Big Data Value Graph: Enhancing Security and Generating New Value from Big Data

Massimiliano Gervasi[1,2], Nicolò G. Totaro[1,2,*], Anna Fornaio[3] and Danilo Caivano[4]

[1] *Department of Innovation Engineering, University of Salento, Lecce, Italy*
[2] *Centre for Applied Mathematics and Physics for Industry (CAMPI), University of Salento, Lecce, Italy*
[3] *Cybersecurity Research Lab - CRLab, University of Salento, Lecce, Italy*
[4] *Department of Computer Science, University of Bari "Aldo Moro", Bari, Italy*

Abstract
The opportunity to extract value from data has become a key capability for organisations, especially with the spread of Big Data. The Data Mesh is a model that proposes a solution for Big Data management. In this paper, we propose that the integration of this model with the Big Data Value Chain can provide benefits for organisations participating in the mesh. The two main benefits relate to security and Value Generation. The new framework resulting from this merging allows to increase the level of security and the Value Generated for participants. Throughout the paper, the concepts of Big Data and Data Mesh will be introduced. Then, our proposed integration and how it can positively affect the levels of security and the results of Value Generation will be explained.

Keywords
Data Mesh, Big Data, Accountability, Integrity, Value Creation

1. Introduction

Data have economic value, and the capacity to extract it is becoming a key feature for organisations [1]. This relevance transformed the opportunity of extracting Value from Big Data into an imperative to establish or remain competitive [2]. Typical characteristics of Big Data frameworks have been identified in the literature. The first aspect to be considered is related to data availability, to which the creation of Value is linked [3]. On the other hand, security paradigms favour data sharing only between organisations working in a common direction [2]. A second aspect is related to the Value Generated according to the paradigm $\text{Value}(A+B) > \text{Value}(A)+\text{Value}(B)$, as reported in [4], applied to the Big Data context. This can also be extended to the cost and resource consumption that a Big Data initiative requires, as sharing resources could lead to cost savings. Applying the previous paradigm and specifying it to costs, we get $\text{Cost}(A+B) < \text{Cost}(A) + \text{Cost}(B)$, as verified in [5, 6, 7].
Focusing on security, there are two fundamental aspects that are worth of note regarding the approach to the confederation of different organisations sharing resources: the first aspect is the need to safeguard through a collective bargaining agreement the methodology of collaboration

ITASEC 2023: The Italian Conference on CyberSecurity, May 03–05, 2023, Bari, Italy
*Corresponding author.
massimiliano.gervasi@unisalento.it (M. Gervasi); nicologianmauro.totaro@unisalento.it (N. G. Totaro); anna.fornaio@unisalento.it (A. Fornaio); danilo.caivano@uniba.it (D. Caivano)

and resource sharing between different organisations [5, 6]. The second concern regards the way to ensure security in such large perimeters with several independent actors.

In response to the proposed requirements and the considerations, a new architecture was identified in the literature, namely, the *Data Mesh*, which will be explored in more detail in one of the next section. We propose to integrate the Data Mesh model with *Big Data Value Chain* model to combine the enhancement of security approaches in data and technology sharing with the identification of processes that can generate value from Big Data. The resulting framework brings two major benefits: first, an enhanced level of security during the generation of Data Products through the adherence to the Data Mesh Governance; then, an increase in the Value Generated for each of them, since both data and technologies are mapped into a network structure where observable quantities (e.g. KPI) can be attributed to both nodes (individual products) *and* processes.

The scope of this work is the definition of a basis for proper modelling data and technology sharing. In Section 2, the state of the art concerning Big Data, its security and the associated value chain is explained. Section 3 introduces the Data Mesh model and briefly explains its operating principles. Section 4 sets out the basis of our proposal and the research questions it addresses. A new relational structure is illustrated in Section 5 to formalize the expected benefits of the proposed framework in terms of security, which are described in Section 6, together with security-related implementation aspects. Conclusions are drawn in Section 7.

2. State of the art

2.1. Big Data and its security issues

In literature, Big Data are often defined according to their characteristics or features called *V's of Big Data*. The original definition included three V's, namely *Volume*, *Variety* and *Velocity* and was first introduced in [8]. Since then, many V's have been added to the domain in order to refine the very definition of Big Data, up to 56, see e.g. [9, 10, 11, 12, 13]. The security of Big Data is a critical issue for organisations using this technology. The amount and variety of data collected makes it necessary to pay more attention to its security. The main focus is on data management and protection, regulatory compliance and risk management. The level of security of Big Data is reflected in one of the V's that characterize it, which is the Vulnerability [13]. This depends on the protection barriers put in place in defence of the data. But the level of security to be ensured depends on another V, the Value. Naturally, the greater the sum of the value of the data held, the greater must be its protection. The value of Big Data also derives from the correlations that can be made with it, given the large volumes available. In [14] is described how it is necessary to find an optimal balance between data quality and data security.

Numerous studies have delved into security analysis for Big Data. In [15] the phases and characteristics of Big Data are studied to analyse what security requirements they need. The five requirements identified are: confidentiality, efficiency, authenticity, availability, integrity. In [16] there is a framework containing the most interesting threats and challenges in Big Data security. Some possible approaches are encryption, access control mechanisms, authentication and authorization, anomaly-based network intrusion detection and Machine Learning techniques. In [16] are also described the main problems and attacks, as well as some solutions adopted to protect infrastructure, privacy, and data integrity. Security systems are implemented to protect the flow of data input and output for each element, so that each data packet in transit meets privacy and security standards, is congruent and integral. The security system provided for both input and output communication flows, together with mechanisms to defend the protocols of the communication itself, is designed to counter any man-in-the-middle attacks. The layers of protection can ensure that the data in transit are indeed the correct data in various ways. For instance, the use of artificial intelligence could allow a check on possible pollution or repetition of data. AI could also check the size of data in transit, to prevent DoS-type attacks. This system of multiple layers of protection is known as defence-in-depth. As described in [17], this methodology consists of applying distinct levels of security mechanisms, creating a redundancy that provides protection even if one of these barriers is breached.

From these studies it can be understood that to ensure data security, even in the Big Data Value Chain, multiple security measures must be implemented. Furthermore, it is important that organisations adopt a security culture that encourages the adoption of secure practices by all employees.

2.2. Big Data Value Chain

Value, although considered as a V's of Big Data, must be understood as an atypical feature insofar as it relates more to the use of the data and not to its characteristics. In this regard, in [18] is distinguished Value from other V's, separating the concept of Big Data from its use and attributing to this nexus the real reason for the inconsistency of Big Data definitions in the literature. In reference to Big Data modelling the processes, we find direct reference to the *Value Chain* of Porter [19], and also to [20, 21, 22, 23]. In [19] the type of resources used in its activities are material resources that organisations can observe and are useful in production. In [24], the set of resources includes intangible ones, as information, that managers can capture from the virtual world and use them to generate Value. In [24], is theorized the **Virtual Value Chain**, whose main activities are: *gathering, organizing, selecting, synthesizing,* and *distributing*. As reported in [23]: "*The Virtual Value Chain framework is perfectly applicable in a Big Data context*", concept confirmed in [25, 26, 27].
From the Virtual Value Chain, in line with the Process View, meaning with a focus on processes that enable the generation of value, and moving from the framework in [23], Wu *et al.* [28] extended Ackoff's Data-Information-Knowledge-Wisdom (DIKW) model [29] in the RDIKW model, specialising the different phases of the Virtual Value Chain according to the capabilities generated by the resources and required in the different phases of the Big Data initiatives. For more details, see Figure 1. Here, of special interest is the division of the initiatives into two phases: **Big Data Capitalization** and **Big Data Exploitation**, which separate into two distinct macro-areas the steps necessary to extract value from the data. Namely, phases of generation and transformation to a usable format, from the subsequent phases dedicated to analytics. The two areas differ in the modelling of value-added processes, in techniques used, in skills required and in costs to be incurred for the activities. Although the distinction is clear, the Exploitation phase is highly dependent on the previous one, in fact the data analytics changes depending on the granularity chosen in the Capitalization phase, as well as the type of ETL and data cleaning adopted. The two areas can also be distinguished because data are obtained at the end

of Capitalization for consumption in Exploitation, but actors involved are not always the same, and this leads to strategic misalignments and high consumption of resources.

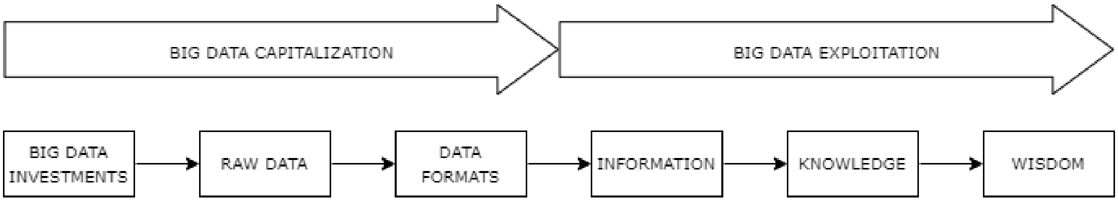

Figure 1: Wu et al. RDIKW model [28]

3. Data Mesh

The Data Mesh was introduced by Dehghani [3] as a shift from centralized architectures based on five themes. Organisation goes from a centralized data ownership to a decentralized model, which links the ownership and responsibility of the data to the domains of source. Architecture moves from monolithic data collection in Data Warehouse and Data Lake [30, 31] to the data connection through distributed network. Technologies moves from technological solutions that treat data as a sub-product of the execution of a code pipeline, to ones that treat data and the code that handles it as a single autonomous unit. Operational model shifts data Governance from a centralized top-down operational model with human intervention to a federated model with computational policies embedded in the network domains. Principle shifts the system of value association away from data seen only as a resource to be collected with its limitations (as low quality, redundancy, difficulty of retrieval, use of different formats and targeting to a specific use) to *data as a product* to be shared. The architecture is based on the presence of several actors, called *domains*, which may represent entire organisations or individual business units. The domains would therefore have lower total costs than the sum of the costs they would have acting individually, as described in [4, 5, 6]. In the case of selling data to actors outside the Data Mesh, domains could generate DPSs according to customer requirements, thus providing a highly customized product. Here we introduce the four fundamental principles on which it is based the Data Mesh [3].

1 - Domain-oriented decentralized data ownership and architecture: the Data Mesh requires the decentralisation of data sharing ownership to the reference domains. Each domain has a team of experts covering every area of data management and of ETL processes, so they can work completely independently.

2 - Data as a product: to be considered a Data Product (DP), data must adhere to some rules of usability, feasibility, and be valuable. Usability characteristics include: meeting predefined quality metrics and formats, being user-friendly and understandable, being easily accessible by any user, and being interoperable and unified with data from other domains. Each individual DP must also fulfil the DATSIS principles [32] being: discoverable, addressable, trustworthy, self-describing, interoperable, and secure. We define DP as *primitives* (DPP) if derived from a domain and *secondaries* (DPS) if obtained from the elaboration of one or more DPs.

3 - Self-serve data infrastructure as a platform: a self-service data platform to enable the domains' teams to store and manage DPs throughout their life cycle, manage the network connection of DPs, and share knowledge that is extracted from the information on the network [3]. As can be seen in [32], users can view DPs on a Data Mesh Catalogue, which presents the DPs stored in the Data Mesh Catalogue Storage, where DPs of any level are stored.

4 - Federated computational Governance: autonomy over data management is granted to each domain, but they must follow common processes and rules in order to obtain data in the predetermined format and at a minimum quality level. Governance federation thus seeks to balance the autonomy and agility of the domains, while guaranteeing the interoperability, security and global conformity of the network.

3.1. Data Mesh Application Examples

As the Data Mesh is a recently theorized model, there is still limited literature on use cases. In [32], the architectures of Netflix and Zalando after the implementation of the Data Mesh are illustrated. In [33], various examples of model implementation with various technologies are reported, including Google Cloud, AWS, Databricks, and Kafka. In particular, in the book there is a summary table for each technology, with a brief description of it, the pros and cons, and a suggestion as to which type of organisation would be preferable to apply each of them. The use case of implementing the Data Mesh for Kafka is also taken up in [34], explaining how Saxo Bank integrated Kafka mesh to achieve particular benefits. The bank had integrated Kafka to be able to have real-time and high-quality data, but it was not enough. The integration of the Data Mesh brought new data Governance and numerous benefits. It has increased data quality through Governance, cataloguing and quality management processes. It reduced time to market by facilitating the data search system. And finally, it has facilitated integration by setting internationally followed standards.

4. Proposal: merging Data Mesh and Big Data Value Chain to enhance secure sharing

Our proposal is to integrate the Data Mesh within the Big Data Value Chain to enhance identifying value along multiple value chains, and simultaneously the overall security of data sharing. As the Data Mesh is a recent approach, the proposal also aims at strengthening relevant security issues not yet addressed in the literature. The relevance of the security aspects in the Data Mesh have been considered in [35], which mention how authorization and authentication mechanisms are required in the Data Mesh. However, as reported in [36], there are other security mechanisms that have to be ensured within a network: confidentiality, integrity, non-repudiation, access control, availability, and accountability. As the data value can increase in the Data Mesh, as we argued above, an associate increase in the level of security for its protection is required. Specifically, the main aspects of security we will focus on are *accountability* and *integrity*. Accountability makes it possible to keep track of all the activities carried out in an environment, starting from their origin, while integrity ensures that data are not modified in an unwanted or unauthorized manner [36].

Two main limitations arise when working with Big Data. The first is the management of potential attacks or malfunctioning, which could make a DP inaccessible. The second issue is harder to handle, as it does not relate to a failure of the information system, but to an improper combination of DPs and technologies along a given value chain. In particular, data quality and accountability can be granted along with technological performance, but their incompatibility (e.g., improper application of methods for specific data types, lack of hypothesis verification) may generate invalid inferences from their combination (insights, other products).

Our proposal relies on a new representation model that allows for the allocation and identification of Data Products creation together with their processing, through the use of Technology Products. This representation relies on a relational network structure, i.e. a directed graph, that helps to locate not only Data Products, but also the processes to extract information and value from them, thus including technologies in the representation. Formally, this approach generates two node attributions, adding *Technology Products* as a dual of the existing Data Products. This approach also supports the traceability of data production, so that the different value chains arising from Data Products can be securely accessed, e.g. relying on pseudonymization protocols. In this way, we also enable the tracking of additional information of the multiple data production processes that can be relevant when assessing the accountability of data *in relation to technologies*, and vice versa.

This data structure will be linked to the Data Mesh principles to foster an architectural design for data and technology sharing and consumption in line with increasing data sources, consumers, diversity of use cases, and access modes. As for security, this provides advantages in terms of: (i.) increasing data veracity, by bridging the gap between the actual origin of data and the details (where and when) of their sharing for analysis, which could be implemented using current distributed ledger technologies; (ii.) allowing autonomy through clear contracts, enforcing the Data Mesh Governance; (iii.) realize Governance requirements such as security, privacy, and legal compliance; (iv.) fostering domain segregation as actions on specific DPs do not affect other ones in different domains; (v.) maintain data integrity, as the network generates new nodes when new Data Products are created instead of changing them, which reduces the risks of corruption of existing Data Products. In addition to security benefits, operational advantages in terms of optimal use of resources have to be underlined: (i.) optimization of continuous change through localization of changes within domains; (ii.) cost reduction for data ownership in a decentralized architecture; (iii.) reduced complexity of data management and burden on domain teams in managing the life cycle of their DPs, as well as computational capabilities necessary for Governance to find, access, build, or deploy a DP; (iv.) opportunity to achieve higher-order value through aggregation and correlation of independent DPs.

5. Big Data Value Graph (BDVG)

Now we discuss more in-depth the proposal to merge the Data Mesh with the Big Data Value Chain, generating a new framework that we call **Big Data Value Graph**. The reference to graphs derives from the relationship between the various points in the new model. The combination of these two elements would make it possible to increase the Value Generated and to identify and reduce the costs associated with this generation. It could also be seen as a generalisation of the current Big Data Value Chain. In fact, it is possible to regain the chain when a single node participates in the mesh.
The Big Data Value Chain is divided into two phases (Figure 1): *Big Data Capitalization* and *Big Data Exploitation*. In the BDVG, the first phase is completely performed within the domains. Here DPs assume the role that was of the data formats in the BDVC, they are the output of the capitalization phase. In the Exploitation phase, the first step consists of aggregating and correlating DPs, to extract new information and produce DPSs at the same time. Further elements in the Exploitation phase correspond to those who collect the extracted information to generate knowledge. This brings better decision-making, as it is based on more reliable and more easily exploitable data, with consequences in reduction of costs and in increase of profits, as described in [37, 38]. The edges in the BDVG are direct, and the terminal nodes in that structure are functional to the representation of wisdom, understood as the ways in which an organisation decides to integrate the knowledge it has obtained into its processes. This is why the multiplicity of nodes in the Graph stops at the level of knowledge, given the individuality of each organisation in the transformation of this into wisdom, thus abstracting it outside the perimeter of the BDVG. The BDVG brings with it many benefits:

1) a **Technology Mesh** which fosters the union between the Data Mesh and the rest of the Graph, allowing, through a series of technologies, the correct interaction between those in charge of the Exploitation phase and those in charge of making DPs available. Just as Dehghani proposes data as a product, in our model we also extend this concept to technologies. This is because technological immaturity can affect the success of an initiative, such as in the creation of assets [23]. We then define **Technology Product** software, algorithms, or more generally services, including those offered by platforms. Among the technologies present, in addition to those needed for the Data Mesh, there will obviously be those needed for the elements dealing with information and knowledge, such as, for example, analytical tools, data visualisation and knowledge graphs. Just as for each DP there is a responsible, in the same way, a responsible is identified for each Technology Product. The set of data and technologies would then constitute the entire catalogue of products that could be shared and used within the mesh or sold to external actors. Shared technologies may also be labelled to be contextualized to individual use cases, each of which may be based on specific assumptions, constraints, and needs. The concatenation process underlying the Technology Mesh and, more generally, the BDVG leads to a compositional structure of the analysis techniques, which is not, in general, commutative; so, the order in which they are used also affects the final result. The implementation of the Technology Mesh also reduces the total cost to be incurred compared to the sum of the individual costs, as these are amortized across several elements, following the principle in [4];

2) an architecture that favours the **collaboration of different actors**, even belonging to separate organisations, in order to ensure access to and use of a wide variety of resources, as desired in the literature [2, 7]. In addition to technology and data, the BDVG also provides for the sharing of knowledge, skills, and competences. For example, the opportunity of having numerous data analysis techniques that are easy to use makes it possible to reduce model risk problems, thus minimising the influence that the use of a model has on the final result. On the other hand, such measurable effects also make it possible to estimate latent constructs such as skill and knowledge with appropriate analysis techniques;

3) a **dynamic architecture**, in order to adapt to the random nature of initiatives operating in Big Data contexts. The requirements negotiations, typical of Big Data initiatives [39], is facilitated through easier access to different and higher-level data and technologies, thanks to the Data and Technology Mesh. They enable adaptation and scalability at different stages of the initiative. The freedom granted by the dynamism of the architecture makes it possible to envisage atomic activities of the initiative, even cyclical ones, in which testing and validation phases are foreseen [39]. At such stages, it is necessary to leave freedom to domain experts [40]. The cyclicity of the different phases is visible in various Big Data Frameworks such as in [41, 42];

4) identify the **Value Generated** and the **associated cost**. In [9, 43], several criteria are find to identify the Value Generated by Big Data. It derives from the benefits that an organisation obtains from the integration of acquired knowledge within its systems, hence wisdom. In the Graph, it is possible to derive the path that led to the generation of wisdom. It is not trivial to calculate the Value Generated by every single element of the mesh that participated in the initiative, as it is very complex to give a weight to the participation in the generation of Value. Easier to identify is the cost associated with each arc of interaction, according to the resources used. By comparing the final Value Generated with the sum of the costs along the route, it is possible to understand whether the initiative ultimately brought added value to the organisation, as well as analysing each point along the route to identify critical points in the event of a failed initiative. In this way, it is also possible to correctly define for each element the suggested retail price to actors outside the mesh. Recalling again the principle in [4], at the end the Value Generated will be greater than the sum of the individual Values, while the total cost will be less than the sum of the individual costs. At this point, the economic benefit of the BDVG can be understood.

To better understand point 4, we can observe Figure 2. It can be seen that at the $DPS_{1.1}$ is immediate to associate the corresponding cost. Thanks to the Graph all the products can be associated with each interaction that took part in its creation. In this way, even for the $DPS_{2.1}$ it is easy to calculate the associated costs, simply adding up the costs of all interactions that led to its generation. In this way, the unit cost associated with knowledge creation can be reduced. In fact, each element does not participate in the creation of a single block of knowledge, but contributes to the creation of several, so that the cost can be amortized in the best possible way.

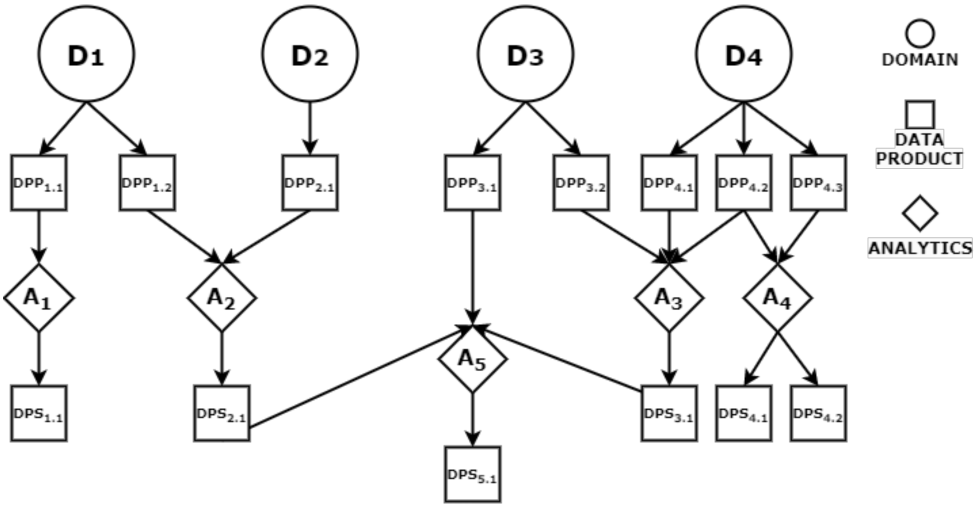

Figure 2: Data Mesh Graph

In Figures 3-4, the BDVG is schematized. Governance, as visible in both images, manages all the steps that take place in the Graph. In the first picture you can see how the Data and Technology Mesh are integrated within the model, enabling the creation of Data Products and Technology Products. In the second image we can see how the Graph has a multi-linear path to the knowledge creation stage. As mentioned earlier, wisdom is outside the perimeters of the Graph, as it is individual to each organisation. This image also shows a special case where knowledge creation is reached via a linear path, which falls under the definition of the Big Data Value Chain.

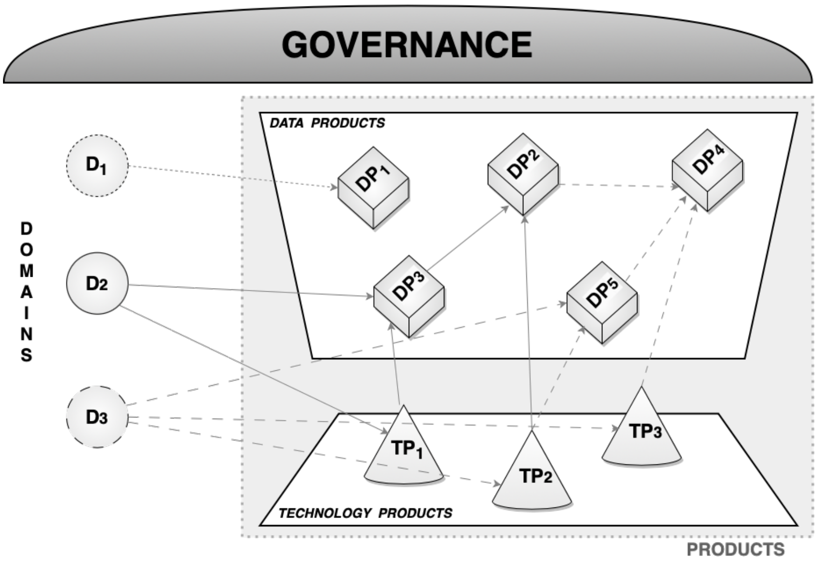

Figure 3: Data and Technology Mesh

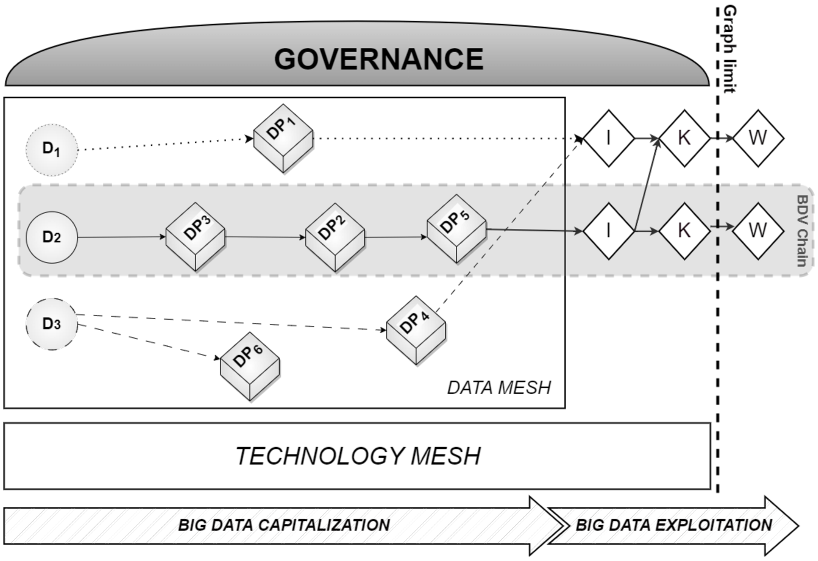

Figure 4: Big Data Value Graph representation

6. Big Data Value Graph Security

6.1. Data and Technology Accountability and Integrity

In this work, we pay special attention to accountability and integrity in the BDVG. The path that led to the creation of a given Data Product can be represented as in Figure 2. The path associated with each individual DP is stored internally, and knowing this path enables the accountability property to be fulfilled. In order to still have accountability of *Data and Technology Products*, and thus study their paths backwards for verification and, potentially, make an anomaly detection within the whole path, we can envisage a registry containing the scheme of all the paths generated, together with the additional attributes and metadata functional to check their validity. This scheme does not contain any DP, but merely stores the steps taken within the platform. This register, preserving every edge of interaction of the BDVG, would ensure accountability for the entire system.

To enhance integrity, it is sufficient to also keep track of the DPs hash. In this way, if the hash remains unchanged, integrity is guaranteed along the path. Otherwise, if this is not respected, the scheme can indicate the timestamp of the moment when the hash changed and, hence, identify when the compromise occurred. To increase the security level of this log, distributed ledger technology (DLT) could be used to increase the difficulty of a cyberattack. In this case, security also depends on the type of technology chosen to implement the BDVG and its nodes. DLT may be useful in the present scenario, depending on the transactions and consensus protocols,as long as more than half of the nodes are not compromised.

Accountability here relate to both data and technologies applied to obtain a product. In this way, it is possible to always keep track of the data used, the technologies applied, *and* their combination. The comparison of different results derived from the combination of the same data with different technologies requires, in general, *partial* order structure: specifically, we cannot assume *a priori* the comparability of different products and processes within the BDVG, which involve multiple domains and value/security dimensions (here, integrity and accountability). The lack of a total order underlying the comparison of data/technology production processes, in fact, lead to a richer view where inequivalent evaluations can be highlighted by means of order-theoretic approaches [44].

6.2. Big Data Value Graph Security Challenges

There is no unique recipe for the Data Mesh, its implementation can be declined following various strategies. One of the elements that the different implementation strategies must certainly have in common is the focus on security. The level and quality of security must meet the standards that this architectural model provides. The integration of the Data Mesh within the Big Data Value Chain forces a focus on new elements and changes in priorities. Data Products have a higher level of quality and usability than data usually stored in databases. Their higher expendability and reliability increases their value, and consequently their attractiveness to cybercriminals. With the same level of security and volume of data, the level of risk associated to attacks on DPs is higher than one associated to normal data. For this reason, the importance of security becomes even more critical. This leads to numerous challenges to be solved to ensure protection within the Big Data Value Graph. Dehghani [3] describes how the Data Mesh must define five basic security points: access control, encryption, confidentiality levels, data retention, regulations, and agreements.

However, the Data Mesh model is a recent creation, which is why the security measures to be taken to protect it are also in their infancy. There are various challenges to overcome, concerning all the properties that must be secured in order to define a model as secure.

7. Conclusions

This paper proposes a new framework for value generation to increase the Value Generated and improve security mechanisms. This solution involves the collaboration of different actors who share resources and expertise. The paper analyses the mechanisms that bring these benefits to member organisations. The focus on future developments could be on implementing additional security systems that ensure the protection of the entire network, also integrating the Data Mesh Governance with a knowledge base that enhances privacy [45]. The main objective of the proposed BDVG is to create more value at a lower cost for its member organisations. In some contexts, one could also think of selling Data and Technology Products to actors outside the mesh, while respecting the rules governing these markets. The sale to actors outside the mesh may also not be envisaged, for instance, in the case of integrating this model to manage public administration [46].

Future work will address in more detail the validation of this framework in relevant scenarios. For this purpose, the identification of the main targets and actors will be a pivotal step: in fact, the present model aims at scalability, so it could be implemented with different types of actors. In operative scenarios, the elements constituting the domains of the mesh could be compartments of the same organisation, nodes of public bodies spread across the territory, or even different organisations. In each case, the advantages would derive from different factors: as for the first two targets, the advantage would be the increase in the level of efficiency and effectiveness in the exploitation of the shared digital resources, together with increased transparency and targeted involvement of stakeholders thanks to the structure provided by domains. In the latter case, medium-small organisations could join to to share resources and enable synergistic functionalities and services that they could not provide individually.

References

[1] V. Fast, D. Schnurr, M. Wohlfarth, Data-driven competitive advantages in digital markets: An overview of data value and facilitating factors, Lecture Notes in Information Systems and Organisation 48 LNISO (2021) 454.

[2] J. M. Cavanillas, E. Curry, W. Wahlster, New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe, Springer Nature, 2016.

[3] Z. Dehghani, Data Mesh: Delivering Data-Driven Value at Scale, 1.ed - preview version, O'Reilly Media, Inc., 2022.

[4] J. Dong, C.-H. Yang, Business value of big data analytics: A systems-theoretic approach and empirical test, Information and Management 57 (2020).

[5] D. Lavie, The competitive advantage of interconnected firms: An extension of the resource-based view, Academy of Management Review 31 (2005).

[6] J. Dyer, H. Singh, The relational view: Cooperative strategy and sources of interorganizational competitive advantage, The Academy of Management Review 23 (1998).

[7] J. Zeng, K. W. Glaister, Value creation from big data: Looking inside the black box, STRATEGIC ORGANIZATION 16 (2018) 105–140.

[8] D. Laney, 3-D Data Management: Controlling Data Volume, Velocity, and Variety, Technical Report, META Group Res. Note, 2001.

[9] J. Gantz, D. Reinsel, Extracting value from chaos, Technical Report 2011, IDC IView, 2011.

[10] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, P. Tufano, Analytics: The real-world use of big data, Technical Report 2012, IBM Global Business Services, 2012.

[11] S. Arockia, The 17 V's Of Big Data, International Research Journal of Engineering and Technology 4 (2017) 329–333.

[12] S. Dhamodharavadhani, G. Rajasekaran, Unlock different V's of big data for analytics, International Journal of Computer Sciences and Engineering 6 (2018) 183–190.

[13] A. Hussien, How Many Old and New Big Data V's Characteristics, Processing Technology, And Applications (BD1), International Journal of Application or Innovation in Engineering & Management 9 (2020) 15–27.

[14] M. Talha, A. A. El Kalam, N. Elmarzouqi, Big data: Trade-off between data quality and data security, in: Procedia Computer Science, volume 151, 2019, pp. 916–922.

[15] H.-M. Chen, R. Kazman, J. Garbajosa, E. Gonzalez, Toward big data value engineering for innovation, in: Proceedings - 2nd International Workshop on BIG Data Software Engineering, BIGDSE 2016, 2016, pp. 44–50.

[16] R. Bhatia, M. Sood, Security of big data: A review, in: PDGC 2018 - 2018 5th International Conference on Parallel, Distributed and Grid Computing, 2018, pp. 182–186.

[17] M. T. Rahman, M. S. Rahman, H. Wang, S. Tajik, W. Khalil, F. Farahmandi, D. Forte, N. Asadizanjani, M. Tehranipoor, Defense-in-depth: A recipe for logic locking to prevail, Integration 72 (2020) 39–57.

[18] O. Ylijoki, J. Porras, Perspectives to Definition of Big Data: A Mapping Study and Discussion, Journal of Innovation Management 4 (2016) 69–91.

[19] M. E. Porter, Competitive advantage: Creating and sustaining superior performance, Free Press, New York and London, 1985.

[20] E. Curry, The big data value chain: Definitions, concepts, and theoretical approaches, in: New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe, Springer Nature, 2016, pp. 29–37.

[21] R. Louati, S. Mekadmi, Toward a conceptualization of big data value chain: From business problems to value creation, in: Research Anthology on Big Data Analytics, Architectures, and Applications, IGI Global, 2022, pp. 319–335.

[22] A. Z. Faroukhi, I. El Alaoui, Y. Gahi, A. Amine, An adaptable big data value chain framework for end-to-end big data monetization, Big Data and cognitive computing 4 (2020).

[23] O. Ylijoki, J. Porras, A recipe for big data value creation, Business Process Management Journal 25 (2019) 1085–1100.

[24] J. Rayport, J. Sviokla, Exploiting the virtual value chain, Harvard Business Review 73 (1995) 75–85.

[25] H. Miller, P. Mork, From data to decisions: a value chain for big data, IT Professional 15 (2013) 57–59.

[26] A. Braganza, L. Brooks, D. Nepelski, M. Ali, R. Moro, Resource management in big data initiatives: processes and dynamic capabilities, Journal of Business Research 70 (2017) 328–337.

[27] M. Janssen, V. Van der, A. Wahyudi, Factors influencing big data decision-making quality, Journal of Business Research 70 (2017) 338–345.

[28] X. Wu, L. Liang, S. Chen, How big data alters value creation: through the lens of big data competency, Management Decision 60 (2022) 707–734.

[29] R. Ackoff, From data to wisdom, Journal of Applied System Analysis 16 (1989) 3–9.

[30] M. E. M. El Aissi, S. Benjelloun, Y. Loukili, Y. Lakhrissi, A. E. Boushaki, H. Chougrad, S. Elhaj Ben Ali, Data lake versus data warehouse architecture: A comparative study, in: WITS 2020: Proceedings of the 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems, volume 745 of *Lecture Notes in Electrical Engineering*, Springer Nature, 2022, pp. 201–210.

[31] J. Kachaoui, A. Belangour, Challenges and benefits of deploying big data storage solution, in: ACM International Conference Proceeding Series, 2019.

[32] I. Machado, C. Costa, M. Y. Santos, Data-driven information systems: the data mesh paradigm shift, in: 9th International Conference on Information Systems Development (ISD2021 Valencia, Spain), 2021.

[33] M. Jacek, B. Sven, S. Marian, S. Mariusz, Data Mesh in Action, 1.ed - preview version, Manning Publications Co., 2023.

[34] D. Joshi, S. Pratik, M. P. Rao, Data governance in data mesh infrastructures: The saxo bank

case study, in: Proceedings of the International Conference on Electronic Business (ICEB), volume 21, 2021, pp. 599–604.

[35] I. A. Machado, C. Costa, M. Y. Santos, Data mesh: Concepts and principles of a paradigm shift in data architectures, in: Procedia Computer Science, volume 196, 2021, pp. 263–271.

[36] J. Andress, The basics of information security: Understanding the fundamentals of InfoSec in theory and practice, The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice, Syngress, 2011.

[37] F. Provost, T. Fawcett, Data science and its relationship to big data and data-driven decision making, Big Data 1 (2013) 51–59.

[38] E. Brynjolfsson, L. Hitt, H. Kim, Strength in numbers: How does data-driven decision-making affect firm performance?, in: International Conference on Information Systems 2011, ICIS 2011, volume 1, 2011, pp. 541–558.

[39] H.-M. Chen, R. Kazman, J. Garbajosa, E. Gonzalez, Big data value engineering for business model innovation, in: Proceedings of the Annual Hawaii International Conference on System Sciences, 2017, pp. 5921–5930.

[40] M. Akerman, C. Lundgren, M. Bärring, M. Folkesson, V. Berggren, J. Stahre, U. Engström, M. Friis, Challenges building a data value chain to enable data-driven decisions: A predictive maintenance case in 5G-Enabled manufacturing, Procedia Manufacturing 17 (2018) 411–418.

[41] D. Kriksciuniene, V. Sakalauskas, B. Kriksciunas, Process optimization and monitoring along big data value chain, in: W. Abramowicz (Ed.), Business Information Systems Workshops, BIS 2015, volume 228 of *Lecture Notes in Business Information Processing*, Springer, 2015, pp. 75–86.

[42] D. Dutta, I. Bose, Managing a big data project: The case of ramco cements limited, International Journal of Production Economics 165 (2015) 293–306.

[43] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Byers, Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, 2011.

[44] M. Angelelli, Tropical limit and a micro-macro correspondence in statistical physics, Journal of Physics A: Mathematical and Theoretical 50 (2017) 415202.

[45] M. T. Baldassarre, V. S. Barletta, D. Caivano, A. Piccinno, M. Scalera, Privacy Knowledge Base for Supporting Decision-making in Software Development, in: Sense, Feel, Design: INTERACT 2021 IFIP TC 13 Workshops, Bari, Italy, August 30–September 3, 2021, Revised Selected Papers, Springer, 2022, pp. 147–157.

[46] C. Catalano, P. Afrune, M. Angelelli, G. Maglio, F. Striani, F. Tommasi, Security testing reuse enhancing active cyber defence in public administration., in: ITASEC, 2021, pp. 120–132.