Clinical Risk Management Data Safety

Published 27 September 2018

Information and technology
for better health and care

Copyright © 2018 NHS Digital & SCSC
The Health and Social Care Information Centre is a non-departmental body created by statute, also known as NHS Digital.

Executive Summary

Health Information Technology (IT) Systems are defined as products that comprise hardware, software or a combination of both. It is therefore natural for the Clinical Risk Management (CRM) process to consider, as part of its patient centric hazard identification and risk assessment process, hazards to patients arising from hardware and software - both in normal operation and in the cases where the hardware and/or software exhibit unintended behaviour.

However, the role of data in influencing the safe operation of systems is just as important as the hardware and software. For example, it is possible that an IT System is built and assured to a very high level but if the data used by the system, such as a patient's blood group, is erroneous then the consequences can be just as severe as a hardware or software failure.

In recent years, a working group that draws representation from industry, government and academia has been studying data safety risks and has determined that there is little published guidance on the management of data safety risks not only in Healthcare but across most industries. The lack of guidance is becoming particularly noticeable as many types of data are now used to specify, deploy, configure, operate, test and justify safety systems; moreover, the volume of data in systems is also growing at an unprecedented rate. This is particularly an issue with clinical systems that typically store large volumes of patient-centric data used to inform clinical decisions. It is no longer safe to assume clinicians will, or can, spot errors in such vast stores of data - systems are no longer just there as productivity tools but are increasingly depended on for clinical decision making.

This document has therefore been prepared to provide guidance on the management of data safety risks in the manufacture, deployment and use of Health IT Systems.

The document draws from the Data Safety Guidance and shows how the requirements in the DCB0129 and DCB0160 Healthcare Standards can be interpreted and addressed from a data safety perspective.

Copyright © 2018 NHS Digital & SCSC

References

These documents provide additional information and are specifically referenced within this report.

Ref	Doc Reference Number	Title	Version	Status
1.	DCB0129	Clinical Risk Management: its Application in the Manufacture of Health IT Systems - Specification	4.2	Approved
2.	DCB0160	Clinical Risk Management: its Application in the Deployment and Use of Health IT Systems - Specification	3.2	Approved
3.	SCSC-127B	Data Safety Guidance	3.0	Published
4.	ISO 31000	Risk Management – Guidelines	2018	Published

Acknowledgements

The content of this guidance report is largely drawn from the work of the Data Safety Initiative Working Group and encapsulated in their Data Safety Guidance document [Ref. 3]. Full credit is therefore given to their work and contributions in informing this sector specific guidance.

Copyright © 2018 NHS Digital & SCSC

Abbreviations and Acronyms

A&E	Accident and Emergency
ALARP	As Low As Reasonably Practicable
DCB	Data Coordination Board
CRM	Clinical Risk Management
CRMP	Clinical Risk Management Plan
CSCR	Clinical Safety Case Report
CSO	Clinical Safety Officer
DSAL	Data Safety Assurance Level
GP	General Practise
INR	International Normalised Ratio
IT	Information Technology
ODR	Organisational Data Risk
PAS	Patient Administration System
SCSC	Safety Critical Systems Club
SDM	Safety Data Management

Copyright © 2018 NHS Digital & SCSC

Contents

Tables

Copyright © 2018 NHS Digital & SCSC

1 Introduction

1.1 Background

Health IT Systems are defined as products that comprise hardware, software or a combination of both. It is therefore natural for the CRM process to consider, as part of its patient centric hazard identification and risk assessment process, hazards to patients arising from hardware and software - both in normal operation and in cases where the hardware and/or software exhibit unintended behaviour.

Assessing safety risks in hardware and software is a practice common to many industries and several strategies for managing these risks have been developed over the years. These strategies have typically been embodied in a variety of standards and best practice guides including the Data Coordination Board (DCB) Healthcare Standards [Ref. 1] and [Ref. 2] for those organisations manufacturing or deploying Health IT Systems.

However, the role of data in influencing the safe operation of systems is just as important as the roles of hardware and software. For example, a Health IT System may be built and assured to a very high level but if the data used by the system, such as a patient's blood group, is erroneous then the clinical impact can be just as severe as a hardware or software failure.

In recent years, a working group that draws representation from industry, government and academia has been studying data safety risks and has determined that there is little published guidance on the management of data safety risks not only in Healthcare but across most industries. The lack of guidance is becoming particularly noticeable as many types of data are now used to specify, deploy, configure, operate, test and justify safety systems. Moreover, the volume of data in systems is also growing at an unprecedented rate. This is particularly an issue with clinical systems that typically store large volumes of patient-centric data used to inform clinical decisions. It is no longer safe to assume clinicians will, or can, spot errors in such vast stores of data - systems are no longer just there as productivity tools but are increasingly depended on for clinical decision making.

This document has therefore been prepared to provide guidance on the management of data safety risks in the manufacture, deployment and use of Health IT Systems.

The content of this guidance is largely drawn from the work of the Data Safety Initiative Working Group and encapsulated in their Data Safety Guidance document [Ref. 3]. Full credit is therefore given to their work and contributions in informing this sector specific guidance.

The document draws from the Data Safety Guidance and shows how the requirements in the DCB0129 and DCB0160 Healthcare Standards can be interpreted and addressed from a data safety perspective.

1.2 Audience

This guidance is applicable to all involved in the establishment and operation of the CRM process for Health IT Systems.

Copyright © 2018 NHS Digital & SCSC

1.3 Scope

This guidance is intended as a supplement to the Implementation Guidance for the following standards:

- DCB0129 - Clinical Risk Management: its Application in the Manufacture of Health IT Systems [Ref. 1] and
- DCB0160 - Clinical Risk Management: its Application in the Deployment and Use of Health IT System [Ref. 2].

1.4 Assumptions and Constraints

It is assumed that the reader is familiar with the DCB0129 [Ref. 1] and DCB0160 [Ref. 2] standards and their associated Implementation Guidance.

Copyright © 2018 NHS Digital & SCSC

2 Guidance

2.1 Data Safety Issues

Data can give rise to risks that may be different to other system elements such as hardware and software. An awareness of data-related issues promotes a more comprehensive assessment of safety risk. Typical issues that arise in the Healthcare context are as follows:

Fluidity: Hardware and software can undergo significant amounts of product assurance and once assured, may change relatively infrequently. Where change is required to hardware or software, it is carefully managed and the impact on the safety case assessed. This is not always the case for data, which is often much more fluid and may undergo limited assurance before use; indeed, the ease with which data can be changed is one motivation for the move towards data-driven Healthcare systems.

Reuse: Data used in one Healthcare system is often also used in a different system or system context, for example, when exchanging data between Trusts. Just because data was valid for use in a particular system or context, it does not immediately follow that it can be reused again in a different system or context. Many considerations associated with data reuse are similar to those of software reuse, for example establishing the similarity of requirements, role and required integrity/assurance level. One consideration that is different is that of timeliness: data that was valid for use in a particular system at a particular time is not necessarily valid for reuse in the same system at a different time.

Ageing: All safety-related data has a lifetime and this needs to be explicitly managed. This can involve, for example, refreshing, purging, deletion or alerting. For example, data relating to the current medication of a patient will age over time. Without explicit events to trigger updates to the data, the data can become out-of-date and hence inappropriate and possibly misleading. For example, if a system records an anticoagulant as a patient's current medication, which they no longer take, then this could mislead clinicians and affect the outcome in the treatment of an acute stroke patient.

Transformation: Data is often filtered, mapped, fused or aggregated as it moves through systems, sometimes creating new data sets as a result. Data properties are not necessarily preserved by these transformational processes. A key problem is the lack of provenance for the data (i.e. information about where it came from and how it was produced); the data may be moved through several systems and therefore its integrity may become lost.

Ownership: The transformation or transmission of data can result in a lack of clarity regarding who has ownership of, and responsibility for, the data (if anyone). It is important that responsibility for errors in the data can be tracked, for example, to determine whether they were present in the initial data or whether they arose as part of the transformation process. In Healthcare, sometimes complex and lengthy data supply chains of data are established. For example, data can be captured at many points of patient encounters and the information exchanged amongst a variety of Health IT Systems such as General Practise (GP), pharmacy, hospital and community systems as wells as on to secondary use systems. This means there can be a lack of ownership and provenance for data.

Archiving and Retrieval: Safety-related data needs to be available when required. There is thus a need to think about data accessibility over the complete system lifetime. It is also important to consider what properties of the data need to be preserved through archiving and retrieval and how this affects the choice of storage medium. For example, it may be natural to focus on the availability of Patient Administration System (PAS) data but availability of say, data from a pathology system, may be just as important as delays in processing a blood result could result in adverse patient outcomes.

Copyright © 2018 NHS Digital & SCSC

Biasing: This is a systemic inaccuracy in data due to the characteristics of the process employed in the creation, collection, manipulation, presentation and interpretation of data. Hence this is usually an unintentional distortion in the data set, which may be due to how the set has been selected or originated; it may also arise through the injudicious use of default data values (see Defaulting).

Aliasing: This is an effect that causes different data to become indistinguishable when accessed; that is, there is only one value or record visible when there should be several. This could be due to the way the data is filtered, sampled, indexed, stored or retrieved. These data issues are typically related to loss of resolution leading to similar data points appearing to be identical. For example, when merging patient records as a result of multiple clinical setting encounters, different medication administration events on the same day could be erroneously treated as a single event.

Defaulting: Many systems use default or initial values for data items; sometimes in data sets and sometimes embedded in software. Often these default values are designed to be neutral e.g. “0” or unrealistic e.g. “VOID” so as to avoid impacting the processing or ever being used. However, this is not always the case and defaults may be used when in fact they should always be replaced by real data. There are therefore issues of default values being used by mistake, and it is easy to see how default values could cause major issues in healthcare, e.g. default blood group.

Sentinels: A sentinel value is a data value used to indicate that a special action needs to be taken, typically indicating the end of a record or a data set. The sentinel value should be one that is not allowable in the data set itself, but often is not properly considered and may use common sequences (e.g. five zeroes). Sentinels can cause problems in two ways: 1) where they are not recognised and so, e.g. processing continues past the sentinel, and 2) where the data itself somehow contains the sentinel value and so processing is erroneously interrupted. This could cause corruption in medical data where two records could be mistakenly read as one.

Disassociation: This effect is in some senses the opposite of aliasing e.g. there are several records when there should only be one. This could occur for example if two records are created for the same individual using names entered differently. It could also arise if different systems use different indexing methods and the association between the indexes becomes lost or corrupted. In healthcare this is common as individuals will present themselves at different clinical settings such as General Practices, Accident & Emergency (A&E) departments, hospital outpatient departments etc. and multiple records for the same patient will often be created, even sometimes within the same system.

Masking: This issue can arise if a significant proportion of a data set is of a poor quality. This poor data can hide errors in the way that the system handles the good quality data, i.e. it is all rejected. For example, if a faulty laboratory test machine is known to occasionally generate erroneous results, which are discarded, it could hide genuine data processing errors that will result in correct data being discarded.

Migration: In Healthcare data is often physically migrated between different contexts of use and often attracts the same issues as previously discussed such as Reuse, Transformation and Ownership. However, it is so widespread in the industry that it warrants explicit consideration as a data safety issue. Migration might at the simplest level involve re-hosting the same data as a result of hardware storage technology upgrade (tech refresh) or could involve movements between different versions of the same system and different systems both within and external to the Health Organisation. Even with unmodified data sets errors can be introduced that may be difficult to detect given the potentially large volumes of data involved.

Copyright © 2018 NHS Digital & SCSC

2.2 Data Types

There are many different data types that can have safety implications and Appendix E of [Ref. 3] provides a comprehensive list. Five data types are discussed in this report as presented in the following table. These are the data types where there is significant reliance in the Healthcare context and most likely to be encountered in Health IT Systems.

Type	Description	Explanation	Typical containers
Verification	Data used to test and analyse the system.	This is data comprising the test values and test data sets used to verify the system. It may include real data, modified real data or synthetic data. It includes data used to drive stubs, and any data files used by simulators or emulators.	Test data sets, Stub data, Emulator and Simulator files.
Infrastructure	Data used to configure, tailor or instantiate the system itself.	Data used to set up and configure the system for a particular installation, product configuration, or network environment.	Network configuration files, Initialisation files, Hardware pin settings, Network addresses, Passwords, etc.
Performance / Resilience	Data collected or produced about the system during trials, pre-operational phases and live operations.	Data produced by and about the system during introduction to service and live service itself. Includes fault data, diagnostic data, failover data and back-up data. This may be the results of various phases of introduction and may include trend analysis to look for long-term problems.	Field data, Support calls, Bug reports, Non-Compliance Reports, Defect Reporting System data, Back-up data.
Dynamic	Data manipulated and processed by the system during operations.	This is the data processed, transformed or produced by the system that has end-user meaning. It may be displayed and used within the system or may be for transfer and distribution to other systems or downstream users. It is data that has some real domain meaning.	May be manipulated within the system in data structures or transferred into or out of the system through interfaces.

Copyright © 2018 NHS Digital & SCSC

Type	Description	Explanation	Typical containers
Justification	Data used to justify the safety position of the system.	Data used to justify, explain and make the case for starting or continuing live operations and why they are safe enough. Often passed to external bodies (regulators, Health and Safety Executive, Independent Safety Auditors) for their review.	Clinical Safety Case Report, Certification case, Regulatory documents, COTS Justification file, Design Justification file.

Table 1: Safety Data Types

2.3 Data Properties

When considering how *data can give rise to harm*, it is more useful to define this as "*where failure to preserve a critical property of data results in a hazardous system state that in combination with an external event can give rise to harm".* For example, missing data from patient allergy records gives rise to a hazardous system state. Harm can occur if this information is then relied upon (the external event in the definition) – for example, administering penicillin when a record of an allergy to this drug has been lost.

The following table shows a non-exhaustive list of typical data properties.

Property	Description
Integrity	the data is correct, true and unaltered
Completeness	the data has nothing missing or lost
Consistency	the data adheres to a common world view, e.g. units
Continuity	the data is continuous and regular without gaps or breaks
Format	the data is represented in a way that is readable by those that need to use it
Accuracy	the data has sufficient detail for its intended use
Resolution	the smallest difference between two adjacent values that can be represented in a data storage, display or transfer system
Traceability	the data can be linked back to its source or derivation
Timeliness	the data is as up to date as required
Verifiability	the data can be checked, and its properties demonstrated to be correct
Availability	the data is accessible and usable when an authorized entity demands access
Fidelity / Representation	how well the data maps to the real-world entity it is trying to model

Copyright © 2018 NHS Digital & SCSC

Property	Description
Priority	the data is presented/transmitted/made available in the order required
Sequencing	the data is preserved in the order required
Intended Destination/Usage	the data is only sent to those that should have them
Accessibility	the data is visible only to those that should see them
Suppression	the data is intended never to be used again
History	the data has an audit trail of changes
Lifetime	when does the safety-related data expire
Disposability	the data can be permanently deleted when required

Table 2: Data Safety Critical Property Types

In general, the objective of the risk management process is to assess which of these properties are critical to each data set in the scope of the analysis and establish mitigations to reduce the risk of occurrence. Not all properties will be critical in all cases. For example, precision of an International Normalized Ratio (INR) reading may be more critical than an individual's estimated weekly alcohol intake. The important point is to identify those properties that are significant for the data set in a given context of use and to ensure those properties are preserved with a sufficient level of confidence.

2.4 Assessing Applicability of Managing Data Risk

As data safety risk assessment is a relatively new discipline, Manufacturers and Health Organisations will naturally want to understand to what extent the guidance applies to them and how much effort will be required in addressing the risks. Patient safety is of course a key reason for investing effort in managing safety risk, but the level of investment is also driven by the maturity of the organisation in managing data safety, the organisation's appetite for risk and the wider corporate level implications that may arise from patient safety incidents and accidents.

The Organisational Data Risk (ODR) assessment form has therefore been developed as a useful tool to capture a high-level perspective of the risk posed to an organisation by data safety issues within a specific enterprise. The ODR is based on the ISO31000 standard for risk management [Ref. 4]. The ODR is effectively a multiple-choice questionnaire covering different aspects of data safety that in totality provide a holistic assessment of the corporate risk for a given enterprise.

The ODR firstly asks the assessor to establish the context of the risk assessment to ensure that the enterprise being considered, and the scope of the assessment is well defined. The ODR questions will then establish:

- The risk tolerance of external stakeholders;
- The level of risk that is allocated to the organisation;
- The applicable regulatory environment within which the enterprise will operate;
- The maturity of the organisation in terms of their attitude to not simply risk, but specifically data-driven risks;

Copyright © 2018 NHS Digital & SCSC

- Responsibilities for data ownership through the use cases of the system;
- The data-driven specifics about failure consequences and the issues raised by data complexity, boundary complexity and system complexity for the enterprise.

ODR assessment is discussed in more detail in Appendix B of [Ref. 3].

Part of the ODR assessment relates to assessing the organisation's maturity in managing data safety risks; responses are aimed at establishing the depth of awareness of data safety and the associated management processes within the organisation. However, measuring the level of awareness of processes and concepts in an organisation is not always straightforward. There may be sufficient high-level knowledge of this for the purposes of the ODR, but it still may be an area that warrants further investigation.

To support this a separate questionnaire has been developed to specifically assess the data safety culture within an organisation as a whole or as directed towards a particular project, service or other activity. However, here the focus is on a personal view rather than a project or company view, so the questionnaire would be completed by all or a significant subset of staff. Responses can be aggregated to give an overall data safety culture value. A key aspect of this approach is that it can be periodically repeated to determine trends: for example, if overall scores are declining, this may suggest that further training and briefings will be required. More detail on the data safety culture questionnaire is provided in Appendix C of the Data Safety Guidance [Ref. 3].

An organisation may be dependent on a third-party supplier for all or part of a Health IT System solution. It may also therefore be useful for the organisation to assess the maturity of that supplier in managing data safety risk especially during procurement phases where there may be a choice of supplier. A supplier data maturity questionnaire has therefore been developed for this purpose and is provided in Appendix D of the Data Safety Guidance [Ref. 3].

2.5 Managing Data Safety Risks

The following sections provide specific guidance for those Manufacturers and Health Organisations responsible for implementing the DCB0129/DCB0160 standards. Each of the key requirements from the given standard is presented along with specific guidance for meeting those requirements from a data safety perspective. Throughout, the guidance is supported by practical examples based on a Manufacturer building a new Health IT System and the Health Organisation that will be deploying it. These examples are by no means exhaustive but should give insight into how to apply the guidance in practice.

2.5.1 Establishing the Context

The guidance in this section relates to the following requirements in the relevant DCB standards from the perspective of managing data safety:

Standard	Section	Requirement
DCB0129	4.2.1	The Manufacturer MUST define the clinical scope of the Health IT System which is to be delivered.
DCB0129	4.2.2	The Manufacturer MUST define the intended use of the Health IT System which is to be delivered.
DCB0160	4.2.1	The Health Organisation MUST define the clinical scope of the Health IT System which is to be deployed.

Copyright © 2018 NHS Digital & SCSC

DCB0160	4.2.2	The Health Organisation MUST define the intended use of the Health IT System which is to be deployed.
DCB0160	4.2.3	The Health Organisation MUST define the operational environment and users of the Health IT System which is to be deployed.

Table 3: DCB Requirements Relating to Establishing the Context

From a data perspective, fundamental to fulfilling these requirements is establishing the context within which the use of data in system development, enhancement, introduction, integration or operation is occurring. This will typically involve activities to:

- Describe the organisational context;
- Describe the system context;
- Plan the assessment; and
- Identify Data Artefacts.

This should establish the risk appetite: essentially, how much effort is devoted to making data risks As Low As Reasonably Practicable (ALARP). In turn, this will inform the nature and scope of assessments that are conducted during system development and, furthermore, its introduction into operational service.

An ODR assessment form is recommended at this stage to understand in broad terms the level of risk that will need to be managed. This is also a convenient means of describing the organisational and system context.

It is essential to identify the data artefacts that are potential sources of safety hazards. Like other components of a safety-related system, the safety dependency of data is dictated by the context in which it is used and the causal links that become established where loss of one or more of the required properties can contribute to hazardous system states. For example, a given data set (say Configuration Data) could be used in a number of separate contexts such as:

- prototyping a system to demonstrate solution feasibility of a safety-related system;
- development testing of a safety-related system; or
- live operational use of a safety-related system.

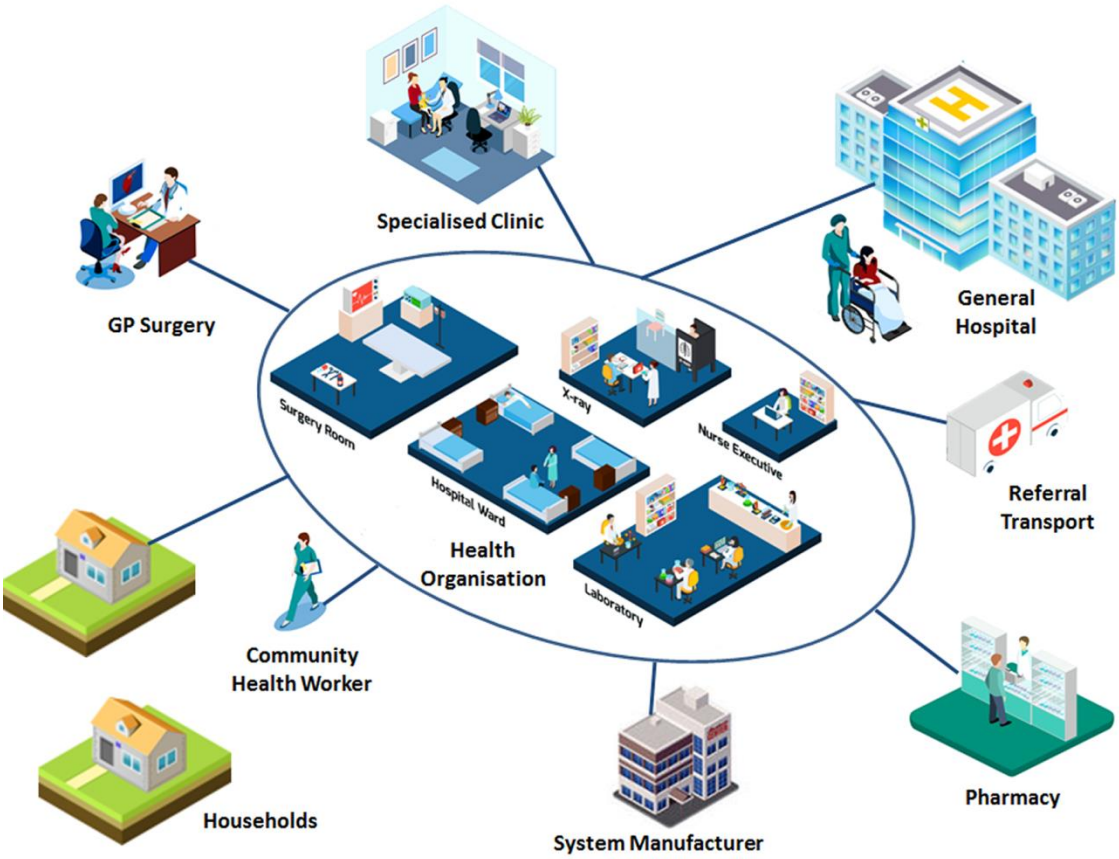
In these cases, the same data set could be used but the context of its use changes the safety significance and therefore the level of assurance that it may require. Hence not only is the type of data under consideration important but also when in the organisation's process lifecycle the data will be used and relied upon. It follows that the assessed integrity level of a data set is also predicated on where and when in the lifecycle the data set will be applied. It is recommended that these considerations are addressed in the Clinical Risk Management Plan (CRMP) by modelling the organisation's lifecycles and explicitly documenting where a specific data set will be used and therefore subject to further assurance techniques.

Example

A Manufacturer is building a new integrated health and social care system to support holistic care for community health services. The system supports clinical workflows for aspects such as referrals, tracking clinical encounters, appointment scheduling, outcome measures through to letter and report generation. The system follows a typical development lifecycle as

Copyright © 2018 NHS Digital & SCSC

a series of phases: business modelling, requirements, analysis & design, implementation test and deployment.



The system is being targeted to meet the requirements of a Health Organisation who are procuring a solution to help their clinicians maintain a high level of quality of care in the face of increasing volumes of patients and pressure to reduce staff costs.

The Manufacturer decides to use an ODR assessment form to understand in broad terms the level of risk it will have to manage in developing and supporting this system. This will allow the Manufacturer to **Describe the organisational context** and **Describe the system context**.

The following list identifies the questions in the ODR, along with the assessments for the Manufacturer's system.

Q1	How severe could an accident be that is related to the data? Could it be caused directly by the data?

Failings in the system could give rise to non-optimal treatment plans for a patient that might delay detection of a more serious condition or prolong the recovery for a known condition. The system is not solely relied upon however, and there are other people and systems in play involved in checking data. On balance, this question is assessed as: **1c**; Score **4**.

Q2	What would be the impact on the organisation, client or public if an accident occurred related to the data?

Unfortunately, accidents in the health domain are relatively frequent and grievances are usually resolved by financial settlements through the courts. The Manufacturer believes

Copyright © 2018 NHS Digital & SCSC

through their contractual arrangements that the Health Organisation would be liable for any claims even if it was attributed to an error in the Manufacturer's system's handling of data. On balance, this question is assessed as: **2b**; Score **2**.

Q3	How much responsibility does this organisation have for data safety?

The Manufacturer is responsible for building the system in compliance with DCB0129 and so is responsible for executing the associated safety management system to manage risk. The Manufacturer however plans to sell the product with a condition of use that places end responsibility for patient safety on its clients. On balance, this is question is assessed as: **3b**; Score **2**.

Q4	What legal and regulatory environment will this work be subject to?

The work will be contracted under UK law and subject to the DCB standards for Health IT systems. However, there is no special regulator who is currently empowered to intervene in the delivery of healthcare systems, i.e., the standards are not currently enforced through law. The Health Organisation however will make compliance with the standards a contractual requirement. On balance, this question is assessed as: **4c**; Score **4**.

Q5	How mature is this organisation regarding data safety?

The Manufacturer has a good understanding of data as a source of safety risk. Many of their systems are data intensive to support clinical decision making. There is good support and funding for the identification and resolution of data-related risks. On balance, this question is assessed as: **5b**; Score **2**.

Q6	How widely is the data used and who by?

The data will be used in multiple clinical settings and by many clinicians and other support staff. There are several data supply chains and public web access to data. On balance, this question is assessed as: **6c**; Score **4**.

Q7	What is the scale, sophistication and complexity of the data and its manipulation?

The data is complex and although transmitted through industry standard data structures, these require knowledge of the associated abstract clinical data model. Some data manipulations are required to map between different encodings for data held in the various heterogeneous systems. Some legacy systems transfer data in unstructured format. On balance, this question is assessed as: **7c**; Score **4**.

Q8	How well defined and understood are the boundaries and interfaces for this data scenario?

The boundaries of the supply are well understood and although the interfaces are complex and of mixed formats, these will be defined and agreed formally through Interface Control Documents. Most of the integrating systems are established COTS based systems but some of the legacy systems still need to be investigated and working assumptions have been made by the Manufacturer. On balance, this question is assessed as: **8c**; Score **4**.

The final score is **26**, which corresponds to **ODR2**. The Manufacturer therefore concludes that there is low to medium risk that loss of properties of data in the system can contribute to

Copyright © 2018 NHS Digital & SCSC

or give rise to harm. The Manufacturer has an internal policy for engagements based on the ODR level that dictates how the organisation shall **Plan the assessment**; this policy dictates the amount of proportional effort it needs to spend on Safety Data Management (SDM) and the level of rigour to be employed. In this case, the policy dictates, amongst other requirements, that a separate section covering SDM is required in its CRMP.

The Manufacturer aims now to **Identify Data Artefacts** that are potential sources of safety hazards. The Manufacturer also knows that the safety dependency of data is dictated by the context in which it is used so it now develops an understanding of when in its process lifecycle the data will be used and relied upon. The Manufacturer decides to build an early prototype to show to clients to help elicit requirements definition. To support this, the Manufacturer plans to create a test data set that comprises a typical range of scenarios that the system will encounter. This form of data is identified as **Verification Data.** The system also needs to be configured to support deploying Health Organisations' policies. This data is **Infrastructure Data**[1] and for the prototyping phase, the Manufacturer plans to use largely default values.

In later phases when the system functionality is specified and the system is being built, the Manufacturer plans to create a test data set that will be key to demonstrating the correct functioning of the system and hence acceptance by the deploying Health Organisation. This still involves the use of **Verification Data** and **Infrastructure Data** but there will be far greater dependency on these data sets than the prototyping case.

The Manufacturer therefore documents the planned use of each of the data types during the entire delivery lifecycle in the Clinical Risk Management Plan.

The procuring Health Organisation will have a different perspective of the IT system that they will deploy into their organisation. They will already have many integrated systems in live operations and as part of establishing the context for the system's deployment they will need to consider many different types of data sets:

- **Infrastructure Data**: how the system will be configured in the specific environment;
- **Verification Data**: the test data sets to be used to support certain deployments such as integration testing and training;
- **Dynamic Data**: the data entered or fed into the system and the data presented to the user, generated in the form of reports or data passed to other systems.

Post acceptance, the procuring Health Organisation decides to run a series of user training sessions for clinicians. Once users are trained, the system will be integrated into live operations. The Health Organisation identifies the Infrastructure, Verification and Dynamic Data types to be used during these phases. The Health Organisation also realises that the system will form part of a data supply chain as many external organisations and departments within their own organisation engage in the procurement and use of safety-related data. For example, it will receive referral data from many other GP systems, it will receive outcome measures from hospitals and clinical data acquired from remote workers visiting patients in the community and from the patients themselves using the system's online portal. The

[1] In [3] **Infrastructure Data** is a specific type of the more general class of data defined as 'Configuration' data. Other forms of Configuration data are 'Behavioural Data' and 'Adaptation Data' but for the purpose of this guidance, the focus is on Infrastructure Data only. This guidance will use the expression 'configuration data' from time to time to refer to Infrastructure Data.

Copyright © 2018 NHS Digital & SCSC

system also produces data for other external systems such as electronic prescriptions for pharmacies.

The Health Organisation understands that it is important when establishing context that stakeholder roles within a data supply chain are clearly understood and defined in the CRMP:

- The Commissioning User: an organisation or unit of an organisation that has the need for the data;
- The Data Provider: the organisation or unit of an organisation that will fulfil that need for data;
- The Data Acquirer: the organisation employed by or collaborating with the Data Provider to carry out physical collection of data.

The Health Organisation sees that by using the new system it will become a Commissioning User as it will require and be a **Consumer** of data from GP systems, hospital systems, systems used by remote workers in the community and the system's portal capturing data entered by the patients themselves, each of these acting as Data Providers. Those healthcare professionals (and the patient themselves) gathering patient data through physical inspections and measurement are the Data Acquirers.

The Data Provider acts both as a **Consumer** (from the Data Acquirer) and **Producer** (to the Commissioning User) of data. Similarly, an organisation that augments data sets is both a **Consumer** and **Producer** of data in the supply chain. Data augmentation may be, for example, to add GP practice postal addresses to a data set based on the given Practice codes.

The Health Organisation defines the data supply chain relevant to the system including the roles and interfaces involved in its CRMP This will therefore show where there are dependencies on Dynamic Data used and produced by the system.

Questions the Health Organisation will need to address when establishing the context are:

- Have all the dependent interfaces been identified?
- Have the roles of Commissioning User/Data Provider/Data Acquirer been established and acknowledged?
- What 'service levels' or contracts exist for the delivery of the data?
- What level of assurance do Data Providers/Data Acquirers provide for their data?

2.5.2 Risk Identification

The guidance in this section relates to meeting the following requirements in the relevant DCB standards from the perspective of managing data safety:

Standard	Section	Requirement
DCB0129	4.3.1	The Manufacturer MUST identify and document known and foreseeable hazards to patients with respect to the intended use of the Health IT System in both normal and fault conditions.

Copyright © 2018 NHS Digital & SCSC

DCBI0160	4.3.1	The Health Organisation MUST identify and document known and foreseeable hazards to patients in both normal and fault conditions through the introduction and use of the Health IT System

Table 4: DCB Requirements Relating to Risk Identification

The Risk Identification phase for data comprises an analysis of all relevant data items and data sets and considering whether harm can arise if some or all of the properties for those data items/data sets were lost. As discussed in Section 2.5.1 this assessment must be conducted based on knowledge of the context and project phase in which the data is being used. For example, a data set used in a proof of concept system cannot give rise to harm, but the same data set used in live clinical operations could potentially cause patient harm.

To support the hazard identification assessment, it is sometimes useful to consider the use of guidewords. Guidewords help hazard identification by providing a structured qualitative set of expressions to be considered for each data property. For example, the following guidewords can be used when considering the integrity property of a data set: ‘correctness’, ‘truth’, ‘original’, ‘trustworthy’, ‘coherency’, ‘stability’, ‘perfect’, ‘unquestionable’, ‘faithful’, ‘certain’, ‘ordered’. A series of guidewords have been developed to support this form of analysis for data related hazard (see HAZOP Guidewords: Appendix F of the Data Safety Guidance [Ref. 3]).

Example

The Manufacturer of the Health IT System decides that during the prototyping phase there is little safety dependency of the test and configuration data sets as no clinical decisions will be made based on their content; the data is simply being used to support the elaboration of requirements.

In later phases however, when the system functionality is specified and the system is being built, the Manufacturer will want to create a test data set that will be instrumental in demonstrating the correct functioning of the system. This still involves the use of **Verification Data** and **Infrastructure Data** but there is far greater dependency on these data sets than the previous case. For example, if the verification or configuration data is not sufficiently diverse or insufficiently models real world scenarios, it is possible that erroneous and unsafe functional behaviour is present in the system during live operation despite this system having passed factory and site acceptance testing.

The Health Organisation will likewise need to conduct risk identification relevant to their deployment context. Hazards arising from data sources that are to be delivered into the new system from existing systems need to be assessed for data risks.

From the Health Organisation’s perspective, one key focus for hazard identification is in the use of **Dynamic Data,** i.e. The data that will be delivered into the new system from existing system data sources, and the data presented to the user. For the interactions identified in the supply chain, the Health Organisation needs to consider the risks associated with loss of properties of the data it will receive. Questions that the Health Organisation will need to consider and address more formally in the CRMP are as follows:

Copyright © 2018 NHS Digital & SCSC

- Which data sets or items being received from other systems have properties (such as timeliness, completeness, consistency, fidelity etc.) that are significant to patient safety?
- What data presented to the user has properties (such as availability, format, resolution, etc.) that are significant to patient safety?
- What existing barriers or mitigations (physical, technical, procedural) exist to reduce the risk of loss of data properties?

2.5.3 Risk Analysis

The guidance in this section relates to meeting the following requirements in the relevant DCB standards from the perspective of managing data safety:

Standard	Section	Requirement
DCB0129	4.4.1	For each identified hazard the Manufacturer MUST estimate, using the criteria specified in the Clinical Risk Management Plan: • the severity of the hazard • the likelihood of the hazard • the resulting clinical risk
DCB0160	4.4.1	For each identified hazard the Health Organisation MUST estimate, using the criteria specified in the Clinical Risk Management Plan: • the severity of the hazard • the likelihood of the hazard • the resulting clinical risk

Table 5: DCB Requirements Relating to Risk Analysis

When analysing data risks as with other forms of risks such as software and hardware risks, an assessment of likelihood and severity will need to be conducted against defined safety criteria. There will be no quantifiable failure rates for data and so to help support such an assessment, it is useful to consider the different dimensions of how data can contribute to accidents:

- Proximity: how directly a data failure will lead to an accident;
- Dependency: how dependent the application is on the dataset;
- Detection: the likelihood of being able to detect a data failure prior to an accident;
- Prevention: the ability of the systems architect/developers to guard against errors;
- Correction: the ability of the system to work around or correct errors.

By considering the likelihood of these different factors a holistic picture of the data contribution to risk can be established. Table 6 shows a more structured approach to the assessment:

Copyright © 2018 NHS Digital & SCSC

	Likelihood of Data Causing Accident		
Concern	**Very High / High**	**Medium**	**Low / Very Low**
Proximity	A known use of the data is highly likely to lead to an accident.	A possible use of the data could lead to an accident.	All currently foreseen uses of the data could lead to harm only via lengthy and indirect routes.
Dependency	Data is completely relied upon.	Data is indirectly relied upon.	Little reliance on data.
Detection	Low or no chance of anything else detecting an error.	Some other people/systems are involved in checking the data.	Many other people/systems are involved in checking the data.
Prevention	Difficult or impossible to guard/barrier against errors.	Possible to guard/barrier against errors.	Easy to guard/barrier against error.
Correction	Difficult or impossible to correct or workaround errors.	Possible to correct or workaround errors.	Easy to correct or workaround errors.

Table 6: Aspects of Data Contribution to Accidents

It is possible to use the table in different ways depending on the organisation's attitude to risk. For example, for the data use under consideration an assessment is made against each concern and:

- the average of the concerns is taken to provide the overall risk likelihood;
- the highest rated concern dictates the overall risk likelihood.

The assessment method chosen should be documented in the CRMP.

The likelihood as determined above along with the assessed severity of the resulting impact can then be used to determine the categorisation of the risk using the Clinical Risk Matrix that will have been established as part of the risk assessment of hardware and software hazard (see Section 4.4.3 of the Implementation Guidance to DCB0129 [Ref. 1] and DCB0160 [Ref. 2]).

If the organisation requires more structured guidance in the assessment process, then consideration should be given to using Data Safety Assurance Levels (DSALs). Using

Copyright © 2018 NHS Digital & SCSC

DSALs provides a structured approach to risk assessment; it simplifies the data risk categorisation process and allows the organisation to draw from current best practice in terms of applying appropriate risk reduction methods and techniques commensurate for the level of risk. This will not only demonstrate that a systematic approach to the assessment has been adopted but will also help justify the investment spent in risk reduction measures.

Appendix A - Structured Data Risk Analysis, describes structured methodology and it is highly recommended that organisations adopt this approach. Section A.1 provides specific guidance and examples of the structured approach to Risk Analysis.

2.5.4 Risk Evaluation

The guidance in this section relates to meeting the following requirements in the relevant DCB standards from the perspective of managing data safety:

Standard	Section	Requirement
DCB0129	5.1.1	For each identified hazard, the Manufacturer MUST evaluate whether the initial clinical risk is acceptable. This evaluation MUST use the risk acceptability criteria defined in the Clinical Risk Management Plan.
DCB0160	5.1.1	For each identified hazard, the Health Organisation MUST evaluate whether the initial clinical risk is acceptable. This evaluation MUST use the risk acceptability criteria defined in the Clinical Risk Management Plan.

Table 7: DCB Requirements Relating to Risk Evaluation

Once classified, hazards related to data can be evaluated in the same way as other hazards such as those arising from software and hardware using the Clinical Risk Matrix. The evaluation of risk for data however may need to consider many factors that are unique to data as opposed to software and hardware. For example, the data may be provided entirely, or may be transposed or modified by a 3rd party and so the evaluation may need to consider the level of confidence the organisation has on those 3rd parties and whatever data assurance processes may exist in the data supply chain.

A more structured evaluation process is recommended using the concept of DSALs and is described in more detail in **Appendix A - Structured Data Risk Analysis**. Section A.2 provides specific guidance and examples of the structured approach to Risk Evaluation.

Copyright © 2018 NHS Digital & SCSC

2.5.5 Risk Treatment

The guidance in this section relates to meeting the following requirements in the relevant DCB standards from the perspective of managing data safety:

Standard	Section	Requirement
DCB0129	6.1.1	The Manufacturer MUST identify appropriate clinical risk control measures to remove any unacceptable clinical risk.
DCB0160	6.1.1	The Health Organisation MUST identify appropriate clinical risk control measures to remove an unacceptable clinical risk.

Table 8: DCB Requirements Relating to Risk Treatment

As stated in the DCB standards implementation guidance, risk controls or risk treatment can take many forms:

1. changes to the design or the inclusion of protective measures in the Health IT system;
2. product verification and validation (for example, testing). A testing programme should address each of the hazards and thus provide a practicable demonstration that the claimed risk reduction has been achieved;
3. business, administrative and implementation procedures;
4. user, operator and other stakeholder training and briefing;
5. information for patient safety, including warnings.

For items 1 and 2, it is recommended that the organisation refers to the Data Safety Guidance [Ref. 3] for specific guidance on method and techniques for risk reduction associated with data risks. This is particularly more effective when used in the context of DSALs described in **Appendix A - Structured Data Risk Analysis**. Incorporating DSALs as part of the analysis helps to systematically determine and justify which method and techniques are appropriate for the level of risk. Section A.3 provides specific guidance and examples of the structured approach to Risk Treatment.

Copyright © 2018 NHS Digital & SCSC

Appendix A – Structured Data Risk Analysis

A.1 Risk Analysis

The DSAL has been developed to provide a more structured and systematic method of categorising risk so that guidance can be provided on appropriate techniques that can be employed to manage the risk. The DSAL metric is not a statistical measure of likelihood, or a literal numeric measure of integrity. Instead, the DSAL is qualitative and determines the level of rigour required in demonstrating that critical properties of data are sufficiently well preserved. As such, DSALs share a common theoretical basis with concepts like Safety Integrity Levels and (Item / Function) Development Assurance Levels.

Table 9 presents a classification system allocating DSALs to safety-related data items using a function of likelihood and consequences. The system described below is not prescriptive and can be tailored depending on the context; the reasons for any such tailoring should, of course, be documented and agreed amongst relevant stakeholders. Furthermore, where this table suggests a low DSAL because (for example) a work-around is simple to implement it is important to ensure that the work-around (or similar) is ultimately implemented.

The table augments the likelihood table already introduced in Section 2.5.3 that uses the following categories to help examine how data can contribute to accidents:

- Proximity: how directly a data failure will lead to an accident;
- Dependency: how dependent the application is on the dataset;
- Detection: the likelihood of being able to detect a data failure prior to an accident;
- Prevention: the ability of the systems architect/developers to guard against errors;
- Correction: the ability of the system to work around or correct errors.

		Likelihood of Data Causing Accident		
	Concern	**Very High / High**	**Medium**	**Low / Very Low**
	Proximity	A known use of the data is highly likely to lead to an accident.	A possible use of the data could lead to an accident.	All currently foreseen uses of the data could lead to harm only via lengthy and indirect routes.
	Dependency	Data is completely relied upon.	Data is indirectly relied upon.	Little reliance on data.

Copyright © 2018 NHS Digital & SCSC

	Detection	Low or no chance of anything else detecting an error.	Some other people / systems are involved in checking the data.	Many other people / systems are involved in checking the data.
	Prevention	Difficult or impossible to guard / barrier against errors.	Possible to guard / barrier against errors.	Easy to guard / barrier against error.
	Correction	Difficult or impossible to correct or workaround errors.	Possible to correct or workaround errors.	Easy to correct or workaround errors.
Severity or Impact of data related accident				
Catastrophic	Multiple patient deaths. Permanent life-changing incapacity and any condition for which the prognosis is death. Permanent life-changing incapacity for multiple patients. Severe injury or severe incapacity from which recovery is not expected in the short term for multiple patients.	**DSAL4**	**DSAL4**	**DSAL3**

Copyright © 2018 NHS Digital & SCSC

Major	A single death. Permanent life-changing incapacity and any condition for which the prognosis is death for a single patient. Permanent life-changing incapacity for a single patient. Severe injury or severe incapacity from which recovery is not expected in the short term for a single patient. Severe injury or severe incapacity from which recovery is expected in the short term for multiple patients. Severe psychological trauma for multiple patients.	**DSAL4**	**DSAL3**	**DSAL2**

Copyright © 2018 NHS Digital & SCSC

Considerable	Severe injury or severe incapacity from which recovery is expected in the short term for a single patient. Severe psychological trauma for a single patient. Minor injury or injuries from which recovery is not expected in the short term for several patients. Significant psychological trauma for multiple patients.	**DSAL3**	**DSAL2**	**DSAL1**
Significant	Minor injury or injuries from which recovery is not expected in the short term for a single patient. Significant psychological trauma for a single patient. Minor injury from which recovery is expected in the short term for multiple patients. Minor psychological upset; inconvenience for multiple patients.	**DSAL2**	**DSAL1**	**DSAL0**

Copyright © 2018 NHS Digital & SCSC

Minor	Minor injury from which recovery is expected in the short term for a single patient. Minor psychological upset; inconvenience for a single patient.	**DSAL1**	**DSAL0**	**DSAL0**

Table 9: DSAL Safety Matrix

If a Clinical Risk Matrix is being used that is consistent with the matrix recommended in Section 4.4.3 of the Implementation Guidance to DCB0129 [1] and DCB0160 [2], then this table can be presented as follows:

Likelihood						
	Very High	**DSAL1**	**DSAL2**	**DSAL3**	**DSAL4**	**DSAl4**
	High	**DSAL1**	**DSAL2**	**DSAL3**	**DSAL4**	**DSAL4**
	Medium	**DSAL0**	**DSAL1**	**DSAL2**	**DSAL3**	**DSAL4**
	Low	**DSAL0**	**DSAL0**	**DSAL1**	**DSAL2**	**DSAL3**
	Very Low	**DSAL0**	**DSAL0**	**DSAL1**	**DSAL2**	**DSAL3**
		Minor	**Significant**	**Considerable**	**Major**	**Catastrophic**
		Severity				

Using this DSAL oriented matrix simplifies the categorisation of data safety risk and allows recommended mitigations to be systematically determined.

There are many instances where the system architecture could justify the movement of a dataset from one DSAL for another. For example, there may be cases where multiple independent data items fulfil the same (or similar) usage and assessors may choose to reduce the DSAL requirements on each item to reflect the inherent redundancy (and associated risk reduction) that brings. Additionally, the same dataset may be reused by multiple functions with different levels of risk, in this case it would be recommended to assign the highest required level of integrity to the dataset so that it meets the requirements of the most demanding use case. As discussed, the implementation of a classification system is the responsibility of the Manufacturer or Health Organisation, but any such manipulation of DSALs should be carefully considered and appropriately documented and agreed in the CRMP.

Example


System Manufacturer

The Manufacturer had previously identified phases where the use of specific types of data could give rise to hazards. In the first, the prototyping phase, the Manufacturer sees no use of the data that can give rise to credible clinical risk and assessing the DSAL for that data set as **DSAL0**. In the second phase of the development lifecycle, where **Verification Data** and **Infrastructure Data** is being used to demonstrate the correct functioning of the system, the Manufacturer considers that loss of any of the data properties of **Integrity, Completeness, Consistency, Continuity, Format, Accuracy, Resolution, Timeliness, Availability, Fidelity / Representation, Sequencing,**

Copyright © 2018 NHS Digital & SCSC

Intended Destination / Usage of this data could give rise to hazards. For example, if the verification data set selected is not representative of the eventual diversity experienced in practice, then it is possible that the system may contain latent software errors that could give rise to harm. However, the Manufacturer acknowledges that the system will be subject to further testing and trials in the clinical setting and so there will be other opportunities to detect errors in the system:

- the likelihood that the data use gives rise to an accident is **Low** as other systems and processes are in place that would detect errors;
- the severity is **Considerable**; failings in the system could give rise to non-optimal treatment plans for a patient that might delay detection of a more serious condition or prolong the recovery for a known condition.

The Manufacturer therefore assesses these data types as **DSAL1** in this context of use.

From the Health Organisation's perspective, the focus for risk assessment is in the use of **Dynamic Data**. For the interactions identified in the supply chain, the Health Organisation needs to consider the risks associated with loss of properties of the data it will receive and present to the user. Questions the Health Organisation will need to consider and address more formally in the CRMP are as follows:

- How likely is it that there would be a loss of the given data property?
- How would such a loss of a property be detected?
- How would such a loss be isolated to prevent further risks of harm?
- What recovery action would be required to resolve the issue to maintain patient safety?

In considering the receipt of outcome measures data received from a hospital, the Health Organisations considers that it is likely that some credible errors would not be readily detected by their new system; if the hospital system confused a result or there were errors in the precision of data then there would be few chances to catch these once received by the system.

- the Health Organisations assess the likelihood of this loss of property as **Medium**;
- The impact of such errors, although not realistically likely to lead to death, could result in delays to treatment that could result in serious injury and hence **Considerable** impact.

The data received from this data source is therefore classed as **DSAL2** in this context of use.

Copyright © 2018 NHS Digital & SCSC

A.2 Risk Evaluation

Once all data items and/or data sets have been classified under the DSAL scheme, an evaluation is conducted as to whether the resulting classification is acceptable for the data at each relevant point of use in the project/service/operational lifecycle. Such an evaluation may consider what existing assurance and mitigations exist to manage the risk of loss of critical properties to the data item or data sets.

Consideration should also be given to any arrangements in place with other parties who may produce or consume the data. For example, if a data set bears a high DSAL then a producer of that data should be aware that this places an assurance obligation on them to ensure data properties are maintained to a level of confidence commensurate with the risk. Such obligations would ideally be embodied in the contractual arrangements with 3rd parties. If no such arrangement or assurance can be guaranteed, then further mitigation will be required to reduce the DSAL for that dataset. For example, additional validation of received data or crosschecks with other independently held sources of the same data.

Example

The Manufacturer has determined there is some, albeit low, risk (**DSAL1**) associated with its use of data at a specific point of its lifecycle. The Manufacturer evaluates this risk and considers that the risks should be reduced further by taking some reasonably practicable steps.

Likewise, the Health Organisation has identified **DSAL2** data and needs to ensure further risk reduction activities are undertaken.

A.3 Risk Treatment

Where the DSAL of a data item/data set is evaluated as being too high then mitigations must be introduced to reduce the DSAL to an acceptable level. There are many methods for the treatment of data safety risks that will differ based on the type of data and the DSAL level. For example, some techniques may only be applicable to Dynamic Data as opposed to Configuration Data and the higher the DSAL, the more rigorous the technique should be in assuring that the data property is maintained.

Specific guidance on how data-related risks may be treated is presented in the Risk Treatment tables in Section 6.4 of the Data Safety Guidance [Ref. 3], based on the data type and the DSAL of the data. These methods and techniques represent good practice in mitigating the risks associated with safety-related data. The Risk Treatment tables indicate where a particular method/approach is applicable to a given lifecycle data type, and for each DSAL, whether the method/technique is:

- No recommendation for or against (-);
- Recommended (R);
- Highly Recommended (HR).

The lifecycle data types considered in the referenced version of the guidance are as follows:

- Verification (V);
- Infrastructure (I);
- Dynamic (D);
- Performance (P);
- Justification (J).

Copyright © 2018 NHS Digital & SCSC

The methods/approaches are not intended to be prescriptive, but they should be sufficiently well-defined to allow interpretations to be applied to the given context in which the guidance will apply. Methods/techniques employed are expected to be more rigorously applied as the DSAL level increases. For example, the depth, level of coverage and effort/resources employed for analysis techniques must be proportionate to the DSAL - sampling may be appropriate for lower DSALs where full coverage will likely be expected for higher DSALs. Assurance methods and approaches must be considered for each stage of the data lifecycle as appropriate for the given DSAL. Strategies for dealing with large data sets must be fully justified with respect to the DSAL.

The CRMP must document:

- planned compliance with the Risk Treatment tables;
- the interpretation for the given method/technique (e.g. depth of checking);
- justification in the case where a technique is not to be adopted.

The overall safety justification, the Clinical Safety Case Report (CSCR), for the given project/service/operational context must then provide evidence of compliance against the plan.

It is important to note that the Risk Treatment tables are not intended to be exhaustive. As well as only considering certain data types each Risk Treatment table only considers one particular aspect of data safety. No claim is made that the collection of Risk Treatment tables provides complete coverage either across the system lifecycle or across all possible approaches to one part of the lifecycle. Also, these tables merely identify techniques using a few key words; in almost all cases, further information on the technique should be readily available from a range of sources. Despite these limitations, the guidance they contain should prompt considerations that lead to the use, and justification for the use, of an appropriate set of methods and approaches for any given system.

Example

Having decided that further risk reduction is necessary, the Manufacturer needs to select assurance methods and technique that are appropriate for **DSAL1** data and in doing so demonstrate that reasonably practicable steps have been taken to reduce the risk. The Manufacturer therefore refers to the Risk Treatment tables in [Ref. 3] for guidance.

For **DSAL1** Verification Data, the Risk Treatment tables show that the following are recommended (R) or highly recommended (HR).

In the table, each applicable technique has been given a unique identifier (e.g. SD1) so it can be referenced later:

Copyright © 2018 NHS Digital & SCSC

Ref	Technique	R/HR	System Design
SD1	Sanity /Reasonability Checks	R	Dedicated processing implemented to check that data is within reasonable tolerances and/or logically/semantically consistent with what the data represents. For example, range checks, date checks, record counts, record sizes, special values (e.g., NaN) etc.
SD2	Syntax Checks	R	Semantic checking of data values and sequences based on defined rules.

Ref	Technique	R/HR	Data Design
DD1	Governance Model	R	A governance model is established that defines aspects such as data ownership, processing roles and responsibilities (who can do what to the data), processing authorisations and permissions.
DD2	Data Flow Diagram	HR	To describe the data flow in a diagrammatic form.
DD3	Data Model	HR	To articulate how data is organised.
DD4	Client Sign-off	R	Agreement from the client that the data is appropriate.
DD5	Data Dictionary	HR	A data dictionary is a collection of descriptions of the data objects or items in a data model for the benefit of data users.

Ref	Technique	R/HR	Data Implementation
DI1	Review / Inspection	HR	Manual review / inspection of data possibly involving data visualisation tools.
DI2	Ground-Truth Check	R	Inspection against physical measurements (e.g., lengths, positions, heights) taken in the real world.
DI3	Auditing	R	A period of comprehensive internal and external testing of the data quality process.
DI4	Authorisation	R	A security model is established to control who is authorised to create, view, edit, delete the data.

Copyright © 2018 NHS Digital & SCSC

DI5	Defined Confidence / Trust Levels	R	Criteria are established to provide an objective measurement of the confidence or trust in a given dataset.

Ref	Technique	R/HR	Data Migration
No relevant technique			

Ref	Technique	R/HR	Data Checking
No relevant technique			

Ref	Technique	R/HR	Test Data
TD1	Using Informal / ad-hoc means	R	Data is generated by simple means (eg. spreadsheets, scripts, basic assumptions). There is no formal checking or review of the method of generation.
TD2	Using Manual means	R	Simple test data can be produced by manual means, although this may be prone to human error.
TD3	Using Initial Runs of New System	R	This method is often used where the system is breaking new ground and there is no prototype or legacy system to produce test data. Initial operations may differ from eventual usage, and so the test data suite must also evolve.
TD4	Derived from Real Data	R	Where real data is available this is usually a good basis for generating test data (e.g. by modification to increase the test space coverage).
TD5	Produced by Client	R	Ideally the client is involved in producing or at least checking the test data.
TD6	Client Sign-Off	R	Where possible, the client should formally agree and sign-off the test data as appropriate.
TD7	Error Seeding	R	This is where errors are deliberately inserted into the dataset to demonstrate the effectiveness of data validation.

Copyright © 2018 NHS Digital & SCSC

TD8	Data Reuse	R	Reusing data for one project that was created and thoroughly assured for another project. This can be effective but the read-across should be established.
TD9	Feedback testing	R	To check output data by comparing

Ref	Technique	R/HR	Media - Paper
MP1	Photographic Copies	R	Photocopy and store separately.
MP2	Scan to Electronic Format	R	Retain both paper and electronic copies.
MP3	Indexing / Cataloguing	R	To support efficient accessibility.

Ref	Technique	R/HR	Media - Electronic
ME1	Regular Refresh / Rewrite	R	Of magnetic media or flash memory.
ME2	Suitable Physical Environment	R	Store media in a clean, low-humidity environment at a steady temperature, cool but not cold.
ME3	Copies at Different Locations	R	Physically separate to cover natural disasters, accidental or malicious damage.
ME4	Backups / Duplication	R	Backups are essential. Frequency of backup is dependent on the rate of change of the data. The number of generations of backup to be kept should be commensurate with the impact of data loss.
ME5	Sample Restores	R	Sample restores are performed at intervals to ensure that the backups are readable and retrievable.

Table 10: DSAL1 Recommendations for Verification Data

Copyright © 2018 NHS Digital & SCSC

From these tables the Manufacturer decides on a series of activities to implement the recommendations that are applicable to the endeavour being considered. These activities are expressed as a series of requirements that can be placed on the Manufacturer's delivery organisation and tracked through to completion. Each requirement addresses one or more of the guidance techniques as listed in the Guidance Reference column.

Ref	Requirement	Technique Reference
R1	The verification data shall be carefully controlled in the Manufacturer's configuration management system. There shall be a configuration management plan that shall define who has responsibility for the data and who is authorised to create and amend it.	DI4, DD1
R2	The verification data shall be held on an industry standard file share that is regularly backed up with copies moved periodically to offsite storage. The Backup / Recovery plans shall include periodic sampling of restores.	ME1, ME2, ME3, ME4, ME5
R3	The data shall be modelled as a series of patient "journeys" that cover the entire lifecycle of data from first encounter through to archival and deletion of data. The complete set of journeys shall be chosen to exercise all the functionality of the system. The modelling shall include a data dictionary, data flow diagrams and a data model.	DD2, DD3, DD5
R4	To model data from external systems, the Manufacturer shall use manual data entry and spreadsheet based records to hold the data.	TD1, TD2
R5	The Manufacturer already has a set of clinical standing data that was used for another system and derived from real data. This data includes data such as encounter codes, clinical terms, consultant names, surgery and hospital address etc. and this shall be reused for this system. The Manufacturer's Clinical Safety Officer (CSO) has reviewed the data and agreed on its suitability for reuse.	TD4, TD8
R6	Some of the verification data sets shall include errors deliberately inserted to check the effectiveness of data validation.	TD7
R7	The controlled verification data set shall be subject to review and analysis against defined confidence/trust criteria. Scripts shall be written to check for syntax and semantic consistency of the data and provide a basic sanity check. The scripts themselves shall be validated and verified before use.	SD1, SD2, DI1, DI2, DI5

Copyright © 2018 NHS Digital & SCSC

Ref	Requirement	Technique Reference
R8	The project shall be subject to a delivery quality assurance audit.	DI3
R9	Data loaded from external system into the system and displayed to the user shall be crosschecked against the original source data, using manual spot-checks.	TD9
R10	The level of rigour employed in verifying all the above requirements shall be commensurate with the DSAL criticality and so an ISO9001 compliant quality management system shall be adopted.	All

Table 11: Requirements Established to Implement DSAL1 Recommendations for Verification Data

The following guidance recommendations have not been adopted by the Manufacturer for the reasons given. Note that some may however become relevant in the future so actions are set, where appropriate, to review the applicability of the recommendation when the given condition is met.

Ref	Technique Reference	Justification	Action
E1	TD3	The data will be used before any initial run of the system.	Review when data from initial runs is available.
E2	DD4, TD5, TD6	There is no contracted client at the moment as the system is a new developed, so it will not be possible to get the client to create or signoff data.	Review when contracting with a client.
E3	MP1, MP2, MP3	There are no paper based resources for this system.	No further action

Table 12: Justification for the Non-adoption of Recommendations

Copyright © 2018 NHS Digital & SCSC

Document Information

Author	Paul Hampton/Mark Thomas	**Version issue date**	27.09.2018

Revision History

Version	Date	Summary of Changes
0.1	04.08.2016	First draft
0.2	06.09.2016	Second draft after DSIWG review
0.3	09.09.2016	Updated after NHS Digital review
0.4	15.03.2017	Aligned with Data Safety Guidance v2.0
0.5	31.05.2017	Updated to reflect comments from NHS Digital and the DSIWG Working Group
0.6	30.10.2017	Updated after NHS Digital review
1.0	18.01.2018	Format updated prior to formal issue
2.0	27.09.2018	Updated to reflect the transition from SCCI0129 & SCCI0160 to DCB0129 & DCB0160 respectively

Reviewers

This document has been reviewed by the following people:

Reviewer name	Title / Responsibility	Date	Version
Sean White	Senior Safety Engineer, NHS Digital	27 September 2018	2.0
Clinical Safety Group	NHSD Clinical Safety Governance	17 January 2018	1.0

Approved By

This document has been approved by the following people:

Name	Title	Date	Version
Sean White	Senior Safety Engineer, NHS Digital	27 September 2018	2.0
Stuart Harrison	Head of Safety Engineering, NHS Digital	27 September 2018	2.0

Document Control

The controlled copy of this document is maintained in the NHS Digital corporate network. Any copies of this document held outside of that area, in whatever format (e.g. paper, email attachment), are considered to have passed out of control and should be checked for currency and validity.

Copyright © 2018 NHS Digital & SCSC