Data Quality - the Foundation for Effective Modelling & Simulation

Simon Albert
Babcock International Group
Sedgemoor Building | Monxton Road | Andover
SP11 8HT
UNITED KINGDOM

simon.albert@babcockinternational.com

Tim King
Babcock International Group
Radar Road | Leicester
LE3 1UF
UNITED KINGDOM

tim.king@babcockinternational.com

ABSTRACT

In applying emerging and disruptive modelling and simulation (M&S) capabilities to improve delivery of services to UK MOD, Babcock International has recognized that data quality issues are a significant barrier to the success of those capabilities.

Many current approaches rely on technology as the foundation for data quality. This reliance fails to address the significant influence of other factors such as people and processes. These other factors have been the root cause of previous major data quality failures, such as a single wrong data value causing NASA to lose the Mars Climate Orbiter [1].

To take a more sustainable, systemic and systematic view of data quality, Babcock has applied existing international best practice (including the ISO 8000 series [2] for data quality and the ISO 55000 family of standards [3] for asset management) to a wide range of M&S exploitation activities. These activities have included mixed reality, digital twin, artificial intelligence and machine learning to enable military customers to improve equipment reliability, force planning and agility of decision making. The result has been a robust approach to data quality to underpin effective M&S for militaries and suppliers across NATO.

By adopting this approach to data quality, the exploited M&S has more reliably delivered benefits including improved cost effectiveness and readiness of military capability pan-domain.

1.0 INTRODUCTION

1.1 The modelling and simulation context

Digital technology is the foundation of effective complex modelling and simulation. Such technology continues relentless growth, driving the capability of modelling and simulation at the same pace [4]. This growth soon brings ground-breaking research themes into widespread exploitation [5]. These themes affect Defence as much as any other sector. Defence examples include a wide range from optimising the deployment of assets, to informing engineers delivering maintenance, to sustaining operational performance and through to providing insight and foresight when planning and controlling at tactical, operational and strategic headquarters across commands and domains. Modelling and simulation is routinely enabling decisions throughout the military system lifecycle: from complex asset design such as for the new Dreadnought submarine programme [6], to fleet design of land equipment, through to in-service asset performance [7] and procurement.

In parallel, the proliferation of dashboards and visualizations of current and historical data, enabled by increasingly powerful analytics tools (PowerBI, Qlickview) is driving a greater awareness of the importance of data quality. This awareness is evident in the UK MOD publishing the Data Strategy for Defence [8], which shows senior leaders across Defence have begun to understand the significance of data quality and the need for change (building on a wide body of hierarchical strategic direction that emphasizes the compelling nature of digital transformation [9] [10] [11] [12] [13] [14] [15]). Leaders providing clear strategic direction is extremely important but the practical challenges of change are faced by a wide range of teams, departments and functions across organizations. Such change requires a coherent, informed set of guidance to tackle the detailed realities and complexities of data quality.

1.2 A standards-based approach to data quality

The ISO 8000 series defines "data" as the "reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing". The series also recognizes the most significant impact from data comes when the data is in digital format. This definition is the starting point for taking the definition of quality from ISO 9000 [16] to create the definition of "data quality": "the degree to which a set of inherent characteristics of data fulfils requirements". These definitions provide a rigorous foundation from which the ISO 8000 series specifies a comprehensive framework for managing and sustaining the quality of digital data, including mitigating the risks of poor data quality. This framework is appropriate for all types of data and all types of sector, as demonstrated by significant existing exploitations of the standard (for example, [17] [18]). The framework is no less applicable to any application of modelling and simulation, where such applications can only ever be as capable as the quality of the input data allows.

In delivering practical implementations of ISO 8000, a key consideration is to understand the role of data hazards, which cover both the issues (current and known) and the risks (future and uncertain) inherent in any data set. Any implementation should seek to enhance the quality of the data by reducing both the probability and impact of data hazards negatively affecting the purpose being served by the data. Such purposes include the role of data as an input to modelling and simulation and the consequential effects on the reliability of the outputs of that modelling and simulation.

Whilst the primary organizational drivers in the Defence context typically relate to readiness, performance and cost effectiveness of assets, the data hazards are what impact the ability of organizations to achieve their objectives and realise benefits. Each organization needs a systematic approach to sustain data quality and, thus, to secure the necessary foundations for achieving objectives and delivering benefits.

1.3 Scope of this paper

This paper describes three case studies in which, when supporting UK MOD to deliver effective military capability, Babcock International has faced the challenges of data quality and experienced the impact on the effectiveness of modelling and simulation. These challenges have led to the development of practical approaches to the application of the principles and framework specified by the ISO 8000 series.

The paper addresses three key themes as separate sections, each with a supporting case study:

- probability [see Section 2.0], exploring the proposition that the likelihood of an information item having an impact on modelling and simulation is dependent on the condition of the item;
- impact [see Section 3.0], exploring the proposition that the importance of an information item is determined by the organizational need and is, therefore, always relative;
- mitigations [see Section 4.0], exploring the proposition that interventions need to be scoped, prioritized and implemented by using a programmatic approach.

2.0 THEME 1: PROBABILITY

Proposition: The likelihood of an information item having an impact on modelling and simulation is dependent on the condition of the item.

In the British Army, the portfolio of land equipment has extensive breadth, scale and age. The portfolio covers in-service equipment and those still under procurement. This breadth, scale and age drive a significant complexity for applying modelling and simulation to deliver insight and foresight. The portfolio includes a wide range of different types of equipment from vehicles, to radios, to weapons and through to specialist equipment. The portfolio involves a total expenditure over 10 years at over GBP 40 bn[1] [19]. This expenditure does not include a valuation of the existing in-service equipment, some of which dates back 50 or more years. Equipment is constantly in use at different readiness levels, across three different services, six commands and multiple budget holders.

Managing this portfolio requires dozens of applications and management information systems of different ages and commercial statuses. These systems enable asset management of all the equipment types by using more than 200 fields, which cover data from technical reference data, to configuration, to fleet management allocations and through to component fault reporting. These fields have users ranging from operators and maintainers, through to planners and forecasters, from the lowest level of military hierarchy, through to Army Headquarters, Defence Equipment & Support [20] and the supply chain across Industry.

For this portfolio, data quality varies across data sets in levels of accuracy, completeness, consistency and coherence. This variation prompted the discovery phase of a joint Army and Industry project to understand the situation in more detail. This phase generated anecdotal evidence that highlighted the impact of sub-optimal data: excessive effort needed for data wrangling; manual manipulation of data sets; proliferation of standalone spreadsheets; and over-reliance on static, isolated data sets to answer specific questions. These impacts often cause activities that are hard to repeat in a consistent and reliable fashion. Such activities have a negative effect on morale. The impacts ultimately erode confidence in the data, especially at the staff planning and control levels.

Whilst common to many other organizations of any sector, size and nationality, and by no means unique to the UK Defence enterprise, data quality issues not only pose challenges to effective asset management [21] but also to effective data management. Asset management and data management ultimately have a symbiotic relationship [22] that demands strategic alignment as articulated by relevant international standards including the ISO 55000 family [3] and ISO 8000 [2]. This alignment is against an evolving target, which for the Army has recently become a focus on Operation MOBILISE [23]. This focus includes a need to improve readiness through better equipment availability within the specified budget envelopes.

The appetite to rely more heavily on modelling and simulation is evident from the British Army Digitalisation programme, THEIA [24]. This programme seeks to build on existing pockets of information exploitation capability exist by increasing access to data analytics tools. This increase also demands growth of digital skills across Army personnel at all levels.

The rise of data analytics is casting a spotlight on the condition of the underpinning data across the Defence enterprise. This spotlight has heightened awareness of the risk to decisions from inadequate data quality. Such awareness was a major driver for publication of the first edition of the Data Strategy for Defence in 2021 [8]. Analytics is only going to carry on growing in significance as a consequence of rapid technology developments across all sectors of society and as contemporary events drive demand for even more capability (for example, the extensive reliance on modelling and simulation during the Covid pandemic [25]). This technology strives to meet the complex demands for insight by being wide ranging, including examples such as data integration and digital.

[1] GBP 40,000,000,000 | £40,000,000,000 | ~ USD 45,656,000,000 (at 2022-09-19)

The growth of modelling and simulation has already driven improvements in data quality. These improvements include, in particular, completeness and consistency. Such improvement does not, however, mean that all data is necessarily perfect before modelling and simulation can be effective. Perfection is never achievable; in reality, all data sets contain data hazards. These hazards are often actually uncovered by performing modelling and simulation, generating results and providing feedback on the data quality; in other words, a symbiotic relationship exists between modelling and simulation and data quality improvement.

Decision briefings now routinely include decision support products, which are often provided as a service by Industry partners. Such products rely on the outputs of complex risk-based modelling techniques to look forward, rather than summaries of the current situation generated by dashboards. These outputs range from: specific equipment availability assessments, to capability and fleet design, through to operational deployment modelling. This modelling enables adjustments of the variables to run scenarios routinely using a consistent and repeatable process, often in live simulations.

The Army turned to Industry to work together on a discovery project to develop and pilot a robust, rigorous, repeatable approach to enhance and sustain data quality for the land equipment assets of the Army. The project applied the principles of ISO 8000 to develop a three-step approach that focuses on three core questions [see Figure 1].

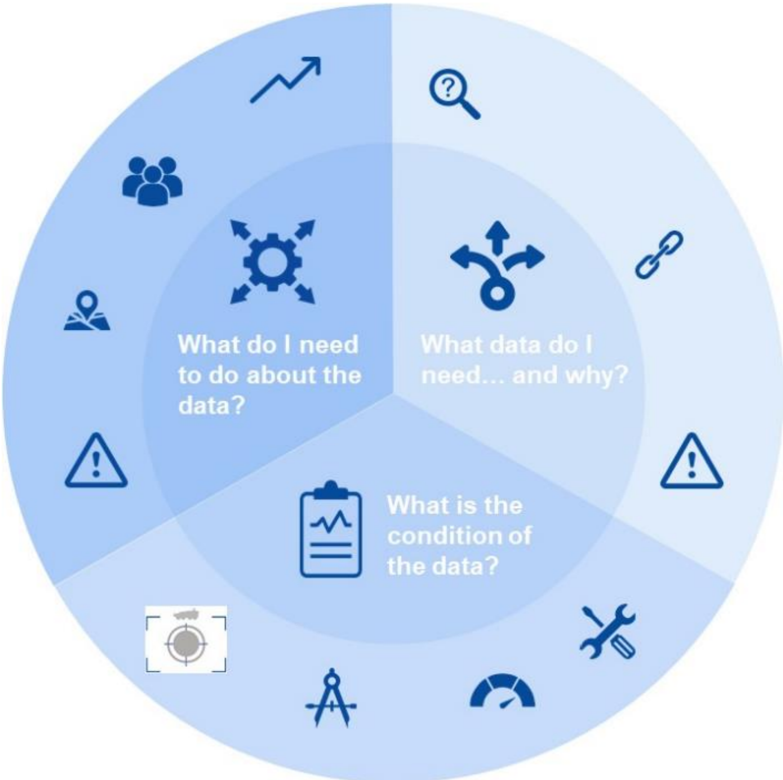

Figure 1: Three-step approach to data quality

The project involved analysis of the available data fields to identify how each is being exploited by users. This analysis provided a first pass of the first question (*What data do I need and why?*). The analysis was supported by capturing the core decision questions routinely addressed by the land equipment enterprise. This capturing involved a prototype three-dimensional visualization of relationships between the information items and the decision questions to reflect the data flows in the enterprise.

The next step was to prioritise the data sets for two specific fields to be the subject of more detailed systematic assessment of the condition of these data sets. This assessment addresses the second question of the approach (W*hat is the condition of my data?*).

The first data field was the taxonomy that classifies the equipment. This taxonomy is reference data that underpins all fleet management. The taxonomy provides alignment and traceability to other systems and is the basis for financial accounting activity.

The second data field was the distance travelled by platforms (mileage). This distance is transactional data that informs activities including: maintenance, repair and overhaul; logistic planning and estimating for spares procurement; and validating equipment safety cases.

The assessment of the data sets was relatively straight forward with the right domain knowledge, analytical skills and supporting tools. The assessment generated detailed results with some specific implications of a sensitive nature but also indicated more general core drivers and levers to effect improvement to data quality through the wider considerations of people, process and systems [see Table 1].

Table 1: Condition assessment findings

	People	**Process**	**Systems**
Taxonomy	• Errors from manual data input resulting in inconsistent data. • Awareness training and education can reduce errors. • Significant variations in the condition increase the effort of data wrangling. • Incomplete data entry is driving high levels of manual checks.	• Complex policy landscape requiring pan-Defence level approach. • Initiatives to change the data require co-ordination and programmatic approaches. • A condition baseline will enable more proactive data quality management.	• Established interconnections between systems reduce freedom to change master data. • Traditional hierarchical approaches drive repetition; an ontological approach will increase opportunities for modelling.
Mileage data	• Input errors by operators, resulting in statistically significant outliers could be addressed by targeted training.	• Impact of individual errors likely mitigated through aggregation to fleet level and historical trend analysis. • Track and report data quality score to drive change.	• Outliers easily identified using statistical techniques, with logic coded in Mileage Dashboard for tracking • Automate data input from usage monitoring systems, where available. • Improved data rules will reduce probability of entry errors.

The assessment provides a baseline for the data condition and a start point for proactive management of data quality. This management involved collaboratively identifying, scoping and prioritising interventions based on whether they were preventative or were to recover and manage the consequences of the data hazard on decision making.

For mileage data, the project concluded that hazards in this data cause an impact on the decisions of procurement, maintenance and safety. This impact is from driving uncertainty and inefficiency into each decision process. Each hazard has a probability of materialising, where this probability can be mitigated by the staffing process. Such mitigations are, however, inconsistent and inefficient to deliver. The uncertainty was addressed by using statistical analysis to determine appropriate business rules to achieve a quality confidence threshold of ~98% for a pilot data set piloted. This threshold delivered appropriate improvements to the reliability of decision making. The business rules were not sufficient to address all concerns and, instead, also need the support of enduring interventions that include increased automation (prevent) and proactive data quality tracking (recover) to further reduce the probability of the hazards occurring. These interventions vary in feasibility and cost effectiveness as determined by appropriate benefits realization mapping.

For taxonomy data, the project concluded that hazards in this data cause impact on users across the enterprise, affecting every aspect of asset management at fleet and platform level. This impact has resulted in current mitigations that mostly rely on people, with only some interventions in place within systems. These current mitigations need to be replaced by a systematic approach that provides more focus on process and policy, leading to consequential changes to systems. Such an approach is feasible but harder to achieve due to the impact on and cost of changing applications and systems. The project proposed a short-term solution of cleansing data where possible but adopting a more enduring set of changes as part of the Business Modernisation for Support programme, which will deliver enhanced asset management capability to the whole of Defence [26].

Whilst there are undoubtedly many parallels with the data and asset management challenges faced by the Army with land equipment, the next section describes the experience of seizing the opportunity to lay reliable data foundations to underpin the delivery of a smaller number of highly complex assets in the maritime domain.

3.0 THEME 2: IMPACT

Proposition: The importance of an information item is determined by the organizational need and is, therefore, always relative.

The UK Royal Navy has, in the first instance, set a challenge for physical boat building and traditional programme management in initiating the acquisition of the new Inspiration Class Type 31 frigate. This challenge centres on being exacting on both the budget envelope and the delivery timeline. The acquisition is proceeding with a design and build based on the Arrowhead 140 solution provided by Babcock International. This programme also poses a challenge for data management.

At the heart of the challenge is integrating over 400,000 parts per ship and thousands of individual processes for every step of the design and build phases. This integration has necessitated constructing a data model from the bottom upwards. This model has generated a complete, holistic view of data requirements across the programme. The data model is underpinned by an information and knowledge management solution that connects a suite of key applications consistently and coherently to share and reuse business and operational critical data. The challenge has led to the same starting point as the approach to data quality for land equipment: the question "*What data do I need, and why?*".

In the case of the Type 31 frigate, there were two macro answers to this question originating in organizational need: firstly, to assist with the build process; secondly, to enable through-life support of the ship in service. These two answers both lead to investment in establishing robust data quality and effective data management. This investment enables the organization to realise operational benefits and reduce risk during the life of the build programme.

The data audit for the Type 31 frigate programme identified 72 different data classes across 20+ systems and applications. This audit drove a solution: using open architecture principles to create a modular system architecture; automating data transfer through a web of 20+ APIs; increasing access to digital data on the shop floor; and delivering training for data stewards. This solution is generating the mitigations to reduce the risk of the quality of data not being sufficient for demands today and in the future.

There are three fundamental aspects to the data architecture [see Table 2]. Each aspects offers a different opportunity to intervene and to improve data quality, whether for the immediate purposes of the short term ("today") or to support the complete lifecycle of the asset ("tomorrow").

Table 2: Data architecture framework

	Aspect of data architecture		
	Transactional and observational data	**Reference and master data**	**Data models and schemas**
Purpose of the aspect	To capture "raw data" that represents the topic under consideration	To codify knowledge that repeats across data sets	To control the technology that manages the data
Data quality consideration	Pragmatic	Semantic	Syntactic
Definition	The utility of the data set content	The meaning of the data in the data set – consistent & structured	The format of each data set and of the data values in each set
Function	Usability	Interpretability	Processability
Impact on the data value chain	What insight and foresight can we generate from this data?	Can we connect all related data?	Can we perform processing of data effectively and efficiently using all available tools?
Examples of delivering benefit from effective data capability	Asset management can address effects relevant to sustaining capability	Asset management can address considerations that depend on multiple interacting factors	Asset management can build upon existing application functionality to identify trends and patterns in data

Different data has different value to different users at different times. Data is not static: it, instead, flows through an organization and across the data lifecycle. This flow involves users exploiting data, while also changing, adding to, aggregating and segmenting the data. The flow will also involve fusing of data sets with new or disparate other sets to form the inputs to models, which generate different insights and enable decisions that were not always evident when initially collecting and curating the data. The data lifecycle involves ebbs and flows to the value (importance) of the data at any point in time.

For the new fleet of frigates, the primary focus during the build phase is on curating the primarily technical reference data and creating the data foundations. This focus is, however, to support the strategic intent of improving the probability of being able to deliver cost effective in-service support. The build phase is generating relatively little operational or performance data. The in-service phase will change this balance in the data generation and will result in significant volumes of operational and transactional data being generated by in-service use. In-service demands will only be satisfied by modelling and simulation fusing the transactional data with the initial reference data. Such modelling and simulation will, thus, only deliver effective future insights and decisions through activity today to build quality into the reference data and to lay appropriate data foundations.

Driven by the need for data of the right quality, establishing the information and knowledge management solution has driven a digital transformation. This solution means affecting all aspects of the data value chain: challenging and adapting internal processes; automating build processes, enabling the use of commercial-off-the-shelf assembly equipment; shifting comprehensively from paper to digital with deployment of information technology infrastructure from shop floor through to planning and control functions; identifying 10% of the ~1000 strong workforce as data stewards, with their responsibilities being to be active analysts rather than data storers; and implementing data management systems and protocols that are also ready for deployment elsewhere in the business. The solution is delivering benefits to all types of user, during both the build phase (today) and the in-service phase (tomorrow) [see Table 3].

Table 3: 'Benefits summary'

	When benefit expected	
Asset users	**Today (build)**	**Tomorrow (in-service)**
Designers	• Improved record and management of technical requirements for build • Faster assurance checks • More accurate asset modelling to simulate design changes	• Faster, more accurate ability to improve and adjust ship design
Builders	• More accessible reference data to enable build • Improved process planning and simulations to inform risk management • Build efficiency improvement • Contract compliance	• Fulfil contractual requirements • Packaged service provision • Reduced risk of rework (estimated at 2-3 times more effort)
Operators	• Improved assurance contract management	• Enhanced reference data baseline to fuse with operational performance data
Organization	• Return on investment • Explainable and adaptable for international customers • Programme assurance	• Transferable, repeatable processes and foundations for export • Competitive advantage for future work • ISO 8000 accreditation

Whilst there is rarely a greenfield data site and investment has been required to achieve digital transformation, the estimate of the return on investment is a multiple of ~2.5 to 1. This return has come from many items including, for example, 136 hours per week being saved just by removing the need for manual transfer of information from one system to another. The investment has also mitigated the risk of downstream costs and not realising the benefits of portable IT investment. The return depends on the data architecture as the foundation. This architecture has also set the conditions for applying new analytical approaches.

A bottom-up approach has been essential to mapping the data flows between systems. This approach enables understanding of the scope of the problem but does not help with prioritising effort and resources. The mapping also cannot guarantee that collection of performance data will generate data of sufficient quality or fidelity to enable high-confidence outputs from modelling and simulation. These outputs are fundamental to ensuring the reliability of the dependent engineering decisions and interventions.

4.0 THEME 3: MITIGATIONS

Proposition: Interventions need to be scoped, prioritized and implemented by using a programmatic approach.

In response to the challenge to improve availability of the 4.5-inch naval gun [27], the responsible engineering team chose a strategy to create a digital twin. This strategy had the objective to generate actionable insight directly from data arising from the gun operating in extremely challenging environments. This insight would enable planning and delivering support service activities both to reduce cost and to improve availability.

Stoppages were affecting performance and availability of the gun. These stoppages seemed, according to evidence in the form of engineering experience and user feedback, to originate in vibrations on the loading mechanism and reports of a specific switch coming loose. This evidence was not, however, being supported by analysis of the available engineering failure data. This analysis found 15 out of 15 of the electronic controllers went back to the customer on the basis of no fault found, despite each controller having a self-reporting failure function. The analysis also found, of the 11 recorded failures for a core component, 70% of the faults remained unknown. Failure data was available with only limited supporting error codes and lacked inputs such as noise and calibration as a result of the conversion of some control unit data to a digital signal.

In this case, the engineering support team had no access to some necessary data (answering the first question: *What data do I need and why?*), while the condition of what was available was below the required quality threshold to inform engineering decisions (answering the second question: W*hat is the condition of my data?*). The team shifted focus to the third question: *What do I do about it?*

As an answer to this question, the team developed a data acquisition solution consisting of 196 analogue channels to capture data from sensors positioned at specific, targeted locations on the asset. These sensors have sampling rates of up to 1 million times per second. This solution also deploys slow-speed digital photography to inform analysis of precise asset function, frame by frame. The solution generates an increased scale and breadth of data that enables identification of component performance failure in minutes. This identification saves the team weeks-worth of traditional fault-finding methods and also predicts failures before stoppages occur.

The team has used the new data set (created automatically by the sensors at the right quality thresholds) to serve as the foundation for a digital twin of the actual physical asset, in other words providing "a dynamic, digital representation of the real entity" [28]. This digital twin has been enabled by a top-down ontological approach that supports both increased modelling and simulation capability and representing a fleet of digital twins of corresponding real-life assets. The digital twin can generate performance predictions according to the condition of one physical asset, providing the proxy information for other assets in the fleet and simulations of scenarios for the through-life performance of those other assets.

Whilst the organizational needs were driven by availability and cost reduction, the resulting engineering interventions were driven by the need for data quality. These interventions have created a digital twin approach that Babcock International is now adapting for a different asset class in the land equipment domain. This adaptation will generate outputs that enable more accurate modelling and drive more insightful simulations of the land equipment fleet.

A programmatic approach was the necessary foundation for effective investment in developing, trialling and operationalising the engineering solution for the navy gun. This approach was enabled by appropriate contractual and commercial mechanisms to provide a combination of sufficient time and resources to address the data quality hazards involved. This combination realized the tangible and intangible benefits in an appropriate timescale.

Whilst investment in addressing data quality is rarely quick, simple or free, this paper describes the practical experience that implementing the principles of ISO 8000 does result in this investment delivering benefits for organizations. Such investment was at the heart of the seminal quotation from Philip Crosby [29]: "Quality is free. It's not a gift, but it is free. What costs money are the 'unquality' things — all the actions that involve not doing jobs right the first time." This quotation is no less true by replacing "quality" with "data quality" [30].

5.0 CONCLUSIONS

5.1 The key challenge

Modelling and simulation practitioners are the guardians of what is typically perceived by the wider organization as a vital but complex black box. This guardianship includes a critical responsibility: wherever possible and feasible, to reduce the risk of "garbage in, gospel out". This risk is, in other words, the wider organization accepting superficial elements of credibility rather than questioning the end-to-end coherence of the data value chain.

George Box provided the significant insight that "All models are wrong, some are useful" [31]. This insight can be further enhanced by the addition of "and all data is a model". This addition emphasises that no data set can be perfect.

This paper has described the practical experience of addressing data quality hazards across projects in an asset and engineering focused Defence company. This experience emphasises, as follows, three key features relevant to modelling and simulation.

5.2 Value

Enhancing data quality directly underpins organizational drivers to improve readiness of assets and cost effectiveness of service delivery. These drivers are satisfied by decisions that are enhanced by insight and foresight generated by effective modelling and simulation. Data quality is the start point and enduring theme of the digital threads that enable a transformed approach to any organization completing its mission. Data only contain value to the extent to which each organization treats the data as an asset to meet identified needs and exploits this asset to fulfil those needs.

5.2 Systematic

The volume, velocity and variety of data sets continue to grow with no obvious end point in sight. Each data set consumes resource to assess condition and address hazards in the data. Such resources are never available without limit.

Organizations need to prioritise what data merits targeted interventions on the basis of rationale justification. This prioritization is done systemically, by applying programmatic approaches including: sponsorship by senior leaders; resource and risk management; and benefits tracking. These approaches are further enhanced by specifying, apportioning and implementing data stewardship responsibilities across the organization. This stewardship seeks to deliver data that are fit for purpose, recognising that imperfections (hazards) are inherent in data.

5.2 Systemic

Neither the approach in this paper nor the underpinning standards are complete, closed solutions to addressing data quality.

Each organization faces a unique combination of need and availability of the resources of people, processes, and systems. This uniqueness drives development of an approach that is proportionate and relevant. The approach will typically be more than just isolated interventions because data interconnections are a complex web. Such an approach is systemic across the organization, delivering benefits and underpinning digital transformation.

In summary, as demonstrated by the pioneers of applying quality approaches within manufacturing [16], achieving quality for data is not an end point but a journey of continual improvement.

REFERENCES

[1] *Phase I report*. Mars Climate Orbiter mishap investigation board. 1999-11-10. Available from: https://llis.nasa.gov/llis_lib/pdf/1009464main1_0641-mr.pdf [viewed 2022-09-19]

[2] International Organization for Standardization. *ISO 8000 (all parts), Data quality*

[3] International Organization for Standardization. *ISO 55000, Asset management — Overview, principles and terminology*. 2014-01

[4] WESTGARTH Tom, CHEN Wen, HAY Graham and HEATH Robert. *Understanding UK Artificial Intelligence R&D commercialisation and the role of standards*. UK Department for Digital, Culture, Media and Sport, London. 2022-05. Available from: https://www.gov.uk/government/publications/understanding-uk-ai-rd-commercialisation-and-the-role-of-standards [viewed 2022-09-19]

[5] KING Timothy, COLACI Raffaella and HAENISCH Jochen. *SIMBASE - fusing product data with modelling and simulation technologies*. In proceedings, 14th PDT Europe, Amsterdam. 2005-09-26 to -28

[6] *Dreadnought class*. Online. Royal Navy. Available from: https://www.royalnavy.mod.uk/the-equipment/submarines/future-submarines/dreadnought-class [viewed 2022-09-19]

[7] *Complex Asset Performance Analytics*. Online. Babcock International Group. Available from: https://www.babcockinternational.com/case-study/capa-t23/ [viewed 2022-09-19]

[8] UK Ministry of Defence. *Data strategy for Defence*. Edition 1. 2021-09. Available from: https://www.gov.uk/government/publications/data-strategy-for-defence [viewed 2022-09-19]

[9] UK Department for Digital, Culture, Media & Sport. *National data strategy*. Policy paper. 2020-12-09. Available from: https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy [viewed 2022-09-19]

[10] UK HM Government. *Global Britain in a competitive age: The Integrated Review of Security,*

Defence, Development and Foreign Policy. CP 403. 2021-03. Available from: https://www.gov.uk/government/publications/global-britain-in-a-competitive-age-the-integrated-review-of-security-defence-development-and-foreign-policy [viewed 2022-09-19]

[11] UK HM Government. *Defence and Security Industrial Strategy: A strategic approach to the UK's defence and security industrial sectors*. CP 410. 2021-03. Available from: https://www.gov.uk/government/publications/defence-and-security-industrial-strategy [viewed 2022-09-19]

[12] UK Ministry of Defence. *Defence in a competitive age*. CP 411, 2021-03. Available from: https://www.gov.uk/government/publications/defence-in-a-competitive-age [viewed 2022-09-19]

[13] UK Ministry of Defence. *Integrated Operating Concept*. 2021-08. Available from: https://www.gov.uk/government/publications/the-integrated-operating-concept-2025 [viewed 2022-09-19]

[14] UK Ministry of Defence. *Digital strategy for Defence*. 2021-04. Available from: https://www.gov.uk/government/publications/digital-strategy-for-defence-delivering-the-digital-backbone-and-unleashing-the-power-of-defences-data [viewed 2022-09-19]

[15] UK Royal Navy. *Digital and data strategy: Part 1: Ends*. 2019

[16] International Organization for Standardization. *ISO 9000, Quality management systems — Fundamentals and vocabulary*. 2015-09

[17] *Quarters 3-4 of Year 3 of CP5, 16 October 2016 to 31 March 2017*. Network Rail Monitor. Office of Rail and Road. 2017-07-20. Available from: https://www.orr.gov.uk/media/16910 [viewed 2022-09-19]

[18] KING Timothy and CROWLEY-SWEET Davin. *Best practice for data quality enables asset management for rail*. In proceedings, IET/IAM Asset Management Conference, London. 2014-11-27 to -28

[19] *The defence equipment plan 2021*. Online. UK Government. Available from: https://www.gov.uk/government/publications/the-defence-equipment-plan-2021 [viewed 2022-09-19]

[20] Defence Equipment & Support homepage. Online. Available from: https://des.mod.uk/ [viewed 2022-09-19]

[21] International Organization for Standardization. *ISO 55001, Asset management — Management systems — Requirements*. 2014-01

[22] KING Timothy, DIXON Richard and GALEOZZIE Raymond. *Decision support for asset management: A Defence perspective*. In proceedings, IET/IAM Asset Management Conference, London. 2013-11-27 to -28

[23] *General Sir Patrick Sanders, Chief of the General Staff, opens the RUSI Land Warfare Conference with his speech*. Online. British Army. Available from: https://www.army.mod.uk/news-and-events/news/2022/06/rusi-land-warfare-conference-cgs-speech/ [viewed 2022-09-19]

[24] *Army Digitalisation: the THEIA programme*. Online. UK Government. Available from:

https://www.gov.uk/guidance/army-digitalisation-theia [viewed 2022-09-19]

[25] *Data science and AI in the age of COVID-19*. The Alan Turing Institute. 2021. Available from: https://www.turing.ac.uk/sites/default/files/2021-06/data-science-and-ai-in-the-age-of-covid_full-report_2.pdf [viewed 2022-09-19]

[26] *Business Modernisation for Support*. Online. Team Defence Information, Bristol (GBR). Available from: https://www.teamdefence.info/business-modernisation-for-support/page/3/?et_blog [viewed 2022-09-19]

[27] *Annual report and financial statements 2021*. Babcock International Group, London. 2021-08-18. Available from: https://www.babcockinternational.com/wp-content/uploads/2021/08/Annual-Report-and-Financial-Statements-2021-singles.pdf [viewed 2022-09-19]

[28] D'AMICOA Rosario Davide, ERKOYUNCUA John Ahmet, ADDEPALLIA Sri and PENVER Steve. *Cognitive digital twin: An approach to improve the maintenance management*. CIRP Journal of Manufacturing Science and Technology, vol. 38 (2022), pp. 613–630. Available from : https://www.babcockinternational.com/wp-content/uploads/2022/09/Cognitive-digital-twin-An-approach-to-improve-the-maintenance-management.pdf [viewed 2022-09-19]

[29] D'AMICO Greg. *Quality is free*. Online. PrintingNews. 2009-05-27. Available from: https://www.printingnews.com/home/article/10265447/quality-is-free [viewed 2022-09-19]

[30] KING Timothy and SCHWARZENBACH Julian. *Managing data quality: A practical guide*. BCS, London. 2020-05-04

[31] BOX George. *Science and statistics*. Journal of the American Statistical Association, vol. 71 (356), pp 791–799. 1976-12