On the Design of a Privacy-Centered Data Lifecycle for Smart Living Spaces. Presented at the Cham (2020).
[15] Cabitza, F. et al.: Making open data more personal through a social value perspective: a methodological approach. Inf. Syst. Front. 22, 131–148 (2020).
[16] Cavoukian, A.: Privacy by design and the emerging personal data ecosystem. Information and Privacy Commissioner, Ontario (2012).
[17] Christopherson, L. et al.: Toward a data lifecycle model for NSF large facilities. In: Practice and Experience in Advanced Research Computing. pp. 168–175 (2020).
[18] Corti, L. et al.: Managing and Sharing Research Data. A Guide to Good Practice. Sage Publications Ltd, London (2014).
[19] Coyle, D., Manley, A.: What is the value of data? A review of empirical methods. J. Econ. Surv. 38, 4, 1317–1337 (2024). https://doi.org/10.1111/joes.12585.
[20] Curcin, V.: Embedding data provenance into the Learning Health System to facilitate reproducible research. (2017).
[21] DAMA International: DAMA-DMBOK: Data management body of knowledge. Technics Publications LLC, Basking Ridge, New Jersey, USA (2017).
[22] DeStefano, R.J. et al.: Improving Data Governance in Large Organizations through Ontology and Linked Data. In: 2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud). pp. 279–284 (2016). https://doi.org/10.1109/CSCloud.2016.47.
[23] El Arass, M., Souissi, N.: Data Lifecycle: From Big Data to SmartData. In: 2018 IEEE 5th International Congress on Information Science and Technology (CiSt). pp. 80–87 (2018). https://doi.org/10.1109/CIST.2018.8596547.
[24] Elkhodr, M. et al.: Data Provenance in the Internet of Things. In: 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA). pp. 727–731 (2018). https://doi.org/10.1109/WAINA.2018.00175.
[25] European Union: 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/EC (General Data Protection Regulation). Publications Office of the European Union (2018).
[26] Fanzo, J.: Considering Ethics Along the Data Value Chain for Nutrition. Sight Life. 33, 1, 122–126

(2019).
[27] Fatema, K. et al.: Compliance through Informed Consent: Semantic-Based Consent Permission and Data Management Model. Presented at the (2017).
[28] Francisco, A., Taylor, J.E.: Understanding citizen perspectives on open ur-ban energy data through the development and testing of a community energy feed-back system. Appl. Energy. 256, 113804 (2019).
[29] Freund, G.P. et al.: An Analysis of Blockchain and GDPR under the Data Lifecycle Perspective. Mob. Netw. Appl. 26, 1, 266–276 (2021).
[30] Gagliardi, D. et al.: Information and communication technologies and public participation: interactive maps and value added for citizens. Gov. Inf. Q. 34, 1, 153–166 (2017).
[31] Garcia, R.D. et al.: A Blockchain-based Data Governance with Privacy and Provenance: a case study for e-Prescription. In: ICBC. pp. 1–5 IEEE, Piscataway (2022). https://doi.org/10.1109/ICBC54727.2022.9805545.
[32] Geisler, S. et al.: Knowledge-Driven Data Ecosystems Toward Data Transparency. ACM J. Data Inf. Qual. 14, 1, 1–12 (2022).
[33] Georgiopoulou, Z. et al.: GDPR Compliance: Proposed Technical and Organizational Measures for Cloud Providers. Presented at the Cham (2020). https://doi.org/10.1007/978-3-030-42048-2_12.
[34] Huhtala, T.: Transformation Of The Business Model In An Occupational Health Care Company Embedded In An Emerging Personal Data Ecosystem: A Case Study In Finland. (2017).
[35] Iacob, N., Somonelli, F.: Towards a European Health Data Ecosystem. Eur. J. Risk Regul. 11, 4, 884–893 (2020).
[36] International Advertising Bureau Europe: TCF – Transparency Consent Framework - IAB Europe, https://iabeurope.eu/transparency-consent-framework/, last accessed 2024/11/11.
[37] Ismagilova, E. et al.: Security, Privacy, and Risks Within Smart Cities: Literature Review and Development of a Smart City Interaction Framework. Inf. Syst. Front. 24, 2, 393–414 (2022).
[38] Jafarbeiki, S. et al.: ACE: A Consent-Embedded privacy-preserving search on the genomic database. Heliyon. 10, 8, (2024). https://doi.org/10.1016/j.heliyon.2024.e29399.
[39] Javed, I.T. et al.: SecureConsent: A Blockchain-Based Dynamic and Secure Consent Management for Genomic Data Sharing. In: 2024 International Conference on Smart Applications, Communications and Networking (SmartNets). pp. 1–7 (2024). https://doi.org/10.1109/SmartNets61466.2024.10577693.
[40] Kashyap, R.: Has demography witnessed a data revolution? Promises and pitfalls of a changing data ecosystem. Popul. Stud. 75, S1, 47–75 (2021).
[41] Khalid, M.I. et al.: Privacy-First Paradigm for Dynamic Consent Management Systems: Empowering Data Subjects through Decentralized Data Controllers and Privacy-Preserving Techniques. Electronics. 12, 24, 4973 (2023). https://doi.org/10.3390/electronics12244973.
[42] Kumi, S. et al.: A Blockchain-based platform for data management and sharing. Procedia Comput. Sci. 203, 95–102 (2022). https://doi.org/10.1016/j.procs.2022.07.014.
[43] Ladley, John.: Data governance: how to design, deploy, and sustain an effective data governance program. Academic Press, [Place of publication not identified] (2020).
[44] Lassinantti, J. et al.: Relevant social groups for open data use and engagement. Gov. Inf. Q. 36, 1, 98–111 (2019).
[45] Liaw, S.-T. et al.: Quality assessment of real-world data repositories across the data life cycle: A literature review. J. Am. Med. Inform. Assoc. 28, 7, 1591–1599 (2021).
[46] Lim, C. et al.: From data to value: A nine-factor framework for data-based value creation in information-intensive services. Int. J. Inf. Manag. 39, 121–135 (2018).
[47] Line, N.D. et al.: Control, use, and ownership of big data: A reciprocal view of customer big data value in the hospitality and tourism industry. Tour. Manag. 1982. 80, 104106 (2020).
[48] Lis, D., Otto, B.: Data Governance in Data Ecosystems – Insights from Organizations. (2020).
[49] Lytras, M.D., Visvizi, A.: Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research. Sustainability. 10, 6, 1998 (2018).
[50] Mani, Z., Chouk, I.: Impact of privacy concerns on resistance to smart services: does the 'Big

Brother effect' matter? J. Mark. Manag. 35, 15–16, 1460–1479 (2019).
[51] Manna, A. et al.: An Analysis of FIPP Clauses with Respect to Data Lifecycle Phases and Privacy Properties. Presented at the Singapore (2022).
[52] Margheri, A. et al.: Decentralised provenance for healthcare data. Int. J. Med. Inform. Shannon Irel. 141, 104197–104197 (2020).
[53] Meijer, A., Potjer, S.: Citizen-generated open data: An explorative analysis of 25 cases. Gov. Inf. Q. 35, 4, 613–621 (2018).
[54] Nagaraj, A.: A Mapping Lens for Estimating Data Value. Harv. Data Sci. Rev. Special Issue 4, 1–32 (2024). https://doi.org/10.1162/99608f92.82f0de5a.
[55] Neisse, R. et al.: A Blockchain-based Approach for Data Accountability and Provenance Tracking. Presented at the ACM International Conference Proceeding Series (2017). https://doi.org/10.1145/3098954.3098958.
[56] Nokkala, T. et al.: Data Governance in Digital Platforms. In: AMCIS 2019 Proceedings. AIS, Cancún, Mexico (2019).
[57] Ofner, M.H. et al.: Management of the master data lifecycle: a framework for analysis. J. Enterp. Inf. Manag. 26, 4, 472–491 (2013).
[58] Otto, B., Wende, K.: Data Governance. In: Hildebrand, K. et al. (eds.) Daten- und Informationsqualität: Auf dem Weg zur Information Excellence. pp. 265–283 Vieweg+Teubner, Wiesbaden (2008).
[59] Paik, H.-Y. et al.: Analysis of Data Management in Blockchain-Based Systems: From Architecture to Governance. IEEE Access. 7, 186091–186107 (2019).
[60] Pandit, H.J. et al.: Investigating Conditional Data Value Under GDPR. In: Proc. 14th Int. Conf. Semantic Syst.(SEMANTiCS). pp. 1–5 (2018).
[61] Pansara, R.R.: Transactions on Latest Trends in IoT Open Access, Peer Reviewed, Refereed Journal 3246-544X A Double-Blind Peer Reviewed Journal Un-raveling the Complexities of Data Governance with Strategies, Challenges, and Future Directions. Transactions. 1, 3, 2–5 (2023).
[62] Pasquier, T. et al.: Data provenance to audit compliance with privacy policy in the Internet of Things. Pers. Ubiquitous Comput. 22, 2, 333–344 (2018).
[63] Pedersen, K.: What can open innovation be used for, and how does it create value? Gov. Inf. Q. 37, 2, 101459 (2020).
[64] Petersen, S.A. et al.: Value-added services, virtual enterprises, and data spaces inspired enterprise architecture for smart cities. In: Collaborative Networks and Digital Transformation: 20th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2019, Turin, Italy, September 23–25, 2019, Proceedings 20. pp. 393–402 Springer (2019).
[65] Politou, E. et al.: Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. J. Cybersecurity. 4, 1, tyy001 (2018).
[66] Ranaweera, T.A.V.Y. et al.: Ensuring Electronic Health Record (EHR) Privacy using Zero Knowledge Proofs (ZKP) and Secure Encryption Schemes on Blockchain. In: 2023 5th International Conference on Advancements in Computing (ICAC). pp. 792–797 (2023). https://doi.org/10.1109/ICAC60630.2023.10417417.
[67] Rantos, K. et al.: Blockchain-based Consents Management for Personal Data Processing in the IoT Ecosystem. SCITEPRESS - Science and Technology Publications (2018).
[68] Rasheed, M.A. et al.: Use of big data governance in several corporate sectors. VFAST Trans. Softw. Eng. 9, 4, 92–101 (2021).
[69] Rosenbaum, S.: Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access. Health Serv. Res. 45, 5p2, 1442–1455 (2010).
[70] Saad, M.I.M., et al.: Achieving trust in cloud computing using secure data provenance. In: 2014 IEEE Conference on Open Systems (ICOS). pp. 84–88 (2014). https://doi.org/10.1109/ICOS.2014.7042634.
[71] Savona, M.: The Value of Data: Towards a Framework to Redistribute It. SSRN Electron. J. (2019).
[72] Sebastian-Coleman, L.: Meeting the Challenges of Data Quality Management. (2022).
[73] Seneviratne, O.W.: Data Provenance and Accountability on the Web. In: Provenance in Data Science. pp. 11–24 Springer International Publishing, Cham (2020).

[74] Shah, S.I.H., et al.: DaLiF: a data lifecycle framework for data-driven governments. J. Big Data. 8, 1, 1–44 (2021).
[75] Shahmansoori, A., Roedig, U.: Dynamic Recognition of Speakers for Con-sent Management by Contrastive Embedding Replay. IEEE Trans. Neural Netw. Learn. Syst. 1–15 (2023). https://doi.org/10.1109/TNNLS.2023.3317493.
[76] Sharon, T., Lucivero, F.: Introduction to the Special Theme: The expansion of the health data ecosystem – Rethinking data ethics and governance. Big Data Soc. 6, 2, 205395171985296 (2019).
[77] Singh, D., Maniam, J.N.: TOWARDS DATA PRIVACY AND SECURITY FRAMEWORK IN BIG DATA GOVERNANCE. Int. J. Softw. Eng. Comput. Syst. 6, 1, 41–51 (2020).
[78] Spiekermann, S. et al.: The challenges of personal data markets and privacy. Electron. Mark. 25, 2, 161–167 (2015).
[79] Swartz, P., Da Veiga, A.: PoPI Act - opt-in and opt-out compliance from a data value chain perspective: A South African insurance industry experiment. In: ISSA. pp. 9–17 IEEE (2016). https://doi.org/10.1109/ISSA.2016.7802923.
[80] Tith, D. et al.: Patient Consent Management by a Purpose-Based Consent Model for Electronic Health Record Based on Blockchain Technology. Healthc. Inform. Res. 26, 4, 265–273 (2020).
[81] Tokas, S., Owe, O.: A Formal Framework for Consent Management. In: Gotsman, A. and Sokolova, A. (eds.) Formal Techniques for Distributed Objects, Components, and Systems. pp. 169–186 Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-50086-3_10.
[82] Truong, N.B. et al.: GDPR-Compliant Personal Data Management: A Blockchain-Based Solution. IEEE Trans. Inf. Forensics Secur. 15, 1746–1761 (2020).
[83] Ujcich, B.E. et al.: A Provenance Model for the European Union General Data Protection Regulation. In: Provenance and Annotation of Data and Processes. pp. 45–57 Springer International Publishing, Cham, Switzerland, London, UK (2018). https://doi.org/10.1007/978-3-319-98379-0_4.
[84] Vayena, E., Gasser, U.: Strictly Biomedical? Sketching the Ethics of the Big Data Ecosystem in Biomedicine. Presented at the Cham (2016).
[85] Voytek, B.: The Virtuous Cycle of a Data Ecosystem. PLoS Comput. Biol. 12, 8, e1005037–e1005037 (2016).
[86] Werder, K. et al.: Establishing Data Provenance for Responsible Artificial Intelligence Systems. ACM Trans. Manag. Inf. Syst. 13, 2, 22:1-22:23 (2022).
[87] Yazici, I.M., Aktas, M.S.: A novel visualization approach for data provenance. Concurr. Comput. Pract. Exp. 34, 9, e6523 (2022).
[88] Khalid, M. I., Ahmed, M., & Kim, J. (2023). Enhancing data protection in dynamic consent management systems: formalizing privacy and security definitions with differential privacy, decentralization, and Zero-Knowledge proofs. Sensors, 23(17), 7604.
[89] Khalid, M. I., Ahmed, M., Ansar, K., & Helfert, M. (2024, March). Leveraging Blockchain Technologies for Secure and Efficient Patient Data Management in Disaster Scenarios. In World Conference on Information Systems and Technologies (pp. 12-21). Cham: Springer Nature Switzerland.
[90] Khalid, M. I., Ahmed, M. (2023, September). Blockchain Based Dynamic Consent Management Systems for Enhancing Quality of Life for People with Disabilities. In 2023 IEEE International Smart Cities Conference (ISC2) (pp. 01-07). IEEE.