International Journal of Science Annals, Vol. 8, No. 2, 2025
print ISSN: 2617-2682; online ISSN: 2707-3637; DOI:10.26697/ijsa

SOCIAL AND BEHAVIORAL SCIENCES. Economics

ORIGINAL RESEARCH

Data Quality Factors for Big Data Analytics in Occupational Health and Risk Management

Authors' Contribution:
A – Study design;
B – Data collection;
C – Statistical analysis;
D – Data interpretation;
E – Manuscript preparation;
F – Literature search;
G – Funds collection

Lephoto N. [1 ABCDEFG], **Segooa M. A.** [1 ABCDEFG], **Motjolopane I.** [2 ABCDEFG], **Seaba T. R.** [3 ABCDEFG]

[1] *Tshwane University of Technology, South Africa*
[2] *University of Witwatersrand, South Africa*
[3] *Nelson Mandela University, South Africa*

Received: *30.10.2025;* ***Accepted:*** *19.12.2025;* ***Published:*** *25.12.2025*

Abstract

Background and Aim of Study: *Occupational health and risk management (OHRM) in the South African mining sector remains a critical national priority, where the life or death outcomes can be impacted by poor quality-data usage. Big data analytics (BDA) is increasingly used for hazards predictions and timely decision-making.*
The aim of the study: to explore critical data quality factors that influence the reliability and effectiveness of BDA for decision-making to guide occupational health practitioners and risk managers within South African mining sector.

Material and Methods: *The study employed a quantitative survey methodology, informed by the literature review, to identify key data quality factors of BDA impacting OHRM in the South African mining sector. Underpinned by Technological, Organizational and Environmental (TOE) theory and contextual factors within big data quality dimensions and big data sources. Data was collected from 103 OHRM experts determined by the population size of 140.*

Results: *The results reveal the following factors to have influence on data quality for BDA within SA mining OHRM; Environmental factors with a predictive power of 25.0% (β=0.250) at p=0.014; followed by big data quality dimensions with 24.1% (β=0.241) at p=0.008; then, technological factors with 15.9% (β=0.159) at p=0.027; big data sources with 13.2% (β=0.132) at p=0.026; lastly organisational factors was less significant at p=0.228 with 10.0% (β=0.100).*

Conclusions: *This study identifies the factors of data quality, highlighting its role in BDA for decision-making within OHRM. These factors can further be used to provide guidance for SA mining OHRM decision makers to target critical data quality improvement areas for enhanced decision making in the sector.*

Keywords: *big data analytics, data quality, mining safety, occupational health, risk management, South Africa.*

DOI: *https://doi.org/10.26697/ijsa.2025.2.5*

Conflict of interests: *The authors declare that there is no conflict of interests*

Peer review: *Double-blind review*

Source of support: *This research did not receive any outside funding or support*

Information about the authors:

Lephoto Nyakallo (Corresponding Author) *– https://orcid.org/0009-0000-7899-6950; n.lephotoj@gmail.com; Department of Informatics, Tshwane University of Technology, Pretoria, South Africa.*

Segooa Mmatshuene Anna *– https://orcid.org/0000-0002-4190-8256; Doctor of Computing, Senior Lecturer, Department of Informatics, Tshwane University of Technology, Pretoria, South Africa.*

Motjolopane Ignitia *– https://orcid.org/0000-0001-9047-6720; PhD in Information Systems, Associate Professor, Digital Business Wits Business School, University of Witwatersrand, Johannesburg, South Africa.*

Seaba Tshinakaho Relebogile *– https://orcid.org/0000-0002-5773-887X; Doctor of Computing in Informatics, Senior Lecturer, Department of IT Management and Governance, Nelson Mandela University, Gqeberha, South Africa.*

Introduction

The mining sector serves as the pillar of the world's financial resource; however, minimizing risk-related issues and negative environmental effects presents a significant challenge in the industry (Bag et al., 2021). The mining industry leverages various big data sources to prevent occupational hazards, and to ensure a secure working environment (Abd Karim & Sejati, 2021). These big data sources generates vast amount of data, according to Ntlhakana et al. (2021), big data has a radical effect on occupational health and enables the early identification of high-risk patients through the integration of big data source technologies (Brouwer & Rees, 2020). Mining industry professionals make use of this data to inform decision-making processes and mitigate the adverse effects of the occupational health challenges such as occupational hearing loss (Moroe et al., 2019). Failure to address environmental, social and governance challenges may negatively impact the reputation of organisations. Loss of revenue and further increase the risk of none compliance (van Rensburg et al., 2019).Therefore, big data analytics (BDA) and data quality may be of value, as both appear to be drivers of transformation and improvement in the mining industry (Bisschoff & Grobbelaar, 2022).

The use of BDA is expanding with increasing acknowledgement from academia and industry. BDA refers to the systematic examination and analysis of large datasets that exceed traditional analytical capabilities (Hariri et al., 2019), utilizing innovative techniques for data storage, management, analysis, and visualization (Vassakis et al., 2018) of massive and complex datasets, commonly known as Big Data (Kuo et al., 2014). BDA offers potential significant benefits for organisational performance in the mining industry. Furthermore, can enables data driven decision making, which may lead to improvement of organisation's efficiency and profitability (Vassakis et al., 2018). Additionally, applying BDA within Occupational Health and Risk Management (OHRM) in the mining sector may improve effectiveness of the environment using big data-driven innovations beneficial for sustainability (Bag et al., 2021).

Despite noticeable BDA potential on OHRM, data quality remains the main challenge to the accuracy of the outcomes. Poor data quality appears to be disadvantaging organisations to fully benefiting from the value of using BDA (Cai & Zhu, 2015). According to Vassakis et al. (2018) obtaining the insightful outcomes from BDA analysis of accurate and reliable data of is required. As data quality remains essential to leverage accurate and meaningful decision-making, which may influence organizational growth (Segooa & Kalema, 2024) taking into account the conditions of digitalisation in the economy (Pypenko, 2019; Pypenko & Melnyk, 2021).

Data Quality is defined as the degree of data usefulness (Wang et al., 2023), for its intended application and requirements (Declerck et al., 2024). In the realm of big data analytics, data quality is critical for identifying patterns, correlations, and trends within massive amounts of data (Feng et al., 2019), in which impacts the success of the processes that are driven by data, analytics and decision-making systems (Rangineni et al., 2023). According to Bisschoff and Grobbelaar (2022), data quality is critical to obtain accurate insights and protect companies from making poor decisions as a result of poor data quality and includes objectively and correctly describing real situations (Tylečková & Noskievičová, 2020).

High-quality data is critical in the mining sector due to the nature of the environment, which involves managing number of risks, including safety, health, and environmental sustainability. Hence, in South African mining sector, sustainable development entails the investigation for the intersections between the mining companies' goals, their business procedures, and the subsequent effects on the welfare of the community, safety, and health (Bag et al., 2021). Poor data quality leads to inaccurate evaluations of occupational health risks, in which can potentially compromise employees' safety, increase penalties (Mishra & Mishra, 2023), erroneous reporting and noncompliance with various occupational health and safety regulations (Maroun, 2019). In addition, Feng et al. (2022) emphasized the significance of missed organizational learning opportunities within the healthcare field, pointing out concerns related to underreporting, contributing factors, and quality improvement projects. Organizations involved in mining can gain a better understanding of unsafe behaviours and potentially uncover instances of underreporting that impact the accuracy and reliability of data related to occupational health and safety in the mining sector (Kumar & Bhattacharjee, 2023). Moreover, Luo et al. (2023) emphasized how inadequate safety technology training and delayed hazard identification can contribute to underreporting of accidents, affecting quality of data within the OHRM in the mining sector.

The aim of the study. To analyse the factors that influence data quality in big data analytics to improve decision making within the SA mining sector. By exploring these factors, the study is intended to address big data quality challenges as well as their impact in decision making processes for OHRM within SA mining organisations.

Materials and Methods

According to Lim et al. (2013) the information services (IS) theories are considered a foundation of information systems research study, which provides a design and guidance on investigating a phenomena. For the researcher to present unbiased results, the choice of IS theory framework is derived from the topic of the study, research objectives and literature review (Chukwuere, 2021).

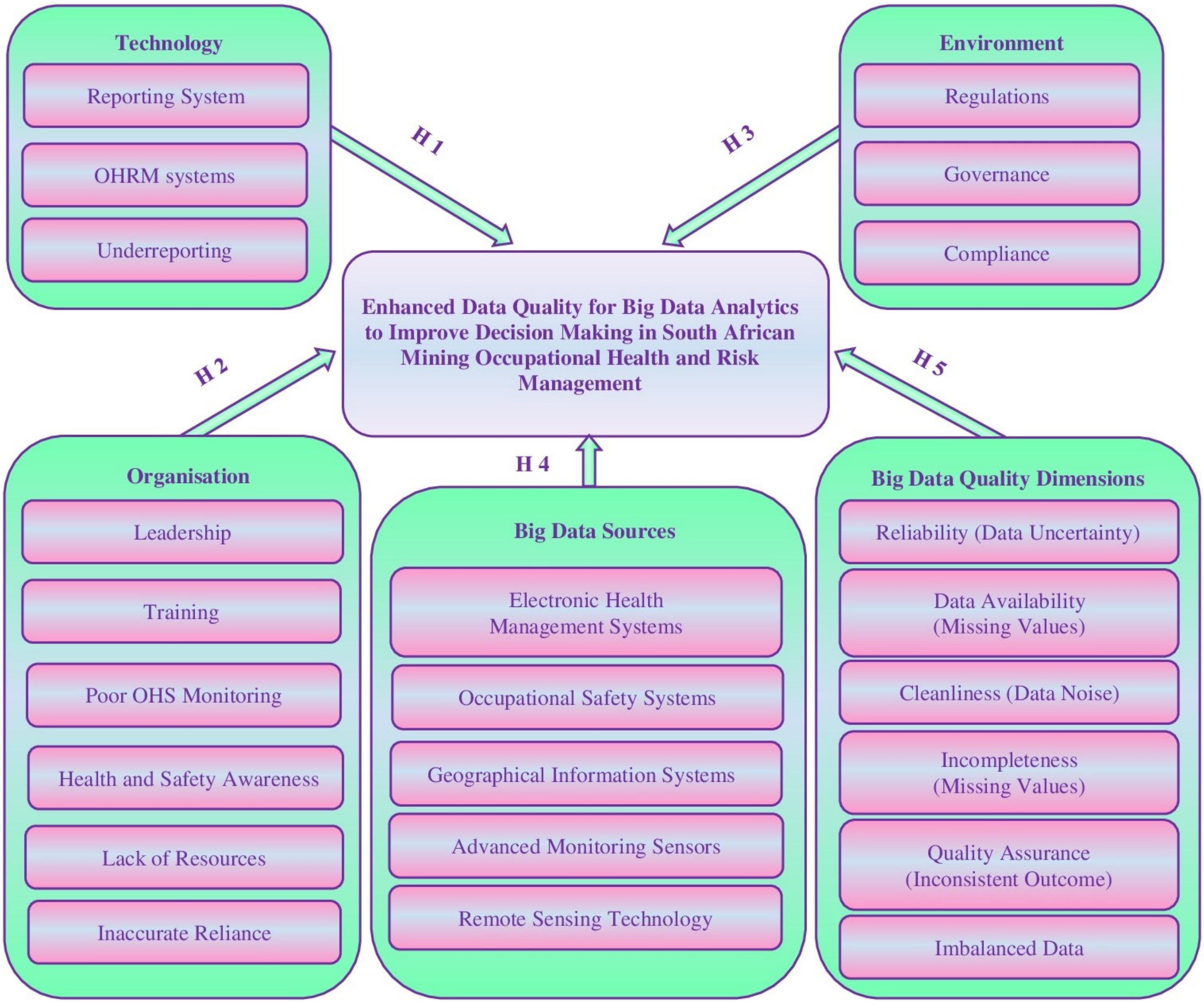
This study integrated TOE IS theory with big data quality dimensions and big data sources, as an underlying theory to expand the existing theoretical body of knowledge, considering the factors identified by the researcher while during the review of the literature.

TOE framework is known for its ability to provide a more comprehensive approach by taking into account technological, organisational and environmental factors (Ullah et al, 2021).

Figure 1 presents conceptual model of enhanced data quality for BDA to improve decision making in SA mining OHRM.

Figure 1

Conceptual Model of Enhanced Data Quality for Big Data Analytics to Improve Decision Making in South African Mining Occupational Health and Risk Management



This study employed quantitative methodology following positivist approach to explore big data quality factors for BDA to improve decision making in the SA mining industry. Positivist approach is usually associated with the quantitative research paradigm, in which the researcher would utilize surveys, questionnaires, or experimental techniques to extract and generalize the results (Kivunja & Kuyini, 2017). The questionnaire was developed and employed as a data collection tool in this study, to discover patterns and factors that influence the quality of data in BDA, for the incidents involving occupational health and safety (OHS), risk hazards and processes for decision making. This study used sampling determinants method by Krejcie and Morgan (1970) to determine the sample size of the study, which guided that the population of 140, requires a sample size of 103. The researcher selected relevant participants and aligned with the field of study and research objective to obtain insights and measure data quality factors identified during the review of the literature in the OHRM within the mining sector. To collect data from the sample size of 103 OHRM participants, the researcher used google forms to create a questionnaire for seamless administration of the responses.

This study considered directly impacted stakeholders from one of the largest gold mine in South Africa as a sample population, specifically selected subject matter experts (SMEs) within the OHRM disciplines such as occupational health, occupational hygiene, safety management, radiation and risk management. As they rely on BDA for decision making and data quality is critical for their prediction. The selection criteria is presented on Table 1.

Table 1
Selection Criteria

Stream	Sample size for Questionnaire
Occupational Safety	21
Occupational Hygiene	11
Radiation	10
Occupational Health	41
Risk Management	20
Total	103

According to Albers (2017), in order to reach a conclusion in a quantitative research study, a numerical data must be gathered and analysed. Data analysis reveals the linkage of the study's contextual setting, main trends and patterns. In this study statistical tests and tools such as Statistical Package for the Social Sciences (SPSS) version 28.0.0.0 from IBM was used for data analysis to obtain conclusions from the collected data. According to Bauer et al. (2021) most of the quantitative studies consists of the basic statistical analytic methods, such as correlation regressions, descriptive statistics, and analysis with or without probabilities, measurements of statistical significance and interactions. The results overall reliability was conducted for the study as presented in Table 2.

Table 2
Overall Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
0.966	0.966	86

The overall reliability statistics based on the Cronbach's alpha coefficient was 0.966 measured on 86 items. This value is acceptable (Taber, 2017) as it is above the minimum value of 0.7.

Results

The study considered South African Mining Occupational Health and Risk Management experts, which included the total responses of 104 participants. Thus, 1 chief safety officer, 9 group technology (GT) systems specialists, 15 occupational health managers, 22 occupational health nursing practitioners, 1 occupational hygiene manager, 12 occupational hygienists, 13 occupational medical practitioners, 1 radiation manager, 8 radiation protection officers, 20 risk management specialists and 2 safety officers as shown in Table 3.

Table 3
Frequencies of Participants' Demographics

Variables		Frequency		
		Person	Percent	Cumulative percent
Gender	Female	56	53.8	53.8
	Male	48	46.2	100.0
	Total	104	100.0	–
Age group	21-30 years	2	1.9	1.9
	21-40 years	28	26.9	28.8
	41-50 years	42	40.4	69.9
	51 years and above	32	30.8	100.0
	Total	104	100.0	–
Education	Matric	10	9.6	9.6
	National Diploma	32	30.8	40.4
	Bachelor Degree	55	52.9	93.3
	Master Degree	6	5.8	99.1
	Doctoral Degree (PhD)	1	1.0	100.0
	Total	104	100.0	–
Location	Free State	32	30.8	30.8
	Gauteng	49	47.1	77.9
	North West	22	21.2	99.0
	South Africa	1	1.0	100.0
	Total	104	100.0	–
Position	Chief Safety Officer	1	1.0	1.0
	GT Systems Specialist (SHERQ/HRM)	9	8.7	9.6
	Occupational Health Manager	15	14.4	24.0
	Occupational Health Nursing Practitioner	22	21.2	45.2
	Occupational Hygiene Manager	1	1.0	46.2
	Occupational Hygienist	12	11.5	57.7
	Occupational Medical Practitioner	13	12.5	70.2
	Radiation Manager	1	1.0	71.2
	Radiation Protection Officer	8	7.7	78.8
	Risk Management Specialist	20	19.2	98.1
	Safety Officer	2	1.9	100.0
	Total	104	100.0	–
BDA Utility	Yes	5	4.8	4.8
	No	99	95.2	100.0
	Total	104	100.0	–

Furthermore, Table 3 presents that OHRM discipline consists more of employees above 31 years of age than 21-30 years of age; this result is valid as mining industry retains its employees due to level of experience mostly in occupational safety and risk. Moreover, the table shows that 9.6% of participants had matric certificates as their highest qualification, 30.8% had national diploma, 52.9% had Bachelor's degree, 5.8% had Master's degree and 1.0% of the participants had PhD, the findings indicates that most participants hold Bachelor's degree with 55.0%. Therefore, this study is valid as OHRM specialists are required to have attended a formal training and education. Table 3 further demonstrates locations, and only 3 provinces out of 9 in South Africa, and South Africa as country, the assumption is that participant might be working in multiple provinces, according to the results Gauteng had the highest responses at 47.1%, followed by Free State with 30.8% and the lowest being North West with 22.1%. Therefore, this study is valid as the sampled mining organisation only operates in 3 provinces in South Africa, which is Free State, Gauteng and North West. On BDA utilization, Table 3 indicates demonstrates that only 5 participants of the total of 104 participants which is 4.8% are not using big data analytics tools in their daily duties. As a result, 95.2% of the participants utilize big data analytics tools for decision-making.

Regression Statistical Analysis

This study considered regression statistical analysis to determine the relationship between enhancing data quality for BDA analytics as an independent variable and number of dependent variables thus, Technological, Environmental, Organisational, Big Data Quality dimensions and Big Data Sources. Linear regression statistical analysis is an analytical method used to determine the influence that an independent variables has on the dependent variable (Wardhani et al., 2021). The results of the statistical analysis are presented in Table 4.

Table 4
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	0.779[a]	0.608	0.587	0.410	0.608	30.337	5	98	0.000

Note. a. Predictors: (Constant), Big Data Sources, Technological, Organisational, Data Quality and Environmental; b. Dependent Variable: Enhancing Data Quality for BDA.

According to Table 4, the correlation value of R Square is 60.8% (0.608), which indicates the contribution between the individual variables towards dependent variable "Enhancing data quality in BDA for effective decision-making". While the correlation value of 0.779 indicates, the overall contribution of individual independent factors towards the conceptual model to enhancing data quality in BDA for improved decision-making in South African Mining Sector OHRM is 77.9%.

Furthermore, the sig. F change value of 0.00, which is below 0.05, indicates that the prediction of the identified big data quality factors for BDA is significant and can be considered to improve decision making in SA mining sector. The regression coefficients are shown in Table 5.

Table 5
Regression Coefficients

Model	Unstandardized coefficients		Standardized coefficients	*t*	Sig.	Collinearity statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	0.413	0.286	–	1.441	0.153	–	–
TechFactor	0.159	0.070	0.171	2.252	0.027	0.692	1.445
OrgFactor	0.100	0.093	0.094	1.068	0.288	0.520	1.923
EnvFactor	0.250	0.101	0.278	2.491	0.014	0.322	3.109
DataQualityFac01	0.241	0.089	0.270	2.708	0.008	0.402	2.488
BigDataSourcesfac01	0.132	0.058	0.170	2.266	0.026	0.714	1.400

Note. TechFactor – technological factors; OrgFactor – organisational factors; EnvFactor – environmental factors; DataQualityFac01 – big data quality dimensions; BigDataSourcesfac01 – big data sources.

Based on the regression coefficients on Table 5, the findings reveal that the factors that influence data quality for BDA within SA mining OHRM are; Environmental factors with a predictive power of 25.0% (β=0.250) at p=0.014 which is the most influential; followed by big data quality dimensions with 24.1% (β=0.241) at p=0.008; then, technological factors with a predictive power of 15.9% (β=0.159) with significance level of

p=0.027; big data sources with 13.2% (β=0.132) at significance level of p=0.026; lastly organisational factors was found to be less significant at p=0.228 with predictive power of 10.0% (β=0.100).
Table 6 presents the results for the tested set hypotheses of the study.

The results indicates that only four of the hypotheses were supported after quantitative data analysis: H1, H3, H4 and H5. While one hypothesis, which is H2 – organisational factors, was rejected.

Table 6
Hypotheses Results

Hypothesis	Sig.	Results
H1 – Technologies, such as OHRM systems and Underreporting have influence on enhanced data quality for BDA to improve decision-making in South African mining OHRM	P=0.027<0.05	Accepted
H2 – Organisational factors, which includes leadership, lack of training and resources, lack of Health and Safety awareness, poor OHS Monitoring, inaccurate Reliance influence enhanced data quality for BDA for effective decision-making in South African mining OHRM	P=0.288>0.05	Rejected
H3 – Environment factors, which includes external governance, legislation and compliance influence the enhancement of data quality for BDA to improve decision-making in South African mining OHRM	P=0.014<0.05	Accepted
H4 – Big data sources, such as Occupational safety systems EHMS, GIS, advanced monitoring sensors, remote sensing technology, have influence on enhancing data quality for BDA within South African mining OHRM to improve decision-making	P=0.026<0.05	Accepted
H5 – Big data quality dimensions, which consist of data availability, cleanliness, imbalanced data, reliability, incompleteness and quality assurance have influence on data quality enhancement for BDA to improve decision-making in the South African mining OHRM to improve decision-making	P=0.008<0.05	Accepted

Discussion

The aim of this study was to identify critical factors of data quality in BDA to improve decision making within OHRM for mining sector. In this section, the researcher discusses the key data quality factors identified during literature review, which informed hypotheses, and further tested in this study.

Big Data Quality Technological Factors

Table 6 shows that H1 (P=0.027<0.05) was accepted, suggesting that Technological factors such as reporting systems, OHRM systems and underreporting have significant influence on enhancing data quality for BDA within South African mining OHRM to improve decision-making. These results are supported by the study conducted by Famure et al. (2019) that Electronic Health Record (EHR) systems have contributed to the emergence of BDA in healthcare by offering chances for quality improvements, which are crucial components for enhancing data quality in occupational health and safety. Consistently, the study conducted by Yang et al. (2021) underscores the importance of robust reporting systems and information technology in identifying causes of safety issues and accidents within the coal mine industry, emphasizing the role of technological advancements in enhancing safety practices and data quality within the industry. Moreover, the study by Zhou et al. (2018) further supports the outcomes highlighting the critical importance of robust OHS systems and risk management within the mining sector. Additionally, the research highlights the significance of OHS management practices in fostering organizational safety culture, risk management, and incident prevention, by managing risks and implementing safety measures effectively, mining organizations can enhance data quality, reduce negative occurrences, and cultivate a safe working environment for employees (Stojanović et al., 2024).

Big Data Quality Organisational Factors

Table 6 shows H2 (P=0.288>0.05), was rejected which indicates that organisational factors such as leadership, lack of training, resources, lack of awareness within Health and Safety, poor monitoring of OHS, and inaccurate Reliance do not significantly have influence on data quality enhancement within BDA for effective decision-making in South African mining OHRM. According to Sarstedt and Mooi (2018), the overall parameter that it is greater than 0.05 is considered to be not significant.

These organisational factors were identified in accordance to the literature conducted by Johnson et al. (2021) who revealed that data quality improvement is a top management function through an empirical investigation of BDA capabilities implementation. Whilst, Haas (2020) established that leadership has a critical role in shaping safety culture and impacting health and risk management processes at the operational level, on the study highlighting the need for developing effective decision-making models in occupational health and safety. Moreover, research by Hermanus (2007) identified resource limitations in small mining companies contribute significantly to health and risk management concerns.

Furthermore, Franke and Hiebl (2022) acknowledged the need for skilled data analytics resources to effectively examine big data and derive meaningful insights for informed decision-making in mining. In support, Alnefaie et al. (2022) identified the vital role played by data specialists play in processing big data to facilitate decision support, and identifying data sources and required competencies can significantly influence data quality in mining. In addition, According to Nazari et al. (2020), training and knowledge development are essential to overcome BDA challenges and leverage its benefits effectively. Similarly, Muhunzi et al. (2023) found that training healthcare professionals to leverage BDA effectively may improve patient outcomes and reduce healthcare costs. Moreover, Andrews et al. (2019) supported that stuff training along with data quality initiatives are critical for improving healthcare delivery processes, and for accurate process mining outcomes.

Big Data Quality Environmental Factors

The study accepted H3 (P=0.014<0.05) – environmental factors, which include external governance, legislation and compliance are significantly influencing the enhancement of data quality for BDA within the South African mining OHRM to improve decision-making, as shown in Table 6. These results are supported by Muthelo et al. (2022), who focused on investigating occupational health and safety practices and compliance within South African mining sector, specifically in the province of Limpopo, utilizing principal component analysis. By identifying key attributes associated with compliance with health and safety standards, this study indirectly underscores the importance of regulatory adherence in upholding data quality within the OHS context of the mining sector (Muthelo et al., 2022). Moreover, Donkor et al. (2023) further emphasized the significance of complying with safety regulations to mitigate risks and safeguard workers' well-being, which can ultimately impact data quality by ensuring precise reporting and monitoring of occupational health and safety metrics. Moreover, Chikosi and Mutezo (2023) identified that occupational health and safety risks are frequently known challenges within the mining industry, which includes the inefficient organisational governance systems. In addition, it is important to implement effective data governance to manage and control data use, enhancing data quality, availability, and integrity within organizations (Aseeri & Kang, 2022). South African mining sector is a very well regulated and governed entity more especially within the areas of occupational health and risk management. According to Rikhotso et al. (2022), each regulatory compliance is associated with the cost, which corresponds to the requirements such as medical examination, risk assessment and reassessment, workplace inspections, training programs, workplace control, PPEs and labelling, disposal, offenses and penalties, and keeping records.

Big Data Quality Sources

Table 6 it shows that H4 (P=0.026<0.05) was accepted which indicates that big data sources such as occupational safety systems EHMS, GIS, advanced monitoring sensors, remote sensing technology, have influence on data quality enhancement in BDA within the South African mining OHRM for effective decision-making. These results are consistent with the study conducted by Abd Karim and Sejati (2021), indicating that the mining industry leverages various big data sources such as OHRM systems to prevent occupational hazards, and to ensure a secure working environment (Andri Estining Sejati, 2021). Furthermore, the study conducted by Montisci et al. (2022) identified the variety of big data sources, such as systems for injury-reported incidents, clinical examinations, and electronic health records. Moreover, mining industries integrated multiple big data sources such as remote sensing technologies, geographical information systems (GIS) and machine learning to enhance safety and risk management decision-making (Musiałek & Maksymowicz, 2024; Li et al., 2021). Additionally, according to Ntlhakana et al. (2021) mining industries are using electronic health management systems to maintain employee's records and to proactively monitor occupational health diseases, which includes hearing loss and respiratory conditions in the mining environment.

Big Data Quality Dimensions

As presented in Table 6, this study accepted the significance of H5 (P=0.008<0.05) – big data quality dimensions, which consist of data availability, cleanliness, imbalanced data, reliability, incompleteness and quality assurance have influence in enhancing data quality for BDA in the South African mining OHRM to improve decision-making. This outcome was supported by Arikekpar and Bestman (2023), who identified accuracy, completeness and timeliness as main components of data quality. In addition, Abburi (2024) identified consistency and accessibility as key dimensions to ensure that data is fit for purpose. Furthermore, findings by Cresswell et al. (2024) further identified features such relevance and reliability as relatively defined with major data quality components such accuracy, timeliness and representativeness. According to Luo et al. (2023) there are persistent data availability issues impacting the implementation of appropriate risk management strategies for effective decision-making within responsible customs departments guided by risk assessment outcomes. In addition, Hermanus (2007) identified the reliability issue in occupational health data as a challenge where there is a lack of reporting systems and criteria that are well-established such as within developing countries which includes South Africa (Gheorghe et al., 2022) further supported the outcomes through comparison of the inconsistent number of loss-of-life cases and incidents as evidence in the assessment of data quality for underreporting within occupational health and safety.

Conclusions

This paper has presented and explored the critical data quality factors that impact decision making within the mining OHRM. The identified factors been technologial, environmental, big data quality dimensions and big data sources. These findings suggest that SA mining industry is well regulated by environmental factors such as external governance and compliance. Therefore, there is a full reliance on big data sources to capture data and support effective decision-making within OHRM, despite persistent data quality challenges. Furthermore, this study imparts big data quality dimensions and sources as crucial factors in BDA for effective decision-making in South African mining sector, Occupational Health and Risk Management. This study identifies that technological factors that hinder high quality data usage in the South African mining sector includes reporting systems, OHRM Systems and unnderreporting that is caused by lack of integrations within the systems to be influencing data quality for BDA. However, there is a need to further analyse this factors individually following a qualitative method to gather indepht insights and investigate the level of signficance for the organisational factors, as it was found not sufficient and that led the hypothesis H2 to be rejected.

Acknowledgments

The authors would like to acknowledge the sampled organization and the study participants for their contribution in this study.

Ethical Approval

The study obtained ethical clearance from the institution's Ethics Committee under Ref. No. HREC2024=08=001 (ICT).

Funding Source

The present study was funded by Tshwane University of Technology.

References

Abburi, C. K. (2024). Optimizing big data quality management for national-scale projects: Strategies and frameworks. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(4), 4811–4815. https://doi.org/10.2139/ssrn.5118713

Abd Karim, A. T., & Sejati, A. E. (2021). Evaluation of mining safety management system implementation in PT. ANTAM UBPN Sultra. *Jurnal Ekonomi, 26*(2), 223–238. https://doi.org/10.24912/je.v26i2.747

Albers, M. J. (2017). Quantitative data analysis – In the graduate curriculum. *Journal of Technical Writing and Communication, 47*(2), 215–233. https://doi.org/10.1177/0047281617692067

Alnefaie, A., Albogami, S., Asiri, Y., Ahmad, T., Alotaibi, S. S., Al-Sanea, M. M., & Althobaiti, H. (2022). Chimeric antigen receptor T-cells: An overview of concepts, applications, limitations, and proposed solutions. *Frontiers in Bioengineering and Biotechnology, 10*, Article 797440. https://doi.org/10.3389/fbioe.2022.797440

Andrews, R., Wynn, M. T., Vallmuur, K., ter Hofstede, A. H. M., Bosley, E., Elcock, M., & Rashford, S. (2019). Leveraging data quality to better prepare for process mining: An approach illustrated through analysing road trauma pre-hospital retrieval and transport processes in Queensland. *International Journal of Environmental Research and Public Health, 16*(7), Article 1138. https://doi.org/10.3390/ijerph16071138

Arikekpar, F., & Bestman, A. E. (2023). Data quality management and responsiveness of indigenous oil and gas companies in rivers state, Nigeria. *Journal of Office and Information Management (JOIM), 7*(1-2), 301–313. https://www.rsu-joim.org/journal2.html

Aseeri, M., & Kang, K. (2022). Big data-oriented organizational culture and business performance: A socio-technical approach. *Problems and Perspectives in Management, 20*(4), 52–66. https://doi.org/10.21511/ppm.20(4).2022.05

Bag, S., Srivastava, G., Gupta, S., & Taiga, S. (2021). Diffusion of big data analytics innovation in managing natural resources in the African mining industry. *Journal of Global Information Management, 30*(6), 1–21. https://doi.org/10.4018/JGIM.297074

Bauer, G. R., Churchill, S. M., Mahendran, M., Walwyn, C., Lizotte, D., & Villa-Rueda, A. A. (2021). Intersectionality in quantitative research: A systematic review of its emergence and applications of theory and methods. *SSM – Population Health, 14*, Article 100798. https://doi.org/10.1016/j.ssmph.2021.100798

Bisschoff, R., & Grobbelaar, S. (2022). Evaluation of data-driven decision-making implementation in the mining industry. *South African Journal of Industrial Engineering, 33*(3), 218–232. https://doi.org/10.7166/33-3-2799

Brouwer, D. H., & Rees, D. (2020). Can the South African milestones for reducing exposure to respirable crystalline silica and silicosis be achieved and reliably monitored? *Frontiers in Public Health, 8*, Article 107. https://doi.org/10.3389/fpubh.2020.00107

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal, 14*, Article 2. https://doi.org/10.5334/dsj-2015-002

Chikosi, L. C., & Mutezo, A. T. (2023). Managerial ownership, board gender diversity, occupational health, and safety risk management in an emerging economy. *Journal of Governance and Regulation, 12*(1), 230–241. https://doi.org/10.22495/jgrv12i1siart3

Chukwuere, J. E. (2021). Theoretical and conceptual framework: A critical part of information systems research process and writing. *Review of International Geographical Education Online, 11*(9), 2678–2683. https://doi.org/10.48047/rigeo.11.09.234

Cresswell, K., Claire, R., Avsar, T. S., Hudson, T., Purcell, R., Peeters, R., … Colvin, H. (2024). PD137 reviewing the health technology assessment and regulatory policy landscape on acceptability standards for real-world evidence - Initial findings. *International Journal of Technology Assessment in Health Care, 40*(S1), S147–S147. https://doi.org/10.1017/S026646232400374X

Declerck, J., Kalra, D., Vander Stichele, R., & Coorevits, P. (2024). Frameworks, dimensions, definitions of aspects, and assessment methods for the appraisal of quality of health data for secondary use: Comprehensive overview of reviews. *JMIR Medical Informatics, 12*, Article e51560. https://doi.org/10.2196/51560

Donkor, P., Siabi, E. K., Frimpong, K., Mensah, S. K., Siabi, E. S., & Vuu, C. (2023). Socio-demographic effects on role assignment and associated occupational health and safety issues in artisanal and small-scale gold mining in Ghana. *Heliyon, 9*(1), Article e13741. https://doi.org/10.1016/j.heliyon.2023.e13741

Famure, O., Anderson, B. K., Atienza, J., Lena, E. R., & Singh, S. K. (2019). Standardization and alignment of data capture practices to clinical processes in the evaluation of living kidney donor candidates. *Healthcare Management Forum, 32*(4), 202–207. https://doi.org/10.1177/0840470419843672

Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y., & Liu, Q. (2019). Big data analytics and mining for effective visualization and trends forecasting of crime data. *IEEE Access, 7*, 106111–106123. https://doi.org/10.1109/ACCESS.2019.2930410

Feng, T., Zhang, X., Tan, L., Su, Y., & Liu, H. (2022). Near-miss organizational learning in nursing within a tertiary hospital: A mixed methods study. *BMC Nursing, 21*(1), Article 315. https://doi.org/10.1186/s12912-022-01071-1

Franke, F., & Hiebl, M. R. W. (2022). Big data and decision quality: The role of management accountants' data analytics skills. *International Journal of Accounting & Information Management, 31*(1), 93–127. https://doi.org/10.1108/IJAIM-12-2021-0246

Gheorghe, G. C., Manrique-Hernandez, E. F., & Idrovo, A. J. (2022). Injuries and fatalities in Colombian mining emergencies (2005–2018): A retrospective ecological study. *Revista Brasileira de Medicina do Trabalho, 20*(4), 591–598. https://doi.org/10.47626/1679-4435-2022-799

Haas, E. J. (2020). The role of supervisory support on workers' health and safety performance. *Health Communication, 35*(3), 364–374. https://doi.org/10.1080/10410236.2018.1563033

Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: Survey, opportunities, and challenges. *Journal of Big Data, 6*(1), 1-16. https://doi.org/10.1186/s40537-019-0206-3

Hermanus, M. (2007). Occupational health and safety in mining – Status, new developments, and concerns. *Journal of the Southern African Institute of Mining and Metallurgy, 107*(8), 531–538. https://hdl.handle.net/10520/AJA0038223X_3264

Johnson, D. S., Sihi, D., & Muzellec, L. (2021). Implementing big data analytics in marketing departments: Mixing organic and administered approaches to increase data-driven decision making. *Informatics, 8*(4), 66. https://doi.org/10.3390/informatics8040066

Kivunja, C., & Kuyini, A. B. (2017). Understanding and applying research paradigms in educational contexts. *International Journal of Higher Education, 6*(5), 26–41. https://doi.org/10.5430/ijhe.v6n5p26

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*(3), 607–610. https://doi.org/10.1177/001316447003000308

Kumar, D., & Bhattacharjee, R. M. (2023). Reducing workplace unsafe behaviour using risk classification, profiling, risk tolerance approach. *Heliyon, 9*(3), Article e13969. https://doi.org/10.1016/j.heliyon.2023.e13969

Kuo, M. H., Sahama, T., Kushniruk, A. W., Borycki, E. M., & Grunwell, D. K. (2014). Health big data analytics: Current perspectives, challenges and potential solutions. *International Journal of Big Data Intelligence, 1*(1), 114–126. https://doi.org/10.1504/IJBDI.2014.063835

Li, H., Huang, W., Zha, Z., & Yang, J. (2021). Application and platform design of geospatial big data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B4-2021, 293–300. https://doi.org/10.5194/isprs-archives-XLIII-B4-2021-293-2021

Lim, S., Saldanha, T. J. V., Malladi, S., & Melville, N. P. (2013). Theories used in information systems research: Insights from complex network analysis. *Journal of Information Technology Theory and Application, 14*(2), 5–46. https://aisel.aisnet.org/jitta/vol14/iss2/2

Luo, N., Yu, H., You, Z., Li, Y., Zhou, T., Jiao, Y., Han, N., Liu, C., Jiang, Z., & Qiao, S. (2023). Fuzzy logic- and neural network-based risk assessment model for import and export enterprises: A review. *Journal of Data Science and Intelligent Systems*, *1*(1), *2-11.* https://doi.org/10.47852/bonviewJDSIS32021078

Maroun, W. (2019). Exploring the rationale for integrated report assurance. *Accounting, Auditing & Accountability Journal, 32*(6), 1826–1854. https://doi.org/10.1108/AAAJ-04-2018-3463

Mishra, P. C., & Mishra, P. K. (2023). Challenges and opportunities of big data analytics for human resource management in mining and metal industries. *Journal of Mines, Metals and Fuels, 71*(10), 1747–1753. https://doi.org/10.18311/jmmf/2023/35858

Montisci, A., Palmieri, V., Vietri, M. T., Sala, S., Maiello, C., Donatelli, F., & Napoli, C. (2022). Big data in cardiac surgery: Real world and perspectives. *Journal of Cardiothoracic Surgery,* *17*(1), Article 277. https://doi.org/10.1186/s13019-022-02025-z

Moroe, N., Khoza-Shangase, K., Madahana, M., & Nyandoro, O. (2019). A proposed preliminary model for monitoring hearing conservation programmes in the mining sector in South Africa. *Journal of the Southern African Institute of Mining and Metallurgy, 119*(7), 671-679. https://doi.org/10.17159/2411-9717/18/016/2019

Muhunzi, D., Kitambala, L., & Mashauri, H. (2023). Big data analytics in the healthcare sector: Opportunities and challenges in developing countries: A literature review. *Health Informatics Journal, 30*(4). https://doi.org/10.1177/14604582241294217

Musiałek, M., & Maksymowicz, M. (2024). Remote sensing in mining: A brief overview and examples. *Górnictwo Odkrywkowe – Opencast Mining, 65*(3), 4–14. https://doi.org/10.5604/01.3001.0055.0491

Muthelo, L., Mothiba, T. M., Malema, N. R., Mbombi, M. O., & Mphekgwana, P. M. (2022). Exploring occupational health and safety standards compliance in the South African mining industry, Limpopo Province, using principal component analysis. *International Journal of Environmental Research and Public Health, 19*(16), Article 10241. https://doi.org/10.3390/ijerph191610241

Nazari, E., Ebnehoseini, Z., Agharezaei, Z., & Tabesh, H. (2020). Knowledge, attitude, and challenges of big data analytics based on IT staffs point of view in a developing country. *Frontiers in Health Informatics, 9*(1), Article 36. https://doi.org/10.30699/fhi.v9i1.225

Ntlhakana, L., Nelson, G., Khoza-Shangase, K., & Dorkin, E. (2021). Occupational hearing loss for platinum miners in South Africa: A case study of data sharing practices and ethical challenges in the mining industry. *International Journal of Environmental Research and Public Health, 19*(1), Article 1. https://doi.org/10.3390/ijerph19010001

Pypenko, I. S. (2019). Digital product: The essence of the concept and scopes. *International Journal of Education and Science, 2*(4), 56. https://doi.org/10.26697/ijes.2019.4.41

Pypenko, I. S., & Melnyk, Yu. B. (2021). Principles of digitalisation of the state economy. *International Journal of Education and Science, 4*(1), 42-50. https://doi.org/10.26697/ijes.2021.1.5

Rangineni, S., Bhanushali, A., Suryadevara, M., Venkata, S., & Peddireddy, K. (2023). A review on enhancing data quality for optimal data analytics performance. *International Journal of Computer Sciences and Engineering, 11*(10), 51–58. https://doi.org/10.26438/ijcse/v11i10.5158

Rikhotso, O., Morodi, T. J., & Masekameni, D. M. (2022). Health risk management cost items imposed by occupational health and safety regulations: A South African perspective. *Safety Science, 150*(2), Article 105707. https://doi.org/10.1016/j.ssci.2022.105707

Sarstedt, M., & Mooi, E. (2018). *A concise guide to market research: The process, data, and methods using IBM SPSS Statistics*. Springer. https://doi.org/10.1007/978-3-642-12541-6

Segooa, A. M., & Kalema, B. M. (2024). Big data analytics artefact for outcome-based funding prediction in South African public universities. *International Journal of Service Science, Management, Engineering, and Technology, 15*(1), 1–27. https://doi.org/10.4018/IJSSMET.334220

Stojanović, A., Milošević, I., & Nikolić, Đ. (2024). Developing a novel quantitative approach to evaluate the organizational factors affecting occupational health and safety in the mining industry. In *XIX International May Conference on Strategic Management – IMCSM24 Proceedings*, *20*(1), 60–68. https://doi.org/10.5937/IMCSM24006S

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tylečková, E., & Noskievičová, D. (2020). The role of big data in Industry 4.0 in the mining industry in Serbia. *System Safety: Human – Technical Facility – Environment, 2*(1), 166–173. https://doi.org/10.2478/czoto-2020-0020

Ullah, F., Qayyum, S., Thaheem, M. J., Al-Turjman, F., & Sepasgozar, S. M. E. (2021). Risk management in sustainable smart cities governance: A TOE framework. *Technological Forecasting and Social Change, 167*, Article 120743. https://doi.org/10.1016/j.techfore.2021.120743

Van Rensburg, H. M. M. J., Gous, A. G. S., Vosloo, J. C., & Van Heerden, M. (2019). Improving data management for environmental reporting in the gold mining industry. *South African Journal of Industrial Engineering, 30*(3), 163–173. https://doi.org/10.7166/30-3-2236

Vassakis, K., Petrakis, E., & Kopanakis, I. (2018). Big data analytics: Applications, prospects and challenges. In G. Skourletopoulos, G. Mastorakis, C. Mavromoustakis, C. Dobre, & E. Pallis (Eds.), *Mobile Big Data*: *Lecture Notes on Data Engineering and Communications Technologies, 10* (pp. 3–20). Springer. https://doi.org/10.1007/978-3-319-67925-9_1

Wang, J., Liu, Y., Li, P., Lin, Z., Sindakis, S., & Aggarwal, S. (2023). Overview of data quality: Examining the dimensions, antecedents, and impacts of data quality. *Journal of the Knowledge Economy, 15*, 1159–1178. https://doi.org/10.1007/s13132-022-01096-6

Wardhani, M. K., Setioko, B., & Pandelaki, E. E. (2021). Village spatial planning based on its potential as guidelines for guided and sustainable village development. *Tataloka, 23*(2), 171–182. https://doi.org/10.14710/tataloka.23.2.171-182

Yang, L., Birhane, G. E., Zhu, J., & Geng, J. (2021). Mining employees' safety and the application of information technology in coal mining: Review. *Frontiers in Public Health, 9*, Article 709987. https://doi.org/10.3389/fpubh.2021.709987

Zhou, L. J., Cao, Q. G., Yu, K., Wang, L. L., & Wang, H. B. (2018). Research on occupational safety, health management and risk control technology in coal mines. *International Journal of Environmental Research and Public Health, 15*(5), Article 868. https://doi.org/10.3390/ijerph15050868

Cite this article as:

Lephoto, N., Segooa, M. A., Motjolopane, I., & Seaba, T. R. (2025). Data quality factors for big data analytics in occupational health and risk management. *International Journal of Science Annals, 8*(2), 36–46. https://doi.org/10.26697/ijsa.2025.2.5

The electronic version of this article is complete. It can be found online in the IJSA Archive https://ijsa.culturehealth.org/en/arhiv