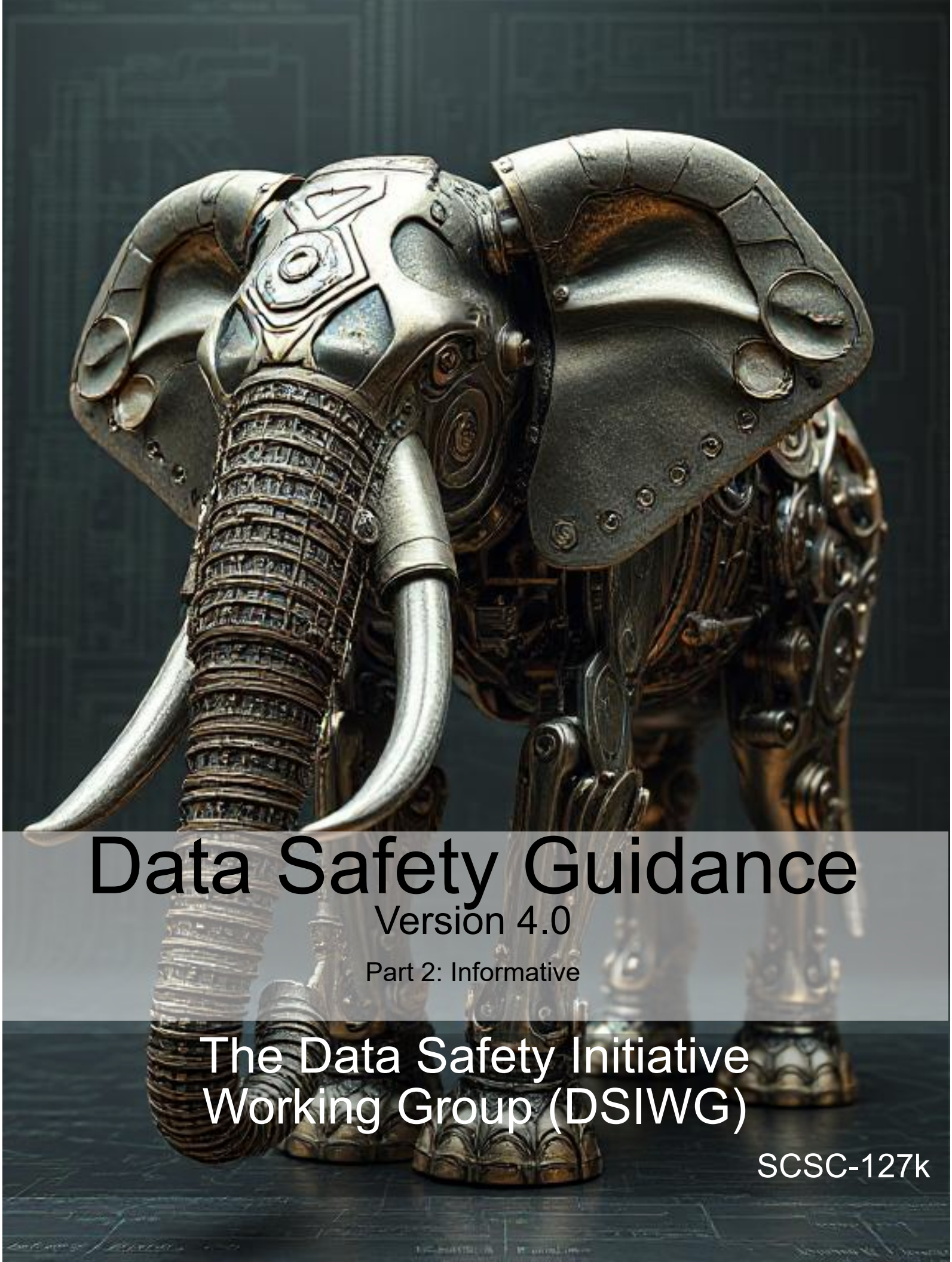

Data Safety Guidance

Version 4.0

Part 2: Informative

The Data Safety Initiative Working Group (DSIWG)

SCSC-127k

SCSC Publication Number: SCSC-127J

This document was prepared using the LaTeX 2_ε typesetting system.

Editing and typesetting by Mark Templeton and Tim Rowe.

Cover design by DeepAI.org and Tim Rowe, based on an original by Paul Hampton.

The Safety-Critical Systems Club (SCSC) is the professional network for sharing knowledge regarding safety-critical systems. It brings together:

- engineers and specialists from a range of disciplines working on safety-critical systems in a wide variety of industries;
- academics researching the arena of safety-critical systems;
- providers of the tools and services that are needed to develop the systems; and
- the regulators who oversee safety.

Through publications, seminars, workshops, tutorials, a web site and, most importantly, at the annual Safety-critical Systems Symposium (SSS), it provides opportunities for these people to network and benefit from each other's experience in working hard at the accidents that don't happen. It focuses on current and emerging practices in safety engineering, software engineering, and product and process safety standards.

This document was written by the Data Safety Initiative Working Group (DSIWG), which is convened under the auspices of the SCSC. The document supports the DSIWG's vision, which is to have clear guidance that reflects emerging best practice on how data (as distinct from software and hardware) should be managed in a safety-related context. This update takes account of the consensus that a process-based guidance document will complement existing safety management processes, making it more usable. It was formally released at SSS'25, 4–6 February 2025 details of which may be found at https://scsc.uk/e1099

Comments on this document are actively encouraged. These can be emailed to:

comments@data-safety.scsc.uk.

Alternatively, a comments submission form is available at:

data-safety.scsc.uk/comments.

Data Safety Guidance

The Data Safety Initiative Working Group [DSIWG]

February 2025

Change History

Version	By	Status	Date
1.0	The DSIWG Team	First draft for external review	31-JAN-2014
1.1	The DSIWG Team	(Internal edition for DSIWG use only)	09-DEC-2014
1.2	The DSIWG Team	For publication at SSS'15	23-JAN-2015
1.3	The DSIWG Team	For publication at SSS'16	29-JAN-2016
2.0	The DSIWG Team	For publication at SSS'17	30-JAN-2017
3.0	The DSIWG Team	For publication at SSS'18	26-JAN-2018
3.1	The DSIWG Team	For publication at SSS'19	01-FEB-2019
3.2	The DSIWG Team	For publication at SSS'20	11-FEB-2020
3.3	The DSIWG Team	For publication at SSS'21	09-FEB-2021
3.4	The DSIWG Team	For publication at SSS'22	08-FEB-2022
3.5	The DSIWG Team	For publication at SSS'23	07-FEB-2023
3.6	The DSIWG Team	For publication at SSS'24	13-FEB-2024
3.7	The DSIWG Team	For publication at SSS'25	04-FEB-2025

Changes Since the Last Edition

The main changes in this edition are:

- the inclusion of a process flow diagram at Section **??**;
- the inclusion of an appendix (Appendix **??**) on AI and autonomy; and
- the inclusion of an appendix (Appendix **??**) on the RADISH tool.

The inclusion of the new Appendix **??** and Appendix **??** means that subsequent appendix numbers have changed.

Hyperlinks (in the electronic edition) to abbreviations, acronyms and defined terms have been substantially expanded and further abbreviations, acronyms and defined terms have been added.

The body of text the guidance, but not yet the appendices, has been edited for clarity and consistency.

Discursive definitions of terms and abbreviations not used in the document have been removed, and new discursive definitions and definitions have been addded.

The definition of hazard has been expanded to include harm, not just accidents.

An inconsistency in the worked example regarding whether there is a contract for the system has been corrected.

To assist users of earlier 3.x versions of the guidance in ensuring that their existing data safety arguments have not been impacted by this update, a version of this document is available which has been annotated with change bars. To avoid clutter, minor changes that should not affect the meaning of the text have not been marked with change bars. The annotated version is available at http://scsc.uk/scsc-127J

Future work

MCA Ltd has continued to work with the DSIWG to develop a prototype software tool to assist in the automation of the processes described in this guidance document. A working version of the tool has been developed and organizations that could benefit from the use and further development of the tool are urged to contact MCA at Mission Critical Applications Limited (radish@mca-ltd.com).

A number of improvements to the guidance are currently planned. These improvements are intended to clarify the application of the data safety process and include:

- further detail on the assurance of communications and data flows,
- data safety considerations associated with distributed datasets and Blockchain,
- addition of new treatments to the tables in section 3,
- review of the tables of treatments, with the aim of making them easier to use,
- further explanation of some treatments, where their use or benefit is not immediately apparent,
- reordering of parts of the document to improve readability, especially as regards likelihood,
- further detail on tool assurance,
- harmonisation of language and guidance on how organizations may expand the tables to incorporate their own internal processes.
- guidance on the application of the data safety culture questionnaire,

Several of these changes are likely to cause parts of the document to be re-ordered – they have therefore been deferred to the next major update, in version 4.0 of the guidance.

If you or your organization are interested in learning more about the work of the DSIWG or joining any of the sub-groups, please visit the SCSC website, where more information including contact details may be found on the "Working groups" section of the site.

Related working groups

The SCSC sponsors initiatives to develop methods and techniques through a number of working groups. These groups each address safety aspects peculiar to their domain, including data aspects when appropriate. The current list of working groups includes:

- Assurance Cases,
- Security Informed Safety,
- Safe AI Working Group,
- Safer Complex Systems.

This page is intentionally blank

Foreword

> *Data is here. Data is growing. Data is causing harm.*

Data is here: Data is becoming ever more important in our lives: influencing, managing and even controlling many critical aspects. The use of artificial intelligence (AI) systems is a new, exciting, but potentially hazardous use of data. Large language model (LLM) based systems are trained on vast amounts of data, and it is this data which enables them to be useful.

Some of this data is related to our personal safety and well-being. Consider, for example, the importance of data defining the layout of railway signals, data that indicates the position of underwater obstructions in nautical channels or data that is used to train a vision recognition system to detect tumours in medical images. Organizations now make significant decisions (including safety-related decisions) based solely on data held in systems. Hence, organizations need to safely manage, control and process their data. In particular, they must actively manage key data properties that preserve safety.

Data is growing: There are at least two reasons why the use of data has grown and, equally important, why it is expected to continue to grow. The first relates to the rapid expansion of the area loosely termed "Big Data", including the use of large datasets to support machine learning and AI applications. The second is the growing use of systems of systems, where data is the lifeblood that connects together disparate elements and allows a cohesive capability to be built. Put simply, the need to address data-related issues is a pressing problem and will continue to be so.

Data is causing harm: Strictly speaking, data can neither cause nor prevent harm. However, mistakes in data, or the inappropriate use of data, within safety-related systems have been factors in a number of documented accidents and incidents. Examples include aircraft attempting to take off from the wrong runway (and consequently crashing), ships running aground, and patients being exposed to higher than planned doses of radiation.

Against this background, the DSIWG was established under the auspices of the SCSC. The DSIWG's aim is to develop clear, cross-sector guidance that reflects emerging best practice on how data (as opposed to software or hardware) should be managed in a safety-related context. For the most part, this guidance is based on well-established techniques, and it has been designed to be compatible with current safety standards and to integrate with existing safety management systems. What is new, however, is the explicit and relentless focus on data, making it a "first-class citizen" within system safety analyses. Because of this focus, this guidance should help organizations identify, analyse, evaluate and treat data-related risks, thus reducing the likelihood of data-related issues causing harm in the future.

Quick Start Guide

> *Data really powers everything that we do.*
> ***Jeff Weiner***

This section provides a single-page introduction to data safety guidance (DSG). For first-time readers this should help place individual sections within an appropriate context. It should also help returning readers quickly navigate the document's contents.

- Systems are changing. The role of data is becoming more prominent. Hence, data needs to be considered as a "first-class citizen" in system safety analyses. This will help mitigate organizational and system-level risks associated with the use of data.
- A data safety management process has been developed. This is based on four phases:
  - establish context;
  - identify risks;
  - analyse risks; and
  - evaluate and treat risks.
- The underlying principles and an overview of the process are described in **??**.
- Normative definitions and abbreviations are described in section 2.
- The objectives associated with, and the outputs produced by, each phase are described in **??**.
- The activities of each phase (and associated tailoring information) are described in **??**.
- Additional guidance information for each phase is described in section 3.
- A worked example is provided in **??**.
- A collection of appendices provide more detail, including:
  - A discussion illustrating how the underlying principles link to the objectives (Appendix **??**);
  - An organizational data risk (ODR) assessment questionnaire (Appendix 4);
  - A data safety culture questionnaire (Appendix 5);
  - A questionnaire to help assess the data maturity of a supplier (Appendix 6);
  - A list of data categories (Appendix 7);
  - A collection of hazard and operability study (HAZOP) guidewords (Appendix 8);
  - The suggested contents of a data safety management plan (DSMP) (Appendix 9);
  - A summary of accidents and incidents in which data was potentially a causal factor (Appendix 10);
  - A discussion of topics loosely related to system lifecycles (Appendix 11);
  - Considerations regarding machine learning (ML) (Appendix **??**);
  - A discussion of the risks of AI and autonomy (Appendix **??**);
  - An introduction to the concepts of both dark and dazzle data (Appendix **??**);
  - The concepts of black swan, dragon king, perfect storm and Pudding Lane data (Appendix **??**).
  - Considerations for the assurance and qualification of data-handling tools (Appendix **??**).

- An introduction to the RADISH tool, that has been developed to assist in the application of the guidance within this document (Appendix **??**).
- Issues that may arise when migrating, porting, importing or exporting data (Appendix 12).
- Some of the data issues that made management of the Covid-19 virus difficult (Appendix **??**);
- Examples of ways that data safety assurance levels (DSALs) may be customised, with particular focus on likelihood (Appendix 13);
- Lists of acronyms, definitions and glossary entries (Appendix 14); and
- A collection of references (Appendix 15).

This page is intentionally blank

Contents

List of Tables

This page is intentionally blank

List of Figures

This page is intentionally blank

1 Introduction (Informative)

> *We're entering a new world in which data may be more important than software.*
> ***Tim O'Reilly***

1.1 Aim and Scope

This guidance document aims to:

- describe the data safety problem;
- provide methods for identifying and analysing levels of risk; and
- recommend methods and approaches for evaluating and treating those risks.

It has been written for a wide readership. Its target audience is all those who have an interest in or a responsibility for safety-related data within systems, including managers, developers, safety engineers, assurers (including independent safety auditors), regulators, and operators.

The document is also intended to cover a number of different sectors. It identifies a wide spectrum of safety-related data that exists in many forms within systems, from specification and requirements data to maintenance and disposal data, and everything in between. In particular, this document is not just concerned with numerical or well-structured data used during system operation.

While they are considered mature enough to be useful, the contents of the document represent current thoughts on what is a complex and evolving area. Furthermore, to allow it to be produced within a reasonable timescale, this edition focuses on key items. It is not intended to be exhaustive. For example, this guidance document does not consider issues relating to staff competence or organizational structure.

1.2 Intended Relationship to Other Documents

This document is intended to be used as a supplement to existing standards and norms that are relevant to the scope of the work being undertaken. It may be used to provide a deeper insight into the risks that data poses to the project team's outputs, allowing them to produce credible improvements to the safety argument. Where a standard or norm sets out specific data-related objectives then, unless agreed otherwise with the regulator or safety duty holder, they shall take precedence over the guidance provided herein.

In the longer term, the hope is that future standards and norms will take up relevant concepts, approaches and methods from those in this document. The DSIWG also hopes that organizations will include the concepts, approaches and methods in their own safety management processes.

1.3 Normative, Informative and Discursive Text

Three types of text are used within this guidance document:

Normative text, which is prescriptive. Typically, this text is restricted to describing objectives and outputs.

Informative text, which is descriptive text that is closely linked to the normative text. Typically, this text provides a suggested way by which compliance with the normative text may be achieved, but alternative means of compliance are possible.

Discursive text, which contains discussions that are relevant to the general topic of data safety, but which are not closely linked to the normative text. A discussion on the relationship between data and software is an example of such text. Descriptions of historical incidents and accidents are another.

Each section and appendix of this guidance document contains a single text type. The relevant type is indicated in the section or appendix title.

1.4 Compliance

There may be occasions when it is desirable or necessary to make a claim of compliance against the objectives listed in this document. Such a claim may be required, for example, if this document is explicitly included as a normative reference from a formal standard. Alternatively, it may be required as part of an organization's internal processes.

To facilitate compliance claims, the following terminology is used within the normative parts of this guidance document:

SHALL denotes items where evidence of compliance must be provided in order to claim compliance with this guidance document.

SHOULD denotes items where, in some circumstances, there may be valid reasons for not complying with a particular item. The full implications of non-compliance must be understood, documented and approved in order to claim compliance with this guidance document.

MAY denotes items that are optional. These may be advantageous in some circumstances but not in others. Organizations are free to adopt any approach to these items without the need for further justification.

The terms have their normal English meanings in discursive and descriptive sections.

This page is intentionally blank

2 Definitions (Normative)

> *It is odd how learned persons fail to see that new terms and definitions are apt to mean new doubts and litigation.*
> ***Frederick Pollock***

artefact, data

An item, or collection of items, that provides a useful perspective on data generated, processed or consumed by a system.

owner, data

The individual or organization responsible for a particular data artefact or collection of data artefacts.

property, data

A characteristic that can be exhibited by a data artefact.

response

The way in which an identified risk is addressed; possible responses include avoid / eliminate, treat, or accept as sufficiently low.

safety assurance level, data

An indication of the level of rigour with which relevant data properties should be demonstrated for appropriate data artefacts.

stakeholder

An individual or organization that has some relationship to the system, possibly including a power of veto.

treatment

An action taken to reduce or control risk. This might be mitigation of the risk or elimination of the hazard.

A more comprehensive list of definitions, including descriptive definitions, is included at chapter 14.

This page is intentionally blank

3 Guidance (Informative)

> *I wanted to separate data from programs, because data and instructions are very different.*
> ***Ken Thompson***

3.1 Establish Context

3.1.1 Interface control

The interfaces between data owners, and indeed data ownership itself, can be much more complicated for data than for hardware or software, where the owner can be clearly identified. Indeed, when combining items from various sources it is possible to create data for which there is not an "owner" in any traditional sense. In such circumstances it may be appropriate for the overall system owner to take responsibility for the collected data and, where appropriate, pass specific, formally recorded requirements on to original data suppliers.

The data owners or lack of data owner throughout the lifecycle of data within the system should be identified, including where data is merged or modified through the system operation. This will help gain a greater understanding of how data safety issues can be controlled within the assessment at a particular organizational level.

3.1.2 Organizational Data Risk Form

section 4 presents an ODR assessment form capture a high-level perspective on the risk posed to an organization by data safety issues within a specific project. How the assessment integrates with an organization's existing risk (or safety) management processes is the responsibility of the implementing organization. Ihe form could be used to help tailor the data safety process. The following paragraphs describe the connections between the ODR and the International Standards Organization (ISO) 31000 [1] standard for risk management.

Establishing the context of a risk assessment ensures that the system being considered and the scope of any assessment is well defined. This helps prevent an overrun of the assessment's boundaries and allows which items are in or out of scope to be explicitly communicated to all stakeholders. In addition, it is the role of this activity to determine the criteria that a system will be judged on. The ODR assessment links directly to the sub-tasks identified by ISO 31000 for establishing the risk assessment context and introduces aspects to guide the assessor in assessing data-specific risks.

Questions 2, 3 and 4 of the ODR align directly with activity 6.3.3 from ISO 31000: establishing the external context of the risk assessment. They guide the assessor in assessing the risk appetite of external stakeholders, the level of risk that is allocated to the organization, and the regulatory environment within the project will operate.

Question 5 establishes the internal context of the risk assessment (Activity 6.3.3 from ISO 31000), helping the assessor determine the maturity of the organization in terms of their attitude to risk and specifically to data-driven risks.

Question 6 explores data ownership through the use cases of the system. This is related to the legal frameworks explored in question 4 but also acts to lay the foundations of activity 6.3.4 from ISO

31000, defining risk criteria, which requires an assessor to identify "the nature and types of causes and consequences that can occur and how they will be measured". Questions 1, 7 and 8 expand on this, considering data-driven specifics of failure consequences and the issues raised by data complexity, boundary complexity, and system complexity for the project.

Finally, the scoring system of the ODR provides a heuristic for defining the risk criteria (Activity 6.3.4 from glsiso 31000) which determines how to combine these different aspects of risk into a single, high-level estimate of the data-related risks associated with a given project. This means that the ODR can, for example, provide some guidance on data safety assurance principle "4 + 1"; that is, it provides some guidance on the amount of effort that should be directed towards the management of data safety issues.

While the completion of an ODR fits within the context establishment activity, it also augments the ongoing communication and consultation activity both by providing a standardised format for capturing the relevant information and securing endorsement.

3.1.3 Data Safety Culture Questionnaire

The ODR assessment includes assessing the organization's maturity in managing data safety risks; the questions are aimed at establishing the depth of awareness of data safety and the associated management processes within the organization. However, measuring the level of awareness of processes and concepts in an organization is not always easy. 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, chapter 5 includes a separate data safety culture questionnaire to explore the specific area of assessing the data safety culture for a particular activity, whether this be for the organization as a whole or for a particular project, service, or activity. 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. This approach 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.

3.1.4 System Definition

The system under consideration should be understood and documented, including interfaces and safety-related data aspects. The process of documenting the system of interest furthers the understanding of stakeholders and approvers so they can make sensible judgements about the system. It also formally declares assumptions that are being made in the system assessment and clearly defines the limits of the assessment. In addition, different levels of risk may be associated with composites of safety-related data, which may be easier to manage than individual data artefacts, or where independence cannot be demonstrated or maintained. Hence, the partitioning of datasets should also be considered during this phase.

3.1.5 Supplier Data Maturity

As already noted, some usage scenarios involve data being supplied by subcontracted organizations. Some formal process will typically be used to select these suppliers. chapter 6 includes a questionnaire to help ensure that the supplier has suitable processes in place to manage data safety-related issues.

3.1.6 Data Categories

The full set of data categories which can have safety implications is large: to date more than twenty categories (and one meta-category) have been identified.

Table 3.1 gives the current view of the categories of safety-related data that contribute to, are used by, produced by, or affected by safety-related systems. They are roughly organized into a number of categories, which aim to cover all aspects of the system lifecycle. The list in Table 3.1 is not exhaustive. A more detailed version of this table is given in chapter 7.

Table 3.1: Categories of safety-related data: concise definitions

No.	Category	Description
	Context	
1	Predictive	Data used to model or predict behaviours and performance
2	Scope, assumption and context	Data used to frame the development, operations or provide context
3	Requirements	Data used to specify what the system has to do
4	Interface	Data used to enable interfaces between this system and other systems: for operations, initialisation or export from the system
5	Reference or lookup	Data used across multiple systems with generic usage
	Implementation	
6	Design and development	Data produced during development and implementation
7	Software	Data that is compiled (or interpreted) and executed to achieve the desired system behaviour
8	Verification	Data used to test and analyse the system, specifically to determine whether it has been built as intended
	Configuration	
9	ML	Data used to train the system
10	Infrastructure	Data used to configure, tailor or instantiate the system itself
11	Behavioural	Data used to change the functionality of the system
12	Adaptation	Data used to configure to a particular site
	Capability	
13	Staffing	Data related to staff training, competency, certification and permits
	The Built System	
14	Asset	Data about the installed or deployed system and its parts, including maintenance data
15	Performance	Data collected or produced about the system during trials, pre-operational phases and live operations
16	Release	Data used to ensure safe operations per release instance
17	Instructional	Data used to warn, train or instruct users about the system
18	Evolution	Data about changes after deployment
19	End of Life	Data about how to stop, remove, replace or dispose of the system

Continued on next page

Table 3.1: Categories of safety-related data: concise definitions (continued)

No.	Category	Description
20	Stored	Data stored by the system during operations
21	Dynamic	Data manipulated and processed by the system during operations
22	Twinning	Data used to create and maintain a digital counterpart of a physical object or process
Compliance and Liability		
23	Standards and regulatory	Data that governs the approaches, processes and procedures used to develop safety systems
24	Justification	Data used to justify the safety position of the system
25	Investigation	Data used to support accident or incident investigations (i.e., potential evidence)
Meta-Property		
+1	Trustworthiness	(Meta) data which tells us how much the system can be trusted

3.2 Identify Risks

3.2.1 Historical Accidents and Incidents

Ideally, data safety risks would be identified and mitigated before they lead to an accident or incident. However, this is not always the case. Historical occurrences can provide an indication of the data safety risks present in planned or existing systems. In particular, accidents and incidents can be analysed to identify potential contributory causes relating to data.

To support this type of analysis a number of previous accidents and incidents have been collected in section 10. These include cases which relate to a number of data properties (see section 3.2.3), for example the properties of completeness, integrity and timeliness. They also highlight the importance of the adaptation data category and dangers associated with the inappropriate use of default data values.

Most of the current collection of accidents and incidents fall into three categories: aviation, maritime, and medical. However, the lessons that can be learned span a much wider range of application areas.

3.2.2 Ways that Data Can Cause Problems

There are some risk-inducing issues that are different or more prevalent for data than for other system elements. An incomplete collection of examples is provided below. This list may provide a quick way of identifying risks, which could be especially useful at an early stage of a project:

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 can be carefully managed and the impact on the safety case appraised. This is not always the case for data, which is often much more fluid. Indeed, the ease with which data can be changed is one motivation for the move towards data-driven systems. This fluidity means that it is not always possible to revisit safety cases when data changes. For example, the safety case for an autonomous vehicle cannot be updated every time that the vehicle acquires new knowledge during operation. Instead, the fact that data can change, along with any associated safety impacts, may need to be captured in the system

safety case. Fluidity can also provide a temptation for unscrupulous operators to falsify data, for example, after an incident has occurred. Rigorous configuration control procedures can help protect against this type of behaviour.

Reuse: For the purposes of this discussion, "reuse" is interpreted as use of the same data in a different system or system context (e.g., lifecycle phase). Just because data was valid for use in a particular system, it does not immediately follow that it can be reused again in a similar system. Many considerations associated with data reuse are similar to those of software reuse, for example, similarity of requirements, similarity of role in the system, and similarity of 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: As highlighted above, all safety-related data has a lifetime and this needs to be explicitly managed. This can involve, for example, purging, deleting and alerting. Ageing can occur as a result of changes external to the system (for example, records of the positions of other aircraft, newly discovered drug interactions, or new software patches), or it can result from internal changes (e.g., valves gradually becoming less responsive, configuration data becoming out of date, or data schemas evolving over time).

Transformation: Data is often filtered, mapped or aggregated as it moves through systems, sometimes creating new datasets as a result. Data properties are not necessarily preserved by these processes. Sometimes data is filtered too much or only some of the data is selected (either deliberately or inadvertently through accident or unintended bias), such as by the selection of only the test runs that succeeded. The main issues are loss of heritage / history / source information. Data can also appear to become something else. Without careful management, the integrity may become lowered to the lowest common denominator and this needs to be recognized. Additional checks (e.g., validation checks, credibility checks) or assurance measures may be needed to ensure that required integrity / assurance is maintained.

Archiving and retrieval: Safety-related data needs to be available when required. Data accessibility needs to be considered over the complete system lifetime. It is also important to consider what properties of the data need to be preserved and how this affects the choice of storage medium.

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. It is usually an unintentional distortion in the dataset. One example of this is the confirmation bias that may be applied to safety claims, and another example is that synthetic autonomous vehicle training databases can have issues with artificial data if it is not realistic. Although there is no perfect way of checking for this within the system, completeness, statistical and validity checks on datasets may help.

Falsification / misinformation: This issue arises where data is created, modified, or deleted either accidentally or deliberately to mislead or misinform potential consumers of that data. Examples from policing and criminal justice might include notes being taken with fabricated times or dates or a crime from the wrong individual in a database being accidentally added or removed. Another example might be a supplier falsifying quality records for materials or goods. There have been many cases of misinformation related to the Covid-19 pandemic (for example on social media) where people have been deliberately misled, and sometimes this has led to harm (for example, bleach being drunk as a cure). Possible mitigations include digitally signing transmitted data, strong access controls, independent fact-checking, and audit records.

item[Defaulting:] Many systems use default or initial values for data items; sometimes in datasets and sometimes embedded in software. Often these default values are designed to be neutral (e.g., "0") or

unrealistic (e.g., "VOID"). There are essentially two cases:

1. initialisation data which may persist and be mistakenly taken as a real value when in fact it should have been changed; and
2. data that is used when no meaningful value has been assigned (e.g., during data migration or data exchange between systems).

These issues can often be managed through good design of data structures, for example by the inclusion of a validity flag.

Sentinels: A sentinel value is a data value that is used to indicate a special action needs to be taken, typically indicating the end of a record or a dataset. The sentinel value should be one that is not allowable in the dataset 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 recognized and so, for example, processing includes or continues past the sentinel; and
2. where the data itself somehow contains the sentinel value and so processing is erroneously interrupted.

Sentinels can be a particular risk in long-lived systems and datasets. As with defaulting, the management or elimination of this issue may often be achieved through improved data structures.

Aliasing: This is an effect that causes different data to become indistinguishable when accessed; that is, there is only one record when there should be several – for example, two patients with similar names inadvertently sharing a single set of medical records. This could be due to the way the data is filtered, sampled, indexed, stored, or retrieved. The data issues are typically related to loss of resolution leading to similar data points appearing to be identical. Methods to maintain resolution, including use of unique indexes, may be beneficial.

One specific case of aliasing concerns homophones – words which sound the same but are different such as "flower" / "flour" and "bare" / "bear" [2]. If these words are read out over a phone there may be confusion as to which is intended. Location services based on a number of words such as W3W don't avoid all homophones, and so there may be issues when the words are given verbally in an emergency. The issue is discussed further in section 10.5.

Disassociation: This effect is, in some senses, the opposite of aliasing: 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 slightly different names. It could also arise if different systems use different indexing methods and the association between the indexes becomes corrupted. Again, methods to maintain data resolution can be beneficial.

Masking: This issue can arise if a notable proportion of a dataset is of a poor data quality, for example, if sensors producing the data are faulty or measurements are taken from the wrong source. This poor quality data can mask errors in the way that the system handles the good quality data. One way of protecting against this issue is the generation and use of test sets of appropriate size and quality, although for some applications this may be a non-trivial task.

Incompleteness: Not all the data that is needed is always available; there may be known, and sometimes, unknown gaps or missing data points. The missing data points ("dark data" – see **??**) can be critical and in some cases, more important than the data that is available. Incomplete data can arise, for example, from limitations on how much data can be physically captured (e.g. sampling frequencies, storage / time constraints) or from the unintentional or deliberate darkening of data (e.g. for privacy, security, political or commercial reasons).

item[Volume:] Data can be so large and unstructured that it is not manageable in practicable timeframes. For example, video records of rail track could take days to inspect manually.

Interpretation: Data can be misinterpreted – too much or too little deduced from available data, or data extrapolated incorrectly to derive unsound results. An example is ML data, especially real or recorded data that may not contain critical edge / corner cases.

Distribution: Data can be decentralised, decomposed or distributed across many sources (e.g. channels, databases, websites) and needs to be consistently integrated to make a coherent picture. In the health sector many IT systems often have to work together feeding in different parts of a patient medical record to make a complete health picture. If parts of this distributed data are missing (for instance diagnostic test results) then it is difficult if not impossible to obtain the complete picture, and mistakes may be made. There are several parts to this problem:

Integration – multiple elements of data have to be brought together in a coherent way, addressing aspects such as which data should supercede or replace other data;

Communication – having correct and current information about what data is available and its location so it can be requested; and

Contingency: – what to to if part of the distributed data is unavailable or late.

Further examples of how data may cause issues in many scenarios is given in the book *Data-Centric Safety: Challenges, Approaches, and Incident Investigation* [3].

3.2.3 Data Properties

Data properties are used to establish what aspects of the data (e.g., timeliness, accuracy) need to be guaranteed in order that the system operates in a safe manner.

James Inge's work [4] produced a useful taxonomy of data categories, and went on to look at faults in data. He concluded that a rigid taxonomy of data categories was unhelpful due to various properties or characteristics of the data which vary independently. In short, it is the combination of data category with the required data properties that facilitates safety analysis.

Data categories were discussed in the preceding phase. Table 3.2 documents a non-exhaustive collection of data properties. Typically, it is the loss of one of these data properties that presents a hazard. This notion of "loss" is dependent on the intended use; for example, what is "timely" for one use may not be for another.

The "Goldilocks" property addresses appropriate sizing and quantity of data. A number of issues have been found to arise when there is too much or too little data. While it is particularly relevant to communications links, it may have relevance to other areas such as databases and when people are involved in reviewing or checking data. The property is named "Goldilocks" as it refers to the need to have not too much, not too little, but just the right amount of data[1].

[1] A system where the property was lost involved a high-speed data bus that connected several safety-critical systems. A transceiver of that bus failed and transmitted random noise. The receivers employed parity checks and cyclic redundancy check, but the system had been designed to eliminate occasional data errors. When random noise filled the bus, several apparently valid messages were created every second, resulting in potentially lethal behaviour.

In a HAZOP carried out during 2020, based upon the HAZOP guidewords in this guidance, the facilitator realised that certain failure modes had not been identified by the HAZOP team. In addition to the issue of system overload already discussed, those omissions also concerned system behaviour following data rejection. In these cases, bad data was detected and rejected, but the consequences of data rejection over an extended period had not been considered.

The Goldilocks property is related to the data volume problem discussed in section 3.2.2, however, given the importance of data sizing and the experience of real-world incidents this is now a separate property.

3.2.3.1 The analysability property

The analysability property recognizes that data is now highly complex and extensive, often large scale, and distributed. And yet, for safety purposes we need to make sure it is of suitable quality and able to support system goals. We therefore need to ensure that it is possible to analyse it for key characteristics and establish meaningful results using tools or other means. This property is related to explainabilty and may be performed by the same set of tools or techniques.

3.2.3.2 The explainability property

Explainability describes the ability to establish what the purpose and effect data (and especially changes to data) has on a system and explain this to relevant stakeholders in terms that they can understand. It is particularly important for learning and AI-based systems. This could be, for example, ML training data or system configuration data. This property is related to analysability and may be performed by the same set of tools or techniques.

Table 3.2: Properties of data

Property	Abbreviation	Description
Integrity	I	The data is correct, true and unaltered
Completeness	C	The data has nothing missing or lost
Consistency	N	The data adheres to a common world view (e.g., units)
Continuity	Y	The data is continuous and regular without gaps or breaks
Format	O	The data is represented in a way which is readable by those that need to use it
Accuracy	A	The data has sufficient detail for its intended use
Resolution	R	The smallest difference between two adjacent values that can be represented in a data storage, display or transfer system
Traceability	T	The data can be linked back to its source or derivation
Timeliness	M	The data is as up to date as required
Verifiability	V	The data can be checked and its properties demonstrated to be correct
Availability	L	The data is accessible and usable when an authorized entity demands access
Fidelity / representation	F	How well the data maps to the real-world entity it is trying to model
Priority	P	The data is presented / transmitted / made available in the order required
Sequencing	Q	The data is preserved in the order required
Intended destination / usage	U	The data is only sent to those that should have access to it
Accessibility	B	The data is visible only to those that should see it

Continued on next page

Table 3.2: Properties of data (continued)

Property	Abbreviation	Description
Suppression	S	The data is intended never to be used again
History	H	The data has an audit trail of changes
Lifetime	E	When does the safety-related data expire
Disposability / deletability	D	The data can be permanently removed when required
Goldilocks	G	The data is just the right size – not too much and not too little
Analysability	Z	The data (including any metadata) is of a suitable size, type and format to enable it be usefully analysed
Explainability	X	The data can be meaningfully explained,by a suitable mechanism, to those who need to understand it

Table 3.3 illustrates where the data issues discussed in Section 3.2.2 can result in loss of one or more data properties (x = potential loss of property).

Table 3.3: Properties that can be lost through issues

	Integrity	Completeness	Consistency	Continuity	Format	Accuracy	Resolution	Traceability	Timeliness	Verifiability	Availability	Fidelity / representation	Priority	Sequencing	Intended destination / usage	Accessibility	Suppression	History	Lifetime	Disposability / deletability	Goldilocks	Analysability	Explainability
Issue	**I**	**C**	**N**	**Y**	**O**	**A**	**R**	**T**	**M**	**V**	**L**	**F**	**P**	**Q**	**U**	**B**	**S**	**H**	**E**	**D**	**G**	**Z**	**X**
Fluidity								x		x		x						x			x	x	x
Reuse		x				x		x	x	x		x			x		x	x	x	x	x		x
Ageing		x				x			x			x			x	x	x	x	x	x			x
Transformation	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Ownership								x		x	x				x	x		x					x
Archiving / Retrieval															x	x	x	x	x	x		x	
Biasing	x	x	x	x		x			x	x		x	x	x								x	x
Falsification / misinformation	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Defaulting	x	x	x	x		x						x										x	
Sentinels	x	x		x		x								x								x	
Aliasing	x	x		x		x	x	x		x	x	x										x	x
Disassociation	x			x		x		x	x	x	x	x	x	x	x	x	x	x	x	x		x	x
Masking	x	x	x	x		x				x		x										x	x

Continued on next page

Table 3.3: Properties that can be lost through issues (continued)

Issue	I	C	N	Y	O	A	R	T	M	V	L	F	P	Q	U	B	S	H	E	D	G	Z	X
	Integrity	Completeness	Consistency	Continuity	Format	Accuracy	Resolution	Traceability	Timeliness	Verifiability	Availability	Fidelity / representation	Priority	Sequencing	Intended destination / usage	Accessibility/indexAccessibility Property	Suppression	History	Lifetime	Disposability / deletability	Goldilocks	Analysability	Explainability
Incompleteness	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Volume										x											x	x	x
Interpretation	x	x	x		x	x	x		x	x		x										x	x
Distribution	x	x	x	x				x			x		x	x	x	x		x		x		x	x

3.2.4 HAZOP Guidewords

A HAZOP [5] provides a structured approach for identifying hazards. It involves a multidisciplinary team collaborating to identify potential hazards and operability problems. Structure and completeness are supported through the use of guideword prompts, for example, considering the implications if software components perform functions early, late or not at all. These prompts are intended to stimulate imaginative thinking, to focus the study and to elicit ideas and discussion.

Table 3.4 lists a set of guidewords for a data-focused HAZOP, based upon the properties defined in Table 3.2. The intent here is to assess the impact of each guideword on the property under consideration. For example, the first row considers *loss* of integrity, *partial loss* of integrity, and so on. The list is non-exhaustive. Other guidewords may be useful for particular systems or may be used to ensure the data safety assessment is fully integrated within the system safety assessment. A more detailed version of this table, including specific HAZOP data considerations, is available at section 8.

Table 3.4: HAZOP guidewords: concise guide

Property	HAZOP Data Guidewords
Integrity	Loss, partial loss, incorrect, multiple
Completeness	Loss, partial loss, incorrect, multiple, insufficient
Consistency	Loss, partial loss, incorrect, multiple, too early, too late, loss of sequence
Continuity	Loss, partial loss, incorrect, late, loss of sequence
Format	Loss, partial loss, incorrect, multiple
Accuracy	Loss, partial loss, incorrect, multiple, insufficient
Resolution	Loss, partial loss, incorrect, multiple, insufficient
Traceability	Loss, partial loss, incorrect, multiple, too early, too late, loss of sequence
Timeliness	Loss, partial loss
Verifiability	Loss, partial loss, incorrect, multiple, too early, too late, loss of sequence
Availability	Loss, partial loss, multiple, too early, too late
Fidelity / representation	Loss, partial loss, incorrect, multiple, too early, too late
Priority	Loss, partial loss, incorrect, multiple, too early, too late
Sequencing	Loss, partial loss, incorrect, multiple
Intended destination / usage	Loss, partial loss, incorrect, multiple, too early, too late, loss of sequence
Accessibility	Loss, partial loss, incorrect, multiple, too early, too late
Suppression	Loss, partial loss, incorrect, too early, too late, too much, too little
History	Loss, partial loss, incorrect, multiple
Lifetime	Loss, too early, too late, incorrect, multiple, loss of sequence
Disposability / deletability	Loss, partial loss, incorrect, too early, too late
Goldilocks	Loss, partial loss, incorrect, too early, too late, too much, too little (insufficient), spurious
Analysability	Loss, partial loss, incorrect, too much, too little, loss of sequence, loss of tooling
Explainability	Loss, partial loss, incorrect, too early, too late, too much, too little, loss of sequence, loss of skills, loss of tooling

3.3 Analyse Risks

3.3.1 Establishing DSALs

Section **??** presented a method for assigning DSALs, by combining severity and likelihood through a risk matrix, such as **??**. This section describes approaches that may be used to determine the severity and likelihood appropriate to the loss of any given property. In principle, Table 3.5 and Table 3.6 are used together with **??** to determine a DSAL. However, the methods to determine severity and likelihood, and to combine them to determine a DSAL, should not be followed rigidly, but should be defined based upon the overall safety requirements of the system being assessed. Examples of situations that may require such alternate approaches are described in section **??**, while potential approaches to customisation are described in chapter 13. It is essential that any customised approach is recorded in the DSMP.

The likelihood of a data safety related risk is qualitatively determined by consideration of the significance of a data error, along with the defences currently in place against such errors. These factors may be addressed by considering the following characteristics:

Proximity: how directly a data failure will lead to an accident;

Dependency: how dependent the application is on the dataset;

Preventability: the ability of the systems architect / developers to guard against data errors;

Detectability: the likelihood of being able to detect a data failure prior to an accident; and

Correctability: the ability of the system to work around or correct data errors.

For example, errors that are easy to guard against are associated with low likelihoods. Conversely, errors that are difficult to detect are associated with high likelihoods. Table 3.5 illustrates how aspects of these characteristics map to three, qualitative likelihoods. The table assumes the actions implied are taken. For example, a low likelihood for the "prevention" characteristic is only valid if the easy guard / barrier is actually implemented. Similarly, it is assumed that if an error is found, under the "detection" characteristic, an appropriate response is implemented.

Table 3.5: Calculation of likelihood

	Likelihood		
	High	**Medium**	**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.
Prevention	Difficult or impossible to guard against errors.	Possible to guard against errors.	Easy to guard against error.
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.
Correction	Difficult or impossible to correct or workaround errors.	Possible to correct or workaround errors.	Easy to correct or workaround errors.

Consideration of different characteristics is likely to result in different likelihoods. The overall DSAL is that associated with the lowest likelihood of any characteristic. Taking the lowest likelihood, not the highest, might seem non-intuitive. However, in doing so, it is important to only apply those rows of the table which are valid for the issue under consideration. For example, if it would be easy to implement a guard against a given error, but that guard is not actually in place, it would not be valid to claim that the likelihood is low based on the prevention row. In theory it should always be possible to determine likelihood based upon proximity and dependency. However, the benefits of prevention, detection and correction will not always be present.

Risk severity is estimated against a five-point scale, as indicated in Table 3.6.

Table 3.6: Definition of severity

Severity	Description
Minor	Minor injury or temporary discomfort for one or two people. Minor environmental impact.
Moderate	Minor injuries affecting several people or one serious injury. Some environmental impact.
Significant	Minor injuries affecting many people or a few serious injuries. Significant environmental impact.
Major	Serious injuries affecting a number of people, or a single death. Major environmental impact.
Catastrophic	Several deaths. Possibly affects the general public or has wide and catastrophic environmental impact.

As noted earlier, DSALs have some commonality with things like item development assurance levels (IDALs) and functional design assurance levels (FDALs)). However, this commonality does not extend across all aspects. For example, there is an accepted calculus of FDALs in which two independent lower-integrity

functions can be used to replace a single higher-integrity function. There are two reasons why this type of calculus is not appropriate for DSALs:

1. The definition of a DSAL already caters for interactions. For example, using two independent data artefacts to provide similar information to a system function reduces the "dependency" of each artefact.
2. These types of consideration are most closely related to system architecture, from which data artefacts, associated data properties and risks are derived. Hence, rather than applying any calculus at the DSAL tier it is more appropriate to apply this to, for example, FDALs, with DSALs changing as a consequence of the revised system definition.

3.3.2 Analysing DSALs

When considering the possibility of data affecting software or software affecting data, the degree of contribution and existing mitigation position are important. The mitigations should be proportionate and full credit for existing mitigations may reduce or remove the need for additional work. A particular case is the use of strong checksums to "wrap" the data. If these are preserved through processing and can be checked later on, then undetected corruption can be largely discounted.

The issue with data affecting software is that the data used by the system might affect the software execution in a way that could credibly lead to hazards, but only where this data-induced effect is not easily detected or mitigated by other means. If this is the situation then appropriate measures need to be put in place *within the data* to mitigate this risk. Alternatively (or additionally), if the software within the system can be modified, mitigations could be placed *within the software* to achieve or enhance the mitigation needed:

Mitigation within the data: In many cases, this is the best or only option is to improve the quality of the data to avoid the issue. This can be done by introducing a DSAL for the data, related to the severity of the hazard which may be induced, and thereby addressing the issue at cause. In this case, the DSAL is bearing a large amount of responsibility. This may mean that a greater number of the "recommended" risk treatment methods and approaches (identified in the following phase) need to be implemented.

Mitigation within the software: If the data cannot be assured to a DSAL (e.g., if it is supplied by a third party or legacy system), sometimes changes can be made to the software in the system to improve the situation. These mitigations could be functional (e.g., introduction of better range checking, rejection of illegal combinations of data values). This requires not only specific software changes but also associated verification of these new features, and any software changes will have to be implemented to an appropriate software assurance level. However there may be no particular functional mitigations that can be targeted at the particular issue (e.g., testing for illegal combinations of values might be too complex). In this case, development (or re-development) to a suitable software assurance level of the complete set of software within the system should be considered

The issue with software affecting data is that the software may affect (e.g., corrupt, delete) the data in a way that key safety properties may be lost and that such loss may not be easily detected. In general, the key data properties should be considered to see if any important ones for this data may be jeopardised. If so, a software assurance level from an appropriate standard or guideline should be introduced that mitigates this risk of undetected property loss. This software assurance level should be determined by the hazards that could be caused and may be localised to the software that can cause the problem.

There are specific functional and architectural approaches that may reduce or avoid the need for a software assurance level, including use of strong checksums and digital signatures, as well as techniques such as storing multiple copies, independent channels and so on. The software performing the check of the data property will itself need to be developed to the introduced software assurance level. The key is to establish what the software in the system is doing to the data: if the operations are simple and non-changing, then the risk is lower; if the operations are complex involving transforming the data, using the data to calculate and insert new values, or reformatting the data, then the risk is higher.

The effectiveness of the functional mitigations in reducing impact to the particular data properties needs to be determined. For example, a strong checksum may be very effective at detecting unwanted change and therefore lower the software assurance level (from the perspective of data-related requirements). However, a checksum may not help at all if the issue is one of timely message delivery. More information on the use of checksums in the aviation domain is available in [6]; this information is likely to be applicable to other domains as well.

3.4 Evaluate and treat Risks

3.4.1 Risk Treatment

There is a range of approaches that could be used to treat data-related risks. One option might be to redesign part of a system, either to remove the risk or to incorporate safety devices. Alternatively, ways of mitigating (or, more generally, treating) the risk could be devised. Another option could be to conduct further analysis, for example to better understand the likelihood of a risk occurring. In extreme cases, risk evaluation may lead to a recommendation to cancel a project.

It is apparent that some of these approaches involve repeating activities (or part of activities) discussed in earlier sections of this document. This type of repetition is to be expected given the iterative nature of risk management.

Discussion is an important part of risk evaluation, allowing a variety of different perspectives to be brought to bear. Documentation is also important, partly to allow these discussions to occur on an even footing and partly to ensure that decisions and supporting rationale are recorded.

3.4.2 Mitigating Data Safety Risks

A range of methods and approaches can be used to mitigate the identified data safety risks. Since mitigation can be a complex process, requiring collaboration with all system engineering elements, a collection of high-level mitigation measures is provided. These may particularly assist those attempting to explain the process to non-practitioners or those conducting assessments in less highly regulated environments. These high-level mitigation measures may prove sufficient for practitioners assessing systems which do not have a high safety criticality.

For practitioners conducting assessments in highly regulated environments, or for highly safety-critical systems, appropriate mitigation measures should be derived from the high-level table. To assist these practitioners, suggested methods and approaches are provided in a series of more detailed tables. These methods and approaches have been developed through cross-industry collaboration, but they may not be complete, especially for different types of system.

The practitioner should always consider whether the mitigation measures used to mitigate the data safety risks are sufficient for their purposes. In highly safety-critical systems, each data safety risk can be linked to a system-level hazard. Each hazard should be tracked in accordance with the existing system safety method, and mitigations should be reviewed for their feasibility, potential to introduce new or amended hazards, and effectiveness.

3.4.2.1 High-level mitigation measures

Table 3.7 presents a number of generally applicable mitigation measures. Each of these mitigation measures should be reviewed to establish whether it is relevant to the system under assessment.

For each DSAL, the tables indicate whether the method / approach is:

- Highly recommended (HR);
- Recommended (R); or
- Neither recommended nor not recommended (-).

Table 3.7: High-level mitigation measures

Ref	Mitigation Measure	DSAL				Example(s)
		1	2	3	4	
M.01	Documentation of data context and suitability for use	HR	HR	HR	HR	Data flow diagram to document and agree how data is handled in the system, recording the impact of design decisions on the data aspects of the safety case
M.02	Definition of data ownership through the data lifecycle in the system	R	R	HR	HR	Governance model, interface control document (ICD)
M.03	Definition and traceability of data requirements	HR	HR	HR	HR	Requirements management, use of test data / test cases
M.04	Recorded trustability of the data source(s)	R	R	HR	HR	Source of the data is trusted (with 'trusted' to be defined in detail for the system), or there are multiple sources of data which are correlated
M.05	Editing limitations	R	R	HR	HR	Encapsulation[2] of data, access limitations
M.06	Diverse and / or redundant manipulation of data	R	R	HR	HR	Data partitioning separation of data that is managed differently (architectural decisions)
M.07	Automatic system checking functionality	-	R	HR	HR	Built in test (BIT), heartbeat functionality
M.08	Monitored, controlled, or redundant manipulation of data	-	-	R	HR	Redundant channels processing the data as hot standby
M.09	Diverse and / or redundant storage of data	R	R	HR	HR	Redundant storage of data, multiple different media types used to back up the data
M.10	Data recovery mechanisms	R	R	HR	HR	Backward recovery, error correcting codes
M.11	Tracking of data	R	R	HR	HR	Digital signatures, sequence numbers, logging of data processing events, using metadata, configuration management
M.12	Recorded derivation of test data	R	R	HR	HR	Test data derived from an established system and supported by field evidence, or from another 'trusted' source
M.13	Documented compliance against the data requirements	HR	HR	HR	HR	Use of test data / test cases

3.4.2.2 Detailed methods and approaches

The following tables detail methods and approaches which may be used by practitioners conducting safety assessments. The tables map the methods and approaches to data categories. These data categories are

[2] Sometimes referred to as "data hiding", encapsulation hides the physical representation of data.

abbreviated using the scheme shown in Table 3.8.

The five data categories presented in Table 3.8 are a subset of those presented in Table 3.1, which presents a comprehensive list of data categories. The five presented in Table 3.8 have been used to populate the tables used for the selection of methods and approaches. Users of the guidance are also encouraged to add to this table as part of the customisation process associated with their own organizations or projects.

Table 3.8: Data category abbreviations

Data Category	Abbreviation
Verification	V
Infrastructure	I
Dynamic	D
Performance	P
Justification	J

The tables also map the methods and approaches to the data properties. The properties which were defined in Table 3.2 have been assigned abbreviations, shown in that table.

These detailed methods and approaches tables have been organized into eight, loosely defined categories:

- System design;
- Data design;
- Data implementation;
- Data migration;
- Data testing;
- Test data;
- Media – paper; and
- Media – electronic.

The method for using these tables when considering any given data artefact is:

- Determine the DSAL applicable to the data artefact, using **??** in association with Table 3.5 and Table 3.6.
- Determine the data category, using the abbreviations in Table 3.8.
- Determine the list of applicable data properties, using the abbreviations in Table 3.2. This will generally provide a list of several applicable abbreviations.

Then consider each row in each table (or each table that you wish to use for this assessment). For each row, if:

- the data category matches the data category for this data artefact; and

- at least one of the data properties listed in that row matches a data property for the data artefact

then the row is likely to be applicable to the data artefact. The result for that row is therefore found from the DSAL column, since if:

- The relevant DSAL entry is "HR", then use of the technique on this row is highly recommended.
- The relevant DSAL entry is "R", then use of the technique is recommended.
- The relevant DSAL entry is "-", then the technique should be considered, as it may be relevant in certain application domains, but generic guidance would not be appropriate for specific contexts.

To give a specific example, consider the row corresponding to the first technique in Table 3.9. This technique will apply to data artefacts where:

- The data category is "dynamic" ("D" from Table 3.8) and
- The data artefact holds one or more of the properties integrity, completeness or verifiability ("I", "C" or "V" from Table 3.2).

In such a case, the technique would be highly recommended if the data artefact were of DSAL 3 or 4, recommended if the DSAL were 2, or should merely be considered if the DSAL were 1.

Many dots have been placed in the table to indicate data categories and data properties which are not applicable to the technique under consideration. The dots enable the tables to be assessed very quickly, as they enable the letters representing specific data categories and data properties to always be presented in the same position within the table. For example, the data property "verifiability" ("V" from Table 3.2 can easily be seen to appear against the first two techniques of Table 3.9, whereas it does not appear in the next seven techniques. This can be seen without actually reading the text, but merely looking at the pattern presented by the letters and dots.

Within each table, the "Serial" column lists a unique serial number for each technique. The serial numbers simply provide a unique reference which aligns with those implemented within the toolset.

3.4.2.3 System Design

Table 3.9: Mitigation methods: system design

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Gglsfmttextdata proper
SD.01	BIT / built in test equipment (BITE)	..D..	-	R	HR	HR	Application tests the data (e.g., at start-up or when requested by an operator).	IC.....V.....

Continued on next page

Table 3.9: Mitigation methods: system design (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data property
SD.02	Cyclic / continuous BIT	..D..	-	-	R	HR	Application applies tests to the data it is processing continuously (e.g., for a live data stream) or periodically (e.g., every nth message, every hour).	IC.Y.....VL.....
SD.03	Backward recovery	..D..	R	R	HR	HR	If a fault in data has been detected, the system resets to an earlier internal dataset, which has been proven consistent.	IC.....
SD.04	Parity checks	..D..	R	R	HR	HR	Within data, e.g., Hamming codes, Reed-Solomon, Hagelbarger.	I.....
SD.05	Automatic error correction	..D..	R	R	HR	HR	Detected errors are corrected automatically.	IC.....
SD.06	Checksums / cyclic redundancy checks (CRCs) / Hashes	..D..	-	R	HR	HR	Digests of datasets are produced, included with the dataset and checked to provide confidence that the data is unaltered.	IC.....G..
SD.07	Digital signatures	..D..	-	R	HR	HR	For non-repudiation and integrity of data.	I.....T.....U.....G..
SD.08	Sequence numbers	..D..	R	R	HR	HR	Data bears sequence numbers so the integrity of a data stream can be checked (e.g., monotonic increase, duplicate detection).	ICN.....PQ.....
SD.09	Automatic repeat request	..D..	R	R	HR	HR	Automatic repeat-request (ARQ) to repeat transmission of data which has not been received correctly.	IC.Y.....G..
SD.10	Auditing facilities	..DP.	-	R	HR	HR	Changes to data properties are audited so the before and after values are recorded and also other related information such as the author and the time of the change.	.....T.V.....H...ZX
SD.11	Logging facilities	..DP.	R	R	HR	HR	Data processing events are logged to allow support staff to monitor the health of the system and provide diagnostic information.	.....T.....H...ZX

Continued on next page

Table 3.9: Mitigation methods: system design (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data property
SD.12	Encapsulation	..D..	R	R	HR	HR	Data is hidden so that it is only accessible through well-defined interfaces.	.....UB.....
SD.13	Multiple stores	..D..	-	-	R	HR	The same instance of a dataset or data items is stored in multiple locations.	.....B.H.....
SD.14	Homogeneous redundancy	..D..	-	-	R	HR	Data is processed using homogeneous redundant channels; detected faults in data of one channel cause processing to switch to another channel.	IC.Y....M.....
SD.15	Heterogeneous redundancy	..D..	-	-	R	HR	Data is processed using heterogeneous redundant channels; detected faults in data of one channel cause processing to switch to another channel.	IC.Y....M.....
SD.16	Data integrity sampling	..D..	HR	HR	R	R	The integrity of subsets of data is periodically checked, in accordance with a given selection criteria (e.g., random, critical records).	IC..O.....L.....
SD.17	Credibility / reasonability checks	V.D..	R	R	HR	HR	Dedicated processing is implemented to check that data is within reasonable tolerances and / or logically / semantically consistent (e.g., range checks, date checks, record counts, record sizes, special values - not a number (NaN)).	I...O.....G..
SD.18	Data correlation	..D..	R	R	HR	HR	Data from a number of sources exists to permit a cross-correlation of the data supplied from one source (the master) with other sources.	ICNY.....
SD.19	Data partitioning	..D..	R	R	HR	HR	To separate data that is managed differently, creating independence so that a whole dataset does not require validation after a change.	.....B.....
SD.20	Syntax checks	VID..	R	R	HR	HR	Semantic checking of data values and sequences based on defined rule sets.	I.N.O.....G..

Continued on next page

Table 3.9: Mitigation methods: system design (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data property
SD.21	Feedback testing	..D..	HR	HR	R	R	To check output data by comparing it with the input source.	IC.Y...T.V.........G..
SD.22	Information redundancy	..D..	HR	HR	R	R	Additional redundant information is supplied from diverse sources. The diverse sources can be checked against each other.	IC.................G..
SD.23	Reverse translation	..D..	-	R	HR	HR	Verifying data output of a process is correct, by attempting to create the source data from the output data and comparing this with the original source.	IC.Y...T.............X
SD.24	Metadata	..DP.	-	R	HR	HR	Auditable data are sent with the data that is about the data (e.g., source, issue state, expiry date).	..N....T.V..PQ....E..ZX

3.4.2.4 Data Design

Table 3.10 addresses design aspects, including the construction of data storage structures and methods.

Table 3.10: Mitigation methods: data design

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
DD.01	Governance model	VI..J	R	R	HR	HR	A governance model is established that defines, e.g., data ownership, processing roles and responsibilities, processing authorizations and permissions.	I....A.T......U.S..D..X
DD.02	Data process definition	VIDPJ	-	R	HR	HR	Documented and agreed process definitions for how data is handled.	.......T......U.....G.X
DD.03	Data flow diagram	VIDPJ	HR	HR	HR	HR	To describe the data flow in a diagrammatic form.	.............U......Z.
DD.04	Data model	VIDPJ	HR	HR	HR	HR	To articulate how data is organized.	..N.O..............ZX
DD.05	Client sign-off	VI.PJ	R	R	HR	HR	Agreement from the client that the system architecture and design are appropriate for the data considered.	......R..V............
DD.06	Data quality Correction mechanisms	...P.	-	R	HR	HR	Design incorporates a data quality management system.	IC.Y..................
DD.07	Configuration management	VIDPJ	HR	HR	HR	HR	A formal process that controls changes to the data and data model.	.......T.........H.....
DD.08	Data dictionary	VIDPJ	HR	HR	HR	HR	A collection of descriptions of the data objects or data items in a data model for the benefit of data users.	..N.O.R....F.........ZX
DD.09	Formal methods	..D..	-	R	R	HR	To specify data (or data formats) in a precise, mathematical manner.	.CN.O..T....PQ......GZX

3.4.2.5 Data verification

A number of issues need to be considered during the implementation phase of a programme, to ensure that at any point it is known how much reliance can be placed on the available data. Table 3.11 addresses the mitigations relevant to this whole programme phase, not just from data capture or generation. It therefore includes aspects of data management, checking and expiry.

Table 3.11: mitigation methods: data verification

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
DI.01	Review / inspection	VIDPJ	HR	HR	HR	HR	Manual review / inspection of data possibly involving data visualization tools.	IC..O.....L.....X
DI.02	Statistics-based sampling	VIDPJ	-	R	HR	HR	More appropriate for real-time large and / or volume data. Could be manual selection, a form of random selection or comparison against statistical norms.	I.NY.A.....Z.
DI.03	Ground-truth check	VIDPJ	R	R	HR	HR	Inspection against physical measurements (e.g., lengths, positions, heights) taken in the real world.	ICN..AR..V.F.....X
DI.04	Auditing	VIDPJ	R	R	HR	HR	A period of comprehensive internal and external testing of the data quality process.	ICNYO....V.....X
DI.05	Tracing	VIDPJ	-	R	HR	HR	Ability to trace data from source across multiple participants in the data supply chain.	.......T.V.....X
DI.06	Defined verification frequency	VIDPJ	-	R	HR	HR	Data should contain an indicator of how often it should be revalidated against other (e.g., real-world) source.	.........V........E....
DI.07	Defined data lifetime(s)	VIDPJ	R	R	HR	HR	Information showing when data validity expires.	.....E....
DI.08	Data quality trend analysis	VIDPJ	-	-	R	HR	Checking that a dataset is consistent with a model of the expected data behaviour (e.g., vibration data increases over time).	IC.Y.....V.F.....Z.
DI.09	Authorization	VIDPJ	R	R	HR	HR	A security model is established to control who is authorized to create, view, edit, delete the data.	.....UBS..D...

Continued on next page

Table 3.11: Mitigation methods: data verification (continued)

Serial	Technique	Data Category	DSAL				Notes	Data Property
			1	2	3	4		
DI.10	Authentication	VIDPJ	R	R	HR	HR	Data is authenticated to validate its provenance.	.....T.V.....X
DI.11	Defined confidence / trust levels	VIDPJ	R	R	HR	HR	Criteria are established to provide an objective measurement of the confidence or trust in a given dataset.	IC.Y....V.F.....X
DI.12	Independent check	VIDPJ	-	-	R	HR	A separate person or system is used to check the data independently.	I.....V.....
DI.13	Update comparison	VIDPJ	-	R	R	HR	Updated data is compared to its previous version (e.g., so the list of changed elements can be compared with a supplier-generated list).	.....T.....H..G..

3.4.2.6 Data migration

Table 3.12 addresses migration of data from one system to another system.

Table 3.12: Mitigation methods: data migration

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
DM.01	Manual load	..D..	R	R	-	-	Data is entered into the system manually relying on human validation and verification.	ICNYOA.....F..........
DM.02	Dedicated translation and loading platform	..D..	-	R	HR	HR	For example, using mature enterprise migration commercial off-the-shelf (COTS) products.	ICNYOA.T...F..........
DM.03	Existing / established system transfer	..D..	-	R	HR	HR	Use of an existing / established proven transfer mechanism.	ICNYOA.T...F..........
DM.04	Client supervision	VIDPJ	-	R	HR	HR	The client provides independent supervision of activities checking processes, inputs and outputs at agreed points.	ICNYOA.....F..........
DM.05	Client sign-off	VIDPJ	-	R	HR	HR	Formal acceptance of the migrated datasets in the target system.	ICNYOA.....F..........
DM.06	Incremental switch-over	..D..	-	R	HR	HR	Users are incrementally switched over to the new system rather than as a “big bang”.	ICNYOA.....F..........
DM.07	Parallel load with existing system	..D..	-	R	HR	HR	Parallel running of the new system alongside the existing system with data crosschecks between the two systems.	ICNYOA.....F..........
DM.08	Shadowing	..D..	-	R	HR	HR	Parallel running of the new system alongside the existing system, only data from the existing system is used operationally, with an experienced user crosschecking between the two systems.	ICNYOA.....F..........
DM.09	End-to-end import-export verification	..D..	-	R	HR	HR	Data is traced and verified at all stages through the entire end to end migration process.	ICNYOA.T...F.........X

Continued on next page

Table 3.12: Mitigation methods: data migration (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
DM.10	End-to-end size compare	..D..	R	R	HR	HR	Data is extracted from new or final system and its size or volume compared with original data as input	IC.Y.....V.........G..
DM.11	End-to-end content compare	..D..	R	R	HR	HR	Data is extracted from new or final system and compared with original data as input, possibly on a sample basis	ICN....T.V.F..........
DM.12	Validation campaign	..D..	R	R	HR	HR	An extensive set of validation checks is performed on the migrated data.	ICNY.....V............
DM.13	Interpretation check	..D..	R	R	HR	HR	Migrated data is checked for misinterpretation in the new system (e.g. due to units, national, or cultural aspects).	..N.O....V.F..U......Z.
DM.14	Data cleanse trial	..D..	R	R	HR	HR	Data cleansing is tried on subsets of the data and special cases, before being applied to the whole set (e.g. removing time-expired, obsolete or repeated data)	IC..............EDG..
DM.15	Metadata preservation	..D..	-	-	R	R	Any meta-data as part of the original dataset which cannot be incorporated or translated into the new system is preserved	IC.Y...T.........H...ZX

3.4.2.7 Data checking

Table 3.13: Mitigation methods: data checking

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
DC.01	Limited / pre-operational deployment	.IDP.	-	R	HR	HR	A period of monitored operation in a specially chosen environment.	ICN..A.....F..........
DC.02	Client sign-off of data	VI.PJ	-	R	HR	HR	Agreement from the client that the data is appropriate.	......R..V............
DC.03	Non-critical trialling	..D..	-	R	HR	HR	Monitored operation in an operational, but non-critical, environment.	.....A.....F..........
DC.04	Beta testing	V....	-	R	HR	HR	Testing with a small group of specially chosen users.	ICN..A.....F..........
DC.05	Parallel running	.IDP.	-	R	HR	HR	Running two systems in parallel and crosschecking between them.	ICN..A..M..FP..........
DC.06	Checklists	.IDP.	R	R	HR	HR	Using a checklist to verify that system behaviour is correct, prior to use. This approach can also be effective for detecting otherwise dormant failures and reduces time at risk.	ICN.O....VLFP..........
DC.07	Widespread distribution to user community	.IDP.	-	R	HR	HR	Large-scale distribution to all users.	ICN.OAR.M.LFP..........

3.4.2.8 Test data

Table 3.14: Mitigation methods: test data

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
TD.01	Using informal / ad-hoc means	V....	R	R	-	-	Data is generated by simple means (e.g, spreadsheets, scripts, basic assumptions). There is no formal checking or review of the method of generation.	ICNY.A.....F..........
TD.02	Using generic testbed	V....	-	R	HR	HR	A testbed is a good way to produce test data. It may require configuration and tailoring for the particular application, and this configuration should be managed.	ICNY.A.....F........G..
TD.03	Using simulator	V....	-	R	HR	HR	Simulators (software or hardware) may be able to produce very good test data, obviously depending on how close and detailed a simulation they can achieve.	ICNYOAR....F........G..
TD.04	Using prototype	V....	-	R	HR	HR	Prototypes are often a good way of generating test data for the real system. However they may not produce data with the appropriate range, accuracy or precision.	ICNY.A.....F........G..
TD.05	Using manual means	V....	R	R	-	-	Simple test data can be produced by manual means, although this may be prone to human error.	ICNY.A.....F........G..
TD.06	Using dedicated platform	V....	-	R	HR	HR	For complex and time-critical systems a dedicated test platform is required which can produce realistic test data for all interfaces and inputs.	ICNY.AR..V.FPQ.........
TD.07	Using existing / established system	V....	-	R	HR	HR	Where a new system replaces an old one, then data can often be extracted from the old system to test the new one. Data formats may change so translation may be required.	ICNYOAR.MV.FPQU.....G..

Continued on next page

Table 3.14: Mitigation methods: test data (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
TD.08	Using initial runs of new system	V....	R	R	R	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, so test data must evolve.	ICNYOAR.MV.FPQ.....
TD.09	Derived from real data	V....	R	R	HR	HR	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).	ICNY.A.....F........G.X
TD.10	Statistical profiling post-production	V....	-	-	R	HR	If a statistical analysis of the data can be produced then greater confidence in the quality of the test data can be obtained.	ICNY.A...V.F.........X
TD.11	Produced by client	V....	R	R	R	HR	Ideally the client is involved in producing or at least checking the test data.	ICNY.A...V.F........G..
TD.12	Client sign-off	V....	R	R	HR	HR	Where possible, the client should formally agree and sign off the test data as appropriate.	ICNY.A...V.F.........X
TD.13	Error seeding	V....	R	R	HR	HR	This is where errors are deliberately inserted into the dataset to demonstrate the effectiveness of data validation.	ICNYOAR.MV.F........G..
TD.14	Data reuse	V....	R	R	HR	HR	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.	ICNY.A.....F........G.X
TD.15	Feedback testing	V....	R	R	R	R	To check output data by comparing it with the input source.	ICNY.A.....F..........

3.4.2.9 Media – Paper

Table 3.15: Mitigation methods: data media handling – paper / physical storage

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
MP.01	Photographic copies	VIDPJ	R	R	HR	HR	Photocopy and store separately.	.C.....B.H.....
MP.02	Scan to electronic format	VIDPJ	R	R	HR	HR	Retain both paper and electronic copies.	.C.....B.H.....
MP.03	Copies held at different locations	VIDPJ	-	R	HR	HR	Meaning of "different" depends on data criticality and similarity of location-based risks.	.....VL....B.....
MP.04	Limited access	VIDPJ	-	R	HR	HR	Control (e.g., by procedure) who can access the data.	.....U.....
MP.05	Secure storage	VIDPJ	-	R	HR	HR	Physical measures to prevent unauthorized access.	.....U.....
MP.06	Manual inspection	VIDPJ	-	R	HR	HR	Used to check data when generated and periodically thereafter.	IC.....
MP.07	Suitable physical environment	VIDPJ	-	R	HR	HR	For example, prevent water ingress, control temperature.	I.....L.....E....
MP.08	Defined handling procedures	VIDPJ	-	R	HR	HR	To ensure that changes to the data can be attributed.	I.....UB.H.....
MP.09	Repair / restoration programme	VIDPJ	-	-	R	HR	To protect against degradation and to ensure availability.	I.....L.....
MP.10	Indexing / cataloguing	VIDPJ	R	R	HR	HR	To support efficient availability.	.....L.....
MP.11	Lifetime planning	VIDPJ	-	-	R	HR	For example, to avoid gradual quality reduction by repeatedly "copying a copy".	.....ED...

3.4.2.10 Media – electronic

Table 3.16: Mitigation methods: data media handling – electronic storage

Serial	Technique	Data Category	DSAL				Notes	Data Property
			1	2	3	4		
ME.01	Regular refresh / rewrite	VIDPJ	R	R	HR	HR	Of magnetic media or flash memory.	I.....E....
ME.02	Suitable physical environment	VIDPJ	R	R	HR	HR	Store media in a clean, low-humidity environment at a steady temperature, cool but not cold.	I.....L.....E....
ME.03	Copies at different locations	VIDPJ	R	R	HR	HR	Physically separate to cover natural disasters, accidental or malicious damage.	.....VL....B.....
ME.04	Backups / duplicates	VIDPJ	R	R	HR	HR	Backups are essential. Frequency of backup depends on rate of change. The number of generations to keep relates to the impact of data loss.	.....L.....
ME.05	Sample restores	VIDPJ	R	R	HR	HR	Sample restores should be performed at intervals to ensure that the backups are readable and retrievable.	.....L.....
ME.06	Multiple copies	VIDPJ	-	R	HR	HR	At least two backups should be kept, preferably in diverse formats.	.....VL.....
ME.07	Copy to latest media format	VIDPJ	-	R	HR	HR	Anticipate obsolescence and plan a smooth transition to new technologies.	.....L.....
ME.08	Media physically secured	VIDPJ	-	R	HR	HR	Access to, and removal of, media should be controlled by suitable procedures. Access permissions should be reviewed at intervals.	.....T.....U..H.....
ME.09	Resilient / redundant format	VIDPJ	-	-	R	HR	This may involve less use of compression, use of error detection and correction protocols, and (at the highest level) two or more redundant data servers.	IC.....L.....
ME.10	Long-lifetime format	VIDPJ	-	-	R	HR	The best formats should be adopted where available.	.....L.....E....

Continued on next page

Table 3.16: Mitigation methods: data media handling – electronic storage (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
ME.11	Easily translatable / convertible format	VIDPJ	-	-	R	HR	Adopt widely-used, well-documented, general-purpose formats in preference to specialist proprietary formats.	....O.....L.....
ME.12	Copy to cloud storage	VIDPJ	-	R	HR	HR	Must specify whether a private cloud or a public cloud shall be used. Cloud storage may not be suitable for highly confidential data.	.........L.....
ME.13	Copy to archiving organization	VIDPJ	-	R	HR	HR	Consider the required level of data integrity and confidentiality; also, the integrity and long-term viability of the archiving organization, and plans in case it ceases to function.	.........L.....

3.4.2.11 Recording the data safety risk mitigation

The DSMP can be used to document:

- the tables of mitigation measures (or methods and approaches) used for the system, context, and planned implementation under assessment;
- any specific mitigation measures identified for the system and their source / justification;
- planned compliance with the tables; and
- confirmation that the mitigation measures are sufficiently complete and consistent.

The overall safety justification for the given project, service, or operational context must then provide evidence of compliance with the plan.

This page is intentionally blank

4 Organizational Data Risk (Informative)

> *Data is becoming the new raw material of business.*
> ***Craig Mundie***

Note: this questionnaire only provides an initial ODR assessment. Further work is required to establish the safety data risks in detail such as determining a DSAL for relevant datasets.

ODR Assessment Form			
This form is used to determine the safety risk related to data for a particular organization and usage. *This form must be completed from the perspective of **one** of the organizations involved; typically this will be the organization using the data or the contractor supplying the system that handles the data. This form needs to be completed for each instance / application / scope / risk profile and should consider a defined boundary for the analysis, e.g. the scope of supply for the contractor or the limit of the data user's operational responsibility. It may be useful for both contracting parties to complete the form from their respective positions to check the data risk responsibilities and apportionment.* *It is anticipated that this form will be used during early phases of a procurement or supply and also for changes to existing supplies. It can also be used to assess existing legacy scenarios.* *Answer the questions as they apply in the context of the scope of supply. Mark the response with the "best" fit for the given scenario. Note that not all elements have to be satisfied. For each response also add a brief justification for that particular selection as opposed to any other choice.* *If the answer to a question is completely unknown at this stage, it is suggested that the middle value or higher is chosen and an explanation added to the justification.* *When all the relevant questions have been answered and justified, add the scores together to give a final total and record the value in the appropriate field. Use this total to determine the final ODR level based on the stated ranges.* *The ODR level determined may be used to support process tailoring and to determine the management regime required to mitigate the risk associated with the data.*			
Data Scenario / Context Name:			
Data Scenario / Context Description:			
Scope / Data Boundary and Perspective:			
Completed By:		Date Completed:	

ODR Assessment Form
Answer each question using the response that forms the best match for the particular scenario. Not all statements have to be satisfied and some judgement is required; it is expected that the majority of statements in the selected response can be satisfied with some interpretation. The use of multiple criteria in each question enables a smaller and manageable set of questions to be posed to provide a holistic view of the overall risk.

QUESTION 1 – SEVERITY AND PROXIMITY			
How severe could an accident be that is related to the data? Could it be caused directly by the data? *This question considers the safety consequence, the proximity and contribution of the data to the accident sequence.*			
1a	All currently foreseen usages of the data could not contribute to an accident. The data is not relied upon for safe operation. Negligible environmental impact.	1	❑
1b	A possible use of data could contribute to a minor accident, but only via lengthy and indirect routes. Could lead to minor injury or temporary discomfort for 1 or 2 people. Many other people / systems are involved in checking the data. Some aspects of safe operation rely very indirectly on the data. Minor environmental impact only via indirect routes.	2	❑
1c	A use of the data could lead to a significant accident resulting in minor injuries affecting several people or one serious injury. Several other people / systems are involved in checking the data. There is a dependency on the data for safe operation. Environmental impact is possible.	4	❑
1d	A likely use of the data could directly lead to a serious accident resulting in serious injuries affecting a number of people, or a single death. One human or independent check is involved for all data. There is major dependency on the data for safe operation. Major environmental impact is possible.	8	❑
1e	An intended use of the data could lead to an accident resulting in death for several people. The accident could be caused by the data with little chance of anything else detecting and mitigating the data issues. The accident could affect the general public or cause catastrophic environmental impact.	16	❑
Justification:			

QUESTION 2 – ORGANIZATIONAL AND SOCIETAL IMPACT			
What would be the impact on the organization, client or public if an accident occurred related to the data? *This question considers the tolerability within this industry sector and the general public. How much would it affect the organization or society? Would a claim be likely? Would it generate press interest? Would a formal investigation ensue?*			
2a	Little interest, accidents happen all the time in this sector; very high societal tolerability. Negligible chance of claims or investigations. No adverse publicity likely.	1	❑
2b	Some concern from the client, but accidents happen occasionally; high societal tolerability. Small chance of claim against the organization. Local or specialist press interest. Minor investigation or audit.	2	❑
2c	Public would be concerned, accidents are rare in this sector; some societal tolerability. Significant chance of claim against the organization. Regional press interest. Client inquiry or investigation likely.	4	❑

QUESTION 2 – ORGANIZATIONAL AND SOCIETAL IMPACT			
2d	Public would be alarmed and consider the accident a result of poor practice; little societal tolerability. Claims very likely. National press or media coverage a possibility. Legal or independent inquiry may follow.	8	❑
2e	Public would be outraged and consider such an accident unacceptable; almost no societal tolerability. Multiple claims / fines from regulators or courts are likely. International press or media coverage. Official and / or public enquiry possible.	16	❑
Justification:			

QUESTION 3 – RESPONSIBILITY			
How much responsibility does this organization have for data safety? *This question considers how much legal and other responsibility and ownership the organization has for data safety aspects within this scenario. What liabilities for consequential losses / 3rd party claims does the organization have via the contract or other means? What is the scale of the organization's contribution to the overall scope?*			
3a	The organization is not responsible for any data safety aspects. No liabilities for accident claims related to the data lie with the organization. Client or other party has accepted full data safety responsibility. The organization is fully covered and indemnified by the client or a 3rd party.	1	❑
3b	The organization is a small part of a large consortium. It has minimal liability for data safety via the contract. It is partly covered by explicit client or 3rd party protections. All safety data is managed by subcontractors, the organization only reviews and monitors.	2	❑
3c	The organization is a significant part of the consortium team. It has some share of the data safety responsibility. Specific data safety liabilities to the client via the contract are mentioned. There are no indemnities in the organization's favour. All key safety data obligations are explicitly flowed down to subcontractors.	4	❑
3d	The organization is prime for a small programme or has the bulk of the data safety responsibility within a team. Specific accident-related liabilities in the contract are significant. The organization provides some indemnities to others via the contract. Some significant data safety obligations are not flowed down to subcontractors.	7	❑
3e	The organization is priming a major programme or has total data safety responsibility. Specific accident-related liabilities in the contract are large (or unlimited). The organization provides explicit indemnities in favour of the client / 3rd parties for accidents. Safety data obligations have not been discussed or are not flowed down to subcontractors.	12	❑
Justification:			

QUESTION 4 – LEGAL AND REGULATORY FRAMEWORK			
What legal and regulatory environment will this work be subject to? *This question considers the legal and regulatory obligations that this work will have to conform to. How well is the legal framework defined and understood? Is there an established standards culture? Is there a regulator and certification process?*			
4a	Well-understood and tested legal framework, one jurisdiction. Highly regulated sector with one overseeing body. Well-established industry guidelines and standards for safety data. Formal certification processes.	1	❑
4b	Understood and established legal framework, a few related jurisdictions. Regulated sector, more than one overseeing body. Industry guidelines and standards for safety data. Some formal certification processes.	2	❑
4c	Some understanding of legal position, several jurisdictions. Partially regulated sector, several possible overseeing bodies. Some industry guidelines and standards that refer to data. Informal certification processes.	4	❑
4d	Complex, poorly defined legal position, multiple different jurisdictions. Largely unregulated sector with no established overseeing body. Some industry guidelines and standards that mention data. Some informal certification processes.	6	❑
4e	Very complex, untested and unclear legal position, many diverse jurisdictions. Unregulated sector with no overseeing body. No industry guidelines or standards for data. No certification processes.	10	❑
Justification:			

QUESTION 5 – ORGANIZATIONAL MATURITY			
How mature is this organiZation regarding data safety? *This question considers the maturity of the organization in relation to awareness and management of the risks associated with safety data. Are staff trained, managed and resourced to enable proper handling of data safety risk?*			
5a	Explicit recognition of data as a source of safety risk. Formal and established processes and procedures in place for the identification and control of safety data. Staff trained and fully aware of safety data risks. Senior management fully aware and supportive of data safety management activities. Management of safety data risks fully supported and funded.	1	❑
5b	Awareness of data as a source of safety risk. Informal processes and procedures in place for the identification and control of safety data. Staff awareness of safety data risks. Senior management awareness of data safety management issues. Good support and funding for management of safety data risks.	2	❑
5c	Some awareness of data as a source of safety risk. Some ad-hoc processes and procedures in place for the identification and control of safety data. Some staff awareness of safety data risks. Some senior management awareness of data safety management issues. Some support or partial funding for management of safety data risks.	4	❑

QUESTION 5 – ORGANIZATIONAL MATURITY			
5d	Little awareness of data as a source of safety risk. Minimal processes or procedures in place for the identification and control of safety data. Little staff awareness of safety data risks. Little senior management awareness of data safety management issues. Little support or minimal funding for management of safety data risks.	7	❑
5e	No recognition of data as a source of safety risk. No processes or procedures in place for the identification or control of safety data. No staff training or awareness of safety data risks. Senior management not aware or in denial of safety data risks. No support or funding for management of safety data risks.	10	❑
Justification:			

QUESTION 6 – OWNERSHIP AND USAGE			
How widely is the data used and who by? *This question considers how much usa and what type of users there are likely to be of the data. How complex is the data supply chain? In what geographies is it used? How many owners and interfaces are there?*			
6a	Minimal or infrequent use. One data owner, a specialist highly trained user group. Single organization or recipient use only.	1	❑
6b	A number of operational data users. Simple linear supply chain. More than one data owners. Specialist user or limited public access. Small-scale operation. No general web access. Few user organizations or recipients.	2	❑
6c	Regional use. Some public or mainstream use. A few supply chains. A few data owners. Some web access. Several user organizations or recipients.	4	❑
6d	National use. Public or mainstream use. Several supply chains. Several data owners. Web access. Some or varied user organizations or recipients.	7	❑
6e	International use. Extensive public or mainstream use. Extensive web access. Many complex supply chains. Many and diverse data owners. Many and diverse user organizations or recipients.	12	❑
Justification:			

QUESTION 7 – SIZE, COMPLEXITY AND NOVELTY			
What is the scale, sophistication and complexity of the data and its manipulation? *This question considers the nature of the data, its lifecycle and how easy it is to detect errors in the data.*			
7a	Simple data structures. Mature and established data storage and manipulation techniques and technologies. One or two interfaces. No timeliness aspects. No transformations. Data is easily verifiable. Data is easily traceable to original source.	1	❑
7b	Varied data structures. Mainstream data storage and manipulation techniques and technologies. Several interfaces. Few timeliness aspects. Few data transformations. Data is verifiable. Data is traceable to original source.	2	❑
7c	Complex with some unstructured data. Current data storage and manipulation techniques and technologies. Multiple interfaces. Some timeliness aspects. Some data transformations. Data is difficult to verify. Data is difficult to trace back to original source.	4	❑
7d	Complex, varied or partially unstructured data. Novel storage and manipulation techniques and technologies. Multiple complex interfaces. Time critical. Complex data transformations. Data is very difficult to verify. Data is very difficult to trace back to original source.	7	❑
7e	Highly complex, varied or unstructured data. Highly novel storage and manipulation techniques and technologies. Many and complex, ill-defined or dynamic interfaces. Highly time critical. Many and complex data transformations. Data is infeasible to verify. Data is impossible to trace back to original source.	10	❑
Justification:			

QUESTION 8 – BOUNDARIES AND INTERFACES			
How well defined and understood are the boundaries and interfaces for this data scenario? *This question considers the number, complexity and definition status of the boundaries and interfaces where data is exchanged. How well understood are the boundaries and interfaces? Are standard formats and protocols used? Is data exchange time critical? Are all assumptions and ambiguities relating to the data exchange resolved?*			
8a	One well-understood boundary and few, well-defined interfaces. Standard interface formats and protocols. No timeliness aspects to data exchange. No remaining ambiguities, TBCs or TBDs. No assumptions.	1	❑
8b	A few, understood boundaries and several defined interfaces. Mainly standard interface formats and protocols. Few timeliness aspects to data exchange. Few areas of ambiguity, few TBCs and TBDs. Few assumptions.	2	❑
8c	Several, established boundaries, some defined, some undefined and some ambiguous interfaces. Mixture of standard and non-standard interface formats and protocols. Some timely data exchanges. Some areas of ambiguity, some TBCs and TBDs. Some assumptions.	4	❑
8d	Many, poorly understood boundaries, many undefined or ambiguous interfaces. Mostly non-standard interface formats and protocols. Time sensitive data exchange. Many areas of ambiguity, many TBCs and TBDs. Many assumptions.	6	❑

QUESTION 8 – BOUNDARIES AND INTERFACES			
8e	A large number of unclear boundaries; a large number of unknown and undefined interfaces. Completely non-standard, complex interface formats and protocols. Real-time data exchange. Large areas of ambiguity, a large number of TBCs and TBDs. A large number of assumptions.	10	❑
Justification:			

ODR LEVEL	
Record the total score and use it to determine the ODR level based on the ranges given below. If the first three questions' scores sum up to 6 or less then disregard the scores for the remaining questions.	
Score 14 or less	ODR0
Score 15 to 21	ODR1
Score 22 to 37	ODR2
Score 38 to 47	ODR3
Score 48 and above	ODR4
Total Score for this scenario /context:	
ODR Level for this scenario / context:	

5 Data Safety Culture Questionnaire (Informative)

> *Data is the fabric of the modern world: just like we walk down pavements, so we trace routes through data and build knowledge and products out of it.*
> ***Ben Goldacre***

This form helps an organization appreciate the data safety culture. It can be applied at various levels, including at the project level and at the organizational level.

Data Safety Culture Questionnaire Form			
This form is used to assess the safety culture related to data for a particular programme. **You play a key role in protecting the organization from data safety risks and your views are important. This self-assessment survey is designed to assess our current level of data safety culture within the programme. The output can help us to improve our safety position.** Please tick the box which reflects your view and answer as honestly as possible. Space is provided for explanatory comments. Your response will only be of value if it reflects what you actually believe is the case, rather than what you believe should happen. If you would like to remain anonymous please print and send this form by post. The survey should take no longer than 10 minutes. It is anticipated that this form will be used on a regular basis (e.g. annually).			
Programme Name:			
Data Scenario / Context Description:			
Completed By:		Date Completed:	
Answer each question as you see it - there is no right answer!			

QUESTION 1 – MY VIEW OF OUR SUPPLY							
		Don't Know	Strongly Disagree	Disagree	Maybe	Agree	Strongly Agree
1a	I see data as an important factor in the safety of my programme.	❑	❑	❑	❑	❑	❑
1b	I am familiar with the safety aspects of our data.	❑	❑	❑	❑	❑	❑
1c	I understand how data in our solution can contribute to an accident.	❑	❑	❑	❑	❑	❑
1d	I think we could be blamed if there were an accident due to our data.	❑	❑	❑	❑	❑	❑
Comments:							

QUESTION 2 – WHAT WE'RE DOING							
		Don't Know	Strongly Disagree	Disagree	Maybe	Agree	Strongly Agree
2a	I think that the programme is aware of data safety risks.	❑	❑	❑	❑	❑	❑
2b	I believe we need to implement measures to manage data safety risks.	❑	❑	❑	❑	❑	❑
2c	I think that the programme meets its obligations (e.g. has a Data Management Plan in place and a role with specific responsibilities in this area).	❑	❑	❑	❑	❑	❑
Comments:							

QUESTION 3 – MY ROLE							
		Don't Know	Strongly Disagree	Disagree	Maybe	Agree	Strongly Agree
3a	I know my role relates to the management of data and associated safety risks.	❑	❑	❑	❑	❑	❑
3b	If I had a safety concern about our data I would report it.	❑	❑	❑	❑	❑	❑
3c	I know who the data safety representative is on my programme.	❑	❑	❑	❑	❑	❑
3d	I have received adequate training regarding data safety for my role.	❑	❑	❑	❑	❑	❑
3e	I feel supported in dealing with data safety risks.	❑	❑	❑	❑	❑	❑
3f	I have adequate time to address any data safety issues.	❑	❑	❑	❑	❑	❑
Comments:							

6 Supplier Data Maturity (Informative)

> *The most valuable thing you can have as a leader is clear data.*
> ***Ruth Porat***

This questionnaire may be used for two purposes:

1. To support a procurement process - distributed by an organization looking for a company that can handle safety-critical development, because the system they require to be developed is known to have safety-critical requirements
2. Internal audits - used internally by a company developing systems with safety-related data which needs to assure itself of its capability to fulfil customer needs.

Organization

1. For each software development involving data, is there a designated data safety manager?
2. If so, does the data safety manager report directly to the project manager?
3. Are the management reporting channels for data assurance and software development separate?
4. Is data subject to a formal configuration control process?
5. Is data engineering represented on the system design team?
6. Is data engineering process improvement part of the company quality systems?

Resources, Personnel and Training

1. Are personnel specified as responsible for data safety as a separate role from software and system design and development?
2. Is there a required training programme for data specialists?
3. Is training on data safety issues part of the training for managers or management teams?
4. Is there a formal training programme for data safety design and review leaders

Data Issues Growth Management

1. Is a mechanism employed for maintaining awareness of the state of the art in data safety technology?
2. Is a mechanism employed for comparing the company approach to data safety with external processes for data safety practised elsewhere in the industry?
3. Is a mechanism used for introducing new technologies and processes into system development?
4. Is a mechanism in place for identifying and replacing obsolescent processes related to data safety?

Documented Standards and Procedures

1. Describe any formal procedures adopted at each periodic management review to assess the status of data related to the system.
2. Describe the methods used for ensuring that the data development team understands each data requirement.
3. Is a data risk assessment method used for assessing the use of existing data in new applications?
4. Are data test cases developed formally with a company standard?
5. Is there a document which describes how the customer is to be consulted over data issues?
6. Is particular care taken to capture requirements, design, review and test data for user interfaces?
7. Is there a data risk monitoring and tracking to closure procedure practised?

Process Metrics

1. Are statistics of failures due to data errors during development kept to feedback and learn from in future development?
2. Are data issue action items tracked to closure and reports maintained of causes?
3. Is configuration data separately developed from everyday operational data?
4. Is data test coverage measured and recorded?
5. Are all states, from which configuration data will be required, tested, (including emergency reboot), and results recorded?
6. Are analyses of errors due to data conducted to determine their process related causes?
7. Are the process causes reviewed and changes to processes implemented where appropriate?

Process Control

1. Is regression testing routinely performed when errors are discovered?
2. Is the adequacy of regression testing subject to an assurance process to ensure new errors are not introduced?
3. Is a mechanism used for identifying and resolving system engineering issues that affect data?
4. Is a mechanism used for ensuring traceability between the data requirements and the top-level design?
5. Is the importance of data in the system engineering process reviewed to maintain processes at an adequate level to cope with the expanding role of data in the Internet of Things?

7 Data Categories – Detail (Informative)

> *It's difficult to imagine the power that you're going to have when so many different sorts of data are available*
> ***Tim Berners-Lee***

Table 7.1 provides additional information, in the form of explanations and lists of typical containers, for the identified Data Categories.

Table 7.1: Categories of safety-related data: detailed definitions

No.	Category	Description	Explanation	Typical containers
			Context	
1	Predictive	Data used to model or predict behaviours and performance	Data for studies, models, prototypes, initial risk assessments, etc. This is the data produced during the initial concept phase which subsequently flows into further development phases.	Prototype results, evaluations, analyses
2	Scope, Assumption and Context	Data used to frame the development, operations or provide context	Restrictions, risk criteria, usage scenarios, etc. explaining how the system will be used and any limitations of use.	Concepts of operation, safety case report part 1
3	Requirements	Data used to specify what the system has to do	Data encompassing requirements, specifications, internal interface or control definitions, data formats, etc.	Formal specifications, interface control documents, user requirements documents, safety case report part 1
4	Interface	Data used to enable interfaces between this system and other systems: for operations, initialisation or export from the system	Data that exists to enable exchange between this system and other external systems. Covers start-of-life operations (data import or migration), end-of-life operations and ongoing operational exchange of data between systems.	Protocols, schemas, interface control documents, transition plans, extract-transform-load tool specifications, cleansing and filtering rules
5	Reference or Lookup	Data used across multiple systems with generic usage	Data comprising generic reference information sets used by multiple systems (i.e., not produced solely for this system). Typically updated infrequently, and not specific to this system.	Dictionaries, materials information, sector data reference sets, encyclopedias

Continued on next page

Table 7.1: Categories of safety-related data: detailed definitions (continued)

No.	Category	Description	Explanation	Typical containers
			Implementation	
6	Design and Development	Data produced during development and implementation	Data encompassing the design and development process artefacts: everything from design models and schemas to document review records. It also includes test documents (specification and results) but not the test data itself.	Design documents, review records, hardware, software, test scripts, code inspection reports, safety case report part 2
7	Software	Data that is compiled (or interpreted) and executed to achieve the desired system behaviour	From some perspectives it is helpful to consider software (e.g., source code) as another category of data.	Text files, configuration management systems
8	Verification	Data used to test and analyse the system	Data comprising the test values and test datasets used to verify the system. It might 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 datasets, stub data, emulator and simulator files
			Configuration	
9	ML	Data used to train the system to enable it to learn from the characteristics of the data	Data used to train, set up or adapt the system for a particular purpose or configuration. Might be subsets of real data or synthetically produced. Might have to include or exclude corner cases.	Images for pattern recognition analysis
10	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
11	Behavioural	Data used to change the functionality of the system	Data to enable / disable or configure functions or behaviour of the system.	extensible markup language (XML) configuration files, comma separated variable (CSV) data, schemas
12	Adaptation	Data used to configure to a particular site	Data used to tailor or calibrate a system to a particular physical site or environment, incorporating physical or environmental conditions.	Configuration files

Continued on next page

Table 7.1: Categories of safety-related data: detailed definitions (continued)

No.	Category	Description	Explanation	Typical containers
			Capability	
13	Staffing	Data related to staff training, competency, certification and permits	Data which allows staff to perform a function within the wider context of the safety-related system. This might include training records, competency assessments, permits to work, etc.	Human resources records, training certificates, card systems
			The Built System	
14	Asset	Data about the installed or deployed system and its parts, including maintenance data	Data related to location, condition and maintenance requirements of the system under consideration. This might cover hardware, software and data.	Inventory, asset and maintenance database systems
15	Performance	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 and diagnostic data. This might be the results of various phases of introduction and might include trend analysis to look for long-term problems.	Field data, support calls, bug reports, non-compliance reports, defect reporting and corrective action system (DRACAS) data
16	Release	Data used to ensure safe operations per release instance	Explanation of particular features or limitations of a release or instance. Might include specific time-limited workarounds and caveats for a release.	Release notes, certificate of design (CoD), Transfer documents, safety case report part 2 or part 3
17	Instructional	Data used to warn, train or instruct users about the system	Data that explains to users the risks of the systems and gives any mitigations that might be required to be implemented by users, e.g., by process, procedure, workarounds, limitations of use.	Manuals, standard operating procedures (SOPs), online help, training courses, safety case report part 3
18	Evolution	Data about changes after deployment	Data that covers enhancements, formal changes, workarounds, and maintenance issues. It also covers data produced by configuration management activities, such as baselines or branch data.	Change requests, modification requests, issue and version data, configuration management system outputs
19	End of Life	Data about how to stop, remove, replace or dispose of the system	Data covering all activities related to taking the system out of service or mothballing / storage / dormant phases.	Transition, disposal and decommissioning plans

Continued on next page

Table 7.1: Categories of safety-related data: detailed definitions (continued)

No.	Category	Description	Explanation	Typical containers
20	Stored	Data stored by the system during operations	Data stored or used within the system which has end-user meaning. It might be displayed and used within the system or might be for transfer and distribution to other systems or downstream users. It is data that has some real domain meaning.	Might be stored internally within the system (e.g., in databases or text files), or transferred into or out of the system through interfaces (e.g., Ethernet)
21	Dynamic	Data manipulated and processed by the system during operations	Data processed, transformed or produced by the system which has end-user meaning. It might be displayed and used within the system or might be for transfer and distribution to other systems or downstream users. It is data that has some real domain meaning.	Might be manipulated within the system in data structures or transferred into or out of the system through interfaces
22	Twinning	Data used to create and maintain a digital counterpart of a physical object or process	The digital twin is an up-to-date and accurate model when supplied with accurate and up-to-date data. This might be a model of a physical object's properties and states, including position, status and motion or of a process flow. A digital twin also can be used for monitoring, diagnostics and prognostics to optimize asset performance and use. Intelligent maintenance system platforms can use digital twins to find the root cause of problems.	Tooling / modelling environments and bespoke software implementations
Compliance and Liability				
23	Standards and Regulatory	Data that governs the approaches, processes and procedures used to develop safety systems	Data predominantly in the form of documents that describe and dictate the activities, processes, competencies etc. to be used for a particular development in a particular sector.	Standards documents, guidelines, legal directives and laws
24	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 (e.g., regulators, Health and Safety Executive, independent safety auditors) for their review.	safety case report, certification case, regulatory documents, COTS justification file, design justification file

Continued on next page

Table 7.1: Categories of safety-related data: detailed definitions (continued)

No.	Category	Description	Explanation	Typical containers
25	Investigation	Data used to support accident or incident investigations (i.e., potential evidence)	Data collected or produced during an incident or accident investigation which might be used in investigation reports, lessons learned or prosecutions. This can be process data, trace data, site data (e.g., photographs of crash site) or might be derived (accident simulations, analyses, etc.).	Incident / accident investigation reports and supporting documents
Meta-Property				
+1	Trustworthiness	(Meta) data which tells us how much the system can be trusted	Data which provides assurance or confidence about the other data within or about the system under consideration. This might be some of the data mentioned in the other categories, but might be different.	Data audits, data quality index measures, sign-off sheets, traceability records, model database

This page is intentionally blank

8 HAZOP Guidewords – Detail (Informative)

> *I don't see the logic of rejecting data just because they seem incredible.*
> ***Fred Hoyle***

Table 8.1 expands upon the HAZOP guidewords presented in Table 3.4 by adding data considerations which may be helpful in determining how to apply the guidewords to a specific instance of a property.

Table 8.1: HAZOP guidewords: detailed definitions

Property	Description	HAZOP Data Guidewords	HAZOP Data Considerations
Integrity	The data is correct, true and unaltered	Loss, partial loss, incorrect, multiple	Correctness, truth, original, trustworthy, coherency, stability, perfect, unquestionable, faithful, certain, ordered, unadulterated, unmodified, unchanged, clean, uncontaminated, untainted, proper, flawless, organized, exact, undistorted, faultless, guided, connected, linked, traced, unbiased.
Completeness	The data has nothing missing or lost	Loss, partial loss, incorrect, multiple, insufficient	Whole, complete, entire, finished, done, stable, qualified, certified.
Consistency	The data adheres to a common world view, e.g., units	Loss, partial loss, incorrect, multiple, too early, too late, loss of sequence	Coherent, compatible, congruent, congruous, harmonious, deconflicted, consistent, appropriate, suitable, sound, cleansed.
Continuity	The data is continuous and regular without gaps or breaks	Loss, partial loss, incorrect, late, loss of sequence	Smooth, continuous, regular, gapless, whole, complete, entire, unfragmented.
Format	The data is represented in a way which is readable by those that need to use it	Loss, partial loss, incorrect, multiple	Conformant, suitable, valid, configured, well-formed, set-up, composed, well structured, arranged, compliant, organized, exact, unaliased, migrated, transformed.
Accuracy	The data has sufficient detail for its intended use	Loss, partial loss, incorrect, multiple, insufficient	Accurate, true, correct, undistorted, unbiased, faultless.
Resolution	The smallest difference between two adjacent values that can be represented in a data storage, display or transfer system	Loss, partial loss, incorrect, multiple, insufficient	Exact, untruncated, retention of detail, clarity, determination, distinguishable, clear, within range, distinct, separated, discernible, discriminatable, unconfused, divisible, unaliased, granularity, precision.

Continued on next page

Table 8.1: HAZOP guidewords: detailed definitions (continued)

Property	Description	HAZOP Data Guidewords	HAZOP Data Considerations
Traceability	The data can be linked back to its source or derivation	Loss, partial loss, incorrect, multiple, too early, too late, loss of sequence	Traceable, verifiable, indexed, linked, connected, justified, proven, evidenced, substantiated, continuous, unfragmented, complete, networked.
Timeliness	The data is as up to date as required	Loss, partial loss	Timely, early, ready, expected, unique, appropriate, opportune, ordered, organized, anticipated, seasonable, converging, settling, on-time, latency, lag, lead time, time slots, real-time, determinism, predictable.
Verifiability	The data can be checked and its properties demonstrated to be correct	Loss, partial loss, incorrect, multiple, too early, too late, loss of sequence	Verifiable, provable, checkable, supportable, demonstrable, sustainable, certifiable, defensible, excusable, justifiable, undisputable, irrefutable, validated.
Availability	The data is accessible and usable when an authorized entity demands access	Loss, partial loss, multiple, too early, too late	Ready, available, obtainable, reachable, accessible, serviceable, operable, functional, usable, capable, released, issued, disseminated, distributed.
Fidelity / representation	How well the data maps to the real-world entity it is trying to model	Loss, partial loss, incorrect, multiple, too early, too late	Representative, accurate, faithful, trustworthy, characteristic, normal, standard, real, expected, natural, typical, regular, fit for purpose, validated, separable, associated, correct units / dimensions, stable, unbiased.
Priority	The data is presented / transmitted / made available in the order required	Loss, partial loss, incorrect, multiple, too early, too late	Current, ordered, included, precedence, hierarchy, pre-eminence, retained, ahead, readiness.
Sequencing	The data is preserved in the order required	Loss, partial loss, incorrect, multiple	Ordered, contiguous, unique, ordered, clear, continuous, successive, uninterrupted, sequential.
Intended destination / usage	The data is only sent to those that should have it	Loss, partial loss, incorrect, multiple, too early, too late, loss of sequence	Directed, delivered, copied, sent, transmitted, correct recipient, unintercepted, unseen, integral, received, acknowledged, forwarded, filtered.
Accessibility	The data is visible only to those that should see it	Loss, partial loss, incorrect, multiple, too early, too late	Secure, open, visible, reachable, seen, usable, accessible, obtainable, uncompromised, secure, encrypted, preserved.
Suppression	The data is intended never to be used again	Loss, partial loss, incorrect, too early, too late, too much, too little	Hidden, encrypted, private, confidential, erased, unlinked, unavailable, unaccessible, redacted.

Continued on next page

Table 8.1: HAZOP guidewords: detailed definitions (continued)

Property	Description	HAZOP Data Guidewords	HAZOP Data Considerations
History	The data has an audit trail of changes	Loss, partial loss, incorrect, multiple	Justifiable, traceable, provable, supportable, demonstrable, sustainable, certifiable, defensible, excusable, justifiable, undisputable, irrefutable.
Lifetime	When does the safety-related data expire	Loss, too early, too late, incorrect, multiple, loss of sequence	Expiry date, age, validity, currency, applicability, durability, duration, lifespan, stretch, tenure, half-life, longevity, span, in-date, best-before, window, established.
Disposability / deletability	The data can be permanently removed when required	Loss, partial loss, incorrect, too early, too late	Unavailable, unaccessible, redacted, hidden, filtered, lost, deleted, destroyed, backup, archive, locked, secured, unlinked.
Goldilocks	There is exactly the right amount of data - not too much, not too little and arrives at the right time	Loss, partial loss, incorrect, too early, too late, too much, too little (insufficient), spurious	Manageable data rate, manageable error rate, expected, within acceptable limits, without system overload, size, volume, sufficient
Analysability	The data is of a suitable size, type and representation (including any metadata) to enable it be usefully analysed	Loss, partial loss, incorrect, too much, too little, loss of sequence, loss of tooling	Readability, representation, methods & techniques, statistical, parameters, characteristics, anonymisation, format, media, volume, tools
Explainability	The data can be meaningfully explained,by a suitable mechanism, to those who need to understand it	Loss, partial loss, incorrect, too much, too little, loss of sequence, loss of skills, loss of tooling	Justifiable, traceable, metadata, interpretation, evidenced, supported, argued, logical, inference, derived, tools

This page is intentionally blank

9 Data Safety Management Plan (Informative)

> *Things get done only if the data we gather can inform and inspire those in a position to make a difference.*
> ***Mike Schmoker***

This section gives a suggested DSMP table of contents. It is expected that this will be needed only for aspects not already covered in a safety management plan (SMP), or similar. It can be merged with an SMP, if appropriate. However it may be useful to consider the distinct data perspective by using a DSMP as well as an SMP. Regardless, a close connection should be maintained between the SMP and the DSMP.

DSMP suggested contents:

1. Introduction:

   Scope and context: sets the scene, describes the project, scenario, concept of operations, etc.;

   Boundaries and interfaces: Describes the main interfaces and exchanges, with a scope boundary diagram.;

   Derived requirements: System-level requirements which impact the data safety process;

   Owners: Who owns the data under consideration as it progresses through the system?;

   Producers / consumers: Who are the producers and consumers of the data the system inputs and outputs?;

   Assumptions ;

   References ; and

   Abbreviations and acronyms .

2. Analysis of Assigned DSAL and ODR Level (Implications of the data analyses.):

   Safety integrity level (SIL), etc., implications: What impact does the DSAL have on the required SIL, or similar measure, and vice versa?;

   Development implications: Are there any special development considerations? Derived from the SIL if there is one, otherwise what is deemed necessary for this system.;

   Verification implications: derived from the SIL if there is one, otherwise what is deemed necessary for this system.;

   Assurance implications: derived from the SIL if there is one, otherwise what is deemed necessary for this system.; and

   Process / procedure implications: derived from the SIL if there is one, otherwise what is deemed necessary for this system.

3. Categories of safety data in scope (a list of all the categories to be considered in the system context.).

4. Data requirements analysis:

   Lifecycles: what data lifecycles are to be used?;

   Specific Targets: are there any qualitative or quantitative targets for the data?; and

   Security Considerations: how will security be managed in this context? Are there any security / safety conflicts? Are there any security-related causes of data hazards?

5. Management approach (how will the organization manage the data safety risks?):

- Organization;
- Responsibilities;
- Authorisations; and
- Approvals and signoffs.

6. Justification Approach (how will the safe use of the data be justified, e.g., as part of the safety case report?).
7. Analyses / verifications to be performed (what analyses or checks are to be performed on the data?).
8. Documents to be produced (The list of documents to be produced related to data aspects.).
9. Appendix: DSAL guidelines response (tailored version of the tables from this document. What is considered applicable / useful and what is not?).

Consider putting accidents in Vol 3

This page is intentionally blank

10 Incidents and Accidents (Discursive)

> *Hiding within those mounds of data is knowledge that could change the life of a patient, or change the world.*
> ***Atul Butte***

10.1 General

The following 'war stories' describe incidents and accidents in which data is considered to have been a contributory factor. A data perspective has been taken to demonstrate the need for data to be given equal footing alongside software, hardware and human factors. The items described here have been arbitrarily selected; the collection is not intended to be exhaustive.

Note: The analysis presented here has no legal standing whatsoever. The purpose of this section is not to discredit, contradict or undermine any existing accident or incident analysis; the aim is simply to view these incidents and accidents from a data perspective. Where possible accident reports have been referenced with the role of data highlighted. All references have been taken at face value and not independently verified.

Two tables are presented to assist the reader in finding incidents and accidents of interest:

- Table 10.1 lists the issues and accidents according to the domain where they occurred.
- Table 10.2 lists them in reverse chronological order of incorporation – those entries that have been added to this document most recently appear at the top of the table. This table can also be used to find examples relevant to specific data properties.

Table 10.1: Incidents and accidents (by domain)

Ref.	Title	Ref.	Title
Accountancy			
10.2	Post Office Horizon System		
Air			
10.9	Boeing 737 MAX 8 crashes	10.30	LOT Flight 282
10.17	Turkish Airlines A330	10.32	Comair Flight 5191
10.19	Qantas Boeing 737 Take-Off	10.33	Überlingen Mid-Air Collision
10.20	Qantas Boeing 737 Loading	10.37	Crash into Nimitz Hill, Guam
10.23	Boeing 737-33A at Chambery Airport, France		
Air (Defence)			
10.15	A400M Torque Calibration Parameters	10.24	Loss of Hermes 450
10.22	Loss of MQ-9 reaper		
Air (Other)			
10.12	Loss of Irish rescue Helicopter		
Defence			
10.3	Battle of Agincourt	10.34	Fort Drum Artillery Incident
Internet			
10.6	Java log4j Library Vulnerability		
Maritime			
10.21	Grounding of *Navigator Scorpio*	10.29	Grounding of *The Pride of Canterbury*
10.26	Grounding of *Sichem Osprey*	10.31	Annabella container ship – Baltic Sea
Maritime (Defence)			
10.16	Royal Navy Submarine		
Medical			
10.5	What3Words systemic ambiguities	10.18	Dallas Hospital Ebola Incident
10.7	Immensa False Negative Covid-19 polymerase chain reaction (PCR) Tests	10.25	Advocate Lutheran Hospital
10.8	Covid-19 Test Results Silently Deleted by Excel	10.28	Cedars-Sinai Medical Center Scanner
Oil & Gas			
10.39	Lake Peigneur Drilling Accident		
Policing			
10.5	What3Words systemic ambiguities	10.35	Early release from Washington State Prison
10.14	Interception of Communications		
Rail			
10.11	Cambrian Line Data Loss	10.38	San Bernardino derailment and pipeline rupture
10.27	Near Collision of Trains, Cootamundra		
Space			
10.4	Gemini V	10.13	Loss of Schiaparelli Mars Lander

Continued on next page

Table 10.1: Incidents and Accidents (continued)

Ref.	Title	Ref.	Title
10.10	Loss of Soyuz-2.1b rocket carrying Meteor-M 2-1 weather satellite	10.36	Mars Climate Orbiter

Table 10.2 provides a summary of the items considered in this appendix. More information on each item is then presented in the following paragraphs.

Table 10.2: Incidents and accidents

Ref.	Title	Summary	Domain	Year	Data Property
10.2	Post Office Horizon System	Non-atomic transactions and other errors in accounting software lead to false prosecutions, lost livelihoods, and suicides.	Accountancy	1999	Integrity, completeness, traceability, verifiability, history
10.3	Battle of Agincourt	French misconceptions about English longbow led to devastating losses	Defence	1415	Integrity, completeness, accuracy, fidelity / representation, timeliness
10.4	Gemini V	Erroneous " schoolchild knowledge" that Earth rotates 360° per day sends spacecraft off course	Space	1965	Accuracy, traceability, fidelity / representation
10.5	What3Words systemic ambiguities	What3Words use of homophones and plurals is alleged to have created difficulties in providing emergency services to persons in distress.	Policing and Medical	2021	Integrity, consistency, Aliasing
10.6	log4j Java library vulnerability	Critical zero-day vulnerability affecting Apache Log4j2 java library	Internet	2021	Integrity
10.7	Immensa False Negative Covid-19 PCR Tests	43,000 people with Covid-19 mistakenly given negative PCR results	Medical	2021	Integrity, accuracy, traceability, verifiability
10.8	Covid-19 test results silently deleted by Excel	Importing Covid-19 test results into an Excel file truncated the data after 65536 records	Medical	2020	Integrity, completeness, timeliness, availability, fidelity / representation

Continued on next page

Table 10.2: Incidents and accidents (continued)

Ref.	Title	Summary	Domain	Year	Data Property
10.9	Boeing 737 MAX 8 crashes	On two occasions a single faulty angle of attack (AoA) sensor repeatedly commanded the nose down, leading to the aircraft flying into the sea / ground	Air	2018 & 2019	Integrity, completeness, accuracy, availability, verifiability, fidelity / representation
10.10	Loss of Soyuz-2.1b rocket carrying Meteor-M 2-1 weather satellite	The launch vehicle and satellite were lost because programmers gave coordinates for the wrong cosmodrome	Space	2017	Fidelity / representation, integrity, verifiability
10.11	Cambrian Line Data Loss	Speed restriction data failed to be passed to trains, placing pedestrians on level crossings at risk.	Rail	2017	availability, history
10.12	Loss of Irish rescue Helicopter	At the time of writing, the investigation is continuing; possible controlled flight into terrain; possible issues with terrain / obstacle databases. Four fatalities.	Air	2017	Fidelity / representation, completeness, accuracy
10.13	Loss of Schiaparelli Mars Lander	High rates led to the saturation of the inertial measurement unit (IMU); the lander prematurely believed it was on the ground and released its parachute; the lander was lost. The high rates should have been expected, but were not due to modelling deficiencies.	Space	2016	Fidelity / representation, integrity
10.14	Interception of Communications	Incorrect data was disclosed during an investigation into indecent images. A welfare check was delayed on a child believed to be in crisis.	Policing	2015	Integrity
10.15	A400M Torque Calibration Parameters	A software update apparently wiped the engine torque control parameters. Aircraft crash; four fatalities.	Air (Defence)	2015	Completeness
10.16	Royal Navy Submarine, Trawler *Karen*	A royal Navy submarine snagged the fishing gear of the trawler *Karen*. The trawler was dragged backwards at about seven knots and suffered structural damage.	Maritime	2015	Resolution, integrity
10.17	Turkish Airlines A330	Inaccurate navigation data, relating to runway location, led to touchdown with left main gear off the paved surface. Aircraft written off.	Air	2015	Accuracy, timeliness, verifiability

Continued on next page

Table 10.2: Incidents and accidents (continued)

Ref.	Title	Summary	Domain	Year	Data Property
10.18	Dallas Hospital Ebola Incident	A man suffering from Ebola was mistakenly sent home from a Dallas hospital. He later returned to hospital, was diagnosed but died; two nurses contracted Ebola but survived.	Medical	2014	Completeness, format
10.19	Qantas Boeing 737 Take-Off	Two independent and inadvertent data entry errors meant weight used when calculating take-off performance was 10 tonnes less than actual weight. Tail strike.	Air	2014	Integrity, verifiability
10.20	Qantas Boeing 737 Loading	Default settings meant that children were incorrectly recorded as adults, resulting in incorrect aircraft weight and balance. Take-off safety speed was exceeded by about 25 knots.	Air	2014	Completeness, fidelity / representation
10.21	Grounding of *Navigator Scorpio*	The *Navigator Scorpio* was sailing with out-of-date charts, the planned route was not checked and positional fixes were not taken as often as required. The vessel was grounded, but refloated on the rising tide, with no damage. After the event, false information was added to the navigation chart.	Maritime	2014	timeliness, verifiability
10.22	Loss of MQ-9 reaper	The ground control station (GCS) station was misconfigured following a change from MQ-1 to MQ-9 operations. The misconfiguration was not spotted. It caused any throttle position aft of full forward to command negative thrust. The aircraft decelerated below stall speed and impacted ground in unpopulated area.	Air (Defence)	2012	Consistency, verifiability
10.23	Boeing 737-33A at Chambery Airport, France	Misuse of Electronic Flight Bag results in tail strike	Air	2012	Timeliness, suppression, lifetime
10.24	Loss of Hermes 450	While attempting an automatic landing the unmanned air system (UAS) self-aborted. This abort was due to an incorrect set-up parameter that had been loaded by the crew. The crew elected to intervene rather than let the UAS self-recover. The air vehicle hit a new, unoccupied hangar; it was ultimately deemed “non-repairable”.	Air (Defence)	2011	Integrity

Continued on next page

Table 10.2: Incidents and accidents (continued)

Ref.	Title	Summary	Domain	Year	Data Property
10.25	Advocate Lutheran Hospital	An infant boy died after a series of medical errors: incorrect information was entered into an electronic intravenous order; automatic alerts had been turned off; and a bag was mislabelled.	Medical	2010	Integrity, verifiability.
10.26	Grounding of *Sichem Osprey*	Anti-collision radar thresholds were apparently set incorrectly; there were also sizeable discrepancies between positions plotted on a chart and those displayed on the radar. The vessel grounded at more than 16 knots; no pollution occurred.	Maritime	2010	Integrity, accuracy
10.27	Near Collision of Trains, Cootamundra	Following a signalling system design error, a passenger train had to unexpectedly apply its brakes; it stopped just 5 m short of a goods train.	Rail	2009	Integrity, completeness
10.28	Cedars-Sinai Medical Center Scanner	A software misconfiguration led to 206 patients receiving radiation doses approximately eight times higher than intended; the error persisted for 18 months.	Medical	2008	Verifiability
10.29	Grounding of *The Pride of Canterbury*	An unapproved electronic chart system (ECS) was apparently being used as the primary means of navigation for the passenger ferry *The Pride of Canterbury*. Due to user settings a charted wreck would not have been displayed on this system. The vessel grounded on the wreck, causing severe damage to her port propeller system.	Maritime	2008	Accuracy, completeness, intended destination / usage
10.30	LOT Flight 282	Incorrect data input to the Flight Management System, 'E' rather than 'W', meant loss of instruments. Aircraft had to return to London Heathrow airport (LHR).	Air	2008	Integrity, fidelity / representation
10.31	Annabella container ship – Baltic Sea	Software developed loading plan using incorrect container specifications	Maritime	2007	Integrity, verifiability, fidelity / representation
10.32	Comair Flight 5191	Inaccurate (out-of-date) aerodrome charts led to take-off being attempted from the wrong runway. Aircraft overran the runway; 49 fatalities.	Air	2006	Timeliness, completeness

Continued on next page

Table 10.2: Incidents and accidents (continued)

Ref.	Title	Summary	Domain	Year	Data Property
10.33	Überlingen Mid-Air Collision	Contradictory advice from traffic collision avoidance system (TCAS) and an air traffic controller led to a mid-air collision between two TCAS-equipped aircraft. 71 fatalities.	Air	2002	Consistency, availability, Timeliness
10.34	Fort Drum Artillery Incident	Movement of an artillery site led to errors in targeting. Artillery shells were fired more than a mile off target: 2 soldiers killed; 13 injured.	Defence	2002	Integrity, verifiability
10.35	Early release from Washington State Prison	A software update led to miscalculation of the time an inmate was due to serve in prison. Although the results of the calculation could easily be checked, the problem persisted for 13 years and over 2,000 offenders were released early.	Policing	2002	Integrity, verifiability
10.36	Mars Climate Orbiter	A mismatch in the units used by two software teams led to errors in the Flight Management System and, ultimately, the loss of a multi-million dollar space mission.	Space	1998	Consistency
10.37	Crash into Nimitz Hill, Guam	Controlled flight into terrain (CFIT); the ground-based minimum safe altitude warning (MSAW) designed to alert air traffic controllers had been inhibited. 228 fatalities; 26 serious injuries.	Air	1997	Continuity, fidelity / representation
10.38	San Bernardino derailment and pipeline rupture	Criticality of train weight not recognized, resulting in multiple fatalities	Rail	1989	Integrity, completeness, accuracy, timeliness, verifiability, fidelity / representation, lifetime
10.39	Lake Peigneur Drilling Accident	While drilling a test well, a rig crew inadvertently caused a flood in a nearby salt mine. The previously freshwater lake became a salt water lake and the flow of a river was reversed.	Oil and Gas	1980	Verifiability

10.2 Post Office Horizon System

Post Office Limited is a company wholly owned by the UK government, and provides a variety of counter services to the general public. These include postal services, banking services including currency exchange, issue of international driving permits, driving licence renewals, passport application checks, benefits payments, and various other services. A small number of branches are operated by Post Office Limited

itself ("Crown Post Offices"), but the vast majority are operated under contract ("Branch Post Offices") by independent persons known as subpostmasters (SPMs).

In the 1990s, Post Office Counters Limited (POCL) (the name at that time for what would become Post Office Limited), the Department of Social Security (the government department at that time responsible for the Post Office) and ICL agreed to replace the paper-based accounting scheme at Post Office branches with an electronic system, in particular to allow the payment of benefits by electronic transfer instead of cash. A pilot system known as Pathway was rolled out to a small number of branches in 1996, but was subsequently abandoned due to "greater than expected complexity". However, POCL and ICL decided to continue with a system based on Pathway to automate branch post offices. ICL was acquired by Fujitsu in 1998, and the resulting system, known as Horizon (now known as Legacy Horizon), was rolled out from 1999 and a version that combined management accounting functions and electronic point-of-sale functions, Horizon Online, was rolled out from 2010.

From the outset, Horizon has been a data-driven system "in which any requirement which might change frequently is encoded as data, rather than software code. The code is written and tested to work with all allowed values of the data" [7].

Shortly after the introduction of (legacy) Horizon, there was a sharp increase in SPMs reporting accounting shortfalls; the products that Horizon showed they had sold far exceeded the money the SPMs had taken. However, unlike the paper system, Horizon did not allow SPMs any access to the transaction records, so they were unable to trace the cause of the discrepancy. Under the terms of their contract with Post Office Limited the SPMs were obliged to make good any shortfall unless they could prove they were not at fault. Without access to the accounting trail there was no possible way for them to do that.

Following the roll out of Horizon, Post Office Limited "prosecuted more than 700 SPMs for crimes such as theft and false accounting. Hundreds of SPMs were sent to prison and many more received punishments such as being forced to do community service and having to wear electronic tags. [...] Hundreds were made bankrupt, losing their livelihood, and many struggled after being forced to pay the Post Office to cover shortfalls that didn't exist outside the Horizon system. The lives of the victims and their families were severely impacted, with several suicides linked to the scandal and many cases of illness caused by stress." [8]. The shortfalls were eventually found to have been due to faults in the Horizon system and to Fujitsu staff changing the accounts, apparently on the instruction of the Post Office, without the knowledge of the SPMs.

The false prosecutions resulting from the faults in the Horizon system are at the time of writing subject to a public inquiry. The Criminal Cases Review Commission (CCRC) described the prosecutions as "the most widespread miscarriage of justice the CCRC has ever seen and represents the biggest single series of wrongful convictions in British legal history" [9].

The data issues included:

- Neither SPMs nor Post Office Limited were able to access the full data necessary to identify the source of accounting discrepancies [10, §995] and in particular SPMs' "ability to investigate was itself similarly limited. The expert agreement [...] makes it clear in IT terms (based on the transaction data and reporting functions available to SPMs) that SPMs simply could not identify apparent or alleged discrepancies and shortfalls, their causes, nor access or properly identify transactions recorded on Horizon, themselves. They required the co-operation of the Post Office [10, §1000]".
- Change control processes for the data representing the products and services provided were inadequate [7, §54].

- Transactions within Horizon were not atomic, so transactional integrity was not maintained: if a transaction failed, the payment for the goods or service could be recorded without showing on the SPM's point of sale system. In that case SPMs were instructed to retry the transaction, so the payment would be recorded multiple times for a single transaction. Possible causes of failure of a transaction included the speed with which a button on the point-of-sale terminal was pressed [10, §113].
- Horizon did not maintain correct double-entry bookkeeping even within transactions that were completed normally [7, §128ff]
- Fujitsu, apparently on Post Office Limited's instruction, changed accounting transactions without the knowledge of the SPMs responsible for those transactions [7, §61.4].
- Records were not kept of the occasions accounting transactions were altered by Fujitsu [10, §1013, §1014]
- Post Office Limited staff were given unnecessary top-level security access to the accounting data [10, §390].

10.3 Battle of Agincourt

At the battle of Agincourt (1415), the French cavalry had little regard for the English longbowmen. They appear to have been so proud that they thought little of the power of the English archers. The French underestimated the longbow at Agincourt in several ways. Thomas Walsingham, an English chronicler, described how the French were "utterly astonished" by the effectiveness of the English longbowmen, noting that "the French never thought that such weapons could have such power". This chronicler, among others, demonstrates that the French underestimated the longbow at Agincourt due to their misconceptions about its range, accuracy, and impact on their cavalry. This underestimation ultimately contributed to their defeat at the hands of the English army.

The French assessment of English capability was probably based on their own experience with crossbows, which had more limited range and effectiveness. Thus even when Henry moved his troops closer to the French army, so as to be within longbow range, the French still did not perceive the threat.

Data properties involved: Their assessment was based upon incorrect data, resulting in loss of *integrity* and *completeness*. Although they understood the principles of one type of bow, the French data could be regarded as inaccurate, due to their read-across from one type of weapon to another, resulting in loss of the accuracy and the *fidelity / representation* property. However another view of the problem was that the French used data from what (to them) was old technology – the crossbow – and failed to keep their data up to date as technology advanced, leading to loss of the *timeliness property*.

Links

- Agincourt: https://www.longbow-archers.com/historyagincourt.html (Accessed 11 February 2024)
- Walsingham, Historia Anglicana, Volume II: https://archive.org/details/thomwalsinghamh00walsgoog/page/312/mode/2up?view=theater (Accessed 11 February 2024)

10.4 Gemini V

The Gemini V crewed spaceflight took place in 1965. It lasted eight days – twice as long as Gemini IV, thereby demonstrating that a spaceflight long enough to get to the moon and back was feasible.

There had been various system failures during the mission. However during re-entry, systems behaved correctly, and the crew were able to control the descent as planned. However the spacecraft landed eighty miles short of its intended landing zone.

Post-mission analysis revealed that the computing equipment had not failed. The error had been introduced by a simple data error – an assumption that the Earth turns 360° in 24 hours. The 24 hour day was based upon the position of the Sun at midday, and a simplistic understanding of Earth's rotation suggests that the ball must have rotated $360°$ for the Sun to appear in the same vertical plane. However the Earth follows an orbit around the Sun, making one full orbit per year. Thus on average, every 24 hours, the Earth moves $360 \div 365.24 = 0.99°$ along its orbit, and must over-rotate by a similar amount for the Sun to appear in the same place. In other words, whilst everybody "knows" that the Earth rotates 360° per day, that figure is incorrect.

As the computer had been given a parameter of 360°, instead of the accurate value, the calculations to reach the landing site were inaccurate.

Data properties involved: The approximation was a loss of the accuracy property. Its introduction resulted from a lack of consideration of celestial movement, and so could potentially also be regarded as a loss of both the *traceability* and *fidelity / representation* properties.

Links

- A general outline of the Gemini V mission: https://en.wikipedia.org/wiki/Gemini_5 (Accessed 26 January 2022)
- An explanation of how Earth's orbit results in more than 360° rotation in 24 hours: https://en.wikipedia.org/wiki/Sidereal_time (Accessed 26 January 2022)
- The Gemini V mission report: https://www.ibiblio.org/apollo/Documents/Gemini5MissionReport.pdf (Accessed 26 January 2022)

10.5 What3Words

The What3Words application is intended to provide a means to easily and unambiguously define a location anywhere on the surface of the Earth. The ease with which the application can be used has led to its use by emergency services worldwide, for locating people in need of assistance – in safety terms, it has provided a useful form of mitigation, enabling emergency services to easily locate people, whether simply at the roadside, or somewhere on a mountainside. However there have been a number of claims that What3Words is not suitable for this purpose.

Accoring to the Telegraph (1 June 2021), Mountain Rescue England and Wales (MREW) claimed it had been told to go to 45 locations in the past 12 months which had turned out to be incorrect, whilst a rescue request that indicated a location in Australia actually emanated from a location in China. The article suggested that the problem was a combination of local accents and spelling errors.

However the BBC reported that research by Andew Tierney indicates that the problem is more systemic, and reported:

- The dictionary used by What3Words includes a number of homophones – words that have the same pronunciation, but different spelling.

- Similar sounding words
- Plurals

Cybergibbons illustrated the problem by presenting a list of thirty two pairs of words in the What3Words dictionary that appear to be extremely similar in pronunciation. They also pointed out a further algorithmic deficiency – that plural words can be followed by a word beginning with the letter "s". Their examples include likely.stage.sock and likely.stages.sock, which denote locations on opposite sides of the River Clyde, and illustrate how in mountain rescue situations, even a small error in location could result in a serious delay to emergency provision.

Thus it appears that the system is working as designed, by providing a simple means to carry out geolocation. However a tool that was designed as a social application to help friends meet up has now been applied to a safety-related domain, without being re-engineered for that more demanding purpose. Three words will appear on the screen of the person requiring help, and if those same three words are used by the emergency services, the parties will be able to meet up. However that transfer is carried out by voice, where ambiguity can introduce error. The dictionary appears to contain homophones, plurals are used, and the algorithm permits the selection of pairs of words where the ending of one word cannot easily be distinguished from the start of the following word – each of these failings can lead to errors in communication.

Data properties involved: The ambiguities can lead to loss of *integrity*, whilst the resulting distortion of world view is a loss of *consistency*, and the Data Issue is one of *Aliasing*.

Links

- The Telegraph article: https://www.telegraph.co.uk/news/2021/06/01/rescuers-directed-china-australia-what3words-app-regional-accent/ (Accessed 16 January 2022)
- A BBC article: https://www.bbc.co.uk/news/technology-56901363 (Accessed 16 January 2022)
- Cybergibbons report: https://cybergibbons.com/security-2/why-what3words-is-not-suitable-for-safety-critical-applications/ (Accessed 16 January 2022)

10.6 Java log4j Library Vulnerability

On Dec 10th 2021 a new critical zero-day vulnerability was detected that affected Apache Log4j 2 Java library. It adversely impacted the digital domain and security systems worldwide.

The vulnerability, when exploited, permited remote code execution on the vulnerable server with system-level privileges.

Log4j is a highly configurable logging mechanism for Java ("log4j") that is used for documentation and debugging. Although originally developed for the Apache web server, it has been used part of many commercial applications, including network monitoring tools and even games such as Minecraft.

The exploit was a combination of the Java code that contains different logging functions (typically error(), warn(), info(), debug(), …) and a configuration file. The configuration file specifies which information shall be added to the log-file, the associated format, and how to "interpret" the logged data.

The security risk was that the logging mechanism was by default configured in a way such that it interpreted the logged data, and that the logged data that the user entered could be used to attack the server. For

example if a user were to enter into the name field of a html-page instead of his name a "delete *.*" command, along with certain escape sequences, it might cause huge damage on the server – if this data were logged from the software and interpreted from the configuration file.

Data property involved: *integrity*.

Links

- Apache provides details on security issues with the log4j library, including available fixes, on its website https://logging.apache.org/log4j/2.x/security.html (Accessed: 09/01/2022)
- Further details may be found at https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2021-44228 (Accessed: 09/01/2022)

10.7 Immensa False Negative Covid-19 Tests

Failings at the Immensa lab in Wolverhampton led to an estimated 43,000 people with Covid-19 being mistakenly given negative PCR results; they thought they were in the clear, but were actually positive for Covid-19. This contributed to soaring rates of infection across the South West and Wales.

It took the government almost a month to identify the issue and to stop sending PCR tests there.

It is unknown how much the virus spread in that time, but the effects of this mismanagement are potentially huge. Professor Deepti Gurdasani, a senior lecturer in epidemiology at Queen Mary University, estimates that the false negatives may have caused up to 200,000 further Covid-19 infections, and more than 1,000 avoidable deaths.

On Monday, November 1st 2021 the Good Law Project launched legal proceedings against the Secretary of State, Sajid Javid, over the Immensa testing scandal, citing the lack of a proper system to monitor the accuracy of tests at such labs breached the Department of Health and Social Care's duty to protect life, and the human rights of those affected.

Data properties involved: integrity, accuracy and possibly *traceability* and verifiability.

Links

- BBC programme Inside Science containing discussion with Professor Gurdasani: https://www.bbc.co.uk/programmes/m0010q9x (Accessed: 09/01/2022)
- https://goodlawproject.org/news/immensa-update/ (Accessed: 09/01/2022)
- https://www.gov.uk/government/news/testing-at-private-lab-suspended-following-nhs-test-and-trace-investigation (Accessed: 09/01/2022)
- https://www.theguardian.com/world/2021/dec/21/immensa-lab-month-delay-before-incorrect-covid-tests-stopped (Accessed: 09/01/2022)

10.8 Covid-19 Test Results Silently Deleted by Excel

In early October 2020 the British Government announced that 15841 positive Covid-19 test results had not been reported in the totals for England between 25 September and 2 October. This also meant that

the contacts of those people were not traced, or asked to self-isolate, meaning that the virus might have spread further than it would otherwise have done, and possibly have taken additional lives.

The results were silently discarded as they were imported into the Public Health England database. Results were delivered as a CSV file, which was imported into an Excel spreadsheet "template", using the ".xls" format, which was then in turn imported into the national database. The ".xls" format has a limit of 65536 (2^{16}) rows, and rows beyond this limit in the CSV file were silently discarded. (Data category: dynamic; properties lost: *integrity, completeness, timeliness, availability, fidelity / representation*)

The newer ".xlsx" file format would have increased the row limit to 1048576 (2^{20}) rows before suffering the same problem, but the CSV file format has no limit on the number of rows.

This demonstrates the danger of using COTS software for safety-related functions without fully analysing its limitations. A decision had already been made to replace this system, but had not been acted upon. It is also an example of dark data, where it comes under the *data we don't know are missing: "unknown unknowns",* and the *Missing What Matters* categories.

Finally, it is an example of where an error is known to the system, but not reported (adequately) to the user (Data Category: dynamic; properties lost: *timeliness, availability*).

Links

- https://www.bbc.co.uk/news/technology-54423988 (Accessed: 11/01/21)
- https://www.bbc.co.uk/news/uk-54422505 (Accessed: 12/01/21)

10.9 Boeing 737 MAX 8 Crashes

On October 29th 2018, Lion Air flight 610 was lost with all on board when it flew into the sea. On March 10th 2019, Ethiopian Airlines flight 302 flew into the ground. In each case the aircraft was a Boeing 737 MAX 8 – the latest modernised iteration of the 737 airframe design, and in each, a software system called the manoevring characteristics augmentation system (MCAS) has been established as the principle cause of the crashes.

The MCAS system was introduced because bigger, and hence more fuel efficient, engines are used on the 737 MAX 8, which had to be positioned higher on the wing and further forward (to ensure sufficient ground clearance for the larger engines). This changed the aerodynamic properties of the aircraft and meant the 737 MAX 8 tended to pitch up during high angles of attack. MCAS was introduced to counter this effect by taking input from an AoA sensor and commanding the horizontal stabiliser control surfaces to bring the nose down. Although the aircraft has two AoA sensors, only one sensor gave input to MCAS at any one time, meaning that a single sensor failure could cause the nose to be forced down incorrectly. The design of MCAS allowed this to happen repeatedly, which if left unchecked would eventually force the aircraft into an unrecoverable dive.

MCAS did not compare data from the two sensors in order to detect a discrepancy between them, and hence indicate that the sensor data was not trustworthy (data category: dynamic data; properties lost: integrity, accuracy, fidelity / representation; mitigations unused: redundancy). Furthermore, the function (outside MCAS) to report a discrepancy between the sensors to the pilots was not enabled, reducing the crew's ability to respond appropriately (data category: dynamic; properties lost: *availability*).

Several other data-safety-related failures can also be found in the report on flight 610:

- The MCAS system was not described in the pilot's manual and training materials (Data Category: Staffing, Properties lost: *availability*)
- There was no indication to the pilots that MCASMCAS was active (Data Category: Dynamic; Properties lost: *availability*).
- The AoA sensor fitted to flight 610 was incorrectly calibrated during a previous repair, reporting an angle 21° higher than the correct value. But this was not detected during the repair (Data Category: Justification; Properties lost: *availability*, *fidelity / representation*, *verifiability*).
- The evidence of the testing of the AoA sensor after fitting to flight 610 by the maintenance crew was erroneous (Data Category: Justification; Properties lost: *integrity*, *availability*, *verifiability*)
- 31 pages were missing from the maintenance log-book for flight 610, including the records of the testing of the AoA sensor after fitting to the aircraft (Data Category: Asset; Properties lost: *completeness*, *availability*, *verifiability*)
- During the previous flight of the aircraft on flight 610, the pilots experienced repeated activation of the stick shaker, and other alarms, which were caused by the faulty AoA sensor. However they did not fully report all the issues in the flight logs, and so the maintenance crew's remediation activities did not lead them to suspect an issue with the sensor. (Data Category: Performance, Properties lost: *availability*)
- During flight testing, adaptation data was changed to give MCAS more authority, without a new safety impact assessment (Data Category: Adaptation; Properties lost: *verifiability*).

Links

- http://knkt.dephub.go.id/knkt/ntsc_aviation/baru/ 2018 - 035 - PK-LQP Final report.pdf (Accessed: 20/12/19)
- https://en.wikipedia.org/wiki/Boeing_737_MAX_groundings (Accessed: 20/12/19)
- https://en.wikipedia.org/wiki/Lion_Air_Flight_610 (Accessed: 20/12/19)
- https://en.wikipedia.org/wiki/Ethiopian_Airlines_Flight_302 (Accessed: 20/12/19)
- https://www.businessinsider.in/thelife/boeing-reportedly-made-the-flight-control-system-that-mistakenly-activated-during-2-deadly-crashes-4-times-stronger-while-creating-the-737-max/articleshow/68840690.cms (Accessed: 03/01/20)
- https://www.theverge.com/2019/5/2/18518176/boeing-737-max-crash-problems-human-error-mcas-faa (Accessed: 04/01/20)
- The design, development & Certification of the Boeing 737 MAX, September 2020, The House Committee on Transportation & Infrastructure. https://transportation.house.gov/committee-activity/boeing-737-max-investigation (Accessed 22/01/2021)

10.10Loss of Soyuz-2.1b Rocket Carrying Meteor-M 2-1 Weather Satellite

On November 28th, 2017, the second launch took place from russia's new launch site at Vostochny, carrying the Meteor-M No.2-1 polar-orbiting weather satellite, and 18 small satellites flying as secondary payloads.

The launch appeared successful, but several hours later it was announced that it had not been possible to establish communications with the weather satellite, because it was not in its target orbit. An unconfirmed report claimed that the rocket was in the wrong orientation during its initial burn, and crashed into the Atlantic ocean. The following January, russian deputy prime minister Dmitry rogozin reported that the 2.6bn-rouble ($45m) satellite was lost because the Soyuz-2.1b launch vehicle had been programmed to take off from Baikonur, and not from the actual launch site at Vostochny. This is an example of a problem with data in the Adaptation category, which did not map correctly to the real-world entity that it was modelling, and hence lost the *fidelity / representation* property. Other properties that were lost include *integrity and verifiability*.

Links

- https://www.theguardian.com/world/2017/dec/28/russian-satellite-lost-wrong-spaceport-meteor-m (Accessed: 14/01/20)
- https://www.space.com/39270-russian-weather-satellite-doomed-human-error.html (Accessed: 22/01/20)
- https://www.space.com/38918-russian-satellites-lose-contact-after-launch.html (Accessed: 22/01/20)

10.11Cambrian Line Data Loss

In the United Kingdom, railway signalling is primarily implemented through trackside signals. In 2011, a trial system was installed in the Machynlleth signalling control centre, which enabled suitably equipped trains travelling over the part of the rail network that it controlled (the Cambrian lines) to acquire data relating to speed restrictions and display these to the driver. The Cambrian lines are a collection of rail tracks which run along the Welsh coast and as far inland as Shrewsbury. On 20th October 2017, a driver reported that his train had failed to display speed restriction data.

The initiating event was just after 23:00 hrs on 19th October 2017, when a train automatically requested a movement authority (permission to travel on a specified part of the rail network) which had already been allocated to another train. This error condition is known to occur several times per year and causes an automatic software reset to be invoked. A well-established set of processes are used by the signalling centre staff to return the rail system to normal operation. The staff followed their processes, and once the system was functional, allowed the final three trains of the day to complete their journeys. On the following day, it was the driver of the fourth train of the day who noticed the error and reported the failure.

Subsequent investigation revealed that speed restriction data had been unavailable since the software reset, resulting in six trains completing their journeys without that data, before the driver of the seventh train observed the problem.

The most significant risk identified thus far by the rail Accident Investigation Board is that a number of the speed restrictions which were in place on the Cambrian lines had been invoked to allow pedestrians at level crossings sufficient time to take action, when observing an approaching train. Luckily the failure did not result in an accident – but this was due to luck, not fail-safe systems on the railway.

This incident highlights the importance of the *availability* data property, as the data had been silently unavailable to trains. In addition, much of the digital audit trail relating to the failure was lost during repeated attempts to correct the problem and get the rail network running again, a loss of the *history* data property.

Links

- Interim report: https://assets.publishing.service.gov.uk/media/5bc871d5e5274a0956564a41/Ir012018_181018_Cambrian_TSrs.pdf (accessed 31 December 2018).
- Final report: https://www.gov.uk/government/news/report-172019-loss-of-safety-critical-signalling-data-on-the-cambrian-coast-line (accessed 14 January 2020).

10.12Loss of Irish Rescue Helicopter

On 14 March 2017, an Irish search and rescue (SAR) helicopter apparently suffered CFIT; all four crew members were killed. *Note: At the time of writing the investigation is continuing. The following discussion is based on a preliminary report from the Air Accident Investigation Unit, Ireland*.

The SAR helicopter was responding to a medical emergency on board a fishing vessel. It left Dublin and requested a route to Blacksod to refuel. The flight data recorded on the health and usage monitoring system (HUMS) showed the helicopter was in stable level flight until the final few seconds, when it pitched up rapidly and impacted with terrain at the western end of Black rock.

The helicopter was equipped with an enhanced ground proximity warning system (EGPWS). This is designed to decrease instances of CFIT by increasing pilot situational awareness, including the use of alerts and warnings: it is not intended to be used for aircraft navigation. The EGPWS can provide terrain alerting using look ahead algorithms, which take information from the aircraft (e.g. position., attitude, heading) and use this in conjunction with internal terrain and obstacle databases. Neither the Black rock lighthouse nor the island's terrain were included in the EGPWS databases; these databases had been sourced from external suppliers by the EGPWS manufacturer.

The preliminary investigation also notes that the flight crew were following an operator-specific route guide: a review of such guides has been recommended.

This incident potentially illustrates the importance of the *fidelity / representation* and *completeness* data properties, with respect to the EGPWS database, and the accuracy data property, with respect to the route guides.

Links

- http://www.aaiu.ie/sites/default/files/report-attachments/ rEPOrT%202017-006%20PrELIMINArY.pdf (accessed 29 November 2017).

10.13Loss of Schiaparelli Mars Lander

The Schiaparelli module, also known as the entry demonstrator module (EDM), was part of the European Space Agency (ESA)'s ExoMars 2016 mission. The objective was to validate and demonstrate entry, descent and landing on Mars in preparation for the ExoMars 2020 mission.

On 19 October 2016, the EDM entered the Mars atmosphere at 14:42:07 (UTC). During its entry and descent it constantly transmitted telemetry. Its signal was lost at 14:47:22 (UTC), about 43 seconds before expected

touchdown. On 20 October, a camera on NASA's Mars reconnaissance Orbiter imaged the planned landing site and observed crash debris.

During entry, a parachute was deployed as planned. This triggered oscillations that saturated the IMU. Integration of this saturated value caused a significant error in predicted attitude. As the descent continued, a radar Doppler altimeter (RDA) was turned on. The significant attitude error led to large discrepancies between the IMU and the RDA. The nature of the guidance and navigation control software meant that this discrepancy led to a premature declaration of touchdown. As such, the parachute was jettisoned too early, causing the EDM to crash into the planet's surface.

The investigation determined the rates that saturated the IMU could have been predicted. Limitations in the modelling of parachute dynamics meant they were not. The investigation also noted issues with the persistence of the flag used to denote IMU saturation, as well as inadequate handling of this saturation by the guidance software.

This incident illustrates the importance of the *fidelity / representation* data property, with respect to the modelling, and the *integrity* data property, with respect to the persistence time of the saturation flag.

Links

- http://exploration.esa.int/mars/59176-exomars-2016-schiaparelli-anomaly-inquiry/ (accessed 29 November 2017).

10.14Interception of Communications

In July 2015, it was reported that a public authority was undertaking an investigation into the uploading of indecent images of children and requested details of the account connected to the internet protocol (IP) address used to upload the images. Issues with a new upgrade of the communication provider's system resulted in incorrect data being disclosed. Investigations revealed that a further five requests had resulted in incorrect data being disclosed. Data was acquired in six cases that related to individuals unconnected with the investigations. In one of these cases a welfare check was delayed on a child believed to be in crisis.

Under the Regulation of Investigatory Powers Act 2000, internet service providers and indeed other communication service providers (e.g. mobile phone network providers) are required to provide data to investigatory bodies such as the police. This data can be used to support criminal investigation and prosecutions and in the protection of vulnerable children and adults. The data clearly has the potential to be safety related, but there is no obligation for data providers to treat it as such. In this case the data errors led to a child being exposed to additional risk of harm.

This incident highlights the importance of the *integrity* data property. It also shows the applicability of data safety guidance to areas that are not traditionally encompassed by safety engineering.

Links

- https://www.ipco.org.uk/docs/iocco/2015%20Half-yearly%20report%20(web%20version).pdf (accessed 9 January 2019).

10.15A400M, Torque Calibration Parameters

On 9 May 2015, just minutes into a routine, pre-delivery test flight an Airbus A400M military plane crashed in Spain, killing four of the six crew. Three of the four engines had become stuck at high power and initially did not respond to the crew's attempts to control the power setting in the normal way. Pilots then succeeded in reducing power only after selecting the thrust levers to idle. The engines subsequently remained stuck in this mode. In an attempt to return to the airport, the aircraft struck power lines and crashed.

Although not confirmed, reports suggest the torque calibration parameters for the engines were wiped during a software installation. The torque calibration data is needed to measure and interpret information coming back from the A400M's engines, and is crucial for the electronic control units (ECUs) that control the aircraft's power systems.

This accident highlights the importance of the *completeness* data property, specifically with respect to the torque calibration parameters.

Links

- http://www.bbc.co.uk/news/technology-33078767 (accessed 29 November 2017).
- http://www.reuters.com/article/us-airbus-a400m-idUSKBN0OP2AS20150609 (accessed 29 November 2017).

10.16RN Submarine, Trawler Karen

On 15 April 2015, a submerged royal Navy submarine snagged the fishing gear of the UK registered trawler *Karen*, 15 miles south-east of Ardglass, Northern Ireland. *Karen* had been trawling for prawns on a westerly heading at 2.8 knots when its fishing gear was snagged and it was dragged backwards at about 7 knots. *Karen*'s crew managed to release both winch brakes, freeing the trawl warps; the starboard warp ran out completely but the port warp became fouled on the winch drum, causing the vessel to heel heavily to port and its stern to be pulled underwater. *Karen* broke free from the submarine when the port warp parted; there was structural damage to the vessel but it returned to Ardglass safely under its own power. Evidence of the collision on board the submarine was either not observed or misinterpreted.

The nature of sub-surface operations requires the use of sonar technology to detect collision hazards. Detection in this way is reliant on noise emanating from contacts. In this instance the fishing trawler was detected but misidentified as a merchant vessel rather than a fishing vessel because the submarine's sonar operators did not detect or report hearing trawl noise. Given the number of vessels operating in the area, it is almost certain that the noise levels being generated would have been extremely high, with noise from one vessel masking the noise from another. Such a situation would make it very difficult for the operators to methodically identify and analyse each contact, in particular to identify discrete acoustic classification clues such as trawl noise. As a result the trawler was assessed to be that of a small merchant vessel and the command team's perception would have been that no risk of collision could exist between a submarine at safe depth and a merchant vessel.

review concluded that the submarine was operating near to the limit of its capability. Given that all the submarine's systems were reported to be functioning properly, it was apparent that the submarine's limit of capability had, in reality, been exceeded, with its sonar and command teams becoming cognitively overloaded, leading to degraded situational awareness and poor decision making.

In conclusion, the Maritime Accident Investigation Board (MAIB) report stated, "The collision happened

because the submarine's command team believed *Karen* to be a merchant ship, so they did not perceive any risk of collision or need for avoiding action."

This incident highlights the importance of the *resolution* data property, specifically with regards to resolving a trawler and a merchant vessel. It also highlights the importance of the *integrity*, specifically with regards to the information provided from the sonar team to the command team.

Links

- https://www.gov.uk/maib-reports/collision-between-the-stern-trawler-karen-and-a-dived-royal-navy-submarine (accessed 29 November 2017).

10.17Turkish Airlines A330

During March 2015, an Airbus A330-303, operated by THY Turkish Airlines, suffered a runway excursion accident upon landing at Kathmandu-Tribhuvan (KTM).

Flight TK726 was a regular passenger service from Istanbul-Atatürk (IST) to Kathmandu, Nepal. The flight was the first international flight to arrive that morning. After descending from cruising altitude, it entered a holding pattern. It was subsequently cleared for a VHF omnidirectional range (VOR) / distance measuring equipment (DME) approach to runway 02.

This approach was abandoned at about the Missed Approach Point at 1DME and the aircraft performed a go around. The aircraft circled and positioned for a second approach to runway 02. The aircraft touched down to the left of the runway centre line with the left hand main gear off the paved runway surface. It ran onto soft soil and the nose landing gear collapsed. Following the accident the aircraft was written off.

The aircraft touched down to the left of the centreline because the flight management guidance system (FMGS) navigation database contained threshold coordinates for a proposed displacement of the runway 02 threshold. This was later withdrawn through a notice to airmen (NOTAM), but had not been updated by the airline in the database. Additionally, the coordinates that were initially published were inaccurate, causing the threshold coordinates to be offset to the left of the actual threshold. This had been noticed and reported by a previous Turkish Airlines flight on March 2. The changes had not been performed by the time TK726 landed at Kathmandu.

Among the safety recommendations stated in the accident report were:

- "The operator must ensure that the correct navigation data are uploaded on flight management guidance system";
- "The operator should establish a system of verifying the quality of charts prepared by the service provider";
- "The operator should establish a system of checking the validity of the flight management system (FMS) database"; and
- "Civil Aviation Authority of Nepal must ensure that raw aeronautical information / data are provided by the aerodrome authorities taking into account of its accuracy and integrity requirements for aeronautical data as specified by International Civil Aviation Organization (ICAO) Annex 15 and its Aeronautical Information Service Manual."

This incident highlights the importance of three data properties: accuracy; *timeliness*; and *verifiability*. All of these apply to data describing the runway's location.

Links

- http://www.tourism.gov.np/downloadfile/TUrKISH_AIrLINE_Final_report_finalcopy4.pdf (accessed 29 November 2017 – no longer available from this location, but on 9 January 2019 was still available through archive.org).

10.18Dallas Hospital Ebola Incident

On 26th September 2014, a Dallas hospital mistakenly sent home a man who had the Ebola virus having missed what would have appeared to be an obvious potential case: a Liberian citizen with fever and abdominal pain who said he had recently travelled from Liberia. The man later returned to the hospital, was eventually diagnosed with the illness, but subsequently died. Two nurses that had treated the man also contracted the virus but later recovered.

There have been mixed reports on the cause of the problem, but it is clear that external social phenomena such as the Ebola outbreak, which are outside the hospital's electronic health record (EHR) system and processes, can change the safety significance of data held in the EHR. If the importance of the data is not recognized and elevated appropriately in the support tools and processes, then the risk of unintended harm can increase. This conclusion is reinforced by system vendors who subsequently updated their systems to reflect the Ebola crisis in light of the Dallas incident.

This incident highlights the importance of the *completeness* and *format* data properties, in that information about the Ebola outbreak was apparently either not available, or not available in a usable form, to decision makers.

Links

- http://www.nbcnews.com/storyline/ebola-virus-outbreak/texas-hospital-makeschanges-after-ebola-patient-turned-away-n217296 (accessed 29 November 2017).

10.19Qantas Boeing 737 Take-Off

On 1 August 2014, a Qantas Boeing 737-838 aircraft commenced take-off from Sydney Airport, New South Wales. The flight was a scheduled passenger service from Sydney to Darwin, Northern Territory.

While the aircraft was climbing to cruise level, a cabin crew member reported hearing a "squeak" during rotation. Suspecting a tail strike, the flight crew conducted the tail strike checklist and contacted the operator's maintenance support. With no indication of a tailstrike, they continued to Darwin and landed normally. After landing, the captain noticed some paint was scraped off the protective tailskid. This indicated the aircraft's tail only just contacted the ground during take-off.

The Australian Transport Safety Bureau (ATSB) found the tail strike was the result of two independent and inadvertent data entry errors in calculating the take-off performance data. As a result, the take-off weight used was 10 tonnes lower than the actual weight. This resulted in the take-off speeds and engine thrust setting calculated and used for the take-off being too low. Hence, when the aircraft was rotated, it overpitched and contacted the runway.

The ATSB also identified that the Qantas procedure for conducting a check of the Vref40 speed could be misinterpreted. This negated the effectiveness of that check as a defence for identifying data entry errors. In this case, uncorrected errors affected the integrity of the data used to calculate take-off parameters.

This incident highlights the importance of the *integrity* and *verifiability* data properties, with respect to the data used to calculate take-off performance data.

Links

- http://www.atsb.gov.au/publications/investigation_reports/2014/aair/ao-2014-162.aspx (accessed 29 November 2017).

10.20Qantas Boeing 737 Loading

On 9 May 2014, a Qantas Boeing 737 was preparing for departure from Canberra to Perth. There were 150 passengers, 87 of which were primary school children. These children were all seated together at the rear of the cabin.

A 'name template' was completed by a travel agent on behalf of the school group. This group was travelling from Perth to Canberra and returning back to Perth. Despite being marked as mandatory, the "Gender Description" field in this template was left blank; options for this field were "Adult", "Child" and "Infant".

As per company procedures, two days before the Perth-Canberra leg of their journey this group was 'advance accepted' into the booking system. Since the fields recording the number of children and young passengers in the group were blank, the Customer Service Agent assumed all of the group were adults. No loading-related issues were experienced during this flight.

Two days before the return flight the group was again "advance accepted" as all adults. This meant they had all been assigned an 'adult weight' of 87 kg. They were checked in at Canberra Airport and assigned seats at the rear of the aircraft. During take-off the aircraft appeared nose heavy. Significant back pressure was required to rotate the aircraft and lift off from the runway. The aircraft exceeded the calculated take-off safety speed by about 25 kt. The aircraft rose at a higher initial climb speed than usual, but the crew did not receive any warnings. No further issues were experienced during the flight.

This incident demonstrates the importance of the *completeness* Data y (i.e., ensuring that the mandatory "Gender Description" field was completed) and the *fidelity / representation* data property (i.e., ensuring the calculated aircraft loads and balances reflect the real situation). It also illustrates some potential difficulties associated with the use of default data.

Links

- http://www.atsb.gov.au/publications/investigation_reports/2014/aair/ao-2014-088.aspx (accessed 29 November 2017).

10.21Grounding of Navigator Scorpio

On 3 January 2014, the liquefied gas carrier *Navigator Scorpio* ran aground on Haisborough Sand in the North Sea. The vessel was undamaged, no pollution occurred and after two and a half hours the vessel refloated on the rising tide.

The schedule for the *Navigator Scorpio* was changed close to the time of its departure. This change meant that additional North Sea coastal charts were required. These charts were delivered to the vessel shortly before its departure. However, they were not up to date with the latest corrections and they were not corrected prior to sailing. In addition, the passage plan (i.e., vessel route) was not checked by the master before sailing.

When the master checked the passage plan, which had been drawn up by the second officer (2/O), he suggested a change to a portion of the route. After discussion with the 2/O the route was left unchanged, but with a requirement that position fixes be obtained every five minutes rather than every fifteen. While acting as the sole bridge watch-keeper the 2/O was distracted by further passage planning activities and lost positional awareness. This led to the grounding of the vessel. After the grounding false information was added to the navigation chart to give the appearance that five minute positional fixes had been taken.

This incident highlights the importance of the *timeliness* data property, with respect to both the additional North Sea charts and the master's check of the passage plan.

In addition, the fluidity of the chart data allowed the 2/O to make false post-grounding additions to create an incorrect impression. According to the MAIB's report, such actions are not uncommon. These actions affect the *verifiability* of the chart data, which makes post-accident investigations more complicated.

Links

- https://assets.publishing.service.gov.uk/media/547c6f1740f0b6024100000d/NavigatorScorpio.pdf (accessed 29 November 2017).

10.22Loss of MQ-9 reaper

On 5 December 2012, an MQ-9 reaperMQ-9 reaper remotely piloted aircraft crashed in an unpopulated area three miles north-east of Mount Irish, Douglas County, Nevada. The crash occurred due to a stall, which was the result of an unrecognized reverse thrust condition. The aircraft and a number of pieces of ancillary equipment were destroyed. The total damage to United States government property was assessed at over $9 million.

The investigation board concluded that the throttle settings of the GCS were incorrectly configured. This misconfiguration arose as the GCS was converted from supporting MQ-1 operations to supporting MQ-9 operations. It persisted despite the presence of a checklist, the completion of which should have identified the error. The misconfiguration meant that reverse thrust was commanded whenever the pilot's throttle was in any position except full forward.

This incident highlights the importance of the *consistency* data property, with respect to the differences between the GCS settings and the aircraft it was meant to be controlling. It also highlights the importance of the *verifiability* data properties, with respect to the GCS settings (and, in particular, the limitations of using checklists to verify data).

Links

- http://www.airforcemag.com/AircraftAccidentreports/Documents/2013/120512_MQ-9_Nevada_full.pdf (accessed 29 November 2017).

10.23Boeing 737-33A at Chambery Airport, France

On the 14 April 2012 and prior to departing Chambery Airport in France, the crew of a Boeing 737 used an Electronic Flight Bag (EFB) computer to calculate the aircraft's take-off performance. During the use of the software application the commander omitted to input the aircraft's take-off weight and it defaulted to the previous flight's data. Compounding the issue was that none of the crew undertook a cross-check of the EFB's output and the pilot subsequently employed incorrect speed and thrust information for the take-off. The consequence of using the incorrect information was that the calculation of the required airspeed for rotation was too low and the pilot continued to increase the aircraft's pitch angle to the point whereby the tail hit the runway. There were no injuries sustained in this incident but the aircraft suffered damage.

Following its investigation, which examined the wider employment of computers to derive aircraft performance information, the Air Accident Investigation Branch (AAIB) identified that there had been "a number" of previous accidents and incidents attributable to the "incorrect calculation of take-off performance"; and that due to the potential for degraded climb performance a catastrophic outcome could be envisaged. The AAIB also recognized that "take-off under-performance" is subtle and many other events of this nature may have been experienced but never reported. In its conclusions the AAIB acknowledged that using computers has "brought about improvements in accuracy and ease with which aircraft performance requirements can be made". However, there are "continued vulnerabilities" associated with the use of incorrect data that it is essential to control through "appropriately designed software and hardware". Although there were no injuries in this instance, this incident and the conclusions of the AAIB highlight some important points for "Safety-related Information Systems".

A clear chain of events was established that involved the use of incorrect information as a causal factor leading to an incident, which had the potential to be of a catastrophic nature;

- This was not an isolated incident;
- The crew did not appreciate the criticality of the EFB's information and it was used without validation;
- The AAIB recognized the essential need for appropriate system development.

It is often recognized that data must be up to date, but the explicit need to prevent the use of old data can be omitted from the safety requirements. This incident illustrates the importance of the properties *timeliness*, *suppression* and *lifetime*.

Links

- *Air Accident Investigation Branch April Bulletin 4/2013* [on line] available at http://www.aaib.gov.uk/publications/bulletins/april_2013.cfm (accessed 17 January 2021).

10.24Loss of Hermes 450

On 2 October 2011, a Hermes 450 UAS crashed at Bastion Airfield, Afghanistan. The aircraft was unrepairable.

The aircraft sortie was terminated early due to rising engine temperature. Due to the presence of vehicles and people in the vicinity of the runway, the GPS take-off and landing system (GTOLS) was selected to land the aircraft. Shortly after the approach had been initiated, the landing was self-aborted by the UAS. This abort occurred as a result of an incorrect data parameter in the GTOLS set-up loaded by the crew.

Moments after the self-abort, due to the urgency of the situation and the addition strain on the engine caused by the aborted landing, the crew chose to abbreviate the pre-programmed go-around GTOLS route. Instead, they issued a 'fly to coordinate' command. As the aircraft was climbing, the engine temperature rose rapidly, before the engine failed completely. On its descent the aircraft initially impacted an unoccupied hangar, before striking the ground upside down. It eventually came to rest on an empty aircraft dispersal pan.

The Service Inquiry determined that the cause of the accident was engine failure, as a result of overheating caused by oil starvation. Like many incidents, there were a considerable number of interacting factors. In total, thirteen contributory factors were identified, including the error in the GTOLS data.

This incident highlights the importance of the *integrity* data property, with respect to the GTOLS data loaded by the crew.

Links

- https://www.gov.uk/government/publications/service-inquiry-investigating-the-accident-involving-unmanned-air-system-uas-hermes-450-zk515-on-02-oct-11 (accessed 29 November 2017).

10.25Advocate Lutheran Hospital

A Chicago hospital paid $8.25 million to settle a lawsuit brought by the parents of an infant boy who died at the institution in October 2010 after a series of medical errors.

The mother gave birth to her son 4 months prematurely. She stayed by his side with her husband for the next six weeks while the boy remained in the hospital's care. On 15 October, the baby suddenly died after coming out of a heart operation without any clear complications.

The hospital determined that a pharmacy technician had entered information incorrectly when processing an electronic intravenous (IV) order for the baby. This resulted in an automated machine preparing an IV solution containing a massive overdose of sodium chloride, more than 60 times the amount ordered. The problem would have been identified by automated alerts in the IV compounding machine, but these were not activated when the customised bag was prepared for the baby. That is, adaptation data had been used to change the behaviour of the machine.

Investigations also found that the outermost label on the IV bag administered to the baby did not reflect its actual contents. Furthermore, although a blood test on the infant had shown abnormally high sodium levels, a lab technician assumed the reading was inaccurate. This highlights a different perspective on the dangers of defaulting, in this case a default assumption rather than a numerical default value.

Since the incident, staff have been activating alerts for similar IV compounders used in the system's hospitals and strengthened "double check" policies for all medications leaving pharmacies.

This incident highlights the importance of the *integrity* and *verifiability* data properties, for example with regard to: the information in the IV order; the bag label; and the blood test results.

Links

- http://articles.chicagotribune.com/2012-04-05/news/chi-parents-awarded-825-million-in-infants-death-20120405_1_clear-complications-lab-technician-double-check-policies (accessed 29 November 2017).

10.26Grounding of Sichem Osprey

On 10 February 2010 at 0436 (local), the chemical tanker *Sichem Osprey*, on her way from Panama to Ulsan (South Korea) stranded at more than 16 knots on the north-easterly part of Clipperton Island. An officer of the watch (OOW) and a lookout were on the bridge at the time and no damage had been reported prior to the accident. A 100 metre fore part of the vessel had been grounded. No pollution was observed.

Anti-collision radar alarm thresholds were apparently not set according to the Captain's instructions. There were also sizeable discrepancies between the fixes plotted on the chart and those displayed on the radar.

This incident highlights the role of the *integrity* data property, with respect to the chart plots, and the accuracy data property, with respect to the alarm thresholds which did not reflect the Captain's wishes.

Links

- https://www.nautinst.org/download.cfm?docid=F9DA081F-6C1E-40F0-A71F0A89B10F426C (accessed 5 December 2017).
- http://www.bea-mer.developpement-durable.gouv.fr/IMG/pdf/rET_SICHEM_OSPrEY_05-2010_Site.pdf (in French) (accessed 29 November 2017).

10.27Near Collision of Trains, Cootamundra

On 12 November 2009, a passenger train was being routed into Number 1 Platform road at Cootamundra, New South Wales. The driver of the passenger train received a signal indicating that the route was clear. However, as he approached, he noticed that the last wagon of a freight train was blocking his path. He applied the train brakes and stopped just short of a collision.

The investigation determined that a signalling system design error had allowed the incorrect signal to occur. The error happened despite the staff involved being suitably qualified and experienced. Working against a tight timescale, they were simultaneously developing a control table and associated software, rather than adopting the normal sequential approach. The control table contains information on points, signal and level crossing interlocking logic. The tight timescale also compromised the normal testing process. In addition, the quality control process was somewhat lacking: for example, not all identified queries and issues were appropriately closed out.

This incident illustrates the importance of the *integrity* data property, with respect to the control table data, and the *completeness* data property, with respect to data produced by the testing and the quality control processes.

Links

- http://www.atsb.gov.au/publications/investigation_reports/2009/rair/ro-2009-009.aspx (accessed 29 November 2017).

10.28Cedars-Sinai Medical Center Scanner

A software misconfiguration in a computed tomography (CT) scanner used for brain perfusion scanning at Cedar Sinai Medical Center in Los Angeles, California, resulted in 206 patients receiving radiation doses

approximately eight times higher than intended. This error persisted for an 18 month period, starting in February 2008. Some patients reported temporary hair loss and erythema.

The problem reportedly arose from an error made by the hospital in resetting the CT machine after it began using a new protocol for the procedure in February 2008. The error was not detected until one of the patients reported patchy hair loss in August 2009. "There was a misunderstanding about an embedded default setting applied by the machine," according to a statement from Cedars-Sinai. "As a result, the use of this protocol resulted in a higher than expected amount of radiation."

This incident highlights the importance of the *verifiability* data property, especially with regards to default (and adaptation data).

Links

- http://articles.latimes.com/2009/oct/10/local/me-cedars-sinai10 (accessed 29 November 2017).

10.29Grounding of The Pride of Canterbury

On 31 January 2008, the roll-on roll-off passenger ferry, *Pride of Canterbury* grounded on a charted wreck while sheltering from heavy weather in an area known as 'The Downs' off Deal, Kent. The vessel suffered severe damage to her port propeller system but was able to proceed unaided to Dover, where she berthed with the assistance of two tugs.

The vessel had been in the area for over 4 hours when, while approaching a turn at the northern extremity, the bridge team became distracted by a fire alarm and a number of telephone calls for information of a non-navigational nature. The vessel overshot the northern limit of the identified safe area before the turn was started. The OOW became aware that the vessel was passing close to a charted shoal, but he was unaware that there was a charted wreck on the shoal. The officer was navigating by eye and with reference to an ECS which was sited prominently at the front of the bridge, but he was untrained in the use and limitations of the system. The wreck would not have been displayed on the electronic chart due to the user settings in use at the time. A paper chart was available, but positions had only been plotted on it sporadically and it was not referred to at the crucial time.

Although the voyage management system (VMS) was loaded with electronic navigational charts (ENCs) for the vessel's area of operation, the system had not been approved by the Maritime and Coastguard Agency (MCA) as the owner's policy was for the VMS to be used as an aid to navigation only, with *Pride of Canterbury*'s paper charts being used as the primary means for navigation. relevant admiralty charts were supplied to the vessel for this purpose.

Despite the VMS being unapproved for use as the primary means of navigation, the officers on *Pride of Canterbury* were apparently using it as if it was. Furthermore, many of the officers, including the Chief Officer, who was in charge at the time of the accident, were not fully trained in the use of the system.

This incident highlights the importance of the accuracy data property, with regards to the information displayed on the electronic chart. It also highlights the importance of the *completeness* data property, with regards to training (and training records) and the *intended destination / usage* data property, with regards to the inappropriate use of the VMS data.

Links

- https://assets.publishing.service.gov.uk/media/547c700ded915d4c0d000071/PrideofCanterburyreport.pdf (accessed 29 November 2017).

10.30LOT Flight 282

On 4 June 2007, just after take-off from runway 09r at LHR, the pilots noticed that most of the information on both of the Electronic Attitude Director Indicators and Electronic Horizontal Situation Indicators had disappeared. The aircraft entered instrument meteorological conditions (IMC) at about 1,500 feet above aerodrome level (AAL), and the co-pilot had no option but to fly using the standby attitude indicator and standby compass. He experienced difficulty in following radar headings. The aircraft returned to land at LHR after a flight of 27 minutes.

A single error made by the co-pilot during the pre-flight preparation caused the subsequent problems. This was the use of 'E' instead of 'W' when the longitude coordinates were entered into the FMS.

The airports around London, because of their proximity to the Prime Meridian, can lead flight crews to make coordinate entry errors of this nature. It is of note that the operator's route network is such that there are few destinations to the west of the Prime Meridian and hence the majority of longitude coordinates that need to be entered would be eastings. inertial reference system (IRS) alignment warnings should have alerted the crew but may have been dismissed.

This incident highlights the importance of the *integrity* and *fidelity / representation* data properties, specifically with respect to coordinates.

Links

- https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384859/Bulletin_6-2008.pdf (accessed 29 November 2017).

10.31Annabella Container Ship – Baltic Sea

On the evening of the 25 June 2007 the container ship the Annabella was subjected to heavy seas causing the vessel to pitch and roll heavily. The following morning the ship's crew discovered that, due to the induced stresses, a "stack of seven 30 ft cargo containers" had collapsed resulting in crushing damage to the lowest containers. A number of these containers were transporting Class 2 Dangerous Goods in the form of Butylene gas.

The Marine Accident Investigation Branch (MAIB) concluded that the container stack had been "piled too high both for the particular hold location and the stacking limits of the containers". The MAIB identified that one of the incident's contributing factors had been an incorrect loading plan which had been produced by planning software used by the cargo company. The software application should have taken account of the stability and stowage information pertinent to the vessel (as provided by its manufacturer). The application, however, had unknowingly converted the container's dimensions from 30 ft to 40ft resulting in the wrong stacking limits being detailed. The cargo company passed the loading plan to the shipping terminal prior to it being inputted to the vessel's on-board loading computer. The computer did not recognize the error and the 40 ft limits were applied. Amongst the MAIB's conclusions it was noted that,

although the master is responsible for the final loading plan, appropriate oversight is difficult in practice in light of the "pace of modern container operations". The MAIB made the following recommendations in relation to the information system:

- Loading computer programs should incorporate the full requirements of a vessel's cargo securing manual and be properly approved to ensure that officers can place full reliance on the information provided;
- The availability of a reliable and approved loading computer programme is a factor to be considered in determining an appropriate level of manning for vessels on intensive schedules;
- Cargo planning software should be able to recognize and alert planners to the consequences of variable data, such as non-standard container specifications.

This incident involved incorrect data, that could have been identified if the software had highlighted the unusual aspects of the data such as the container dimensions to the operator. From a data perspective, the properties that needed to be maintained were *integrity*, *verifiability* and *fidelity / representation*.

Links

- Marine Accident Investigation Branch report 21 – *report on the Investigation of the Collapse of Cargo Containers Annabella Baltic Sea 26 February 2007* [online] available at https://assets.publishing.service.gov.uk/media/547c7032e5274a429000007d/Annabellareport.pdf (accessed 17 January 2021).

10.32Comair Flight 5191

On 27th August 2006, Comair flight 5191 crashed during take-off from Blue Grass Airport, Lexington, Kentucky. The flight crew was instructed to take off from runway 22, but instead lined up on runway 26 and began the take-off roll. The airplane ran off the end of the runway and impacted the airport perimeter fence, trees, and terrain. The captain, flight attendant and 47 passengers were killed.

The National Transportation Safety Board determined that the probable cause of the accident was the flight crews' failure to use available cues and aids to identify the airplane's location on the airport surface during taxi and their failure to cross-check and verify that the airplane was on the correct runway before take-off.

The Airport Charts used by the crew were inaccurate. The airport was under construction, and the charts were not kept current with the rapid changes that were taking place during the construction work. The chart did not accurately reflect either the taxiway identifiers or a taxiway that was closed on the day of the accident.

Due to a previously unrecognized software problem, any information the chart provider received after normal work hours on Fridays was not included in their regular updates. Furthermore, the chart provider modified the Blue Grass Airport chart after the accident to include a note that runway 8/26 is "daytime VMC[1] use only", even though this information had been published since 2001. Additionally there was a local NOTAM issued advising of taxiway closures due to construction work. However the crew was not provided with this information in their dispatch paperwork.

This incident highlights the importance of the *timeliness* data property, specifically with regards to the charts.

[1] visual meteorological conditions

The *completeness* data property is also relevant, given that neither the "after hours" changes on Fridays nor the local NOTAM were communicated appropriately.

Links

- http://libraryonline.erau.edu/online-full-text/ntsb/aircraft-accident-reports/AAr07-05.pdf (accessed 29 November 2017).
- http://en.wikipedia.org/wiki/Comair_Flight_5191 (accessed 29 November 2017).

10.33Überlingen Mid-Air Collision

On 1 July 2002, a passenger jet (Bashkirian Airlines) and a cargo jet (DHL Flight 611) collided in mid-air. The collision happened over the south German town of Überlingen; it occurred despite both aircraft being equipped with TCAS.

The two aircraft were in airspace that was controlled from Zürich and were on a collision course. A single controller was on duty and they were responsible for controlling two workstations. This arrangement was against regulations, but was tolerated by management and had become accepted practice. Initially, the controller did not appreciate the dangerous situation that was developing.

Maintenance was being conducted on the main radar system, which meant that the controller was reliant on a backup system. This delayed the presentation of radar information. In addition, a ground-based optical system that would have provided warning of the impending collision was turned off, also for maintenance: the controller was unaware of this.

Less than a minute before the collision the controller became aware of the situation. He instructed Flight 2937 to descend. Seconds after initiating this decent, the TCAS on Flight 2937 requested a climb, with the corresponding system on Flight 611 requesting a descent. Flight 2937 continued to follow the controller's direction, meaning that both aircraft were descending.

Unaware of the TCAS instructions the controller repeated the request for Flight 2937 to descend; he also provided Flight 2937 with misleading information on the relative location of Flight 611. The planes collided, resulting in the deaths of all 69 people on Flight 2937 and both people on Flight 611.

One of the actions resulting from the accident was a clarification from ICAO of how pilots should respond to contradictory information from a controller and TCAS.

This incident illustrates the importance of the following data properties: *consistency*, with regards to the instructions provided to Flight 2937; *availability*, with regards to information from the ground-based optical system; and *timeliness*, with regards to information from the radar system.

Links

- https://en.wikipedia.org/wiki/Überlingen_mid-air_collision (accessed 29 November 2017).

10.34Fort Drum Artillery Incident

Two artillery shells were fired more than a mile off target during an Army firing exercise at Fort Drum in Northern New York in March 2002. The shells landed near a mess tent where a Battalion were having

breakfast. Two soldiers were killed, 13 were injured.

The initial artillery site was unsuitable so the unit had to move to a different location nearly a mile away. The unit then had trouble setting up its digital and wire communications. The movement of the unit was not taken into account when programming the firing coordinates. Also, in what was termed a 'software behavioural shortfall' the system was designed to reset the gun elevation to zero. The correct altitude for the new site was not entered into the safety calculations, and the mistakes were not captured by the data review process.

This incident highlights the importance of the *integrity* and *verifiability* data properties, specifically with respect to the location and elevation data.

Sources

- http://www.apnewsarchive.com/2002/Army-reports-on-Ft-Drum-Accident/id-539bf2ea24b8dd66009c6efee2be926c (accessed 29 November 2017).

10.35Early Release from Washington State Prison

For over 13 years the Washington State Department of Corrections (DoC) had been releasing certain prison inmates earlier than their sentences allowed.

In 2002 the Supreme Court ruled that the DoC was erroneously denying offenders credit for early release time earned during pre-sentence detention. In attempting to address that issue the DoC incorrectly reprogrammed its computer tracking system. This resulted in the early release of offenders with sentencing enhancements. The programming error went undetected for over ten years, with more than 2,000 offenders being released early.

The error was detected when the family of an assault victim hand-calculated the assailant's release date. The family notified the DoC that it appeared as if the assailant would be released earlier than warranted by statute. It took a further three years before the programming error was finally corrected.

This incident illustrates the importance of the *integrity* data property, with respect to calculated release dates. The *verifiability* data property is also relevant, noting that the calculation of the release date was readily verifiable (as shown by the actions of the family of the assault victim).

Links

- http://www.governor.wa.gov/sites/default/files/documents/2016-02-25_DOC_report.pdf (accessed 29 November 2017).

10.36Mars Climate Orbiter

The Mars Climate Orbiter was a spacecraft launched aboard a Delta II rocket by NASA from Cape Canaveral on 11th December 1998. Its intended mission was to study the Martian atmosphere and climate, while acting as a communications relay for other spacecraft on or near Mars.

The plan was that the rocket would place the spacecraft into a transfer orbit to Mars, which would be optimized along the way by a series of four trajectory correction manoeuvres. Insertion into Mars orbit was

to take place at an altitude of 226 km, but during the week after the final correction manoeuvre, calculations predicted that it would be between 150 km and 170 km; revised to 110 km the day before insertion. The orbiter was able to survive atmospheric stresses down to about 80 km. On 23rd December 1999, the spacecraft passed behind Mars, and so out of radio contact, earlier than expected; communications were never regained.

Final calculations placed the spacecraft in a trajectory that would have taken it within 57 km of the Martian surface, but it is likely to have disintegrated before getting to that point.

It transpires that the orbiter's FMS software was designed to work with metric Newton seconds, whereas a FMS data file generated by ground system software used pound-force seconds. A Newton is about 22.5% of a pound-force or a factor of 4.45.

The cost of the mission was stated by NASA to have been $327.6 million in total ($193.1 million to develop the spacecraft, $91.7 million for launch and $42.8 million for mission operations).

This incident highlights the importance of the *consistency* data property.

Links

- http://en.wikipedia.org/wiki/Mars_Climate_Orbiter (accessed 29 November 2017).

10.37Crash into Nimitz Hill, Guam

On 6 August 1997, Korean Air Flight 801 crashed at Nimitz Hill, Guam. This is high terrain approximately 3 miles southwest of Guam International Airport, where the aircraft had been cleared to land. Of the 254 people on board, 228 were killed and 26 survived with serious injuries.

Probable causes of the accident were the captain's failure to adequately brief and execute a non-precision approach and the first officer's and flight engineer's failure to effectively monitor and cross-check this approach. Contributing factors included fatigue and inadequate flight crew training.

Another contributing factor was the intentional inhibition by the Federal Aviation Administration (FAA) of the MSAW system at Guam, and the agency's failure to adequately manage the system.

The MSAW system uses a terrain database. It is designed to alert a controller if an aircraft equipped with a Mode C transponder descends below, or is predicted to descend below, a predetermined safe altitude.

In 1990, the Guam terminal MSAW was installed to provide protection within a 55 nm radius. In 1993, a new software package was produced in which warnings were inhibited within a 54 nm radius; this left a 1 nm annular region within which warnings could be generated. The motivation behind the new configuration was to reduce false alarms. The software became operational in February 1995. A further software update became operational in April 1996. This also had the 54 nm inhibition.

This incident illustrates the importance of the *continuity* data property, with respect to the MSAW coverage, and the *fidelity / representation* data property, regarding the terrain database used by the MSAW system.

Links

- https://www.ntsb.gov/investigations/Accidentreports/reports/AAr0001.pdf (accessed 29 November 2017).

10.38San Bernardino Derailment and Pipeline Rupture

In May 1989 a Southern Pacific Transportation Company freight train derailed in San Bernardino, California. The train derailment accounted for seven fatalities and two serious injuries; however, that accident also damaged a fuel pipe and less than a fortnight later it ruptured causing a further two deaths and three serious injuries.

One of the causal factors of the train's derailment, as reported by the National Transportation Safety Board, was a "failure to determine the weight of the train" and in summary, the operator thought it weighed less than it actually did, resulting in the dynamic braking being insufficient to deal with the downhill gradient it was travelling on. The Company had used a computer to determine the train's weight and because the actual weights had not been entered the system made its calculations based upon estimated weights, which were lower. Clearly this was not a systematic failure of the computation algorithm but again potentially a failure to appreciate the criticality of the weight information and its potential as a causal factor within an accident sequence.

From the criticality of the weight information, derived safety requirements could be developed for the various data sources that were used to derive the weight. Such requirements could be expected to highlight the properties of integrity, completeness, accuracy, timeliness, verifiability, fidelity / representation and lifetime.

T Hardy, *Software and System Safety – Accidents, Incidents and Lessons Learned,* Bloomington, Author House, 2012, ISBN 978-1-4685-7470-8

10.39Lake Peigneur Drilling Accident

Lake Peigneur is located in Louisiana, United States of America. It was a ten-foot deep freshwater lake popular with sportsmen. On 20th November 1980, an exploration rig drilling for oil in the lake bed was evacuated as it began to sink; this was perceived by the crew as a structural collapse. Meanwhile, the nearby Jefferson Island salt mine was being evacuated due to the sudden onset of flooding.

The rig crew had been drilling a test well into deposits alongside a salt dome under Lake Peigneur. By some miscalculation, the assembly drilled into the third level of the nearby salt mine. Fresh water from the lake soon began trickling into the mine. Over the course of the morning, the fresh lake water began dissolving the salt and enlarging the hole until water was literally flooding into the mine.

The whirlpool created as the lake drained into the mine sucked in the drilling platform, eleven barges, trees and soil. The Delcambre Canal, which usually drains from the lake into a bay on the Gulf of Mexico, had its flow reversed. This resulted in Lake Peigneur becoming a salt water lake. Fortunately, no injuries or loss of human life were reported.

Federal experts from the Mine Safety and Health Administration were not able to determine the cause of the accident due to confusion over whether the rig was drilling in the wrong place or whether the mine's maps were inaccurate.

This incident highlights the importance of the *verifiability* data property, specifically with regards to the location of the rig. Note that this property was relevant both when the rig started to drill and also during the post-incident investigation.

Links

- http://en.wikipedia.org/wiki/Lake_Peigneur (accessed 29 November 2017).

This page is intentionally blank

11 Lifecycle Considerations (Discursive)

> *Failure is an amazing data point that tells you which direction not to go.*
> ***Payal Kadakia***

11.1 Usage Scenarios

If safety-related data is incorrect it can become dangerous when used, either by making a computer or control system perform incorrect actions, or by misleading human users into making incorrect decisions. Since the danger can only be determined when the usage of the data is understood, risk assessment should involve both the consumer of the data and the producer.

Figure 11.1: Consumer-focused integrity requirements

The consumer assesses the use of the safety-related data. (In later phases of the data safety management process this information is used to define the required data properties: for example, how accurate a particular safety-related data artefact must be.)

The producer investigates how the safety-related data is collected and what errors might occur. (Building on activities in later phases of the data safety management process, the producer can provide some form of guarantee, or level of confidence, that the safety-related data meets the specific data-related requirements.)

In some cases a producer will be providing safety-related data without any knowledge of a specific user (e.g., mapping data or generic databases that are sold to many users). In these cases the producer will need to make some assumptions about possible users, and then clearly state what level of integrity the data has been produced to. It is then up to the users to check whether the declared integrity matches their need.

11.2 Data in System Lifecycles

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 dataset (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; and
- live operational use of a safety-related system.

In these cases, the dataset is the same but the context of its use changes the safety significance and therefore the level of assurance that it may require. It follows that the DSAL of a dataset is also predicated on where and when in the lifecycle the dataset will be applied.

To illustrate this concept, a number of generic model lifecycles are discussed below. Note that these are not intended to be prescriptive or mandate the use of any particular model. Instead, they are being used to illustrate how the Data Safety Management Plan could articulate these lifecycle considerations.

Development: the diagram in Figure 11.2 represents a typical development lifecycle using an iterative development approach[1]. In this model there are key phases as the system transitions from concept through to testable executable code. The process is iterative in that several cycles of functional elaboration, design, development and test may be run and these typically will focus on the areas of the system that bear most technical risk or comprise the key functional use cases so the client gets early visibility of the system. This early awareness allows feedback to be provided into the next iteration to help steer the solution to the client's actual needs. Traditional waterfall implementation can map onto this model on the basis that there is only one iteration in each phase and all activities in one phase need to be completed before progressing to the next.

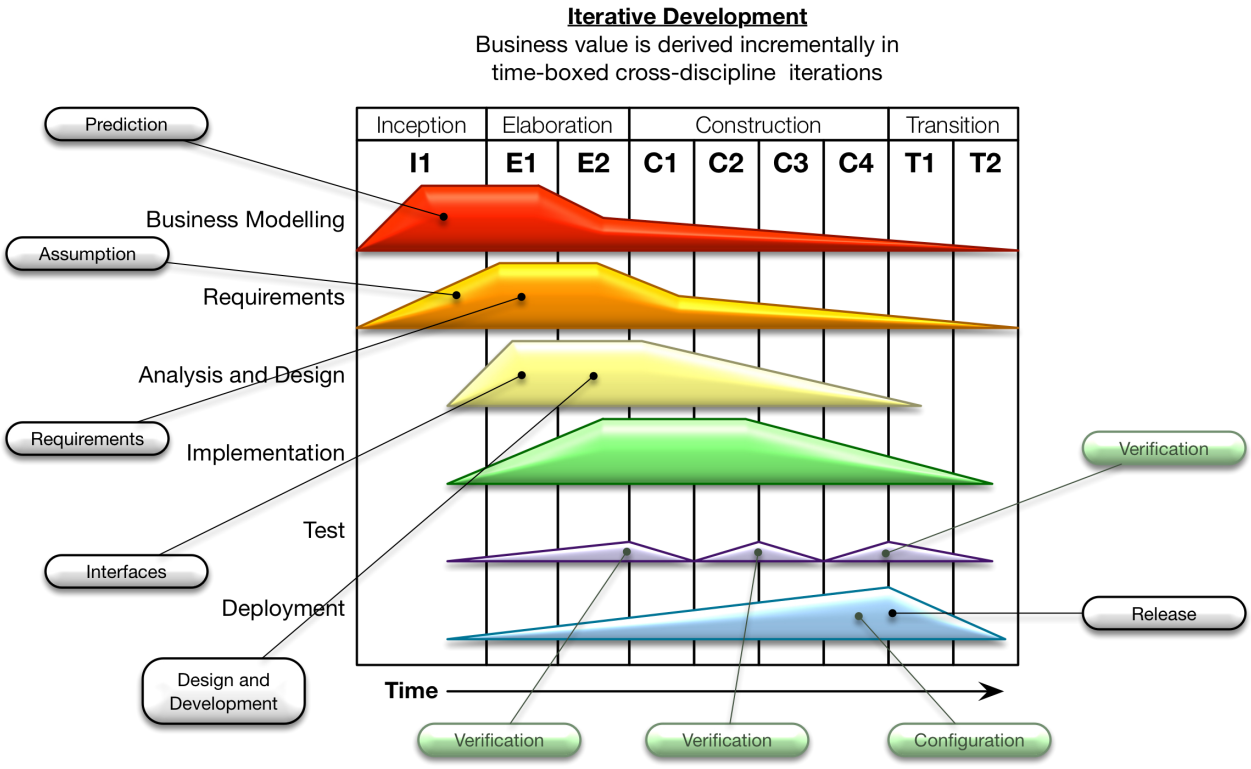

Figure 11.2: Development lifecycle

The model itself may vary depending on the specific needs of the project but the diagram illustrates that different data categories become significant at different points of the process.

Operational Once a system has been developed it will move into an operational lifecycle or indeed, if data safety has not previously been considered for an enterprise, then the system could already be in operational use. These operational lifecycles tend to be cyclical in nature; the diagram in Figure 11.3[2] illustrates a typical model.

[1] The diagram is based on International Business Machines Corporation (IBM)'s Rational Unified Process, an iterative software development process framework. The original diagram is in the public domain.

[2] ITIL is a registered Trade Mark of AXELOS Limited. All rights reserved.

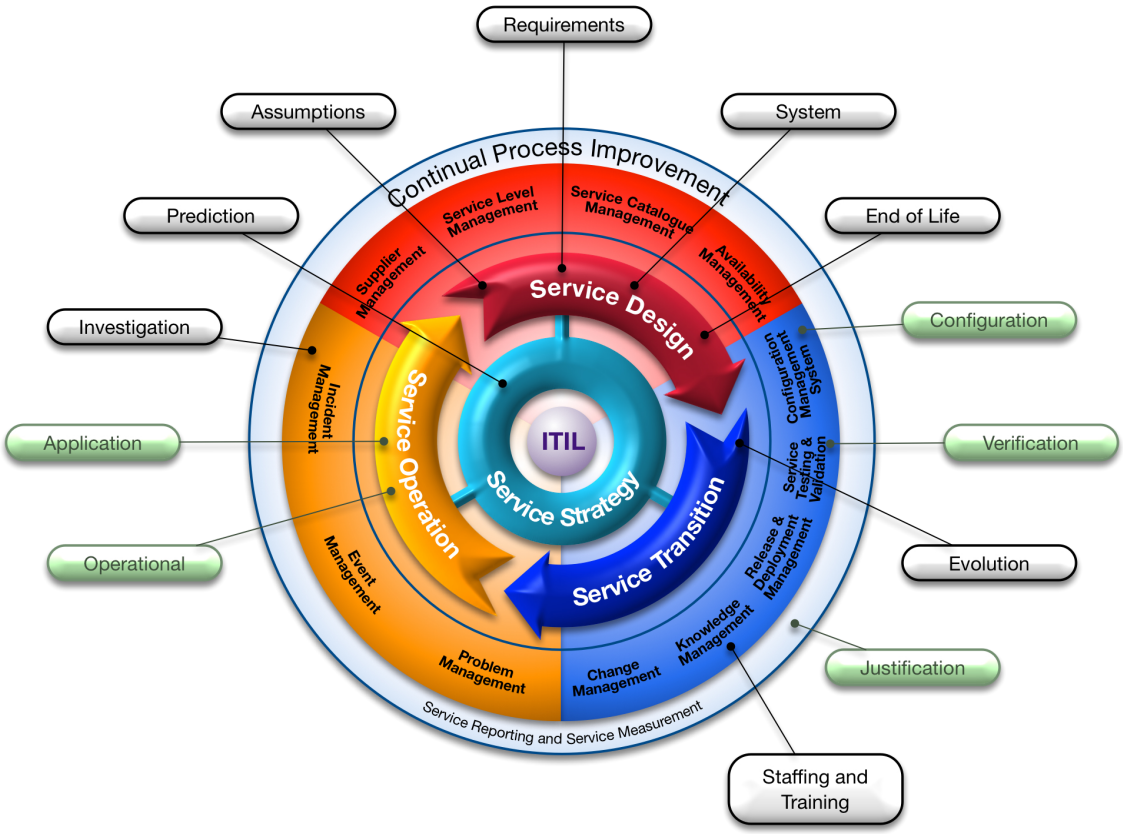

Figure 11.3: Operational lifecycle

Again, specific data will come into play at different periods in the process. Documenting the relationship between process steps and data categories will therefore give clarity as to when a particular assurance technique needs to be applied.

Data supply chains The previous models relate to typical system supply and operate perspectives but there are also other data supply chains where a number of organizations engage in the procurement and use of safety-related data. These processes may include the development and operational lifecycles but a different model is required to fully represent the wider processes that are being employed. The diagram in Figure 11.4 shows such a model representing a data acquisition lifecycle.

This model represents the interactions between three key organizations:

- The commissioning user: the organization that has the need for the data;
- The data provisioner: the organization that will fulfil that need for data; and
- The data acquirer: the organization employed by the Data Provisioner to carry out physical collection of data.

Note that these may be three separate organizations, or they may be separate business units within the same, larger, organization.

In this supply chain, the commissioning user is a consumer of the data and the data acquirer is a producer of data. The data provisioner acts both as a consumer (from the data acquirer) and producer (to the commissioning user) of data. Similarly, an organization that augments datasets is both a consumer and producer of data in the supply chain.

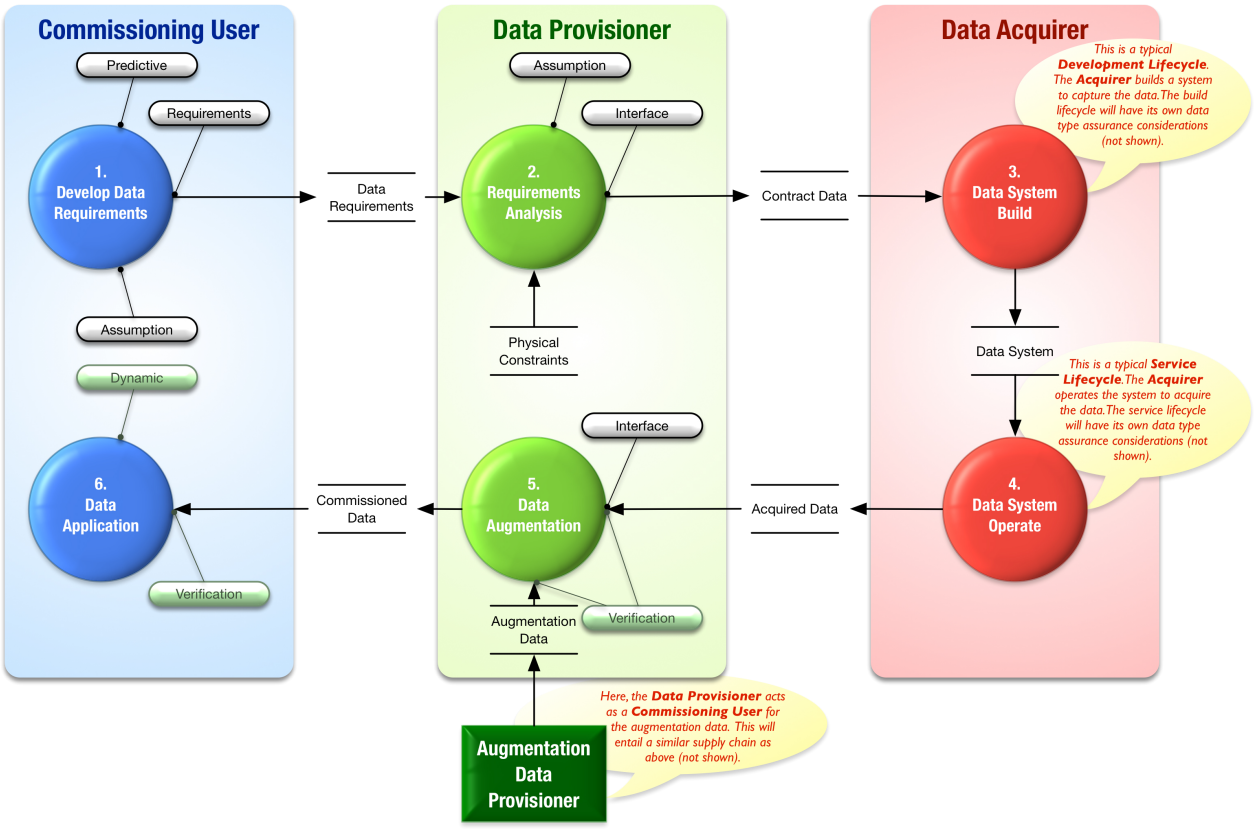

Figure 11.4: Data supply chain

The commissioning user requirement analysis is the key process step where the commissioning user's expectations for data are agreed with the data provisioner. The requirements may be adjusted because of physical constraints (e.g., loss of precision because of physical measuring device constraints) and may include additional requirements to augment the captured data with additional information (e.g., airport codes added to a measurement of a given runway length).

The data provisioner may employ a data acquirer to capture the data (e.g., to carry out a physical survey of a site). The acquisition phase may itself require a specialised system to be built to perform the capture and data refinement to meet the data provisioner's specifications. Such systems will then themselves be subject to the development lifecycle model considerations discussed above. Likewise, the data augmentation phase may require further system development processes or indeed, could trigger an instance of the model again as the data provisioner acting as a Commissioning User.

Acquired and augmented data is then fed into the operational system that has been built for providing the service of generating the commissioned data. This system in its service provision role would then typically follow the operational lifecycle process model discussed earlier.

11.2.1 Tool Assurance

Tools in this context are considered anything that automates all or part of a process, for example, data creation or data transformation. Test tools are also included (i.e., the term is not limited to parts of an operational system).

Tools can impact data safety in different ways, depending on both their function and how they are to be

used. For tools to be considered fit for purpose it is necessary to show that the tool meets its requirements in the context in which it is to be used. The activity to ensure a tool is fit for purpose is usually called "tool qualification".

The first step is to define the purpose for which the tool is required to be fit. Once that is done, and the tool's requirements are specified, there are three main strategies available for qualification:

- Use evidence of a previous certification of the tool by a trusted third party (unlikely to be available in most industry sectors);
- Base tool qualification on the practices used when designing and developing the tool (only practicable for tools developed within the organization); and
- Use one of the available industry-specific guidance documents that admit COTS solutions, e.g., EUROCAE Document ED-215 (RTCA/DO-330) [11].

Further details on tool qualification are presented in Appendix **??**.

11.2.2 Test Data

The generation of suitable test data is critical to verification of a safety system. The test data must include both representative "normal" values based on intended usage and also values which push at, and beyond, normal use to provoke hazards that the system might produce. This latter type of test data is particularly hard to generate; generally it must be credible, yet it must stress the system to react in a way that the preservation of safety properties can be assessed.

In general, all the properties of the test data should be considered and an assessment made as to whether breaking a property (e.g., introducing corrupt or late data) would cause a problem to the system. If it does, then specific test data should be produced to facilitate testing of this potential problem.

Some suggestions for test data for safety-related systems are:

- Use of values on or around boundaries;
- Use of extreme values, way beyond what could be reasonably expected;
- Use of typical "everyday" values / sets;
- Some realistic but unexpected values;
- Try combinations of data values or data items that are problematic together (e.g., inconsistent);
- If possible, use some values known to have caused problems in the past;
- Where appropriate, use values related to timing, rollover or date boundaries;
- Where possible, use white box values (i.e., those derived from an understanding of the system);
- Use a set of values with drift or bias over time;
- Use datasets with particular statistical properties (e.g., distribution, patterns etc.);
- Use data which has incorrect formatting, ordering, or out of sequence, etc.; and
- Try data which has repeated sets of values or pseudo-random characteristics.

Typically very complex test data is derived from recorded live feeds of real data flows. While this data can be extremely useful for regression purposes, it should be recognised that it is unlikely to contain many outlying or boundary data items. Therefore it may need to be modified to test any hazardous situations; this modification can be difficult and may require sophisticated tools to both ensure correct properties and injection of the intended faults (for instance to introduce a statistical bias to the data).

Simulator / emulator derived values can be useful, but again the issue is how realistic the values are: often the accuracy, resolution or timing of simulated values may be different to real data.

Coverage with test data is something to consider. Sometimes the same dataset is used for multiple test scenarios, when in fact it is not stressing all of them to the same degree. Test data coverage can be collected over requirements, code or design, but it is important not to forget hazards: coverage of the hazards and mitigations identified in the hazard log is a key aim.

In general some measure of the quality and suitability of the test data can be useful. This could be based on statistical properties, coverage of hazards or coverage of requirements.

Test data must show continued relevance, through systems evolution and over time. It is good practice to build up extensive regression suites containing coverage of all detected problems to date.

11.2.3 Interfaces with Existing Assessments

11.2.3.1 Data and Software

Although most people feel they have an intuitive understanding of the difference between software and data, upon closer examination the boundary is not always as clear as it may first appear.

Consider, for example, Java bytecode, which is operated on by a Java virtual machine. From one perspective, it could be argued that the Java bytecode is simply data. By extension, it could also be argued that the Java source code is also just data. This type of argument can be extended to suggest that any software can, at least from one viewpoint, be considered as data. Conversely, think about the data used in a 3D printer, perhaps to produce a part for an aircraft. This data could be viewed as a program for the printer; that is, it could potentially be viewed as software. This type of argument can easily be extended across a range of situations, especially those relating to configuration data.

While they are interesting, and potentially important, these philosophical considerations should not detract from the practical issue: there are some aspects of data (using the term in a generic sense) that are often not explicitly addressed in standards. These are a consequence of features that are more readily apparent in data than in software. Examples include:

- It is easy for data to be reused in a range of contexts and despite appearances it is not trivial to translate an assurance argument that the data is fit for purpose from one context to another.
- It is not always clear who owns or is responsible for data, especially when data is shared and processed amongst a collection of disparate systems.
- Data often has a lifetime, that is a time after which it is no longer valid. This may be a strict cut off, or a more gradual degradation in the utility (or applicability) of the data.
- There is often a default value for data. While this can make systems easier to use and hence more productive it can be difficult to identify a default value that is appropriate for all circumstances.

- It can be easy to change data. In some circumstances this can give rise to a temptation to make uncontrolled and potentially untested changes. It can also allow data to be fraudulently changed after an accident.

In summary, data and software are closely related and, as such, need to be considered together in system engineering activities, including system safety analyses. However, data and software emphasise different facets of risk and they are susceptible to different mitigation approaches; this means there is also a need to adopt a data-focused perspective. It also means that software assurance levels cannot be mapped directly to DSALs.

11.2.3.2 Data Safety and Security

When generating high-level processes and techniques to manage the risks posed by data, it is worthwhile understanding the difference between the safety risks posed by accidental failure to preserve Data Properties and the security risks posed by actors maliciously undermining the properties of data.

The relationship between safety and security, as engineering concepts, can be summarised by their relationships to cultural, developmental and aspirational properties of systems development.

Culturally, embedding both safety and security into an organization is seen as a key strategic goal for creating systems that are both safe and secure. Developmentally, safety and security are quality factors, generating transverse requirements that impact the entire system. Most importantly, at the aspirational level, both safety and security have the common goal of preventing harm from accidental and malicious interventions respectively.

For an organization aiming to create systems that are both safe and secure, these connections can be both a benefit and a burden. The shared goal of preventing harm means that both quality factors seek to identify routes to harm through analysis of the system being developed. This can result in shared processes and tools, which in turn can save time and money during systems development. However, safety and security interact in a more volatile way at the functional level. Security failings can undermine the safety case for a system and, conversely, safety requirements can prevent the implementation of standard security solutions. For example, the German government published a report in 2014 into a fire at a steel works caused by a cyber attack that resulted in the control system being placed into an unsafe state and the safety system being unable to intervene (Section 3.3.1 of [12] - in German). In addition, "fail-safe" states can often leave a system with exposed security vulnerabilities.

These links between safety and security infer that there are connections between the sub-categories of data safety and information security: both attempt to take a data-centric view of the system of interest in order to improve the associated quality factor; and both attempt to prevent harm through the preservation of the properties of data within that system.

In the security domain, the three key properties of data considered are confidentiality, integrity, and availability. Confidentiality, the failure of which is termed "information disclosure" in the Microsoft Security Model, [13] is typically not a safety concern as, without malicious intent, information sharing is not inherently unsafe. However, when considering systems where confidentiality is an important property, the interaction between data safety and security cannot be trivially resolved. For example, accidental disclosure of information can form part of a causal chain which leads to harm from a malicious actor.

Data integrity is a critical property for both domains. The Microsoft Security Model describes malicious removal of the property of integrity as "tampering". Whether by accident or through malicious intent, the

potential harm from loss of data integrity can be disastrous to a safety-critical system, from the values of drug dosages to control system parameters.

Data availability is also important to both domains. Loss of availability, or "denial of service" in the Microsoft Security Model, is another property that can be lost accidentally or through malicious intervention. Loss of availability prevents systems from functioning properly and can result in undefined behaviour if not mitigated by design.

Further guidance on the integration of safety and security can be found in a code of practice published by the IET [14]. The Code of Practice is written for engineers and engineering management to support their understanding of the issues involved in ensuring that the safety responsibilities of an organization are addressed, in the presence of a threat of cyber attack.

12 Data Migrating, Porting, Importing and Exporting (Informative)

> *I don't think you'll ever have a perfect world because we humans are prone to error, and so we're always in search of an upgrade*
> ***Henry Rollins***

Migrating, Porting and Import or Export of data between systems is a large and complex topic. This appendix highlights some of the issues that arise, links to the data safety properties and suggests some mitigations.

12.1 Introduction and rationale

All of these activities involve moving data around, with the expectation that the data will be re-used or re-purposed to some extent after the move. Why is this a problem? Because data can be lost, transformed or misinterpreted due to the migration. If some of this data is safety-related then there is the possibility of creating a data hazard. The worst cases are the silent ones: where data is not migrated or modified or changed into something else on migration, and no notification or warning is provided. In general porting of data, import and export actions can be considered subsets of full data migration activities. The following are typical stages of a data migration exercise:

i) Discovery: Identifying what data has to be migrated: the history, locations, store types, formats and the software currently used to manipulate it.

ii) Analysis and Preparation: Filtering and sorting the data, identifying and cleansing the data, fixing any issues, creating missing data that can't be moved, and transforming the data to make it suitable for use in the new system.

iii) Trials: Selected sets or portions of the data are migrated and the results assessed within the new system or context. This may be extended to include more data through incremental, phased stages.

iv) Dry Run: A full migration is performed to a duplicate system or to the new system in a way that it can be backed out (ie. the new system can be reverted to its original state), and use of the current system can be continued if necessary.

v) Parallel Running: Both old and new systems continue running, processing the same data where possible. This is desirable as it allows quick reversion to the old system in case of failure. It also allows (sampling) comparison between the old and new systems outputs.

vi) Monitoring and Checking: Operation of the new system is closely monitored and outputs assessed to establish if these would be similar to those produced by the old system.

vii) Move to New: This is where the old system is switched off and mothballed, and may be subsequently decommissioned and removed.

It is noted that migration is generally between systems (e.g. a legacy system and a new one) but can be between different instances or environments within the same system, involving different software, stores or databases.

Many moves of data are now to Cloud storage, and the implementation may be largely hidden, with only software API interfaces available[1]. In this case the options for comparisons may be more limited.

[1] It is noted that within a cloud environment mini automated migrations of data are happening all the time as new servers and

12.2 Migration Cases

There are at least three cases of migration (these simple diagrams are intended to show the end state of the migration, it is acknowledged that there are intermediate stages, as mentioned in Section 12.1):

1. Migration of data from a single old system to a single new system. This is very common and is the usual upgrade path. The new system, or indeed both systems, may be in the Cloud.

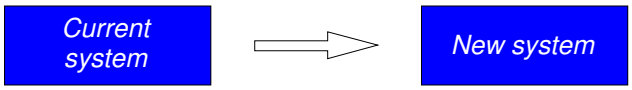

Figure 12.1: Simple migration path

2. Migration of data from multiple disparate systems to one new system. This is less common but can be seen in major technology upgrades where it is seen as beneficial to bring together multiple legacy systems into one new system.

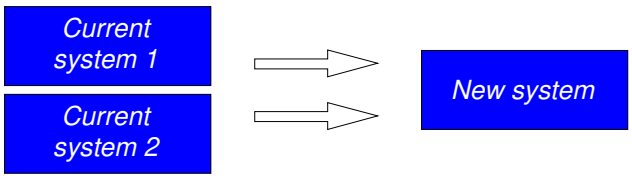

Figure 12.2: Many-to-one migration path

3. Migrating data back to an old system in case of failures or to investigate data issues with the new system

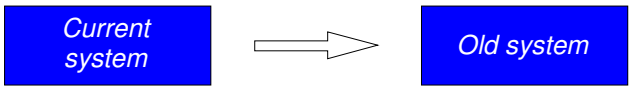

Figure 12.3: Reversion migration path

12.3 Safety issues due to migration

The following table summarises some of the safety issues due to migration, maps these to the data safety properties and gives possible mitigations.

storage are provisioned by the service provider, software is upgraded, etc. There should be some sort of assurance provided in this case to ensure the data is preserved as required.

Table 12.1: Safety issues due to migration

No.	Migration Case	What are safety issues in each case?	Mapping to properties/loss of properties	Possible Mitigations (Table 3.12)
1.	Data format may not map exactly: due to different database formats, word lengths, endian issues, etc.	Data can be lost or discarded, or substituted for defaults. Data may be migrated / imported / exported incorrectly.	Integrity, completeness, format	DM.02, DM.07, DM.08, DM.09, DM.10
2.	Data may not translate exactly: no equivalent data type / record / field type in new system, etc	Data can be lost or discarded, or substituted for defaults. Data may be migrated / imported / exported incorrectly.	Integrity, completeness	DM.02, DM.04, DM.07, DM.08, DM.09, DM.10
3.	Data may not be valid: less numeric range available in new system, etc. Subtype of (2).	Data can be lost or discarded, or substituted for defaults. Data may be migrated / imported / exported incorrectly.	Integrity, completeness	DM.02, DM.04, DM.07, DM.08, DM.09, DM.10
4.	Metadata (data about the data) may be lost, including data dictionary aspects, derivation, history, signoffs, authorisations, logs, etc.	Data may be migrated / imported / exported losing information.	Traceability, history	DM.04, DM.09
5.	Data may have a different meaning or interpretation in the new system context. E.g. if a fluid value is 'gallons' and this is migrated from a UK to a US-developed system without applying an adjustment factor.	Meaning of data can be altered even if migration apparently successful.	Consistency, fidelity / representation	DM.01, DM.04, DM.11
6.	Data may be correctly rejected (detection of error on import / export)	Data is lost. Not necessarily a problem if notified and fixed.	Completeness	DM.07, DM.08, DM.09, DM.10
7.	Data may be wrongly filtered or rejected as incompatible	Correct data is rejected. If data can't be imported what should happen? a. Rejected and migration halted, b. Rejected and notified, c. Rejected silently, d. Substituted?	Integrity, completeness	DM.02, DM.03, DM.07, DM.08, DM.09, DM.10

Continued on next page

Table 12.1: Safety Issues due to Migration (continued)

No.	Migration Case	What are safety issues in each case?	Mapping to properties/loss of properties	Possible Mitigations (Table 3.12)
8.	Two or more data items may map to the same item in the new system – one may overwrite the other with the last one 'winning'.	This is a case of aliasing (See 6.2.2 ref). Data can be lost or discarded. Data may be migrated / imported / exported incorrectly.	Integrity, completeness	DM.02, DM.03, DM.07, DM.08, DM.09, DM.10
9.	There may be situations where the data store has to be 'live' and the migration / import / export process takes some time. In this case some data may be out of date and become inconsistent (stale), and integrity of the whole dataset is lost.	Data may be migrated / imported / exported incorrectly.	Consistency, integrity, timeliness	DM.07, DM.08
10.	If a migration is unsuccessful, data changes may have to be undone and stores reverted to original state. This may not always be possible or completely done.	Original dataset may be corrupted.	Consistency, Integrity	DM.06, DM.07, DM.08
11.	The data 'cleansing' activity conducted before or during migration may introduce faults.	The cleansing activity may miss some data faults or may be over-zealous, deleting good values.	Integrity, completeness	DM.07, DM.09, DM.13

13 DSAL Customisation (Informative)

> *One size does not fit all*
> ***Frank Zappa***

13.1 Introduction

Within the body of this document, a method has been provided for the assignment of DSALs. However it is anticipated that, as with the methods used for the assignment of safety criteria within system safety analysis, some programmes may find it desirable to develop alternate approaches. This section presents some possible methods for the determination of likelihood. These methods are intended to serve as approaches for consideration when developing project-specific criteria, as these approaches are less generic than that presented through Table 3.5, so it is highly likely that customisation will be required.

13.2 Significance factors

A method of determining likelihood from the different characteristics is to implement a scoring scheme that apportions a quantitative value for each of the characteristics, with the sum giving the total likelihood score. The total score is then compared against a scale, and that in turn determines the overall low, medium or high assessment.

For example, consider the modified version of Table 3.5 illustrated in Table 13.1

Table 13.1: Calculation of likelihood – option 1

	Score		
	2	**1**	**0**
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.
Prevention	Difficult or impossible to guard / barrier against errors.	Possible to guard / barrier against errors.	Easy to guard / barrier against error.
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.
Correction	Difficult or impossible to correct or workaround errors.	Possible to correct or workaround errors.	Easy to correct or workaround errors.

For each of the characteristics, the applicable assessment of likelihood is then selected and scored. So for example, if it is "Possible to guard / barrier against errors" for Prevention, then the score for that characteristic would be 1. Each characteristic is then scored and they are then summed to give a total score. The range of possible values will run from 0 (all choices in the far right column, favouring low likelihood aspects) through to 10 (all choices in the far left column, favouring high likelihood aspects).

The overall likelihood in the assessed against a scale such that shown in Table 13.2.

Table 13.2: Likelihood assessment

Low	Medium	High
0–2	3–6	>6

13.3 Weighted characteristics

The method presented in section 13.2 may be enhanced to provide a more holistic overall assessment of the likelihood based on all characteristics. Note that the scheme in section 13.2 allocated equal significance to each of the characteristics. This method could be further refined if necessary to apply weightings to each of the characteristics, as shown in Table 13.3.

Table 13.3: Calculation of likelihood – option 2

	Weighting	Score		
		2	1	0
Proximity	1.5	**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	1.0	Data is completely relied upon.	**Data is indirectly relied upon.**	Little reliance on data.
Prevention	1.3	**Difficult or impossible to guard / barrier against errors.**	Possible to guard / barrier against errors.	Easy to guard / barrier against error.
Detection	0.7	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.**
Correction	0.5	Difficult or impossible to correct or workaround errors.	Possible to correct or workaround errors.	**Easy to correct or workaround errors.**

The total score is then calculated by first multiplying the individual characteristic's score by the weighting before summing all the values. For example, if the selections highlighted in bold were made, then the score would be

$$(1.5 \times 2) + (1.0 \times 1) + (1.3 \times 2) + (0.7 \times 0) + (0.5 \times 0) = 6.6$$

resulting in a *High* assessment rather than the *Medium* that would result with no weighting. This is because Proximity and Prevention are considered (in this particular example) more important than Detection and Correction. The weightings within a project-specific version of this table would be chosen by the organization to suit the particular scenario under consideration.

14 Acronyms, Definitions and Glossary (Discursive)

> *The plural of anecdote is not data.*
> ***Mark Berkoff***

14.1 Acronyms

AAL	above aerodrome level
AI	artificial intelligence
AoA	angle of attack
ARQ	automatic repeat-request
ATSB	Australian Transport Safety Bureau
BIT	built in test
BITE	built in test equipment
CCRC	Criminal Cases Review Commission
CFIT	controlled flight into terrain
CoD	certificate of design
COTS	commercial off-the-shelf
CRC	cyclic redundancy check
CSV	comma separated variable
CT	computed tomography
DME	distance measuring equipment
DoC	Department of Corrections
DRACAS	defect reporting and corrective action system
DSAL	data safety assurance level
DSG	data safety guidance
DSIWG	Data Safety Initiative Working Group
DSMP	data safety management plan
ECS	electronic chart system
ECU	electronic control unit

EDM	entry demonstrator module
EGPWS	enhanced ground proximity warning system
EHR	electronic health record
ENC	electronic navigational chart
ESA	European Space Agency
FAA	Federal Aviation Administration
FDAL	functional design assurance level
FMGS	flight management guidance system
FMS	flight management system
GCS	ground control station
GTOLS	GPS take-off and landing system
HAZOP	hazard and operability study
HUMS	health and usage monitoring system
IBM	International Business Machines Corporation
ICAO	International Civil Aviation Organization
ICD	interface control document
IDAL	item development assurance level
IMC	instrument meteorological conditions
IMU	inertial measurement unit
IP	internet protocol
IRS	inertial reference system
ISO	International Standards Organization
IV	intravenous
LHR	London Heathrow airport
LLM	large language model
MCA	Maritime and Coastguard Agency
MCAS	manoevring characteristics augmentation system
ML	machine learning
MSAW	minimum safe altitude warning
NaN	not a number

NOTAM	notice to airmen
ODR	organizational data risk
OOW	officer of the watch
PCR	polymerase chain reaction
POCL	Post Office Counters Limited
RDA	radar Doppler altimeter
SAR	search and rescue
SCSC	Safety-Critical Systems Club
SIL	safety integrity level
SMP	safety management plan
SOP	standard operating procedure
SPM	subpostmaster
SSS	Safety-critical Systems Symposium
TCAS	traffic collision avoidance system
UAS	unmanned air system
VMC	visual meteorological conditions
VMS	voyage management system
VOR	VHF omnidirectional range
XML	extensible markup language

14.2 Definitions and Glossary

accessibility

Property that the data is visible only to those that should see it.

accuracy

- Closeness of agreement between a test result and the accepted reference value. Note that a test result can be observations or measurements. ISO 19113:2005 [**?**]
- A degree of conformance between the estimated or measured value and the true value. (EU) No 73/2010 [**?**]
- (Temporal) Correctness of the temporal references of an item (reporting of error in time measurement). Correctness of ordered events or sequences, if reported. Validity of data with respect to time. ISO 19138:2006 [**?**]

adaptation data

Data used to customise elements of the system for their designated purpose. Adaptation data is used to customise elements of the system for its designated purpose at a specific location. These systems are often configured to accommodate site-specific characteristics. These site dependencies are developed into sets of adaptation data. Adaptation data includes data that configures the software for a given geographical site, and data that configures a workstation to the preferences and / or functions of an operator. Examples include, but are not limited to:

Geographical Data: latitude and longitude of a radar site.

Environmental Data: operator selectable data to provide their specific preferences.

Airspace Data: sector-specific data.

Procedures: operational customisation to provide the desired operational role.

Adaptation data may take the form of changes to either database parameters or take the form of pre-programmed options. In some cases, adaptation data involves re-linking software code to include different libraries. Note that this should not be confused with recompilation in which a completely new version of the code is generated. Based on ED-153 [**?**]

aeronautical data

- A representation of aeronautical facts, concepts or instructions in a formalized manner suitable for communication, interpretation or processing. (EU) No 73/2010 [**?**]
- Data used for aeronautical applications such as navigation, flight planning, flight simulators, terrain awareness, and other purposes. RTCA/DO-178C [**?**]

analysability

The data (including any metadata) is of a suitable size, type and format to enable it be usefully analysed

artefact, data

(Normative) An item, or collection of items, that provides a useful perspective on data generated, processed or consumed by a system.

assurance level, software

An indication of how much assurance is required (commensurate to risk) before deploying software into an operational system. J Spriggs, based on (EC) No 482/2008 [**?**]

availability

The property of being accessible and usable upon demand by an authorized entity. ISO 27001:2013 [**?**]

completeness

- Property of having every necessary part or element.
- Completeness of the data provided. RTCA/DO-200A

confidentiality

The property that information is not made available or disclosed to unauthorized individuals, entities, or processes. ISO27001:2013 [**?**]

configuration data

- Data that configures a generic software system to a particular instance of its use. (EC) No 482/2008 [**?**]
- Data that configures a generic software system to a particular instance of its use (e.g., data for flight data processing system for a particular airspace, by setting the positions of airways, reporting points, navigation aids, airports and other elements important to air navigation). ED-153 [**?**]Data that configures a generic software system to a particular instance of its use (e.g., data for flight data processing system for a particular airspace, by setting the positions of airways, reporting points, navigation aids, airports and other elements important to air navigation). ED-153 [**?**]

consistency, data

The property that the data adheres to a common world view (e.g., units).

continuity, data

The property that the data is continuous and regular without gaps or breaks.

correctness, data

self-consistency, protection against alteration or corruption and consistency with the functional requirements of the data-driven system. IEC 61508 Part 3 [**?**]

criticality, data

Classification of data by the potential effect of erroneous data on the expected operation that is supported by that data. RTCA/DO-200A [**?**]

data

- A thing given or granted; something known or assumed as fact, and made the basis of reasoning or calculation; an assumption or premiss from which inferences are drawn. Oxford English Dictionary (OED)
- A reinterpretable representation of information in a formalized manner suitable for communication, interpretation or processing. ISO/IEC 2382 [**?**]

data dictionary

The detailed description of data, parameters, variables, and constants used by the system. RTCA/DO-178C

database

A set of data, part or the whole of another set of data, consisting of at least one file that is sufficient for a given purpose or for a given data processing system. RTCA/DO-178C

data-driven system

System which relies upon configuration data or lookup tables to define the functionality of the system. IEC 61508 Part 4 [**?**]

dataset

Identifiable collection of data. Note that a dataset may be a smaller grouping of data which, though limited by some constraint such as spatial extent or feature type, is located physically within a larger dataset. Theoretically, a dataset may be as small as a single feature or feature attribute contained within a larger dataset. A hardcopy map or chart may be considered a dataset. BS EN ISO 19131:2008 [**?**]

disposability / deletability

The property that the data can be permanently removed when required

error, data

- Discrepancy with the universe of discourse. ISO 19138:2006 [**?**]
- Discrepancy between a data value and the true, specified or theoretically correct value or condition. P. Ensor [**?**]

explainability

The property that the data can be meaningfully explained, by a suitable mechanism, to those who need to understand it.

fidelity / representation ,data

The property describing how well the data maps to the real world entity it is trying to model.

format

The property that data is represented in a which is readable by those that need to use it.

Goldilocks

The property that the data is just the right size – not to much and not too little

hazard log

A record of all hazard analysis, safety risk assessment and safety risk reduction activities for the "whole-of-life" of a safety-related system.

hazard, data

Use of data (in the context of a system) that could lead to harm. SCSC DSIWG

history

Property that the data has an audit trail of changes.

information

- Knowledge communicated concerning some particular fact, subject, or event; that of which one is apprised or told - intelligence, news - as contrasted with data. OED
- Knowledge that has a contextual meaning. ISO/IEC 2382 [**?**]

information, aeronautical

Information resulting from the assembly, analysis and formatting of aeronautical data. (EU) No 73/2010 [**?**]

integrity, data

- The assurance that a data element retrieved from a storage system has not been corrupted or altered in any ways since the original data entry or latest authorised amendment. RTCA/DO-200A [**?**]
- The degree of assurance that a data item and its value have not been lost or altered since the data origination or authorised amendment. (EU) No 73/2010 [**?**]

- The degree of undetected (at system level) non-conformity of the input value of the data item with its output value. (EU) No 1207/2011 [**?**]
- The property of protecting the accuracy and completeness of assets, i.e., that which has value to the organization. ISO 27001:2013 [**?**]

intended destination / usage

Property that the data is only sent to those that should have access to it

item, data

Single attribute of a complete dataset, which is allocated a value that defines its current status. (EU) No 73/2010 [**?**]

lifetime

The property of when the safety-related data expire

metadata

Data that represents information about data itself. Note that one should distinguish between "Structural Metadata", which is data about the design and specification of data structures (and is more properly called "data about the containers of data") and "Descriptive Metadata", which is about individual instances of application data, the data content. J. Inge [4]

mitigation

Steps taken to control or prevent a hazard from causing harm and reduce risk to a tolerable or acceptable level. [**?**]

owner, data

(Normative) The individual or organization responsible for a particular data artefact or collection of data artefacts.

priority

The property that data is presented / transmitted / made available in the order required.

property, data

(Normative) A characteristic that can be exhibited by a data artefact.

quality, data

- A degree or level of confidence that the data provided meet the requirements of the user. These requirements include levels of accuracy, resolution, DSAL, traceability, timeliness, completeness, and format. RTCA/DO-200A [**?**]
- A degree or level of confidence that the data provided meets the requirements of the data user in terms of accuracy, resolution and integrity. (EU) No 73/2010 [**?**]

resolution

- The ability of a device to respond to small differences in input and to indicate or represent them accurately in output; a measure of this, expressed as the smallest difference so distinguishable. OED.

- The smallest difference between two adjacent values that can be represented in a data storage, display or transfer system. RTCA/DO-200A [**?**]
- The number of units or digits to which a measured or calculated value is expressed and used. (EU) No 73/2010 [**?**]

response

(Normative) The way in which an identified risk is addressed; possible responses include avoid / eliminate, treat, or accept as sufficiently low.

safety assessment, data

(Normative) The process of explicitly considering data as part of a system safety assessment, via the means of data artefacts, Data Properties and DSALs.

safety assurance level, data

(Normative) An indication of the level of rigour with which relevant data properties should be demonstrated for appropriate data artefacts.

safety requirement, data

(Normative) A requirement to implement an approach specifically designed to achieve, maintain or demonstrate a data property (or data properties) for a given data artefacts (or Artefacts).

sequencing

The property that the data is preserved in the order required.

stakeholder

(Normative) An individual or organization that has some relationship to the system, possibly including a power of veto.

suppression

Property that the data is intended never to be used again

tailoring

Adaptation of processes etc. to be appropriate for a specific system and context

timeliness

- A measure of the time delay between a change in the real world and the associated database update being available to the user. P. Ensor [**?**]
- The difference between the time of output of a data item and the time of applicability of that data item. (EU) No 1207/2011 [**?**]

trace data

Data providing evidence of traceability of development and verification processes software lifecycle data data without implying the production of any particular artefact. Trace data may show linkages, for example, through the use of naming conventions or through the use of references or pointers either embedded in or external to the software lifecycle data. RTCA/DO-178C [**?**]

traceability

Ability to determine the origin of the data. RTCA/DO-200A [**?**]

treat

To apply a treatment

treatment

(Normative) An action taken to reduce or control risk.

validation, data

- The activity whereby a data element is checked as having a value that is fully applicable to the identity given to the data element, or a set of data elements that is checked as being acceptable for their purpose. RTCA/DO-200A [**?**]
- Process of ensuring that data meets the requirements for the specified application or intended use. (EU) No 73/2010 [**?**]

validity, period of

Period between the date and time on which information becomes available and the date and time on which the information ceases to be effective. Based on (EU) No 73/2010 [**?**]

verifiability

Evaluation of the output of an aeronautical data process to ensure correctness and consistency with respect to the inputs and applicable data standards, rules and conventions used in that process. (EU) No 73/2010 [**?**]

verification, data

Evaluation of the output of a process to ensure correctness and consistency with respect to the inputs and applicable data standards, rules and conventions used in that process. Based on (EU) No 73/2010 [**?**]

The normative list of definitions is at section 2. Normative definitions have been repeated here for convenience.

This page is intentionally blank

15 References (Discursive)

> *Data opportunities multiply as the data is transformed.*
> ***Sun Tzu misquoted***

This page is intentionally blank

Bibliography

[1] Risk Management — Guidelines. Standard ISO 31000:2018, International Standards Organisation, February 2018. Second Edition.

[2] Vocabulary.com. homonym vs. homophone vs. homograph : Choose your words. URL https://www.vocabulary.com/articles/chooseyourwords/homonym-homophone-homograph/. Accessed: 17 January 2023.

[3] Alastair Faulkner and Mark Nicholson. *Data-Centric Safety: Challenges, Approaches, and Incident Investigation*. Elsevier, June 2020. ISBN 978-0128207901. Available from https://www.amazon.co.uk/dp/0128207906?ie=UTF8&n=341677031. Accessed 30 January 2021.

[4] J Inge. *Improving the Analysis of Data in Safety-Related Systems*. University of York, September 2008.

[5] Hazard and operability studies (HAZOP studies) — Application guide. Standard IEC 61882:2016, International Electrotechnical Commission, March 2016. Second edition.

[6] P Koopman, K Driscoll, and B Hall. *Selection of Cyclic Redundancy Code and Checksum Algorithms to Ensure Critical Data Integrity*. US Department of Transportation, 2015.

[7] The Honourable Mr.J̃ustice Frazer. Alan bates and others -v- post office limited: Technical appendix to judgement (no.6) "horizon issues". URL https://www.judiciary.uk/wp-content/uploads/2019/12/bates-v-post-office-appendix-1.pdf, 2019. Accessed: 30 January 2024.

[8] Karl Flinders. Post office horizon scandal explained: Everything you need to know. *Computer Weekly*, 2024. Accessed: 30 January 2024.

[9] The ccrc and post office/horizon cases. URL https://ccrc.gov.uk/news/the-ccrc-and-post-office-horizon-cases/, 2024. Accessed: 30 January 2024.

[10] The Honourable Mr.J̃ustice Frazer. Alan bates and others -v- post office limited: Judgement (no.6) "horizon issues". URL https://www.judiciary.uk/wp-content/uploads/2019/12/bates-v-post-office-judgment.pdf, 2019. Accessed: 30 January 2024.

[11] Software Tool Qualification Considerations. Standard RTCA/DO-330, EUROCAE/ED-215, Radio Technical Commission for Aeronautics / European Organisation for Civil Aviation Equipment, January 2012.

[12] Die Lage der IT-Sicherheit in Deutschland 2014. URL https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2014.pdf, December 2014. Accessed: 27 January 2021.

[13] D LeBlanc and M Howard. *Writing secure code*. Microsoft Press, December 2002.

[14] *Cyber Security and Safety Code of Practice*. IET, 2020. ISBN 978-1-83953-318-8.

This page is intentionally blank

16 Acknowledgements (Discursive)

> *Our ability to do great things with data will make a real difference in every aspect of our lives.*
> ***Jennifer Pahlka***

The document contributors would like to thank:

- The SCSC.
- The SCSC Covid-19 Working Group for providing some of the data used in the Covid-19 Appendix.
- Brian Jepson of the SCSC for web hosting support and technical help with the SCSC web site.
- Tim Rowe for editing this edition.
- Paul Hampton and Mark Templeton for managing the publication processes.
- Nick Hales, Mike Parsons, Tim Rowe, Alan Simpson and Mark Templeton for developing the additional text for this edition.
- Martin Atkins and Divya Atkins for driving the development of tooling and promoting data safety.
- Mike Parsons for chairing the Working Group meetings.
- All those who have taken minutes at Working Group meetings.
- All the organisations that have hosted Working Group meetings.
- All the organisations that have provided support to the document's contributors.
- Those that have been unable to attend meetings but have made supporting contributions.

This page is intentionally blank

17 Contributors (Discursive)

> *Without data, you're just another person with an opinion.*
> ***W. Edwards Deming***

This document has had the benefit of contributions from a large number of people, who work for a variety of organisations, which collectively span a range of different sectors. Note that contributions have been made on an individual basis and, in particular, the inclusion of an organisation in the following list does **not** necessarily mean that organisation agrees with the entire contents of the document.

Updates to the most recent version of the document were written by:

- Divya Atkins, Mission Critical Applications
- Martin Atkins, Mission Critical Applications
- Paul Hampton, CGI IT UK Ltd
- Mike Parsons, Ebeni and SCSC
- Tim Rowe, TGR Safety Management Ltd

In addition to the above, contributors to earlier versions upon which this document is based include the following (the organisations listed were correct at the time of their contribution) :

- Mike Ainsworth, Ricardo
- Rob Ashmore, Dstl
- Michael Aspaturian, EDF Energy
- Janette Baldwin, Thales UK
- Dave Banham, Blackberry QNX
- Ian Bingham
- John Bragg, MBDA UK Ltd
- Jennifer Brain, Wood plc
- Eric Bridgstock
- Simon Brown, QinetiQ
- Dermot Martin Burke, BAE Systems
- Dale Callicott, DKCSC Ltd
- John Carter, General Dynamics
- Martyn Clarke, SCSS Ltd
- Steve Clugston, TSC
- Robin Cook, Thales

- Davin Crowley-Sweet, Highways England
- Dijesh Das, AMEC / BAE Systems
- Duncan Dowling, DARD
- Andrew Eaton
- Ashraf El-Shanawany, CRA Risk Analysis
- Paul Ensor, Boeing
- Alastair Faulkner, Abbeymeade
- Ken Frazer, KAF
- Richard Garrett, SQEP
- Paulo Giuliani
- Ian Glazebrook, Atkins
- Rob Green, NATS
- Nick Hales
- Louise Harney, Leonardo
- Ali Hessami, Vega Systems
- David Higgins
- Gordon Hurwitz, Thales
- Pete Hutchison, RPS
- Gavin Jones
- Amira Kawar, Kawar Engineering Consultancy Ltd
- Tim Kelly
- Andrew Kent
- Brent Kimberley, Durham, Canada
- Julian Lockett, Frazer-Nash Consultancy Ltd
- David Lund, David Lund Consultants
- Dave Lunn, Thales UK
- Nasser Al Malki, University of York
- Victor Malysz, Rolls-Royce
- Jim Mateer, SQEP
- John McDermid, University of York
- Paul McKernan, Dstl
- Thor Myklebust, Sintef

- Mark Nicholson, University of York
- Yvonne Oakshott
- Robert Oates
- David Perrin, Virtual PV
- Ashley Price, Raytheon UK
- Andrew Rankine
- Felix Redmill, SCSC
- Sam Robinson, EDF Energy
- Mark Simmonite, Highways England
- Alan Simpson, Ebeni
- Oscar Slotosch, Validas AG
- Dave Smith, Frazer-Nash Consultancy Ltd
- Peter Smith, Highways England
- John Spriggs, NATS
- Carolyn Stockton, BAE Systems
- Mark Templeton, Arcade Experts Ltd
- Andy Williams
- Lesley Winsborrow
- Fan Ye, ESC

This page is intentionally blank

Index of Locations

This page is intentionally blank

Index

DATA IS HERE. DATA IS GROWING. DATA IS CAUSING HARM.

This book has been developed by the Safety-Critical Systems Club Data Safety Initiative Working Group (DSIWG) to provide guidance on how data, as distinct from hardware and software can be managed in a safety-related context.

"If you torture the data long enough, they confess – even to crimes that were never committed."

Nihat Bülent Gültekin

~

This is the seventh minor update since version 3.0. Paragraph numbering within the body of the document remains aligned with that major release. Thus users of any previous 3.x release of the guidance document will find migration to this edition takes little effort.

SCSC

For Everyone Working in System Safety