

Student Name: Shaun James Davies

Student ID Number: 2317135

Programme: Digital Degree Apprenticeship (DDA) Computing -
Software Development

Module: Industrial Based Research

Module Code: ACCA6029

Assignment Title: Industrial Research proposal.

Submission Deadline: 04/09/2026

Table of Contents

Table of Contents.....	2
A Background Rationale	3
Industrial Context.....	3
Literature Review	3
Historical development.....	3
Incident Reporting	3
Competency & Training.....	4
Predictive Analytics & Machine Learning	4
Synthesis & Conclusion	5
The Approach, Techniques, & Methods.....	5
Aim & Objectives	5
Methods & Materials	5
Project Design	6
Work Plan & Tasks.....	6
Resources Required	6
Potential Impact	7
Impact.....	7
Limitations.....	7
Personal Reflection.....	7
Conclusion.....	7
References.....	8
Appendix I	10
Supporting Information	10
Problem Description	10
Problem Consequences	10
Project Need.....	10
Integrated Data & Data Quality.....	10
Project Success Criteria	10
Minimum Viable Product Success Criteria	11
Continued Appendix	11

A Background Rationale

Industrial Context

This project will support Safety, Health, and Environment (SHE) management (SHEM) within the steel manufacturing industry. Producing and manufacturing steel involves risks, potential and actual harm involving the individual staff, communities, environment, and reputation. Incorrect management of SHE matters can provoke interventions from various regulators which can lead to negative reports that affect legal, financial, and reputational status of the company. These risk factors are managed through management systems, standards, procedures, and digital technologies designed to identify, assess and control risk whilst ensuring legal, regulatory, and business compliance whilst progressing with continual improvement. Analysis and reporting of trends can produce essential insight that improves SHEM.

This project will focus on the digital technology in supporting SHEM in two distinctive reporting areas of incident reporting and competency reporting as described in Appendix I Table 1. SHE data management is scattered across many systems within the Tata Steel global ecosystem. There is no centrally owned intelligent reporting platform for all SHE data. For example, incident and competency data both exist in isolation.

The intended stakeholders and users of this project are representatives who manage incident and competency with SHE. This includes directors, senior leaders, managers, and individual staff. Factors that restrict and confine this project are restricted access to systems and its data, inconsistent and poor quality data, cybersecurity and data protection, low user adoption, legal, regulatory and industry standard requirements, fixed project deadlines.

Literature Review

Historical development

Earlier SHEM, as discussed by Duijm [1], within industry was strongly influenced by legal compliance pressures which set minimum standards related to SHE. These requirements did not inherently drive improvement over time, encouraging transition towards performance based management and continuous improvement “i.e. improving these conditions over time” [1]. Continuing, Duijm describes this change related to continuous improvement as a change from “being prescriptive towards being performance based.” They argue that static measures and optimisation do not satisfy elements of continuous improvement, supporting the transition from a focus on compliance to performance SHEM [1]. Subsequently, SHEM became increasingly integrated with formal management system principles such as “Plan-Do-Check-Act,” and “continuous improvement.” Duijm emphasised the use of “early indicators” to identify “precursors” of performance indicators such as incidents. By 2008, an ambition was set for proactive management of SHE, although challenges existed including integration, deriving useful indicators, and understanding of management system complexity [1].

Inferring from Koivupalo [2], the implementation of management system did not result in automatic organisational standardisation. Although the six steel sites had a health and safety management system, the practices were fragmented and varied considerably [2]. This is a broader example of the Duijm highlighted difficulties in coordinating integrated SHEM and emphasises the importance for “a coherent routine for collecting, analysing, storing and using information” [1]. Environmental management through 2014 to 2020 developed with early optimisation of costs of environmental measures with ignorance to “the effect of uncertainties” on management system efficacy, however others consider the uncertainty. The management became more quantitative and analytical to optimise performance measures. The industry progressed towards optimisation, quantitative modelling and multi-objective decision models analysing energy, emissions, and economic drives. This reflects a step towards monitoring outcomes towards analytical and proactive management [3]. AIMarri [4] represents the historical trajectory endpoint where enterprise systems are increasingly used to integrate SHEM through, “real time monitoring, compliance tracking, and risk mitigation processes”. In contrast with earlier SHE history, there is a clear direction of intellectual travel to data environments capable of supporting predictive analytics and more proactive SHEM [4].

Incident Reporting

Incident reporting creates an organisational evidence base to develop understanding of the key factors affecting “safety performance.” Incident investigation systems contain structured historical information about incident characteristics, causal factors, and outcomes. Their purpose is not solely data capture but extends beyond recording events to enabling corrective action and future prevention. Verma [5] provide relevant industrial context by analysing 843 incidents within steel manufacturing operations. Their findings demonstrate incident data can be analysed to identify associations and relationships between incident characteristics and outcomes. In contrast, the authors argue that this resource is often underutilised and insufficiently analysed. They discuss that associations between factors are identified through rule mining which revealed recurring patterns between incident characteristics and outcomes [5]. Verma highlights a directional change from reporting to analytical use of incident data.

Extending incident analysis across SHE, Cao [6] broaden this discussion around environmental incidents through analysis of 5,213 incidents from 2006 to 2015 and 1369 detailed regulatory cases. The findings demonstrate that organised incident data can reveal recurring risk sources, causes and environmental consequences supporting more informed environmental risk management [6].

The analytical potential to Verma can be constrained by the quality of the reporting. The findings by Guan [7] identify 42 facilitators and 28 hindrances to learning from events relating to “learning input... learning process...learning participants... and learning context”. Their review indicates that underreporting and incomplete incident reports are constraints on the information available for the analysis, while weaknesses in investigation, communication and organisational learning can limit how effectively incident data is translated into preventative action [7].

Competency & Training

Competency and training influence safety performance through a variety of aspects Hutchinson [8] establishes workforce training with measures of safety performance across industries by synthesising 100 samples from 90 intervention reports. A study found that training improved employee “compliance...participation...motivation...knowledge...[and] climate” establishing training as a relevant organisational factor in safety performance [8].

Directly connecting competency and steel industry, Pollitt [9] discussed that a set of training influenced a reduction in a key health and safety performance indicator. However, this evidence is observational and the relationship between the training intervention and reduction in the metric should not be interpreted as causality. The value of this source demonstrates the possibility that manageable organisational factors can positively affect safety performance [9].

The effect competency and training have on observable health and safety performance differs by industry according to the risk levels relating to health and safety. Hutchinson described improvements in knowledge and motivation may not equal improvements in incident rates which can indicate that training can improve safety related outcomes [8]. However, training interventions do not generate uniform performance improvements across organisational contexts. Tarigan [10] supports linking training interventions with incident outcomes across 6 studies in the healthcare sector resulting in reduced risk of needle stick injuries occurring [10]. Despite the industry misalignment, this indicates an improved effect on measured metrics when competency and training measures are reviewed.

A major evidence gap and limitation that limits evidence on the effect of competency and training on measured outcomes is the use of indirect indicators that are adjacent to overall performance metrics with Hutchinson including “three studies that reported accidents and injuries” [8]. The evaluation discussed evidence around safety capability and behaviour rather than training efforts systematically reducing incidents.

Predictive Analytics & Machine Learning

The emergence of predictive analytics shifts from retrospective review to “more complex analytical approaches” and prediction with organisational data. Foreman [11] discusses this “recent shift” as not yet mature or a universally embedded SHE practice. To evaluate, the inclusion of 30 steel manufacturing studies demonstrates substantial representation of steel within organisational safety analytic research. Despite this, “only 10 (22%) studies included proactive or leading indicators” such as “safety training history” thereby suggesting further analysis could reveal new relationships between organisational factors and safety outcomes. However, the literature remains focused on method development and validation and practical organisational action remains limited [11].

Maheronnaghsh [12] broaden the evidence base through a systematic review of 42 studies, showing that machine learning is applied across occupational safety and health for classification of injuries and accidents, hazard prediction and regression based pattern identification. This shows a shift towards identifying risk relationships and predictive future safety outcomes and indicates the usefulness of machine learning to potentially support proactive SHEM [12].

Continuing, organisational datasets can support predictive models that estimate health and safety metrics. Oyedele [13] use 168,574 data points from UK power and distribution projects to “advance safety intervention efforts” in relation to prediction of injury occurrence, lost time, and workplace risk with high accuracy. To evaluate, this supports the potential to use historical SHE data to identify future risks [13].

Pandey [14] investigates factors affecting safety performance, namely health and safety related factors such as machine condition, human error, PPE compliance, maintenance and other personnel and operational information. This is used to support prediction of future risk. The study does not integrate an employee competency dataset integration with the investigated areas, leaving scope to investigate whether competency information provides further explanation of SHE patterns [14].

Predictive capability is dependent on the suitability and integrity of organisational datasets, as previously described. Foreman (2024) explores most frequently reported limitations and indicates statistical, or analytical model choice and data quality can mean organisations are not yet “ready for analytics” [11]. Maheronnaghsh conditions successful machine learning use on access to “correct data” [12]. Combining the context from Fu [15], data that is complete, consistent, and compatible with the relationship being investigated can support valuable analytical outcomes.

Synthesis & Conclusion

In summary, industrial SHEM developed from predominantly compliance led and functionally separate practices and tools to an integrated, performance and data driven management. Duijm identifies information management is a requirement for effective SHEM and shows that the industry adopts integrated SHE systems, continuous improvement, and a focus on the proactive rather than reactive. Koivupalo illustrates the fragmented nature of management systems within the steel industry with manufacturing sites varying in local practices. Environmentally demonstrating SHEM, Huang [3] shows the direction towards a further quantitatively sophisticated management. In the most recent source demonstrated above, AlMarri represents the trend towards integrated SHE systems capable of compliance tracking, historical analysis, and predictive capability. The literature suggests that SHEM has become increasingly integrated, analytical, and proactive. However, the literature shows weaknesses hindering fully data driven SHEM related to standardisation, data coherence, and effective integration across organisational systems.

SHE incident data is a resource containing historical evidence of influencing factors and SHE patterns that requires reporting quality and consistent categorisation for useful insight. Verma evaluates that association rules do not establish direct causation. Guan shows that organisational context, participants, and learning processes influence what can be learned from incidents and broader organisational information may be needed to explain why patterns occur. Sufficiently structured and analysed SHE data increases analytical capability. However, Verma, Cao, and Guan indicate that how organisations capture, classify and investigate incidents affects the analytical capability of the historical datasets. To evaluate, incident data alone may not fully explain organisational factors associated with SHE performance and if data is incomplete data quality can be affected.

The literature shows a link between competency, training, and a plausible influence on SHE performance. Importantly the discussed evidence does not reveal a causal relationship. This creates a foundation for investigation on the patterns that exist between relevant competency and incident datasets and whilst recognising other organisational and operational factors influence the measured characteristics and indicator outcomes.

The existing literature demonstrates organisational safety data within the steel industry can jointly produce value with advanced technical predictive capabilities. This value combined with other SHE datasets may provide clear comparison and insight. The reviewed literature lacks predictive comparison of competency data across the wider SHE domain. The integration of staff competency datasets with historical SHE outcomes is not demonstrated in the reviewed literature. This provides the basis for investigating the combined competency and SHE datasets to understand additional relationships relevant to proactive steel industry risk management.

Collectively, the literature demonstrates the value of incident data, workforce capability, integrated information systems, and predictive analytics. However, these elements remain disconnected. In particular, the reviewed steel manufacturing literature does not demonstrate the integration of workforce competency records with historical SHE records and outcomes to investigate whether workforce capability provides additional insight into future SHE risks.

The Approach, Techniques, & Methods

Aim & Objectives

- 1) To design, develop and implement a proof of concept integrated SHE data solution by July 2027 that combines key datasets to support consolidated data management, analytical insight, and investigation of future SHE risks with a software implementation using SHE incident and competency data.
 - a) Design, develop and implement a consolidated data structure that integrates incident and competency data, applying appropriate data quality, standardisation, and governance controls.
 - b) Design, develop and implement an analytical solution that transforms the integrated data into standardised SHE performance measures, trends, and relationships across relevant sites and functions.
 - c) Develop and evaluate a predictive machine learning model using historical incident and competency data to determine whether identifiable patterns provide meaningful indicators of future SHE risk.

The project success criteria and minimum viable product success criteria are provided in Appendix I

Methods & Materials

This software implementation and quantitative analysis are intended to investigate comparisons between SHE datasets. Park [16] demonstrates an application of data analytics in the manufacturing context supporting the use of context specific analysis to investigate industrial data. This will require requirement gathering, implementation, quantitative analysis and stakeholder evaluation. The proposed purpose will produce value for Tata Steel through analytical findings while understanding that these findings may not be directly causal and limit generalisability across the wider Tata Group and alternative industrial contexts.

The datasets for this project will be historical SHE incident and competency data from the systems discussed in the proposal. A set historical timeline will be 5 years. Relevant variables related to this data will be explored such as those related to organisational factors. For example, site name and area, site location, and seasonality. The datasets used were developed for operational use. This may limit their suitability. An appropriate review of the individual records will identify invalid records for appropriate data treatment.

The relevant data preparation methods will be taken to clean, validate and standardise data before creating an integrated model. Data quality rules will be implemented for accuracy, completeness, consistency, validity, uniqueness, and timeliness. Records with missing values, inconsistent classifications or failures to meet standards and industry rules will be evaluated against the defined data quality rules. Validated incidents and competency data will then be standardised and developed through organisational dimensions [15].

The chosen statistical and exploratory analysis will be descriptive, comparative and relationship analysis to understand what patterns and trends exist prior to predictive modelling. Descriptive trends will be investigated through frequencies, distributions, rates, and further trends [17]. Continuing, the trends will be analysed through organisational trends between sites, time periods, and competency characteristics. The associations will be measured to understand the relationships. The appropriate statistical techniques will be chosen after examining the data through the previously mentioned analyses rather than predetermined irrespective of the data. An analytical layer will be used to transform the data into repeatable measures and interactive analytical outputs. This will include SHE measures, trends, comparative views, targets and will be accessible through organisational dimensions.

The predictive machine learning methodology will evaluate if patterns exist to determine future prediction risks of defined SHE outcomes. The validated data will be separated into training and testing data. A variety of models and algorithms will be examined for their suitability in response to evaluating the data. For example, logistic regression can be used to model defined SHE outcomes and develop a suitable predictive baseline. Candidate algorithms are decision trees, random forests, and gradient boosting. Models can be trained on training data, evaluated on test data, and then compared against the logistic regression baseline [18]. More complex models can be considered if they provide meaningful improvements over this baseline. Predictive performance will be evaluated by comparing predicted incidents against actual incidents in the test data. Further evaluations of the predictive capability will be assessed through F1 scores, "ROC curve AUC scores", "confusion matrices" and an accuracy report [14].

The software implementation will require data ingestion, data validation, transformation, modelling and analytical measures, visualisations, and predictive functionality [16] using suggested technologies from the business. Testing of the front-end and back-end will include white box, black box, and unit testing [19]. Finally, the stakeholders of the project will evaluate the usability and accuracy of the outputs against their own industrial context.

Considering ethics, security and governance is key in protecting employee data. Only data necessary to achieve the project objectives will be processed with employee identifying aspects carefully removed. Guidance will be sought from organisational experts, and organisational standards and policies related to data privacy and protection will be adhered to. Where data cannot be provided, datasets will be constructed to demonstrate software functionality. This will introduce further limitations to the project which will need to be extensively documented.

Project Design

Fig 2. details the project design including database connections, integrated data model, and the front-end aspect of the project. The data is anonymised from the data source by other internal company methods to reduce issues around data privacy. Failing that, the data can be recreated and generated in the required format.

Work Plan & Tasks

Figs.3-10 describe the planned work related to this project through stages of design, development, implementation, deployment, and stakeholder engagement.

Resources Required

The resources required to achieve the aims and objectives of this project rely on the infrastructure and support from the business in implementation and deployment. The required software infrastructure is extensively set up through Tata Steel. This infrastructure requires access from the departments that own the software and deployment platforms. This requires extensive networking across the Tata Steel group to successfully deploy this project. The implementation will be developed exclusively without direct implementation support from Tata Steel. The support from them will consist of consultations, advice, individualised and group reviews.

I require engagement and sponsorship from the Learning and Development department, Data Privacy representatives, IT department and professionals, and support from the SHE department and representatives. These key stakeholders provide

essential access to knowledge and technologies. Importantly, budgeting the time available between these stakeholders and sponsors will be an integral requirement to success. I require support from my university and their professionals. To alleviate this time budgeting concern, I will structure my time with each stakeholder and sponsor in a format that allows for their availability to be taken into consideration.

I require no additional funding for this project and a financial budget of £0.

Potential Impact

Impact

The major impact of this project is that it demonstrates how separate SHE datasets can be integrated into a consolidated data structure to derive further value from existing SHE data. This insight has the potential to affect personnel safety and the environment. Successful completion of the project provides a foundation for investigating SHE data rather than separated efforts. With the project, understanding the relationship between competency and incident data can provide evidence to support more targeted allocation of resources. This enables competency data to become an input into SHE resources planning and incident prevention discussions, rather than being reviewed independently. This assists with evaluations investigating whether training records are associated with SHE performance. There is a tendency for SHEM to use incident data reactively. This project creates a proactive focus on SHEM.

Limitations

Using only two datasets limits the evidence that the solution can be successfully implemented across all SHE datasets across the whole business of Tata. The proof of concept project does not represent a complete enterprise wide SHE solution.

The availability and quality of data at all deployment stages of this project restrict the reliability of statistical analyses. Additionally, the analyses and relationships between competency, incident, and all SHE datasets represent associations rather than causation limiting the efficacy of the project. SHE incident occurrence can be influenced by multiple organisational, operational, and human factors with limited direct evidence connecting, for example, safety training with reductions in accidents and injuries [8]. A further limitation is how competency and incident data may be unrelated.

Further research integrating leading and lagging SHE datasets can investigate efficacy. Further research could expand the scope to additional countries, sites, and years to increase the data volume and diversity of data for analysis. This investigation of data could determine significant clusters of incidents and how particular patterns occur rather than a description of what is happening. This project is limited to the context of one organisation and the SHE data environment. Further research across organisations and wider industrial contexts can expand the understanding of SHE metrics.

Personal Reflection

Throughout this proposal I have developed an understanding of the technologies, data and relevant sources that will feed into my project. From gathering this information, I am comfortable to fully deliver my project. I understand the various constraints and benefits of my project, and I look forward to implementing the proof of concept.

Conclusion

Overall, the research shows a clear direction towards developing solutions that foster proactive review of data rather than reactive. The literature indicates a growing trend in using integrated methods to analyse data with advanced analytical technologies. However, there is limited direct evidence of relationships between competency, incident data, or wider consolidation of SHE datasets. Predictive insight using machine learning techniques relies heavily on data being contextually relevant and of high quality. Therefore, a clear gap exists for the software project to deliver value to SHE, Tata Steel and for my personal development.

References

- [1] N. J. Duijm, C. Fie'vez, M. Gerbec, U. Hauptmanns and M. Konstandinidou, "Management of health, safety and environment in process industry," *Safety Science*, vol. 46, no. 6, pp. 908-920, 2008.
- [2] M. Koivupalo, M. Sulasalmi, P. Rodrigo and V. Seppo, "Health and safety management in a changing organisation: Case study global steel company," *Safety Science*, vol. 74, pp. 128-139, 2015.
- [3] D. Huang, C. D. Dinga, Z. Wen and D. Razmadze, "Industrial-environmental management in China's iron and steel industry under multiple objectives and uncertainties," *Journal of Environmental Management*, vol. 310, p. 114785, 2022.
- [4] M. AlMarri, M. Al-Ali, M. Alzarooni, K. Al-Ali, Z. Bahroun and A. AlTeneiji, "Enterprise Resource Planning Systems for Health, Safety, and Environment Management: Analyzing Critical Success Factors," *Sustainability*, vol. 17, no. 7, pp. 1-29, 2025.
- [5] A. Verma, S. D. Khan, J. Maiti and O. Krishna, "Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports," *Safety Science*, vol. 70, pp. 89-98, 2014.
- [6] G. Cao, L. Yang, Z. Ma, J. Wang, J. Bi and L. Liu, "Environmental incidents in China: Lessons from 2006 to 2015," *Science of The Total Environment*, vol. 633, pp. 1165-1172, 2018.
- [7] J. Guan, Y. Zixuan, A. P. Chan, T. Choi and Y. Yang, "Factors affecting learning from incidents: A cross-industry review," *Journal of Loss Prevention in the Process Industries*, vol. 89, pp. 1-17, 2024.
- [8] D. Hutchinson, G. Luria, S. Pindek and P. Spector, "The effects of industry risk level on safety training outcomes: A meta-analysis of intervention studies," *Safety Science*, vol. 152, p. 105594, 2022.
- [9] D. Pollitt, "Corus forges new approach to safety and health," *Human Resource Management International Digest*, vol. 19, no. 1, pp. 7-9, 2011.
- [10] L. H. Tarigan, M. Cifuentes, M. Quinn and D. Kriebel, "Prevention of Needle-Stick Injuries in Healthcare Facilities: A Meta-Analysis," *infection control & hospital epidemiology*, vol. 36, no. 7, pp. 823-829, 2015.
- [11] A. M. Foreman, J. E. Friedel, M. E. Ezerins, R. Matthews, R. E. Nicholson, L. Wellersdick, S. Bergman, Y. Açıkgöz, T. D. Ludwig and O. Wirth, "Establishment-level safety analytics: a scoping review," *International journal of occupational safety and ergonomics*, vol. 30, no. 2, pp. 559-570, 2024.
- [12] S. Maheronnaghsh, H. Zolfagharnasabb, M. Gorgich and J. Duarte, "Machine learning in Occupational Safety and Health – a systematic review," *International Journal of Occupational and Environmental Safety*, vol. 7, no. 1, pp. 14-32, 2023.

- [13] A. O. Oyedele, A. O. Ajayi and L. O. Oyedele, "Machine learning predictions for lost time injuries in power transmission and distribution projects," *Machine Learning with Applications*, vol. 6, p. 100158, 2021.
- [14] S. Pandey, A. K. Singh, S. Parhi and S. K. Jha, "Towards safer steel operations with a multi model framework for accident prediction and risk assessment simulation," *Scientific Reports*, vol. 15, no. 1, p. 13293, 2025.
- [15] Q. Fu, G. L. Nicholson and J. M. Easton, "Understanding data quality in a data-driven industry context: Insights from the fundamentals," *Journal of Industrial Information Integration*, vol. 42, p. 100729, 2024.
- [16] S.-E. Park, S.-J. Lee, H. Bae, K.-H. Kim, E.-J. Kang, J.-S. Kim, Y.-M. Park and M.-J. Park, "Comprehensive issue identification for manufacturing data analytics implementation: Systematic literature review and case studies," *Journal of Manufacturing Systems*, vol. 82, pp. 42-57, 2025.
- [17] C. N. Khan and J. D. Kribs, "Data-Driven Analytics in Smart Manufacturing: A Review of Methods, Applications, and Adoption Constraints," *The International Journal of Advanced Manufacturing Technology*, vol. 144, pp. 3061-3076, 2026.
- [18] H. Tercan and T. Meisen, "Machine learning and deep learning based predictive quality in manufacturing: a systematic review," *Journal of Intelligent Manufacturing*, vol. 33, pp. 1879-1905, 2022.
- [19] G. A. D. Lucca and A. R. Fasolino, "Testing Web-based applications: The state of the art and future trends," *Information and Software Technology*, vol. 48, no. 12, pp. 1172-1186, 2006.

Appendix I

Supporting Information

Problem Description

The data is stored on third-party platforms, separate from other related datasets. Alone, incident datasets cannot determine whether factors such as employee competency, experience, training deficiencies and other elements contribute to incidents. Competency data in isolation cannot determine an outcome related to incident rates. This is not the explicit aim of the project, but it highlights the need for the relationship between the two to be explored.

Problem Consequences

A variety of systems are present in Tata and provide a fragmented and reactive use of data. Data quality management is organised efficiently in some systems but lacks in others, producing an inconsistent data quality across SHE data. In addition to the fragmented nature of the data, the development and gathering of multiple datasets stretches already strained resources. Spread analytic capabilities across these platforms reduce insight and duplicates effort for the transformations of data. The individual measures and aspects of these spread activities limit the understanding of combined performance of SHE. Potential patterns can remain unidentified, and the data driven capacity of the senior leadership will be hindered, leaving reporting predominantly descriptive and retrospective.

Project Need

The lack of combined views and data management efforts across SHE data decreases insight behind the metrics in relation to SHEM. The lack of intelligent transformations, comparisons and data analysis techniques prevents the future use of AI integrations. Longitudinal relationships between key SHE datasets are rarely explored. The information found could impact lives and the environment.

Integrated Data & Data Quality

Integration of datasets may support more sophisticated analysis against “siloes” and “inadequate” models, accessing a broader organisation understanding of measured factors. Fu highlights integrated datasets alone do not produce reliable insight or improve analytical capability. To evaluate further, data insight heavily relies on the underlying data clearly representing investigated phenomena. Fu cautions that inaccurate source data can produce inappropriate analytical results [15]. To summarise, analytical methods and inaccurate data combined can produce a negative effect on analysis.

Data quality should therefore be integrated into the data system by design as a lifecycle characteristic, not as an afterthought prior to analysis. Integrity can be affected at “collection, storage, processing, maintenance and disposal phases.” Limitations in historical data, including people and processes, can propagate into analysis. To evaluate, the analytical quality of industrial SHE data is dependent on the processes and users through which records are generated and maintained. Critically, data suitable for an operational function is not necessarily sufficient for another analytical question. Historical incident and competency records may contain complementary organisational evidence however the data quality for their individual systems may not qualify them for subsequent analysis, namely that of comparative and combined analysis. To evaluate further, as shown in Fig.1, the data may contain deficiencies in “completeness”, consistency, timeliness, validity, or accuracy affecting analysis [15].

Project Success Criteria

The project will result in successful completion where the below is met.

- A functional proof of concept software solution is developed and implemented by July 2027.
- Incident and competency datasets are integrated successfully through a common data structure that can be applied to future datasets.
- Relationships between datasets can be analysed across relevant organisational dimensions such as site name, site area, functional area, and time.
- Standardised SHE measures, trends, patterns, relationships and further analytical outputs are verified and can be demonstrated across relevant organisational dimensions such as site name, site area, functional area, and time.
- Defined data quality implementations to achieve accuracy, completeness, consistency, validity, uniqueness, and timeliness are applied to relevant dimensions.
- Data rules are documented and reproducible through demonstrated data lineage and reconciliation.
- A functional machine learning approach is developed with appropriate training and testing to determine if the available data contains meaningful predictive patterns including the valid outcome that the predictive performance is insufficient.
- Limitations of the project are extensively documented.

Minimum Viable Product Success Criteria

The minimum viable product (MVP) will result in successful completion where the below is met.

- A functional proof of concept software solution is developed and implemented by July 2027 that integrates a defined subset of historical SHE incident and competency data.
- A common structured data model with data quality measures installed.
- A minimum set of visualisations including analytical insight.
- A machine learning model to investigate trends associated with future SHE risk is developed and evaluated.

Continued Appendix

Table 1: System Description

Current System Feature	Incident Reporting Platform	Competency Reporting Platform
System Owner	SHE Systems, Digital Transformation & Analytics Officer	Learning & Development Department
System Stakeholders	SHE representatives, senior leaders, directors, and SaaS product owners.	Business site and functions
Data History	5 years	5 years
Quality Data History	3 years	5 years
Information Flow	User entry. Data enters databases. Email & business intelligence reports produced.	Staff complete course. Learning & Development enter data into platform. Site and function reports produced.
Dependencies	Accurate reporting of data. SHE personnel validating data. Users having correct system access, permissions, and roles.	Accurate reporting of the data. Accurate and maintained data. Competency requirements being defined by subject matter experts. Courses being mapped to the correct roles. HR head count data being accurately transferred into system.
Data Management Tools	SaaS platform. Minimum flexibility.	Limited Oracle database tools.
Business Intelligence Tools	SaaS product owners have an inhouse tool.	Sparse Power BI reports.
Artificial Intelligence Capabilities	None.	None.

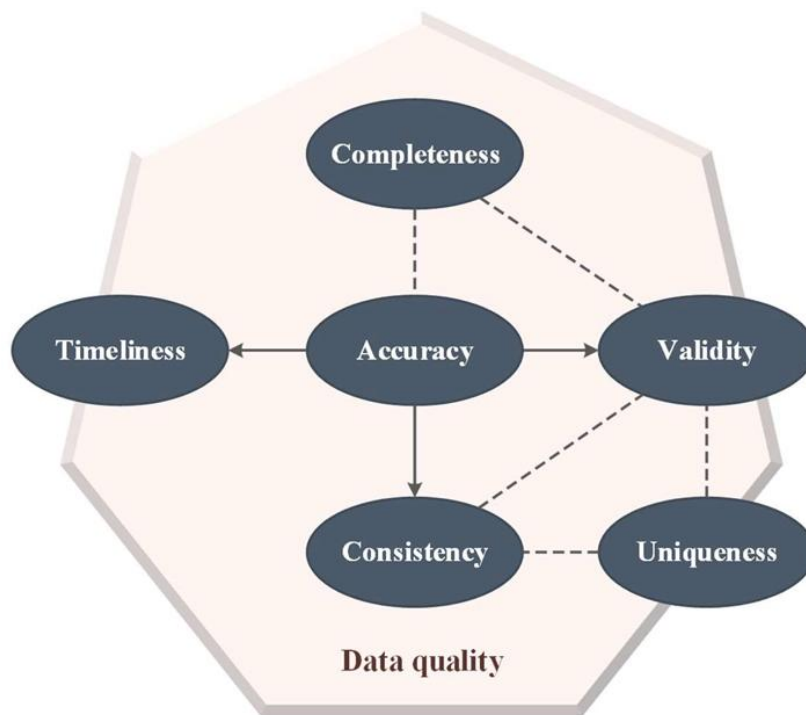
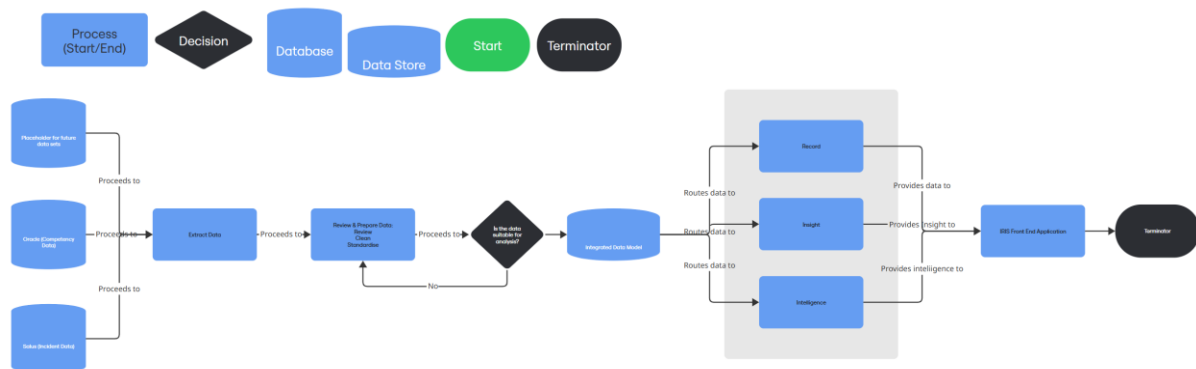


Fig.1. Data quality [15].



https://miro.com/app/board/uXjVHux2c6o=?share_link_id=803039187469

Fig.2. Project design

NO.	TASK	START	END	DAYS	% DONE	WORK DAYS
1	Planning, Design & Stakeholder Engagement	-	-	-	-	-
1.1	Confirm project scope, requirements, deliverables, resources, dependencies, risks and success criteria	01/10/2026	12/12/2026	72	0%	52
1.2	Design system and data diagrams including UML, data flow, user stories, functional requirements and MVP definition	01/09/2026	15/09/2026	14	0%	11
1.3	Engage with Heads of Departments, SHE stakeholders, IT, data owners, information security and policy stakeholders to review requirements and designs	15/09/2026	10/10/2026	25	0%	19
1.4	Obtain approval and sign off with: scope, requirements, data access, security considerations, design, success metrics and KPIs	11/10/2026	12/12/2026	62	0%	45
1.5	Design and build prototype	15/09/2026	10/10/2026	25	0%	19
2	Implementation & System Development	-	-	-	-	-
2.1	Establish the required development environment and technical architecture	12/12/2026	25/01/2027	44	0%	31
2.2	Develop core IRIS platform with the required functionality	14/12/2026	25/01/2027	42	0%	31
2.3	Implement data ingestion transformation, validation and storage processes	14/01/2027	25/01/2027	11	0%	8
2.4	Develop data model, relationships, measure and analytical functionality	17/01/2027	25/01/2027	8	0%	6
2.5	Develop user facing interfaces, dashboards and reporting functionality	23/01/2027	25/01/2027	2	0%	1
2.6	Integrate Record, Insight, & Intelligence sections	23/01/2027	25/01/2027	2	0%	1
2.7	Conduct a series of functional and non-functional testing	23/01/2027	25/01/2027	2	0%	1
2.8	Resolve defects found in testing stage	23/01/2027	25/01/2027	2	0%	1

Fig.3. Gantt table 1

NO.	TASK	START	END	DAYS	% DONE	WORK DAYS
3	Deployment & User Adoption	-	-	-	-	-
3.1	Complete final technical, security and stakeholder checks	25/01/2027	25/02/2027	31	0%	24
3.2	Deploy into the environment suited for Tata identified in planning stage	01/02/2027	06/02/2027	5	0%	5
3.3	Produce user documentation, guides and supporting materials related to user training	07/02/2027	10/02/2027	3	0%	3
3.4	Deliver workshops to relevant users and stakeholders	10/02/2027	27/02/2027	17	0%	13
3.5	Provide initial user support including issues arising during deployment	15/02/2027	27/02/2027	12	0%	10
4	Validation	-	-	-	-	-
4.1	Monitor deployment against predefined requirements and key indicators developed in the planning stage	01/03/2027	09/03/2027	8	0%	7
4.2	Validate data transformations, calculations and outputs through a series of reconciliation methods	01/03/2027	09/03/2027	8	0%	7
4.3	Confirm system performance meets predefined standards	10/03/2027	14/03/2027	4	0%	3
4.4	Implement necessary improvements	15/03/2027	27/03/2027	12	0%	10
4.5	Confirm the platform has reached a stable operational state	01/04/2027	28/04/2027	27	0%	20
5	Review & Evaluation	-	-	-	-	-
5.1	Evaluate project against its aims, objectives and key indicators	01/05/2027	15/05/2027	14	0%	10
5.2	Evaluate the functional and non-functional requirements have been met	01/05/2027	15/05/2027	14	0%	10
5.3	Engage with stakeholders to gather their feedback	01/05/2027	30/05/2027	29	0%	20
5.4	Document limitations, lessons learned and outstanding issues.	01/06/2027	25/06/2027	24	0%	19
5.5	Evaluate recommendations and opportunities for further development	01/06/2027	25/06/2027	24	0%	19

Fig.4. Gantt table 2

		Project Start Date		01/09/2026		Display Week		1																													
		Project Lead		Shaun J Davies																																	

Fig.5. Gantt table 1: Planning, Design & Stakeholder Engagement

		Project Start Date	01/09/2026	Display Week	15																																											
		Project Lead	Shaun J Davies																																													
		START	END	DAYS	% DONE	WORK DAYS	Week 15 07/12/2026							Week 16 14/12/2026					Week 17 21/12/2026					Week 18 28/12/2026					Week 19 04/01/2027					Week 20 11/01/2027					Week 21 18/01/2027					Week 22 25/01/2027				
NO.	TASK						S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S						
1	Planning, Design & Stakeholder Engagement																																															
1.1	Confirm project scope, requirements, deliverables, resources, dependencies, risks and success criteria	01/10/2026	12/12/2026	72	0%	52																																										
1.2	Design system and data diagrams including UML, data flow, user stories, functional requirements and MVP definition	01/09/2026	15/09/2026	14	0%	11																																										
1.3	Engage with Heads of Departments, SHE stakeholders, IT, data owners, information security and policy stakeholders to review requirements and designs	15/09/2026	10/10/2026	25	0%	19																																										
1.4	Obtain approval and sign off with: scope, requirements, data access, security considerations, design, success metrics and KPIs	11/10/2026	12/12/2026	62	0%	45																																										
1.5	Design and build prototype	15/09/2026	10/10/2026	25	0%	19																																										
2	Implementation & System Development																																															
2.1	Establish the required development environment and technical architecture	12/12/2026	25/01/2027	44	0%	31																																										
2.2	Develop core IRIS platform with the required functionality	14/12/2026	25/01/2027	42	0%	31																																										
2.3	Implement data ingestion transformation, validation and storage processes	14/01/2027	25/01/2027	11	0%	8																																										
2.4	Develop data model, relationships, measure and analytical functionality	17/01/2027	25/01/2027	8	0%	6																																										
2.5	Develop user facing interfaces, dashboards and reporting functionality	23/01/2027	25/01/2027	2	0%	1																																										
2.6	Integrate Record, Insight, & Intelligence sections	23/01/2027	25/01/2027	2	0%	1																																										
2.7	Conduct a series of functional and non-functional testing	23/01/2027	25/01/2027	2	0%	1																																										
2.8	Resolve defects found in testing stage	23/01/2027	25/01/2027	2	0%	1																																										

Fig.6. Gantt table 1: Implementation & System Development

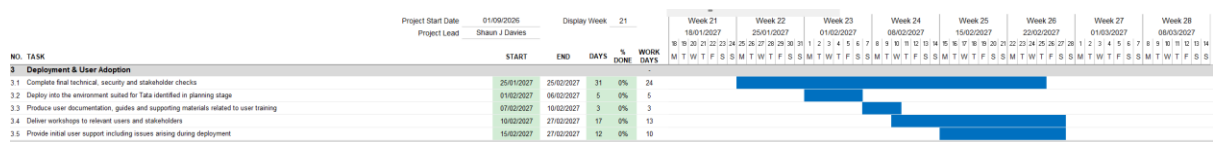


Fig.7. Gantt table 2: Deployment & User Adoption

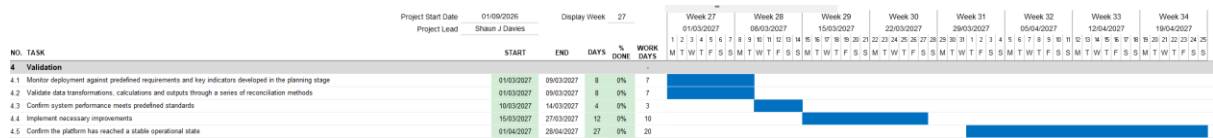


Fig.8. Gantt table 2: Validation

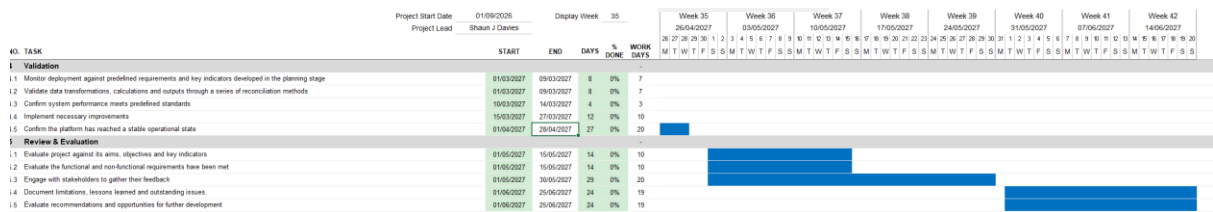


Fig.9. Gantt table 2: Validation, Review & Evaluation

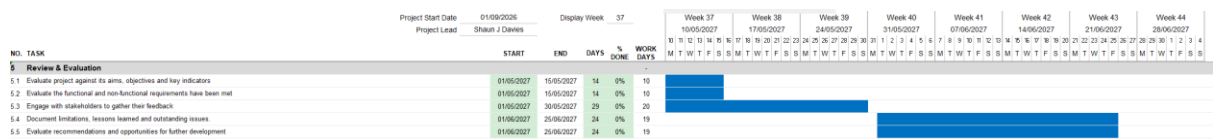


Fig.10. Gantt table 2: Review & Evaluation