Big Data Features As Crucial Concept which Effects Data Interoperability

Sana Ibrahim Shatwan
Graduate school of electrical and computer engineering
Altınbaş university
Istanbul, Turkey
Sanaibrahimwe261189@gmail.com

Weiam S. Elsaghair
Graduate school of electrical and computer engineering
Altınbaş university
Istanbul, Turkey
Weiam2005@gmail.com

DOI: 10.56201/rjmcit.v9.no2.2023.pg39.51

Abstract

Big data has become an exciting term in the last decades, and it conducts a significant rate of data with different features of velocity, variety, value, volume, and veracity. Those features describe big data identity. Think about the data generated by social media websites and apps, such as Facebook, Twitter, and WhatsApp……Etc. This data is crucial because of the amount, variety, and growth acceleration, but it is exhausting to manage. Big data interoperability indicates the capability to interchange and share the data, information, and knowledge between organizations and devices that knowledge and information were extracted from different data sources. This complex mutual process grows an organization's performance. According to the previous, big data and its five features are essential because of its usability and analytics. Organizations need to extract the expected benefit from preserving and retrieving data. We develop a data interoperability concept, which is improved as data exchangeability can be enriched. In this study, we will discuss the impact of big data feature's value, volume, variety, velocity, and veracity on the interoperability of the data and data exchangeability feasibility and how these features affect the prediction of future decisions in the institution, as well as the interaction and exchange of data between the organizations.

***Keywords**: Big data, Big data Interoperability, Big data features.*

I. INTRODUCTION

Big data is defined according to different aspects of information, technology and impact. For instance, big data is defined as a massive amount of data, and diverse information assets that need profitable, creative models of information manipulation to increase the clarity of decision-making [1] . Another official qualifier classified big data according to four categories [2].

- Attributes of data: where five dimensions are indicated.
- Technological needs: a structure that can develop adequate storage, exploit and analyze data.
- Thresholds: exceeding the processing capacity.
- Social impact: big data is a civilized, technological, and universal phenomenon.

Throughout data attributes, four features define Big Data: volume, velocity, variety, and value. Any height in volume, velocity, and variety of information assets demands cost-effective, creative forms of information processing for enhanced insight and decision-making [2] .

Then, with technology concerns, Extensive data sets, primarily in the characteristics of volume, velocity, and/or variety, demand a scalable architecture for efficient storage, manipulation, and analysis [2]. also, another definition threshold the data sets and analytical techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis, and visualization technologies [2].

For social impact, big data is a phenomenon that causes three major transformations in how we evaluate data, transforming how we comprehend and manage society:

1. More data.
2. Incomplete data.
3. Correlation overtakes causality [2].

So, as stated before, big data is not a traditional set of data, and it has different features (velocity, variety, value, volume, and veracity). These features give different properties to the data. So, through these characteristics, the data will be described as big data.

Therefore, to earn big data benefits, must be efficient and flexible to deal with heterogeneous systems. Exchanging data needs to conduct another essential concept, big data interoperability.

Big data interoperability has an impact on the whole organization's performance. This big data interoperability is tough and complex to be implemented with heterogeneity. Also, many barriers have appeared as follows:

- Organizational barrier (refers to adjusting the organization's business process, responsibilities, and personal data protection).
- Legal barrier (different organizations mean diverse operating platforms with different strategies and policies).
- Technical barrier (concerns with different technical aspects of applications and I.T. infrastructure like (Semantic, syntactic, data inconsistency, and data quality).

So, organizations must improve their data interoperability by taking into account the different features of big data.

II. Material and Method

- Maryam Ghasemaghaei, Goran calic (2020). Assessing the impact of big data on firm innovation performance: Big Data is not always better data. Journal of Business Research 108(2020) 147-162 [3].

  Maryam and Goran discuss the impact of three big data features (volume, velocity, and variety) on innovation performance (efficiency and efficacy). The results show that instead of focusing on handling big data features as a holistic variable, it is important to negotiate the conceptual and operational differences among those features. The study implies that velocity and variety positively affect a firm's innovation performance, and data velocity has a substantial role over the others. On the contrary, data volume does not play that significant role. And it is crucial to negotiate big data features as separate concerns, which will enhance overall innovation performance. Also, they conducted that big data is not always the best data.

- Maryam Ghasemaghaei (2019). Understanding the impact of big data on firm performance: the necessity of conceptually differentiating among big data characteristics. International Journal of Information Management [4].

This study reveals the effect of essential big data characteristics (Volume, Velocity, and Variety) on firm performance and the mediating role of data veracity and value to stand that relation. The findings present that variety improves data value, whereas volume and velocity do not impact. Also, veracity has a negative effect on data volume, but variety and velocity are affected positively. This means data value has entirely arbitrated the impact of veracity on firm performance. These results support firms' managers to improve the whole firm's performance and competitive advantage.

- Nir Kshetri (2014). Big data's impact on privacy, security, and consumer welfare. International Journal of Information Management. The University of North Carolina [5].

The study investigates the relationship between big data characteristics, privacy, security, and consumer welfare. This relation is examined from different sides (data collection, storing, sharing, and accessibility). Using big data and specific features and linking them with gathering, storing, manipulating, and being accessible by another partner may lead to being stuck with many privacy and security issues. The companies and consumers will be concerned about potential offences and misuse of their information. Also, the consumers will not be comfortable as companies know and use more than they intentionally provide. Another concern that will negatively affect consumer welfare is unskilled and technologically inexperienced consumers. Those can cause many defects (awareness of multiple suppliers to simplify effective search, controlling their online actions) so, according to that, necessity becomes urgent to develop a firm big data policy, which will take into account the sensitivity of consumer information.

- Elaheh Yadegaridehkodi, Mehrbakhsh Nilashi, Liyana Shuib, Mohd Hairul Nizam Bin Md Nasir, Shahla Asadi, Sarminah Samad, Nor Fatimah Awang (2019). Impact of Big Data Firm Performance in Hotel Industry. Electronic Commerce Research and Applications.[6]

This study proposes integrating four organizational dimensions (Technology, human, organization, and environment) to identify key factors and how they will affect the firm's performance. The result presents that I.T. expertise was the only critical factor under the human dimension. At the same time, external pressure was stated as an essential factor in the environmental dimension. Also, management support is an essential factor related to the organization dimension, and technology is the most influential dimension. Moreover, the result demonstrates two concerns as the biggest hurdle in the adoption of big data. Those concerns were security and privacy. This outcome will improve the whole business and the government's decision in concern of big data adoption.

III. STUDY METHODOLOGY

a qualitative method perceives a specific phenomenon, a specific event, study value, and meaning. Conversely, a quantitative method when there is previous knowledge about a research topic contributes to numerical data and usable statistics. According to the usability of qualitative and quantitative methods, the research will use the quantitively method.

- Gathering the data using a Questionnaire.
- Analysis the collected data using SPSS.
- Result documentation.

According to, a used technique with qualitative methods, which will produce trusted and actual results. All of that will give clear evidence of effective and efficient techniques.

IV. DATA AND BIG DATA INTEROPERABILITY CONCEPT

Data interoperability concerns to understand the meaning of information between the sender and the recipient. So, data interoperability offers a suitable data description across systems [7]. According to that, a good practice of big data interoperability layers in detail can be conducted as follows [8]:

1. **Legal interoperability:** in which organizations with various policies and strategies can work together, the legal interoperability could be legislation affected or issues related when manipulating personal data.

2. **Organizational interoperability:** to make different organizations communicate with each other's, taking into account different barriers. Here, breaking down data silos and using open accessibility will increase overall organizational efficiency and effectiveness. Also, dealing with various data types is a crucial concept in this part. Finally, aligning data requirements to be suitable and support the business process.

3. **Semantic interoperability:** heterogeneous data sources are an enormous challenge to applying semantic interoperability. This can be done by converging the semantic gap between datasets, mapping semi-automated meanings, and choosing the correct format.

4. **Technical interoperability:** dealing with data integration patterns and multiple data sources.

V. LEVELS OF INFORMATION SYSTEMS INTEROPERABILITY [9]

a) **Technical Interoperability:** refers to the infrastructure of equipment and software components, platforms, and communications protocols.

b) **Syntactic Interoperability:** Indicates to standardized the data format to improve the exchangeability.
c) **Organizational Interoperability** is about transferring meaningful data and using it with various information systems with different infrastructures.
d) **Semantic Interoperability:** pertains to the content definition and understanding **of** various data similarly with no ambiguous and inaccurate meaning.

VI. INTEROPERABILITY APPROACHES AND ASSESSMENTS [10]

Establishing interoperability has three approaches that have to be relevant to each other:

a) Integrated approach: general format for entire models has to be built. This format has to be admitted by all entities.
b) Unified approach: it refers to a meta-level standard format and provides model mapping.
c) Federated approach: it is about the languages and approach of a task as the standard format.

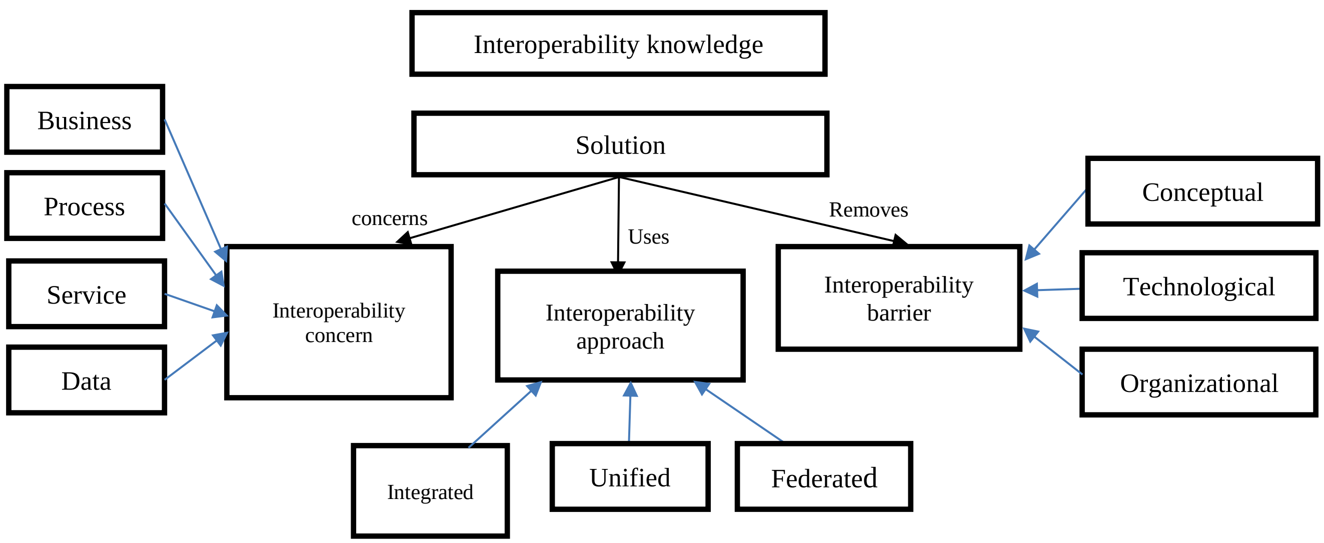

Figure1: Interoperability Basic Concepts and Approaches

1. *Interoperability Assessment:*

For better collaboration, the interoperability needs to be improved and measured by defining metrics. That will help the institutions realize their strengths and weaknesses and avoid enterprise interoperability's adverse effects and barriers.

2. *Types of Assessment: **[11]***

Three types of interoperability assessment have been considered:

a) *The Potential Assessment*

It provides many characteristics of the potential to overcome any hurdles and interoperate with a third partner—Open systems as an example of high potential interoperability. Here in the potentiality interoperability, three barriers (Conceptual (Syntactic/Semantic), Technological, Organizational) are considered according to their impact to evaluate the enterprises. There are five levels of potentiality for each of them: (isolated, initial, executable, connectable, interoperable)—Moreover, the four levels of interoperability (are business, Processes, services, and data).

a) Isolated: Refers to the interoperability incapacity.
b) Initial: Strong efforts are required for interoperability for a better partner.
c) Executable: executing the interoperability is possible, even if the problem rate is high.
d) Connectable: connect ability with the partner; even the risks can occur with a distant partner.
e) Interoperable: indicates to the evolution of different levels of interoperability.

b) *The Compatibility Assessment*

identifying the defeats that may cause problems when the data has to be exchanged. So, the idea is to check the compatibility concerning interoperability barriers. Three concepts have to be conducted to investigate the incompatibility.

i. **Conceptual Compatibility:** could be syntactic or semantic. Syntactic refers to whether the exchanged information has the same syntax or not, and semantic if this information has the same meaning.
ii. **Organizational Compatibility:** authorizes and responsibilities at both sides are considered to check the organization's compatibility.
iii. **Technological Compatibility:** organization I.T. Platform and communication protocols compatibility have to be suitable and acceptable by the partners.

Here we have to deal with many barriers.

Table 1. Shows the interoperability barriers that are relevant to the enterprises' concerns

		The barriers to interoperability		
		Conceptual	***Technological***	***Organizational / Legal***
Interoperability Concerns	***Business***	Visions, strategies, Organization cultures, understanding.	- I.T. infrastructure.	- organization structure. - Legislations. - Business rules.
	Process	Syntax and semantics processes	Process interfaces and supporting tools.	Procedures of work, processes organization.

		The barriers to interoperability		
		Conceptual	***Technological***	***Organizational / Legal***
	Service	Semantics to name and describe services.	Interface, architecture.	Responsibility/ authority to manage services.
	Data	- Data representation and semantics. - Data restriction rule.	Data exchange formats.	Responsibility authority to add/delete, change/ update data.

a) *The Performance Assessment* ***[12]***

Three criterions have to be considered (interoperability cost, interoperability time, and interoperability quality). Those criterions are evaluated during run-time.

The cost: indicates to the cost of updating the application and eliminating the barriers. And it can be (the actual cost and the expected cost).

InteropC= (EC - RC)/EC. (1)

The time: Refers to the difference between the real requested time and when the information has been exchanged.

InteropT=(ET-RT)/ET. (2)

The quality: The quality of the information is considered with the performance measurement. So, two types of quality are conducted:

The quality of exchange: Refers to the data that are successfully posted.

The quality of utilization: Refers to the delivered data in comparison with the demanded.

InteropQ = AE/TE (3)

Conformity: Here, the usability rate of the received information is negotiated.

CE= AC / AR (4)

VII. DATA ANALYSIS & RESULTS

The main aim of regression analysis is to show the significant impact of big data features on interoperability.

The first hypothesis that is tested by the regression analysis is shown below:

1- There is a considerable effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Potentiality

Estimating the Assumption of No Multicollinearity

When two or more independent variables in a regression model have a strong correlation, multicollinearity exists. For revealing the multicollinearity problem, there are a few guidelines

that can be applied. The first is that if the variance inflation factor (VIF) exceeds 10, there is cause for concern. In addition, if tolerance below 0.1 indicates a significant problem and a tolerance below 0.2 indicates a potential problem.
According to Table 2, the VIF for our current model values is well below ten, while the tolerance statistics are above 0.2; as a result, we can reasonably assume that there is no collinearity in the regression data.

Table 2: Multicollinearity Test of Data Interoperability Potentiality Model

Data Interoperability Potentiality		
Variables	**Tolerance**	**VIF**
Value	0.840	1.191
Volume	0.698	1.433
Variety	0.718	1.392
Velocity	0.785	1.273
Veracity	0.590	1.695

When analyzing the first regression model, as in Table 3, using R- Squared (R^2) and ANOVA F, it can be comprehended that the value of R Squared (R^2=0.32), as R^2 points that about 32% of changes in the value of the data interoperability potentiality variable can be attributable to Value, Volume, Variety, Velocity and Veracity. This percentage indicates that the regression model of the current hypothesis has a decent Goodness-of-Fit. The R- Squared finding can be confirmed by looking at the value of ANOVA F, which is (F =3.007), which is considerable at 0.05. Therefore, the current researcher might infer that the regression model of the current hypothesis has a decent Goodness-of-Fit. Based on Table 3, the first hypothesis is not supported, as the Standardized Coefficients (β) of Value, Volume, Velocity and Veracity (β=0.0.322, p>0.05; β=-0.-0.052, p>0.05, β=0.048, p>0.01 and β=0.142, p>0.01) respectively are non-significance. While the Standardized Coefficients (β) of Variety are considerable (β=0.375, p<0.05). That is, only data Variety has a positive effect on the data interoperability potentiality variable.

Table 3: Regression Analysis of Data Interoperability Potentiality

Independent Variable	Dependent Variable: Data Interoperability Potentiality
Value	0.322
Volume	-0.052
Variety	0.375*
Velocity	0.048
Veracity	0.142
R2	0.32
ANOVA F	3.007*
* Standardized Coefficient is significant at the 0.05 level	
** Standardized Coefficient is significant at the 0.01 level, n=38	

2- There is a considerable effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Compatibility

According to Table 4, because the VIF values in our present model are considerably below ten and the tolerance statistics are over 0.2, we can reasonably assume that there is no collinearity in the regression data.

Table 4: Multicollinearity Test of Data Interoperability Compatibility Model

Data Interoperability Compatibility		
Variables	Tolerance	VIF
Value	0.840	1.191
Volume	0.698	1.433
Variety	0.718	1.392
Velocity	0.785	1.273
Veracity	0.590	1.695

To evaluate the second regression model, as demonstrated in Table 5, utilizing R- Squared (R2) and ANOVA F, it can be recognized that the value of R Squared (R2=0.306), as R2 marks that about 31% of the shift in the value of the data interoperability compatibility variable can be attributable to Value, Volume, Variety, Velocity and Veracity. This percentage indicates that the regression model for the current hypothesis has a decent Goodness-of-Fit. The R-squared finding can be backed up by looking at the ANOVA F value (F =2.825), which is considerable at 0.05. Therefore, the current researcher might infer that the regression model of the current hypothesis has a decent Goodness-of-Fit. According to Table 5, the second hypothesis is not entirely supported, as the Standardized Coefficients (β) of Volume, Variety, Velocity and Veracity (β=0.0.133, p>0.05; β=-0.-0.131, p>0.05, β=0.317, p>0.01 and β=-0.033, p>0.01) respectively are non-significance. At the same time, the Standardized Coefficients (β) of the vale are considerable (β=0.353, p<0.05). That is, only data value has a positive effect on the data interoperability compatibility variable.

Table 5: Regression Analysis of Data Interoperability Compatibility

Independent Variable	Dependent Variable: Data Interoperability Compatibility
Value	0.353*
Volume	0.133
Variety	-0.131
Velocity	0.317
Veracity	-0.033
R2	0.306
ANOVA F	2.825*

* Standardized Coefficient is considerable at the 0.05 level
** Standardized Coefficient is considerable at the 0.01 level, n=38

3- There is a considerable effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Performance

According to Table 6, we can reasonably infer no collinearity within the regression data since the VIF values are substantially below ten and the tolerance statistics are over 0.2 for our current model.

Table 6: Multicollinearity Test of Data Interoperability Performance Model

Data Interoperability Performance		
Variables	Tolerance	VIF
Value	0.840	1.191
Volume	0.698	1.433
Variety	0.718	1.392
Velocity	0.785	1.273
Veracity	0.590	1.695

The value of R2(R2=0.454) indicates that about 45% of changes in the value of the data interoperability performance variable can be credited to Value, Volume, Variety, Velocity, and Veracity when evaluating the third regression model, as seen in Table 7, by using R-squared(R2) and ANOVA F. This percentage implies that the model used in the given hypothesis has a good Goodness-of-Fit. By investigating the value of ANOVA F, the R-squared result can be confirmed. The R-squared result can be validated by looking at the ANOVA F value (F =5.305), which is notable at 0.05. As a result, the current researcher might infer that the regression model of the current hypothesis has a decent Goodness-of-Fit. The third hypothesis is not fully supported Based on Table 7, as the Standardized Coefficients (β) of Volume, Variety and Velocity (β=0.032, p>0.05; β=-0.119, p>0.05, β=-0.152, p>0.01) are all non-significance. Simultaneously, the Standardized Coefficients (β) of value and veracity are significant (β=0.444, p<0.01 and β=0.382, p<0.05). Furthermore, data value and veracity have a positive impact on the data interoperability performance characteristic.

Table 7: Regression Analysis of Data Interoperability Performance

Independent Variable	Dependent Variable: Data Interoperability Performance
Value	0.444**
Volume	0.032

Variety	0.119
Velocity	-0.152
Veracity	0.382*
R2	0.454
ANOVA F	5.305*
* Standardized Coefficient is considerable at the 0.05 level	
** Standardized Coefficient is considerable at the 0.01 level, n=38	

VIII. SUMMARY OF HYPOTHESIS

The bellow table showed the summarizing findings of the regression analysis concerning the study hypotheses.

Table 8: Summary of Hypothesis Test

1- There is a considerable effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Potentiality	
Independent Variable	Dependent Variable: Data Interoperability Potentiality
Value	Not supported
Volume	Not supported
Variety	**Supported**
Velocity	Not supported
Veracity	Not supported
2- There is a significant effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Compatibility	
Independent Variable	Dependent Variable: Data Interoperability Compatibility
Value	Supported
Volume	Not supported
Variety	Not supported
Velocity	Not supported
Veracity	Not supported

3- There is a significant effect of Value, Volume, Variety, Velocity and Veracity on the Data Interoperability Performance	
Independent Variable	Dependent Variable: Data Interoperability Performance
Value	Supported
Volume	Not supported
Variety	Not supported
Velocity	Not supported
Veracity	Supported

IX. Conclusion

We can conclude that not all of big data features had a positive correlation between them and data interoperability potentiality, compatibility and performance, as the study proofed that only data value variable has a positive correlation on all of big data interoperability measurements, while it has positive effect on data interoperability compatibility and performance, data variety variable has a positive correlation and a positive effect with only data interoperability potentiality, as for data velocity variable has appositive correlation with only data interoperability compatibility and has no effect on big data interoperability measurements, and data veracity variable has appositive correlation with data interoperability potentiality and performance and has a positive effect only on data interoperability performance, while data volume variable hos no correlation or effect on big data interoperability measurements ,this has a big effect on how the organization could get befits from their big data to develop their work inside and outside their foundation.

from the results of this study, the following conclusion can be conducted:

1- Some Big data features do not correlate with big data interoperability measurements but positively affect them.

2- Some Big data features are correlated with big data interoperability measurements but do not affect them.

References

[1] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.

[2] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of Big Data based on its essential features," *Libr. Rev.*, vol. 65, no. 3, pp. 122–135, 2016, doi: 10.1108/LR-06-2015-0061.

[3] M. Ghasemaghaei and G. Calic, "Assessing the impact of big data on firm innovation performance: Big data is not always better data," *J. Bus. Res.*, vol. 108, no. April 2019, pp.

147–162, 2020, doi: 10.1016/j.jbusres.2019.09.062.
[4] M. Ghasemaghaei, "Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics," *Int. J. Inf. Manage.*, vol. 57, no. August, 2021, doi: 10.1016/j.ijinfomgt.2019.102055.
[5] N. Kshetri, "Big datas impact on privacy, security and consumer welfare," *Telecomm. Policy*, vol. 38, no. 11, pp. 1134–1145, 2014, doi: 10.1016/j.telpol.2014.10.002.
[6] E. Yadegaridehkordi *et al.*, "The impact of big data on firm performance in hotel industry," *Electron. Commer. Res. Appl.*, vol. 40, 2020, doi: 10.1016/j.elerap.2019.100921.
[7] O. Iroju, A. Soriyan, I. Gambo, and J. Olaleke, "Interoperability in Healthcare: Benefits, Challenges and Resolutions," *Int. J. Innov. Appl. Stud.*, vol. 3, no. 1, pp. 262–270, 2013, [Online]. Available: http://www.ijias.issr-journals.org/abstract.php?article=IJIAS-13-090-01.
[8] "SC508DI07171 D05 . 02 Big Data Interoperability Analysis Document Metadata."
[9] R. Rezaei, T. K. Chiew, and S. P. Lee, "A review of interoperability assessment models," *J. Zhejiang Univ. Sci. C*, vol. 14, no. 9, pp. 663–681, 2013, doi: 10.1631/jzus.C1300013.
[10] D. Chen, B. Vallespir, and N. Daclin, "An approach for enterprise interoperability measurement," *CEUR Workshop Proc.*, vol. 341, pp. 1–12, 2008.
[11] G. da Silva Serapião Leal, W. Guédria, and H. Panetto, "Interoperability assessment: A systematic literature review," *Comput. Ind.*, vol. 106, pp. 111–132, 2019, doi: 10.1016/j.compind.2019.01.002.
[12] B. Of, *ICMI-2021*. 2021.