Ministry
of Defence

Data Management
Strategy

TABLE OF CONTENTS

EXECUTIVE SUMMARY

1. Introduction

The Defence Data Management Strategy 2020 sets out the required improvements to the way we manage and use our data that will enable Defence to benefit from the opportunities available through data-driven technologies and ways of working. More effective use of data, information and the systems that manage and process data are vital enablers of both operational advantage and business transformation. New and emerging technologies can provide better capabilities to our operations and greater efficiency in our supporting functions, but success will require us to consider data differently. If we are to deliver improvements at speed and scale, then we must start with managing our data far more effectively than we do today.

2. Current Position

Activity has already begun to transform Defence to deliver core services that will act as the foundations for the exploitation of digital services. The Defence Digital Function is conceptualising approaches to a Single Information Environment and Digital Battlespace capabilities. Our knowledge and understanding of the power of good data is evident through our innovation in equipment and services demonstrated through our growing use and investment in modern data technologies and the delivery of automated services to simplify existing administrative processes. However, Defence needs to grow its data capabilities to ensure we can support the demand and keep pace with our allies[1] and adversaries and continue to be an effective global defence force in the 21st century.

3. Single Information Environment Vision

Driven by the Defence CIO, the Single Information Environment (SIE) vision is for an architected Defence Single Information Environment, ensuring the end user has appropriate access to quality data. A recent SIE review of 100 Defence systems identified the need to develop easy ways to access, exploit and defend our data and to do so through a common technical architecture. However, less than 25% of our systems have data that is automatically discoverable (requires minimal manual intervention) and 33% do not follow international standards for information. This exemplifies the significant challenge in the business and operational exploitation of our data assets with the current data landscape

[1] The new USDOD Data Strategy is expected early 2020. The Canadian Defence Force Data Strategy was released September 2019. Both strategies are seeking similar improvements to the management of data with some objectives aligned, specifically around availability and accessibility of data, improving data governance and data literacy. NATO Data Management Strategy was not available for review.

constraining opportunities to exploit our data through improved services to the front line and our business operations.

4. Quality Assured Data

Defence data must be authoritative, reliable and trusted, with quality assurance by default to fully realise its benefits as an asset to Defence. There are several business drivers for data quality. Increasing the value of Defence data and the ability to use it effectively will improve the decisions we make and improve operational effectiveness. Improving efficiency and productivity will enable our processes to run smoother and faster whether on base or in theatre. Reducing the risk and cost of poor data quality will help in cutting unnecessary waste in the Defence budget and provide better services to the warfighter. This in turn protects and enhances our reputation as an efficient and effective Department. These business drivers support greater efficiency in the way we work and help mitigate reputational damage that can result from inaccurate information and the decisions made from it.

5. Data Driven Culture

If our data is to be more exploitative now and for the future the department needs to think differently about how we manage and use it. A cultural change in how we manage, protect and share data needs to be established across Defence, creating an environment for our existing, new and emerging technologies to deliver greater operational advantage and more trustworthy information and insight to decision-makers.

6. Suitably Qualified Individuals

Data, analysis and the development of data-driven technologies will demand our people to have the right skills with clear development paths. This will be key to achieving game-changing outcomes. We need to ensure our people, whatever their grade or rank, have the right support, training and education opportunities in place with the opportunity to grow their existing skills and experience and where applicable professionalise in modern technology delivery. This training and education will need to encompass all aspects of the data lifecycle, from data entry through to analytics and innovation, recognising that without good, accessible data our aspirations will not be achieved. Creating this data literate workforce requires an understanding of the data skills that all our people need to perform their role.

7. Collaborative Approach

Our delivery organisations need to meet the demand for better data, ensuring our data is managed to the right quality, available to the right people and accessible in the right ways that will enable innovation and deliver the desired outcomes. This will require close collaboration across all organisations. It would be unwise to consider this establishment of a data-driven organisation as a simple undertaking and without its challenges both financially and culturally, but to be successful we must all meet these challenges, recognise the importance of trustworthy and accessible data and the opportunities this will bring.

8. **Conclusion**
Effective data management is a prerequisite to our transformation and future operational success, it cannot be ignored or under-valued. If we continue to manage our data as we do today, for single purposes and with little control and access, the levels of change required will not be realised and we will fail to deliver the data driven Defence capability and run the risk of allowing our adversaries to gain a strategic advantage. We need to change our thinking and our approach to how we manage and use data now, and by doing so deliver the opportunities and outcomes we all seek.

SCOPE

9. This strategy shall apply to all Defence Data, whether this be Core Defence Data or Non-Core Defence Data. The difference between these classifications of data is the extent to which they are shared across Defence. Non-Core Defence Data is generally not shared outside its originating area but should still be treated with the same care and attention as Core Defence Data.

10. **Core Defence Data** is the data that supports business defined critical Defence processes and operational requirements, and which is shared at the Defence level and/or shared externally or between Top Level Budgets (TLBs), Front Line Commands (FLCs) and/or Arm's Length Bodies (ALBs).
Note: Core Defence Data can encompass all data types, including structured, semi-structured, unstructured and images. Core Defence Data is used extensively across the department by many organisations and in support of multiple business processes. This high degree of use and severe impact if managed poorly, identify it as a priority that demands effective management. Examples of Defence Data include People, Organisation, Cost, Location, Asset and Project data.

11. **Non-Core Defence Data** is the data that supports intra-organisational reporting, business and operational process. By its nature it never leaves the Function, TLB, FLC or ALB where it was created and is designed for single/few purposes.

Note: Non-Core Defence Data can encompass all data types, including structured, semi-structured, unstructured and images. Although by its nature it is not a pan-Defence priority, this data could be equally important to a single organisation as Core Defence Data, and therefore should be assessed by the business for its internal value. If deemed sufficiently important, or a priority, it should be managed with the same rigor as Core Defence Data. Examples include system administration, internal reporting data or transactional data created and used solely for the purposes of supporting a single organisation’s business processes.

12. Defence Data priorities will be set by the Defence Functions, TLBs/FLCs and ALBs, working collaboratively to agree an order for improvements. The expectation will be for many priorities to be tackled at the same time, increasing data capabilities across several priority areas.

VISION

A Defence Data Environment (DDE) and culture that delivers timely, accurate and trustworthy data, managed by business Subject Matter Experts (SMEs) and accessible to all who have a right to use it, complimented by strong governance and a skilled workforce that supports innovation and exploitation.

13. Achievement of the Vision will see our data being managed and made available through a network of data service providers that manage the quality and accessibility of our data on behalf of our data owners and stewards, through which defence will be able to see a complete data catalogue for defence.

INTENT

14. The Defence Chief Information Officer (Defence CIO) intent is to create transformative digital, data and information capabilities that enable sustainable military and business advantage, that is secure, integrated, easy to use and delivered at scale and pace to the front line.

15. Defence will be a more data-driven organisation, enhancing its existing data capabilities and establishing new services to create opportunities for wide data exploitation. The desired result will see the removal of the current hurdles relating to data access and poor data quality with the establishment of authoritative data[2] sources, owned and managed by organisations capable of producing and maintaining data of sufficient quality to meet the needs of Defence. Achieving the vision and meeting the intent will see simplified data sharing, improved data quality, stronger accountabilities and responsibilities for data and the acceleration of the establishment of common standards, with the removal of the need to hold duplicate data (unless the consumer is authorised to do so), with its attendant costs and constraints on exploitation and decision-making.

16. Data and the expertise that manage it reside in many organisations across Defence. Not one organisation has the capability to deliver the totality of the requirement, therefore a federative approach to deliver the outcomes will be required with the Functions, Top Level Budgets (TLBs), Front Line Commands (FLCs) and Arm's Length Bodies (ALBs) all delivering capability and contributing to the outcomes. For some solutions, where there is no single organisation capable of delivering the requirement, centralised capabilities may be required. These capabilities will be developed through collaboration and managed by the business SMEs.

[2] The primary characteristics of authoritative data are that the data must be managed to recognised governance and quality standards and accessible by those who have permission to use it.

DEFENCE DATA MANAGEMENT CORE PRINCIPLES

17. A set of Data Management Core principles have been written to underpin this strategy document. The principles are statements of intent with executive level support intended to shape the framework for all Data Management activities such as policies, standards and processes. The following principles shall apply to all Defence data management:

a. All Defence Data is owned;

b. All Defence Data is secure;

c. All Defence Data is managed through life;

d. Only trusted Defence Data is exploited[3];

e. All Defence Data is fit for purpose;

f. All Defence Data will be shared appropriately;

g. There is a single version of the truth;

h. Our people are trained to use data effectively.

A full description of each of the eight Defence Data Management Principles can be found at Appendix A.

18. A lack of adherence to these core principles is considered severe and damaging to the consistent use of Defence Data across the Enterprise. Failure to comply could result in the removal of access to Defence Data and further remedial actions.

CONTEXT AND GOVERNANCE

19. The Defence Data Management Strategy takes its direction from numerous Government strategies that seek improvements in data use[4]. Some are referenced in Diagram One below, however the demand for improved data practices are evident across many initiatives that impact Defence business. Internally the Defence Data Management Strategy takes direction through Defence Transformation and directly the Defence Digital Function's Enabling Warfare in the Information Age (EWITIA). Through its strategic objectives it provides the data management support for these higher-level strategies. Where there is need for Defence-wide action this will be directed through central policy and guidance which will be made available through Joint Service Publication (JSP) 441 - Information, Knowledge, Digital and Data in Defence.

20. The Defence Data Management Strategy 2020 provides the Defence context for and is aligned with the Data Management Association and the DAMA Body of Knowledge (DMBOK), which provides guidance to organisations in implementing data management improvements. The DMBOK is freely available to all Defence staff through the Defence

[3] In particular, for management information and analytical purposes.
[4] UK Gov Transformation Strategy, UK Gov Digital Strategy, UK Gov Open Data Strategy, UK Gov National Data Strategy.

Library Service via DEFNET. Defence Functions, TLBs, FLCs and ALBs are strongly encouraged to develop their data management implementation plans and policies by drawing on the DMBOK for technical guidance. Specific guidance on data governance for Defence, a prerequisite to all data management activities, will be provided through a Defence Data Governance Target Operating Model and supporting Data Quality Guidance provided by the Data Centre of Expertise (CoE)[5], established within Defence Digital[6]. Diagram One below summarises the relationships to higher strategy and supporting guidance.

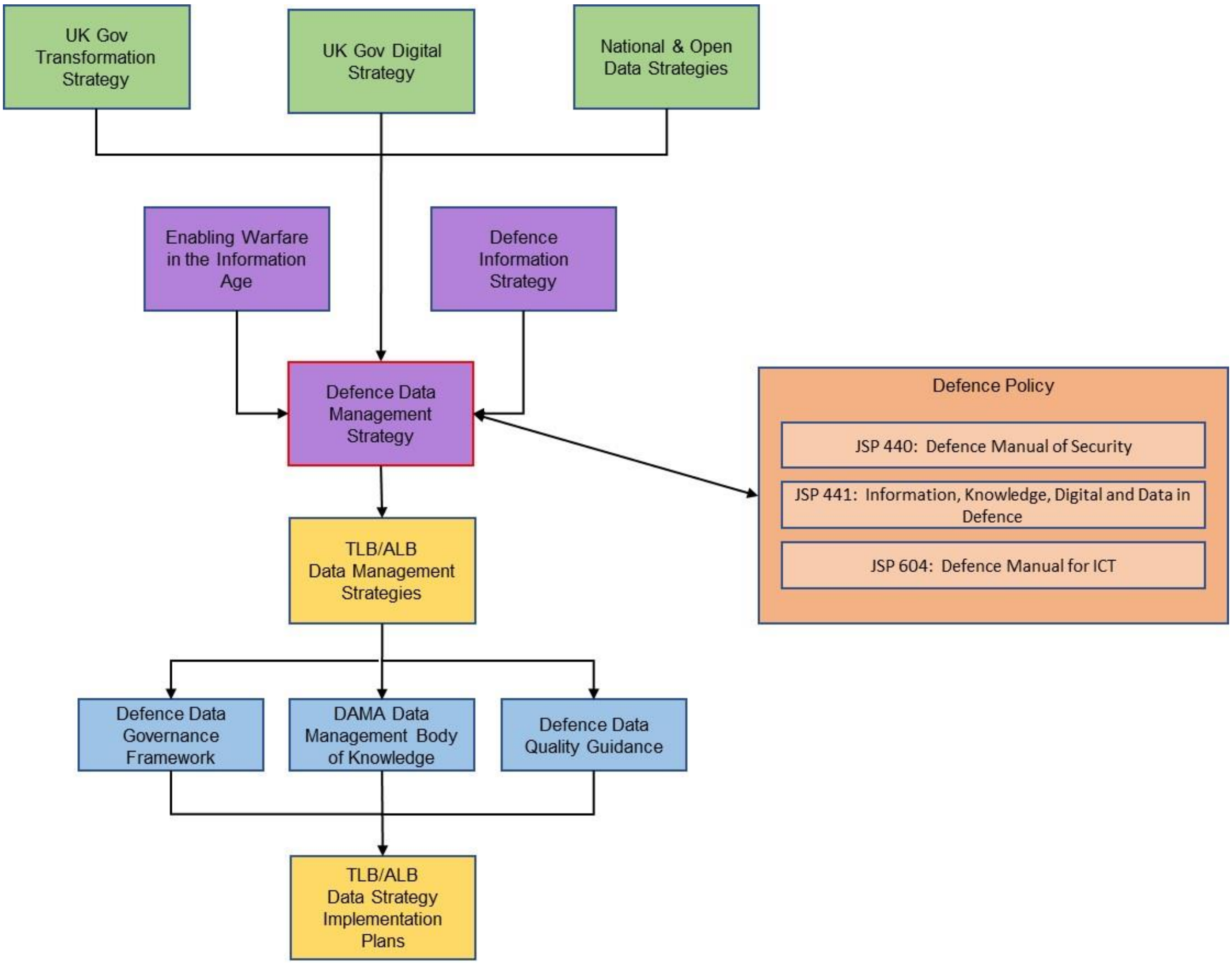

UK Gov Transformation Strategy
UK Gov Digital Strategy
National & Open Data Strategies
Enabling Warfare in the Information Age
Defence Information Strategy
Defence Data Management Strategy
Defence Policy
JSP 440: Defence Manual of Security
JSP 441: Information, Knowledge, Digital and Data in Defence
JSP 604: Defence Manual for ICT
TLB/ALB Data Management Strategies
Defence Data Governance Framework
DAMA Data Management Body of Knowledge
Defence Data Quality Guidance
TLB/ALB Data Strategy Implementation Plans

Diagram One

21. The Data CoE are custodians of this strategy and will oversee the delivery of the Strategic Objectives. It will play a key role in orchestrating the activities and reporting progress to the 2* Data Governance Board (DGB). The DGB will ensure that the key tenets of the Defence Data Management Strategy are applied and adhered to. The board will act as a point of escalation for the empowered Data CoE and Defence stakeholders.

22. Diagram Two below explains the direction, escalation and reporting routes for the DGB. It should be noted that this structure is likely to evolve over time as we implement the Data Management Strategy and identify the need for change.

[5] The Data CoE is part of the Defence Digital organisation

[6] Defence Digital is the new organisation established through transformation of Information Systems and Services (ISS)

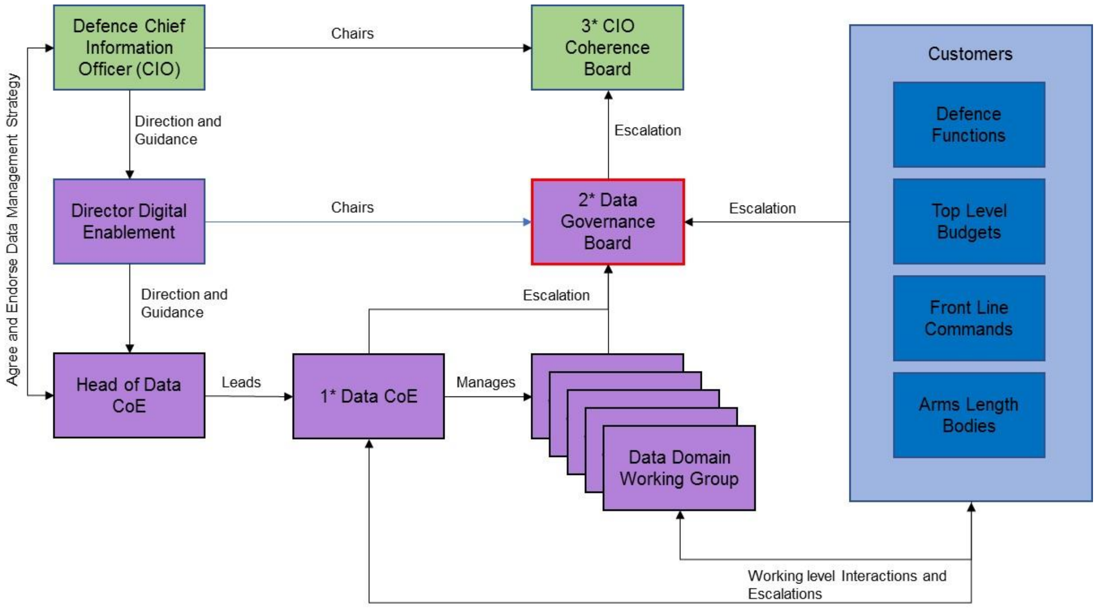


Diagram Two

23. The DGB takes its direction and guidance from the Defence CIO and is Chaired by the 2* Director of Digital Enablement. Under the DGB several Data Domain Working Groups will be established to drive priorities and deliver the outcomes required to establish and maintain a more data-driven Defence. In addition, a Defence Capability Management Working Group will be established that will be chaired by the Data CoE. This group will provide transparency and clarity around project activity and manage the development of new capability where and when required. DGB membership will be made up of senior data subject matter experts from across the Defence Functions, TLBs, FLCs and ALBs. It is essential that people who know the data from a business perspective and have a good working knowledge of the data uses and challenges in their area attend the DGB.

STRATEGIC OBJECTIVES, OUTCOMES AND BENEFITS

24. To realise the potential in our data and the opportunities to innovate and fully exploit it, the department will need to gain the maximum value from the data we own, including the data managed by our allies and partners in industry and do so at pace. This Data Management Strategy describes the data management activities necessary for Defence to become more data-driven and will see every organisation have a defined role in achieving the shared objectives. The foundation activities to a more data-driven Defence have been defined into seven Data Management Strategic Objectives (SOs). These are defined below with detailed descriptions provided later in this strategy.

a. **SO1**: Improve the availability and accessibility of Defence Data;

b. **SO2**: Implement data governance at all levels of the department to ensure the accountabilities and responsibilities for the upkeep of our data are established and upheld;

c. **SO3**: Improve the quality and veracity of our data;

d. **SO4**: Drive the consistent use of decision-making data across the department to improve coherency in the information produced from it;

e. **SO5**: Ensure the integrity, confidentiality and security of data;

f. **SO6:** Improve the knowledge, education and behaviours or our people to ensure data is managed as a strategic asset;

g. **SO7**: Enable the exploitation of new data-driven technologies to meet information, business and operational challenges.

25. Implementing the strategy and delivering these objectives will deliver the following outcomes and benefits. Collectively these outcomes will provide greater opportunities to fully exploit our data.

KEY OUTCOMES	BENEFITS
Authoritative data is available to all who are authorised to use it, wherever they need to use it	• Our people have access to the trusted, authoritative data they need to be effective in their roles • Services can be automated and existing administrative processes simplified • More effort is spent on supporting the warfighter rather than debating if data is correct or not
There are formal governance structures in place across MoD to manage our data as an asset	• Our core datasets are all owned with accountable individuals to ensure that issues are addressed before they impact on business or operational outcomes • Data is actively managed through life, reducing waste and ineffectiveness in what we deliver • The effective management of data is woven in to how we operate so that best practice becomes the norm
Defence data is maintained at an acceptable quality and is trusted by those who use it	• Authoritative data that is accessible, maintained and used consistently will build greater trust in our decision-making, benefiting internal and external customers • There will be a decreased risk of consuming incomplete or incorrect data, thus avoiding fixing problems downstream • Improved data quality and integrity, will lead to increased efficiency in our data driven processes

KEY OUTCOMES	BENEFITS
Valued, coherent data is used effectively across the Department	• The implementation and adherence to data standards will reduce the time spent managing system-to-system interfaces and effort of maintaining non-authoritative datasets • The business will be provided with opportunities to focus on value-adding outcomes rather than maintaining legacy business processes • A single view will exist of our core datasets such as People, Location or Asset removing the issues related to conflicting data in our systems
Data enables us to make better operational and business decisions	• Decisions will be made on authoritative and trusted data • The data presented in reports will be more accurate and less open for interpretation • MoD will gain new insights from our data holdings to inform decision making across the Department, increasing both business efficiency and operational effectiveness
Our people have the knowledge and skills to use Defence data effectively	• Knowing the right thing to do with data drives consistent treatment, improving quality and generating trust to facilitate the wider sharing of data across MoD • Our people are skilled in using data, promoting a corporate culture of data innovation • If appropriate our people will have recognised data management qualifications, demonstrating MoD's commitment to a data literate workforce

IMPLEMENTATION

26. The implementation of the Strategic Objectives will be led by the Defence Digital Function who will establish a collaborative pan-Defence approach to deliver the outcomes, release the value of our data and realise the benefits. Priority will be given to Core Defence Data, the data most exploited (used at the Defence level) but can be equally applicable to TLB/FLC/Arm's Length Bodies (ALBs) specific data; i.e. Non-core Defence Data.

27. Data Literacy training and education programmes for staff will be established to up-skill our existing people and attract specialists to work in Defence and support the drive for information superiority. This requires the ability to assess and develop data literacy skills through embedding a competency framework.

28. New platforms will be built to improve the availability and accessibility of Defence Data. These will be managed by the business SME organisations and supported by Defence Digital through an established Data Governance capability and Target Operating

Model including the 2* Data Governance Board and supporting Data Domain and Capability Working Groups.

STRATEGIC OBJECTIVE ONE

Improve the availability and accessibility of Defence Data

29. Under the authority of the Defence CIO the Data CoE will drive the identification and availability of Defence Data on a priority basis. These data services will be domain specific and service the Defence customer through a 'manage once - use many times' principle. These services will be built through pan-Defence collaboration and managed by appropriate Subject Matter Experts across the department. Foundation business rules to the use of Defence Data will be agreed collectively, driving greater consistency in Defence Data use and transparency in process.

30. Where applicable the centralisation of data services will be undertaken to alleviate constraints around access and quality. This approach is used extensively in large organisations where data can be uncontrolled and incoherent through the instantiation of multiple sources and the development of bespoke data-feeds that are established over long time periods. Through the establishment of centralised services more efficient processes can be realised and sub-optimal business processes reduced, encouraging greater efficiency and effectivity in data use.

31. Master and reference data initiatives will be undertaken following good practice and recognised industry standard approaches to ensure the outcomes are robust and enduring. The Data CoE will lead on establishing these services and work with the TLBs, FLCs and ALBs to ensure the outcomes are beneficial to Defence and are managed to the expected standards. These services will be built through pan-Defence collaboration and managed by appropriate Subject Matter Experts across the department.

32. System-to-system data sharing will follow MOD information exchange policies. Proprietary solutions, particularly those solutions that use bespoke data schemas, are a main cause of ineffective interoperability. Therefore, Defence must develop standard data schemas that support the sharing and interoperability of Defence Data across the department, and with our partners in industry, providing a level of coherence and consistency across applications and reporting. Where possible these standards will be adhered to and deviate only when it is deemed more beneficial to wider Defence needs.

33. Due note must be taken of MOD, Government and international military taxonomies and standards (such as the NATO C3 taxonomy) and where appropriate, these

taxonomies and standards must be adopted, unless there is strong justification for not doing so endorsed by the Defence CIO/CDO.

34. In line with Government Digital Services direction Defence will adopt the Open API Standard[7] (formerly Swagger). The Open Application Programming Interface (API) Standard uses REST style APIs which are called using defined instructions via standard HTTP requests. Adopting a recognised standard is essential to allow the creation of a library of re-usable APIs that can work together. Use of an open API standard, supported by a cloud hosting approach, will:

1. Maximise the reuse of data by decoupling it from the underlying systems and services.
2. Support effective cyber Defence through monitoring access to individual services and assessing excesses call levels.
3. Supports a data-driven approach where data needs to be consistent, available and distributed.
4. Support the war-fighter through presenting the required information to devices when removed from information sources.
5. Creates the opportunity to access valuable data sources that are available outside of Defence and reduce the cost of replicating these sources inside MoD.
6. Support the rapid change an innovation through making data more accessible, removing the hurdles of data access through more conventional means.

[7] https://www.openapis.org/

STRATEGIC OBJECTIVE TWO

Implement data governance at all levels of the department to ensure the accountabilities and responsibilities for the upkeep of our data are established and upheld

35. Data Governance is defined as the service of authority and control (planning monitoring and enforcement) over the management of data assets. Data Governance guides all other data management activities. Its purpose is to ensure that data is managed properly, according to policies and best practices. While the driver of data management overall is to ensure an organisation get value out of its data.

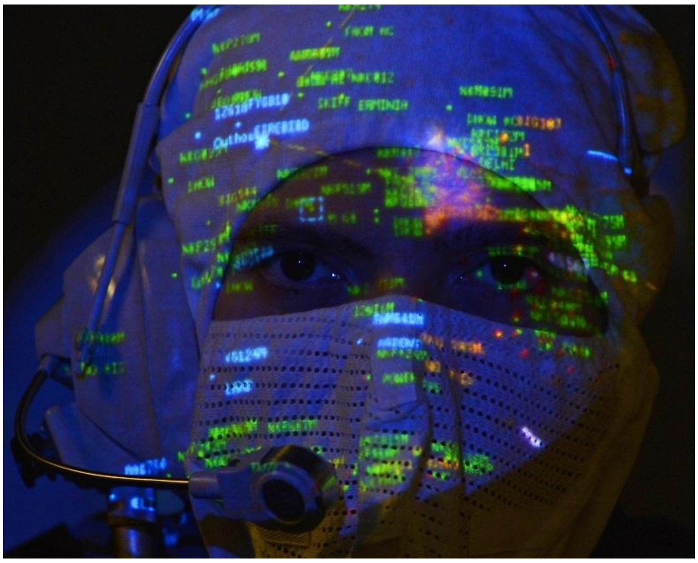
36. Accountabilities and responsibilities for the maintenance and management of Defence Data will be clear with data ownership, stewardship and custodianship introduced into data management and maintenance processes. Through data governance Defence Data quality will be improved as we adopt a Data Quality Management (DQM) approach and embed this as part of normal business practices. These activities will be supported by the Data CoE who will work under the authority of the DGB and the Defence CIO. The Data CoE will ensure that the key tenets of MOD's Data Strategy are applied and adhered to, acting as a point of escalation for the DGB.

37. The Data CoE will have oversight of data governance implementation and provide a Defence Data Governance Target Operating Model to support it, working with the Defence Functions to develop policies that will ensure the management and upkeep of Defence Data.

38. Industry, under current contracting arrangements manage large volumes of data used by MOD. Explicit recognition of this role of industry is necessary to bring focus on the requirement to manage that data that has been inaccessible and/or of poor quality in the past. Data governance must include the requirement to manage the relationship between MOD and industry and policy will be developed to cover how these obligations and responsibilities regarding data will be encapsulated in contracts.

STRATEGIC OBJECTIVE THREE

Improve the quality and veracity of our data



39. All Data Management disciplines contribute to the quality of data, and high-quality data that supports organisational goals should be the goal of all data management disciplines. When data quality is managed appropriately this breeds more trust in its use and the decisions that are made from it. Maintaining high quality data in Defence will require cross-functional commitment and coordination. We will align our Data Quality approach with ISO 8000[8], the global standard for Data Quality. ISO 8000 describes the features and defines the requirements for the standard exchange of Master Data internally or between organisations.

40. Formal data quality processes are similar to the quality processes for other products. They include managing data through a life-cycle, from creation to disposal. This approach requires the implementation of data quality management and the resources to support it. This requires engagement with technical and business professionals to drive the work to apply data quality techniques to data to ensure it is fit for its intended purposes.

41. If data is to be trustworthy, particularly to support the decision-making process, it is vital that data is an accurate reflection of reality (the truth); that it is relevant, timely and accurate and therefore can be trusted.

42. All Defence Data must be subject to data quality controls beginning at creation or the point of entry for each data item and carrying on throughout the life of the data. These data quality controls and measures must be in accordance with Defence data quality assurance policies, standards and protocols, and their effectiveness must be measured according to standard criteria across Defence. Note that these will be developed as part of the implementation of the strategy.

43. Defence, like many organisations, experiences the impact of poor data quality daily. In decision-making this is evident in our inability to understand and provide the answer to basic business questions quickly and accurately. Our inability to count our people accurately, due to a lack of standards applied to the data across the many systems, and our inability to aggregate the data easily are topical examples. Ensuring the right equipment and information reaches the right people in the correct location relies on the existence of accurate and authoritative data being available otherwise our operational readiness will be impacted.

[8] Further information on ISO 8000 can be found in 2017DIN05-011 and JSP 441 - Managing Information in Defence

STRATEGIC OBJECTIVE FOUR

Drive the consistent use of decision-making data across the department to improve coherency in the information produced from it

44. The use of common data definitions, schema and taxonomies is fundamental to ensuring that Defence Data is coherent and interoperable, within Defence, within Government and when on multi-national operations (such as when part of a NATO force). For example, the NATO Stock Number (NSN) uses common data definitions to ensure interoperability across multi-national support chains.

45. For the content and meaning of data to be consistent, wherever it is used it is most important that the data, once extracted from the Authoritative Source (the single source of the truth), is not altered in any way. In particular, local versions of authoritative data should not be created, unless there is a specific business, operational or security reason for doing so, and the duplication of the data is authorised by the relevant Data Policy Owner[9].

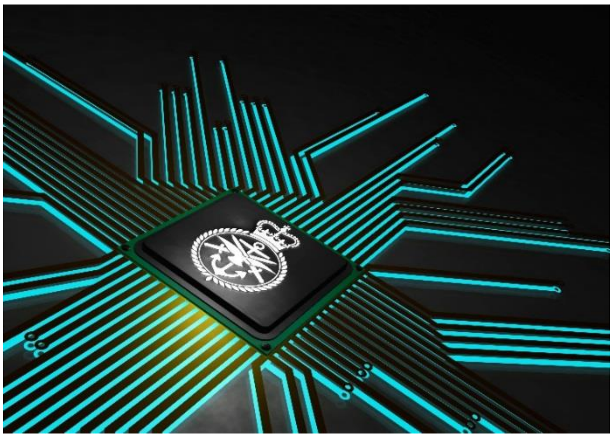
46. Defence Data will be clearly defined when made available to those who consume it. These definitions will be developed collaboratively through the Data Domain Working Groups and will be stored and maintained in a corporate data dictionary. This dictionary will be managed by the Data CoE with the respective Data Stewards maintaining the content. This will include clear instructions on how the data is to be used and explain any limitations on its use.

47. Organisations who consume Defence Data will demonstrate compliance with the standards developed through the Management Information produced from it and in the systems that consume it.

[9] The Data Policy Owner role and other Data Governance roles are defined in the supporting Defence Data Governance Target Operating Model.

STRATEGIC OBJECTIVE FIVE

Ensure the integrity, confidentiality and security of data



48. Defence data must be protected from unauthorised activities[10] in accordance with Defence threat and risk appetite models[11]. Full guidance in this regard is contained in the MOD Technical Information Assurance Architecture[12] and JSP 440, and this guidance must be adhered to. Of equal importance is Cabinet Office and Government Digital Service data security policy which also must be adhered to wherever applicable.

49. Those who are accountable and responsible for data should understand who has access to their data and be able to manage those permissions effectively. Gaining access to data should be appropriately controlled and measures should be in place to ensure access is removed once it is no longer needed. Protecting our data from accidental or unintentional corruption by our own people is a vital element to ensure the quality of our data.

50. The ever-expanding use of Smart Technologies and the vast amounts of data now available in the public domain, combined with an unstable and volatile world, has created an intelligence and operational requirement for access to multi-classification datasets, and these requirements span from TOP SECRET through to open sources in the public domain. The aggregation of this data has become essential to effective intelligence and military operations and directly supports Defence activities.

51. We need to establish multi-classification solutions to data hosting and portability that will allow us to exploit data across more effectively, challenging the existing approach to security and the divisions that exist between public and private domains.

52. The need for multi-nation data sharing, understanding this activity could be tactical or operational, is essential to multi-national working and collaboration. We need greater opportunities to share data with our partners and allies, and in turn benefit from access to their data to enhance our own capabilities.

53. It is recognised that JSP440 needs to support these requirements and a review of the policy to better reflect these requirements is needed. The Data CoE, working under the direction of the DGB, will champion the approach to ensure future versions of JSP440 account for these requirements, remaining cognisant of the levels of protection required to protect against threats to data loss that modern data sharing requirements invite.

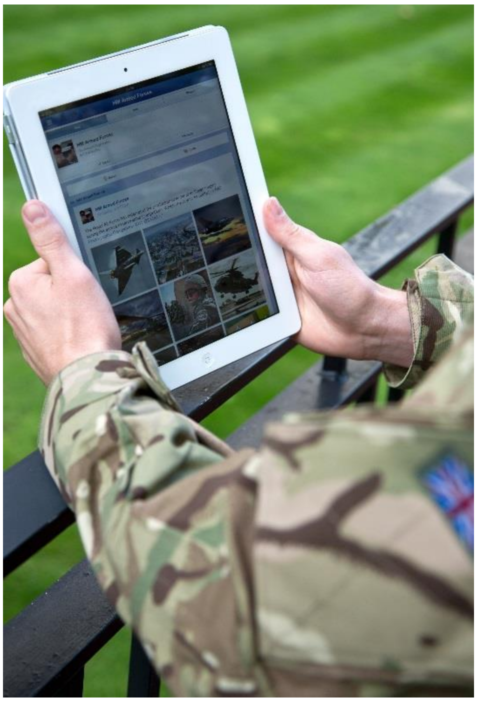
[10] As defined in JSP 604, Part 1 Vol 3, Annex G para 1.

[11] Protection is relative to risk appetite. Data placement within the MoD Trust Model as defined in JSP 604 Part 1 Vol 3 must be a risk decision made by Information Risk Owners.

[12] JSP 604, Part 1 Volume 3.

STRATEGIC OBJECTIVE SIX

Improve the knowledge, education and behaviours of our people to ensure data is managed as a strategic asset



54. In order for Defence to gain information advantage it must have a data literate workforce that can read, write and make decisions confidently using data. The Digital Academy will deliver a range of learning interventions which will improve the maturity of data literacy for all staff across Defence. This will enable Defence staff to have the confidence and capability to make data-enabled decision so that they can:

- Work and use data and technology appropriately to drive insights
- Understand digital, data and technical concepts and apply them to what they do
- Develop and follow MOD processes, systems and services simpler, faster and better

55. The skills and behaviours of our people must improve to meet the need for better management and use of our data, whether that is through the management and maintenance of data in systems or through the exploitation of data through Management Information, Statistics and the development of Automation and Artificial Intelligence capabilities. The empowered Data CoE will establish an active community of people who undertake self-motivated learning which they will then share with their colleagues. Communities of interest for Digital, Data, Artificial Intelligence, Data Science, and Automation will be created and expanded to include all Digital practices as the initiative matures. A culture of collaboration and data sharing will be encouraged as new ways of working are embedded throughout the programme.

56. The concept of a Digital Academy will provide digital and data upskilling opportunities through a range of initiatives and learning interventions aligned to the Defence Digital Upskilling Strategy. Opportunities to professionalise staff in data management will be provided through DAMA UK membership and access to DMBOK which will also form the backbone of data management activities driving a professional and consistent approach to managing our data and driving good practices across the department. DAMA provides members with the option to gain the Certified Data Management Professional (CDMP) qualification. This is something we should encourage relevant staff to work towards as part of our efforts to increase Data Literacy across Defence.

STRATEGIC OBJECTIVE SEVEN

Enable the exploitation of new data-driven technologies to meet information, business and operational challenges

57. Good data management is not an end in itself, but is the foundational layer to data integration, interpretation and data-driven decision making. An effective data management capability enables data consumers such as analysts and data scientists to discover, trust, access, interpret, and reuse data. Defence needs to build analytics data pipelines on a strong data management foundation. Establishing strong data management will ensure that time and effort is spent more on deriving insight from data, than on data wrangling and preparing data for analysis. It will also ensure that trust in analytical products is established not just through the adoption of sound scientific and statistical methodologies, but by ensuring that data semantics[13], quality[14], and provenance[15] issues are understood.

58. Training an Artificial Intelligence (AI) system on error-strewn data, unrepresentative data, or datasets representing bias can lead to poor results due to the dataset not containing clear patterns for the model to explore when making a prediction, or the dataset containing clear but accidental or discriminatory patterns, resulting in poor predictions, incorrect conclusions, false outcomes and bias.

59. Just about any data will have technical errors, logical errors or semantic errors. These can affect certain algorithms and functions that aren't equipped to deal with them. For instance, supervised machine learning involves a pre-processing stage to improve the quality of the training data - the 'ground truth' from which the model 'learns'. If input data contains errors or biases, the AI system will reproduce or even amplify that bias. Data scientists clean the data, by identifying and imputing missing values, parsing dates and numbers, correcting character encodings, matching similar but not identical values, and dealing with outliers and suspicious values. They check that the training data set is representative of operational or administrative data, whether it over or under-represents certain cases or contains spurious correlations. They support the Defence staff that rely on their analysis to know that their data is reliable, secure, and that conditions for processing are understood.

[13] Semantics: Data must be given commonly agreed and understood definitions, in an agreed ontology, and semantic relationships between data, business concepts and data domains is clear.

[14] Quality: Quality dimensions (a combination of accuracy, completeness, uniqueness, timeliness, validity, relevancy, representativeness, sufficiency or consistency) are understood, measured, remediated, and reported – so it meets minimum thresholds for quality for data consumers.

[15] Provenance: Sources of data and their lineage is understood, including the aggregations and transformations.

60. Understanding data quality or provenance is an important part of ensuring that machine learning systems are accurate and work well for users, and that systems are explainable. From the data protection principles of accountability and transparency, the GDPR requires data controllers to provide meaningful information to the data subject about the logic involved in automated processing and profiling, as well as the significance and the foreseen consequences of processing. The quality and provenance of the data used for training algorithms is one important dimension of being able to understand and provide meaningful information to data subjects.

61. Defence will succeed in implementing and scaling AI if it establishes a common data foundation underpinned by:

a. A strategic overview and searchable catalogue of all Defence data assets (with appropriate metadata that helps users to easily find data), organising data collection, consolidating data centres, and implementing quality assurance;
b. Integration of structured data holdings with internal and external semi-structured and unstructured data sources[16], and the technology, process and people to extract statistical and semantic relationships in unstructured data;
c. Defining accessibility rules for sharing the data, and rules and processes for data exchange, while ensuring the full traceability and provenance of the data is coherent and preserved;
d. Clear accountabilities and responsibilities for the maintenance and management of Defence Data driven by data owners and stewards into data management and maintenance processes;
e. A data culture which drives an understanding of data use, and compliance with prevailing personal data legislation (DPA and GDPR);
f. A hybrid cloud architecture by which data can be stored, processed, shared and securely exposed through a library of published and deployed APIs;
g. Validation, verification and accreditation (VV&A) of the data used to train AI, as much as the need for VV&A of models.

62. The Single Information Environment (SIE) will architect the Defence Single Information Environment, and the Data CoE will provide through its services and its data engineers and data scientists (as part of a cross functional and pan-Defence team) access to quality data and a common foundation of reusable data engineering and analytical tools, the adoption of standard policies for application programming interfaces, common algorithm libraries, the reuse of code components and repositories for open source projects, and documentation, frameworks and standards – underpinned by modern cloud services.

[16] Managing unstructured data requires different approaches, priorities and capabilities than structured data. Unstructured data does not have a pre-defined data model and is increasingly heterogeneous and of varied quality.

APPENDIX A – DEFENCE DATA MANAGEMENT PRINCIPLES

Principle	Description
All Defence Data is owned.	All Defence Data will be subject to formal governance processes in accordance with the Defence Data Governance Target Operating Model and comply with supporting Defence level policies. Ownership and accountability for the data and how it is defined and managed must be taken on by the business users who consume it rather than by its custodians in IT.
All Defence Data is secure	All Defence Data must be secured and protected in accordance with Defence data security and protection policies[17], and relevant government security and protection policies[18] and legislation. Data is safeguarded from unauthorised access, whether malicious, fraudulent or erroneous.
All Defence Data is managed through life.	All Defence Data must be formally managed from creation to disposal. This management process is known as Through Life Data Management (TLDM).
Only trusted Defence Data is exploited[19].	Wherever possible, only data from authoritative sources shall be used for exploitation[20] purposes. Competing, unauthorised versions of data must not be created or used unless authorised by the relevant Data Policy Owner[21]
All Defence Data is fit for purpose	Data produced and reported must be of acceptable quality and meet the business need for which it is intended.
All Defence Data will be shared appropriately	Defence Data will be made available for sharing (using recognised standards and consumable formats[22]) with authorised consumers for exploitation purposes.
There is a single version of the truth.	Defence Data provided to data consumers shall not be modified by those consumers unless authorised by the appropriate authorities responsible for the data concerned. Changes required to Defence Data will be made in the authoritative source systems to promote consistent use and preservation of the "single version of the truth". Modification of the data, where authorised, will remove its authoritative nature, unless those modifications are incorporated into the source system through an approved process.
Our people are trained to use data effectively	Relevant training and education on all aspects of the data lifecycle will be available to our people to ensure they have the data skills needed to perform their role.

[17] Data placement decisions within the MoD Trust Model must be made against the risk posture that represents the data as defined by the MoD Technical Information Assurance Architecture. (JSP 604 Part 1, Volume 3).

[18] See https://www.gov.uk/data-protection/the-data-protection-act.

[19] In particular, for management information and analytical purposes.

[20] In particular, for management information and analytical purposes.

[21] The Data Policy Owner role and other Data Governance roles are defined in the supporting Defence Data Governance Target Operating Model.

[22] Examples include (but not limited to): JSON, CSV, XML and ODF.

APPENDIX B – GLOSSARY

Term	Definition
Data Management	Data Management is the development, execution and supervision of plans, policies, programmes and practices that control, protect, deliver and enhance the value of data and information assets. Evidence of good data management includes trustworthy data flowing from authoritative sources to users through communication and information systems, using common formats and open standards.
Data Governance	Data Governance (DG) is defined as the service of authority and control (planning monitoring and enforcement) over the management of data assets.
Core Defence Data	Core Defence Data is the data that supports business defined critical Defence processes and operational requirements, and which is shared at the Defence level and/or shared externally or between TLBs, FLCs and/or ALBs.
Non-Core Defence Data	Non-Core Defence Data is the data that supports intra-organisational reporting, business and operational process. By its nature it never leaves the Function, TLB, FLC or ALB where it was created and is designed for single/few purposes.
Data Governance Board	2* Data Governance Board (DGB) with representation from the Defence Functions, TLBs, FLCs and ALBs.
Data Policy Owner	DOs are responsible for setting the priorities and developing the policies that will ensure the correct actions are carried out. They also monitor progress and report to the DGB on their progress to improve priority data for the users.
Domain Data Steward	Domain Data Stewards are the experts on the data. These are usually people who understand the business processes the data is used in/for and they have a degree of experience in its use throughout the business domain and possibly into other domains.
Executive Data Stewards	Executive Data Stewards (EDS) are accountable to the DO for managing improvements to the data and reporting on progress.
Co-ordinating Data Stewards	Where EDS responsibilities span many different data types it may be necessary to appoint a Coordinating Data Steward (CDS) to organise the work on their behalf.
Data Custodian	Individuals who understand the systems and technologies used by the business to process the data and can advise on resolving problems that arise.
Data CoE	The Data CoE works under the authority and governance of the Defence CIO. Through the Defence CIO's intent the Data CoE will direct all Defence-level data management activities.