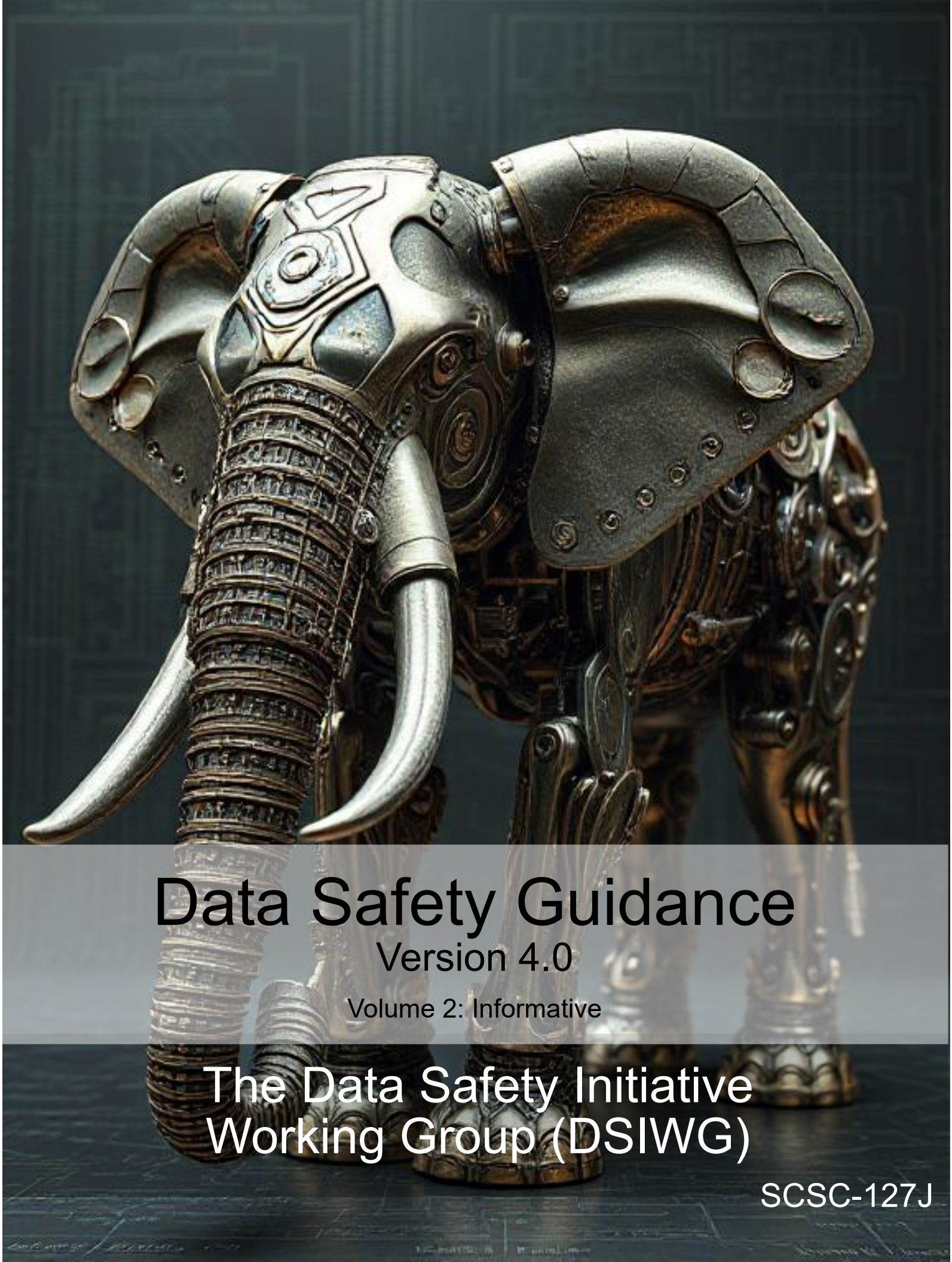


Data Safety Guidance

Version 4.0

Volume 2: Informative

The Data Safety Initiative Working Group (DSIWG)

SCSC-127J

SCSC Publication Number: SCSC-127K

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA. You are free to share the material in any form and adapt the material for any purpose providing you attribute the material to the Safety-Critical Systems Club (SCSC) Data Safety Initiative Working Group, reference the source material, include the licence details above, and indicate if any changes were made. See the license for full details.

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

Editing and typesetting by Mark Templeton and Tim Rowe.

Front cover design by DeepAI.org and Tim Rowe, based on an original by Paul Hampton. Back cover design by vector_corp on Freepik and Tim Rowe

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'26, 10–12 February 2026, details of which may be found at .

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

comments@data-safety.scsc.uk.

While the authors and the publishers have used reasonable endeavours to ensure that the information and guidance given in this work is correct, all parties must rely on their own skill and judgement when making use of this work and obtain professional or specialist advice before taking, or refraining from, any action on the basis of the content of this work. Neither the authors nor the publishers make any representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability or availability with respect to such information and guidance for any purpose, and they will not be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever (including as a result of negligence) arising out of, or in connection with, the use of this work. The views and opinions expressed in this publication are those of the authors and do not necessarily reflect those of their employers, the SCSC or other organizations.

Data Safety Guidance

The Data Safety Initiative Working Group [DSIWG]

February 2026

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
4.0	The DSIWG Team	For publication at SSS'26	10-FEB-2026

Changes Since the Last Edition

This is a major rewrite of the guidance. The document has been split into three volumes, each giving progressively more detail.

The process remains substantially the same. The intention is that conformance with this version does imply conformance with earlier versions , but some material that was informative in earlier versions has been made normative, so conformance with earlier versions of the guidance does not guarantee conformance with this version.

Future work

Proposed future work includes elaboration and expansion of scenarios to enable different routes into this guidance. An appendix of scenarios is included as a starting point at Appendix A.

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;
- further explanation of some treatments, where their use or benefit is not immediately apparent; and

- further elaboration of issues related to machine learning, large language models (LLMs) and generative artificial intelligence (AI).

Related working groups

Related SCSC working groups include:

- 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 AI systems is a new, exciting, but potentially hazardous use of data. LLM based systems are trained on data, and it is this data which enables them to be useful. Although machine learning (ML) systems tend to be trained on specifically curated data, these can also be subject to issues such as bias, poisoning and omission.

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 many 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 AI, particularly LLMs which are trained on vast amounts of data. A second area is "Big Data". A further area 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.

Data is causing harm: Strictly speaking, data can neither cause nor prevent harm. However, mistakes in data or inappropriate uses 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.

The guidance is in three volumes, published as separate documents.

Volume 1 contains the normative material. It describes the "what" of the guidance and states what is needed to ensure conformance with this guidance, if required. It should be of use to those establishing processes and procedures for data safety assurance and for those doing the actual assurance;

Volume 2 (this volume) contains the informative material. It describes the "how" of the guidance and should be of particular value to data safety practitioners;

Volume 3 contains the discursive material. It describes the "why" of the guidance and gives background and supporting material intended to improve practitioner's understanding of the data safety management processes and of issues relating to data safety management.

- The data properties that need to be guaranteed to ensure system safety are described in Vol. 1 Chapter 2
- The underlying principles of the guidance are described in Vol. 1 Chapter 3.
- The objectives and outputs to satisfy the principles are described in Vol. 1 Chapter 4.
- Definitions and abbreviations are described in Vol. 1 Chapter 5.
- The activities of each phase (and associated tailoring information) are described in Chapter 3.
- More detailed guidance is given in Chapter 4
- A worked example is provided in Chapter 5.
- Guidance for the assignment of data safety assurance levels (DSALs) is given in Chapter 6.
- Tool qualification is addressed in Chapter 7.
- Issues relating to AI and ML are considered in Chapter 8 and Chapter 9.
- Additional information on lifecycle data categories is given in Chapter 10.
- Issues that may arise when migrating, porting, importing or exporting data are discussed in Chapter 11).
- Rare events are addressed in Chapter 13.

- The relationship between typical risk-inducing issues and the data properties most likely to be affected are discussed in Chapter 14
- Various supplemental material is given in appendices to this volume.
- Lifecycle considerations are addressed in Vol. 3 Chapter 2.
- The data safety tool RADISH is described in Vol. 3 Chapter 3.
- Some of the data issues that made management of the Covid-19 virus difficult are discussed in Vol. 3 Chapter 4.
- A summary of accidents and incidents in which data was potentially a causal factor is given in Chapter Vol. 3 Chapter 5;

This page is intentionally blank

Contents

1 Introduction **1**
1.1 Aim and Scope . . . 1
1.2 Structure of the Guidance Documents . . . 1
1.3 Intended Relationship to Other Documents . . . 2
1.4 Normative, Informative and Discursive Text . . . 2
1.5 Conformance . . . 2

I The Data Safety Management Process **3**

2 Data Safety Management Process **5**
2.1 Introduction . . . 5
2.2 Overview . . . 6

3 Activities **9**
3.1 Phase 1: Establish Context . . . 9
3.1.1 Overview . . . 9
3.1.2 Describe the organizational context . . . 9
3.1.3 Describe the system context . . . 12
3.1.4 Plan the safety management . . . 13
3.1.5 Identify data artefacts . . . 13
3.1.6 Tailoring . . . 13
3.2 Phase 2: Identify Risks . . . 15
3.2.1 Overview . . . 15
3.2.2 Activities . . . 15
3.2.3 Tailoring . . . 16
3.3 Phase 3: Analyse Risks . . . 16
3.3.1 Overview . . . 17
3.3.2 Activities . . . 17
3.3.3 Tailoring . . . 21
3.4 Phase 4: Evaluate and Treat Risks . . . 21
3.4.1 Overview . . . 21
3.4.2 Activities . . . 21
3.4.3 Tailoring . . . 23

4 Detailed Guidance **25**
4.1 Establish Context . . . 25
4.1.1 Interface control . . . 25
4.1.2 Organizational Data Risk Form . . . 25
4.1.3 Data Safety Culture Questionnaire . . . 26
4.1.4 System Definition . . . 26

4.1.5 Abstraction, Hierarchy and Flowdown . . . 26
4.1.6 Supplier Data Maturity . . . 27
4.1.7 Lifecycle Data Categories . . . 27
4.2 Identify Risks . . . 27
4.2.1 Historical Accidents and Incidents . . . 27
4.2.2 HAZOP Guidewords . . . 27
4.3 Analyse Risks . . . 29
4.3.1 Establishing DSALs . . . 29
4.3.2 Analysing DSALs . . . 31
4.4 Evaluate and treat Risks . . . 33
4.4.1 Risk Treatment . . . 33
4.4.2 Mitigating Data Safety Risks . . . 33

5 Worked Example **57**
5.1 Purpose . . . 57
5.2 Establish Context . . . 57
5.2.1 Background . . . 57
5.2.2 ODR Assessment . . . 58
5.3 Risk Identification . . . 62
5.4 Risk Analysis . . . 63
5.5 Risk Evaluation and Treatment . . . 64

6 DSAL Customisation **71**
6.1 Introduction . . . 71
6.2 Significance factors . . . 71
6.3 Weighted characteristics . . . 72

7 Tool Confidence and Tool Qualification of Data Processing Tools **73**
7.1 Introduction . . . 73
7.2 What the Standards say . . . 73
7.3 Data tools . . . 73

II Additional Data Safety Considerations **75**

8 AI and Machine Learning **77**
8.1 Machine Learning Training Data . . . 77
8.2 Lifecycle Data Categories . . . 77
8.3 Real-world Data . . . 77
8.4 Sensor Data . . . 78
8.5 Data Coverage . . . 78
8.6 Dataset Assurance . . . 78
8.7 Data Diversity . . . 78
8.8 Data Properties . . . 78
8.9 Hallucinations . . . 79
8.10 Dataset Management . . . 81
8.11 Optimization . . . 81
8.12 Compliance . . . 81
8.13 References . . . 81

9 AI and autonomy **83**

9.1 Introduction . . . 83
9.2 Job Displacement . . . 83
9.3 Bias and Discrimination . . . 83
9.4 Errors of Omission . . . 84
9.5 Loss of Human Skills . . . 84
9.6 Security Risks . . . 84
9.7 Privacy Erosion . . . 84
9.8 Control, Autonomy, Ethical and Moral Considerations . . . 85
9.9 Economic Inequality . . . 85
9.10 Dependency . . . 85
9.11 Manipulation and Fake Content . . . 86
9.12 Existential Risk . . . 86
9.13 An AI's view of the risks . . . 87

10 Lifecycle Data Categories **89**

11 Data Migrating, Porting, Importing and Exporting **95**
11.1 Introduction and rationale . . . 95
11.2 Migration Cases . . . 96
11.3 Safety issues due to migration . . . 96

12 Dark Data and Dazzle Data **99**
12.1 Introduction . . . 99
12.2 Dark Data . . . 99
12.2.1 Introduction to Dark Data . . . 99
12.2.2 Dark Data Example . . . 99
12.2.3 Dark Data Varieties and Safety Examples . . . 100
12.2.4 Summary of Dark Data . . . 105
12.2.5 Further Reading . . . 105
12.3 Dazzle Data . . . 106
12.3.1 Introduction to Dazzle Data . . . 106
12.3.2 Dazzle Data Examples . . . 107
12.3.3 Dazzle Data Varieties and Safety Examples . . . 107
12.3.4 Summary of Dazzle Data . . . 111
12.4 Links between dark data / Dazzle Data and . . . 112

13 Data Cygnology **117**
13.1 Introduction . . . 117
13.2 Black Swan Data . . . 117
13.2.1 Managing Black Swan Data Risks . . . 118
13.3 Dragon King Data . . . 118
13.3.1 Managing Dragon King Data Risks . . . 119
13.4 Perfect Storm Data . . . 119
13.4.1 Managing Perfect Storm Data Risks . . . 120
13.5 Pudding Lane Data . . . 120
13.5.1 Managing Pudding Lane Data Risks . . . 121
13.6 Conclusions . . . 121
13.7 Summary Table . . . 121

14 Additional Ways that Data Can Cause Problems **123**

15 Acronyms, Definitions and Glossary **127**
15.1 Acronyms . . . 127
15.2 Definitions and Glossary . . . 129

16 References **135**

III Appendices **139**

Appendix A Scenarios **141**

Appendix B Data Safety Management Plan **145**

Appendix C Interactions Report **147**
C.1 Introduction . . . 147
C.2 Interactions Description . . . 147
C.3 Interactions Analysis . . . 147
C.4 Summary of Findings . . . 147
C.5 Conclusions . . . 147

Appendix D Unintended Behaviours Report **149**
D.1 Identification of Unintended Behaviours . . . 149
D.2 Unintended Behaviour Analysis . . . 149
D.3 Summary of Findings . . . 149
D.4 Conclusions . . . 149

Appendix E Organizational Data Risk **151**

Appendix F Data Safety Culture Questionnaire **159**
F.1 Interpretation of Results . . . 160

Appendix G Supplier Data Maturity **163**

Appendix H HAZOP Guidewords **165**

Index **169**

List of Tables

3.1 Interpretation of ODR . . . 10
3.2 Illustrative DSAL "risk" matrix . . . 17
3.3 Calculation of likelihood . . . 19
3.4 Definition of severity . . . 19
4.1 Calculation of likelihood . . . 30
4.2 Definition of severity . . . 30
4.3 High-level mitigation measures . . . 34
4.4 Lifecycle data category abbreviations . . . 35
4.5 Mitigation methods: system design . . . 37

4.6 Mitigation methods: data design . . . 42
4.7 Mitigation methods: data verification . . . 45
4.8 Mitigation methods: data migration . . . 48
4.9 Mitigation methods: data checking . . . 50
4.10 Mitigation methods: test data . . . 51
4.11 Mitigation methods: data media handling – paper / physical storage . . . 53
4.12 Mitigation methods: data media handling – electronic storage . . . 54
5.1 Worked example: filtered techniques tables . . . 65
5.2 Worked example: derived data safety requirements . . . 67
5.3 Worked example: rejected data safety requirements . . . 68
6.1 Calculation of likelihood – option 1 . . . 71
6.2 Likelihood assessment . . . 72
6.3 Calculation of likelihood – option 2 . . . 72
10.1 Categories of safety-related data: detailed definitions . . . 89
11.1 Safety issues due to migration . . . 97
12.1 Data properties affected by dark data and / or dazzle data issues . . . 112
13.1 Summary of cygnology types . . . 121
14.1 Properties that can be lost through issues . . . 123
A.1 Data Safety Scenarios . . . 141
H.1 HAZOP guidewords: detailed definitions . . . 165

This page is intentionally blank

1 Introduction

> *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.

There are various understandings of the word "safety". Traditionally, these are limited to physical harm to the person: "The state of being protected from or guarded against hurt or injury; freedom from danger" ([1]) or "Freedom from conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment." ([2]). However, data safety needs to consider more general harm such as societal harm and indirect harm caused by misinformation and so on. This guidance concentrates on the former but is applicable more widely, and from time to time other types of harm will be considered, particularly but not exclusively in the examples of incidents and accidents (Vol. 3 Chapter 5).

1.2 Structure of the Guidance Documents

The guidance is in three parts, published as separate documents.

Volume 1 contains the normative material. It describes the "what" of the guidance and states what is needed to ensure conformance with this guidance, if required. It should be of use to those establishing processes and procedures for data safety assurance and for those doing the actual assurance work;

Volume 2 contains the informative material. It describes the "how" of the guidance and should be of particular value to data safety practitioners;

Volume 3 contains the discursive material. It describes the "why" of the guidance and gives background and supporting material intended to improve practitioner's understanding of the data safety management processes and of issues relating to data safety management.

1.3 Intended Relationship to Other Documents

This guidance 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 improve coverage of 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 this guidance. The DSIWG also hopes that organizations will include the concepts, approaches and methods in their own safety management processes.

1.4 Normative, Informative and Discursive Text

Three types of text are used within this guidance document:

Normative text, which is prescriptive and can be used for claims of conformance. Typically, this text is restricted to describing methods 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 conformance with the normative text may be achieved, but alternative means of conformance 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 volume of the guidance document contains a single text type.

1.5 Conformance

There may be occasions when it is desirable or necessary to make a claim of conformance to 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 conformance claims, the following terminology is used within the Volume 1 of this guidance document. The terms have their normal English meanings volumes 2 and 3.

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

SHOULD denotes items where, in some circumstances, there may be valid reasons for not conforming to a particular item. The implications of non-conformance must be understood, documented and approved in order to claim conformance 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.

Part I

The Data Safety Management Process

2 Data Safety Management Process

> *If you can't describe what you are doing as a process, you don't know what you're doing.*
> ***W. Edwards Deming***

2.1 Introduction

To assist organizations with integrating data safety considerations into their existing processes and, in due course, their safety management system, a possible outline data safety management process has been developed, which could be blended with the principles and objectives described in Volume 1 (Vol. 1 Chapter 3 and Vol. 1 Chapter 4).

This is structured around ISO 31000 [3], and takes into account the data safety assurance principles. Links between the process objectives and the assurance principles are described in Vol. 1 Chapter 4. The adoption of ISO 31000 means that a relatively simple process, which focuses on data-related aspects, can be presented here and applied by individual organizations. The ISO 31000 approach was selected as it results in a structure that is easily adapted to align with various existing safety standards. Alignment of this process with an organisation's existing safety processes should therefore be straightforward.

To simplify the presentation, the process is presented as a series of sequential phases. In practice, a degree of iteration is likely to be required (e.g., measures adopted to treat risks may lead to refined system design and revised risk analyses). Likewise, it may be appropriate for some parts of the process to run in parallel (i.e., a subsequent phase may start before a preceding phase has finished).

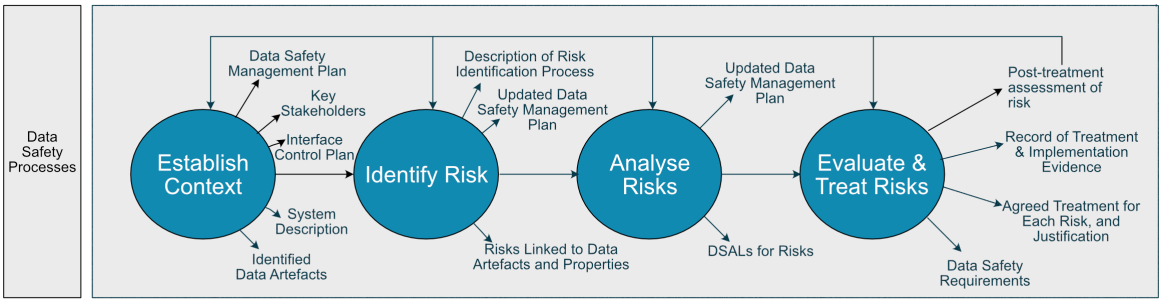


Figure 2.1: Process phases

The data safety management process occurs over four key steps, each shown as a circle in Figure 2.1, and the specific objectives and outputs of each step are described in more detail in Section section 3. The last outcome of the process is a 'post treatment risk assessment'. If the risk reduction is not considered adequate, then the process will iterate again. The whole process is intended to be repeated through the lifecycle of the system and so will be repeated many times and further iterations will be required when there are changes to the system or its data.

Although the data safety management process is based on the structure of ISO 31000, there are some differences between the two. These include:

- the ISO standard's "establish context" phase is concerned with an organization adopting a risk management system; within this guidance document that phase is extended to apply to specific projects;

- the ISO standard considers risk as being synonymous with uncertainty in outcome (i.e., some risks may be beneficial and hence it may be desirable to take actions to increase their likelihood); within this guidance document all risks are considered to have adverse effects; and
- the ISO standard separates the "evaluate risks" and "treat risks" activities; within this guidance document it has been convenient to combine these into a single phase.

ISO 31000 also recommends two activities that run in parallel with risk assessment. These activities are:

1. monitor and review the risk assessment process; and
2. communicate and consult with the stakeholders about the risk assessment process.

Aspects such as "monitor", "review", "communicate" and "consult" are taken to be part of normal project activities; items such as the organizational data risk (ODR) assessment and the data safety management plan (DSMP) are intended to assist from the perspective of data safety.

2.2 Overview

Figure 2.2 provides an overview of how the of the data safety management processes integrate with a typical system lifecycle. The system lifecycle is illustrated in the diagram as a simple model of sequential phases, with each phase shown as a coloured rectangle. Lifecycles may of course be more iterative, involve parallel streams etc., but this simple model suffices to illustrate the concepts. Further examples of system lifecycles are given in Vol. 3 Chapter 2.

In practice, the main safety management activities will occur as part of the normal safety management process, so, for example, when considering risks arising from failures of the software and hardware of a system, issues arising from data should also be considered.

Data, however, pervades all aspects of a systems lifecycle so the safety management process can occur throughout the system lifecycle depending on the type of data that is being assessed. Some data items are relevant for specific phases of the lifecycle, while others apply throughout the entire lifecycle. This guidance also introduces a meta-property data of trustworthiness to cover data that is used to inform on the overall trustworthiness of the system. The specific data categories are covered in more detail in section 10.

Figure 2.2 shows the types of data that are of concern for each process step, including those that apply to all steps, and how these feed into the data safety assessment process.

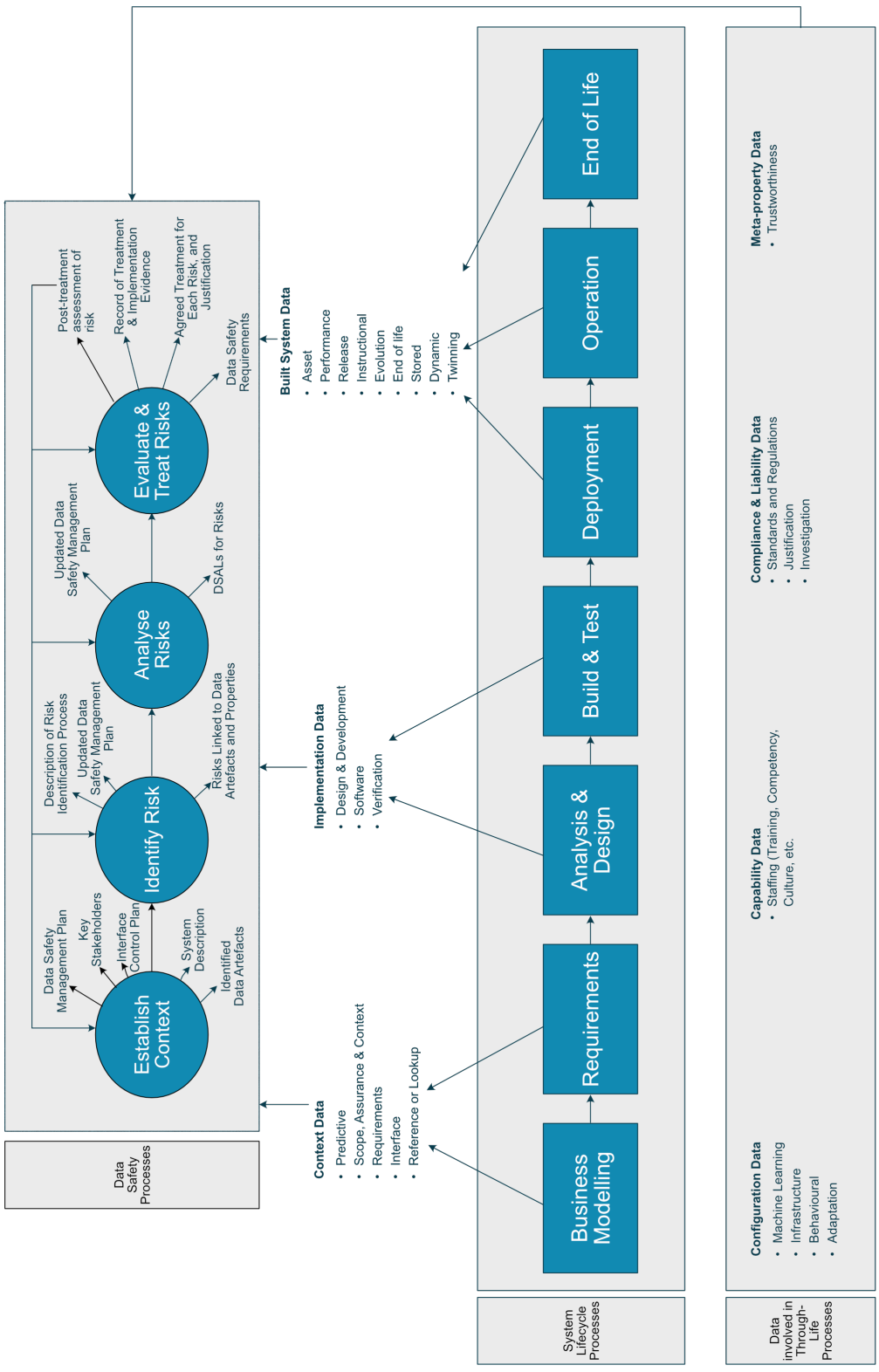


Figure 2.2: Simplified process overview

This page is intentionally blank

3 Activities

> *It is a capital mistake to theorise before one has data.*
> ***Sherlock Holmes - "A Study in Scarlet" (Sir Arthur Conan Doyle)***

3.1 Phase 1: Establish Context

3.1.1 Overview

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.

This phase involves developing:

- an understanding of the context of system within its environment;
- an understanding of the system requirements; and
- an understanding of the system design.

These factors help determine the risk appetite: essentially, how much effort will be devoted to making risks as low as practicable. In turn, this will inform the nature and scope of assessments that are conducted during system development, introduction to service, and operation.

The guidance in this document is general in nature, and so it is likely to need tailoring to align with the organization's attitude to risk, its existing processes, and the relevant sector's regulatory environment. The ODR described below, or a version customised to suit the organization's needs, may therefore provide a structured approach to data safety management.

There are four activities associated with this phase:

- Describe the organisational context;
- Identify risks;
- Analyse risks; and
- Evaluate and treat risks.

3.1.2 Describe the organizational context

This activity includes identifying:

- Stakeholders and approvers;

- Applicable legal and regulatory requirements;
- The organisational data risk maturity and appetite;
- The data safety culture;
- The data safety requirements;
- How data safety requirements will be managed;

An ODR assessment can be helpful in this identification. It results in a rating from ODR0 (the lowest risk) to ODR4 (the highest risk). The intepretation of thse ratings is shown in 3.1. This rating provides an initial, top-level view of the magnitude of data-related risk. As such, it could form the basis for process tailoring; it also gives an indication of the effort that would be proportionate in the management of data safety risks. A detailed questionnaire for conducting an ODR assessment is given in E.

If the rating is ODR0, no further data safety work would normally be needed.

Table 3.1: Interpretation of ODR

ODR Rating	Qualitative Description
ODR4	High risk
ODR3	Medium risk
ODR2	
ODR1	Low risk
ODR0	Very low risk

The ODR assessment can be conducted on the product line or individual product or system, as appropriate for the organization. It is generally not recommended to conduct the ODR without the context of a system type.

The ODR assessment is likely to be most useful for organizations that do not have significant safety engineering experience and that are operating in industrial sectors that are not strongly regulated. Organizations with considerable experience in the development of safety-critical systems in heavily regulated environments might also find it useful for defining data context and for augmenting any existing safety standards which do not explicitly consider data. If this assessment is not conducted, some high-level qualitative estimate of risk may still be required (e.g., to support process tailoring).

3.1.2.1 Stakeholders and approvers

It is important to define how the stakeholders will interact, determine the derived requirements applicable to each stakeholder, and identify how these requirements will be managed through organizational interface control (similarly to systems engineering interface control procedures already in place in many industries). External factors (e.g., economic, social, regulatory) and internal factors (e.g., cultural, processes, strategic) are key to defining appropriate organizational interface control measures. Organizational interface control may be iterative throughout the data safety management process. In particular, organizational interface control may need to be amended to take account of implemented data safety mitigations.

The level of organizational interface control required will depend heavily on the number of stakeholders, complexity of their interactions, and the contractual controls already in place. Many programmes already

require interface control documents (ICDs) to be developed for systems or equipment. The level of detail required in other programme interface control plans may be used as a guide for the requirements of organizational interface control.

The ODR can assist in identifying key stakeholders and necessary approvers (i.e., those who need to formally accept the system). Some approvers might be within the organization, while others may represent external bodies. Approvers could be customers or regulatory authorities.

3.1.2.2 Applicable legal and regulatory requirements

The legal and regulatory context will need to be identified, and where there is a conflict, that context will take precedence over this guidance. Identification of the legal and regulatory context is beyond the scope of this guidance.

3.1.2.3 Organisational data risk maturity and appetite

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.

3.1.2.4 Data safety culture

Appendix F includes a 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.

The data safety culture questionnaire can be applied at an organization level or, more likely, within an individual project team. It can also be used to highlight the importance of data safety related issues within a project team, and before and after measurements could be taken to establish the effectiveness of data safety related training. The Data Safety Culture questionnaire is likely to be of most use for low-risk (i.e., ODR1) systems. Developers of higher risk systems will typically have existing processes to develop, maintain and monitor safety cultures, although the data-oriented questionnaire could help inform those existing processes.

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

3.1.2.5 Data safety requirements management

Data safety will typically be managed as part of the organisations overall safety processes. Organisations will need to ensure that this is the case and that those processes do adequately address data safety.

3.1.3 Describe the system context

This activity is concerned with describing the system under analysis, as well as the key external influences on that system (for example, interfacing systems and human operators). There are many aspects to this activity; for brevity, only aspects directly relevant to data safety are discussed here.

When describing the system, it is often helpful to think in terms of producers and consumers of data. These may be external systems, sub-systems, or a combination of both. It may also be necessary to consider data supply chains, especially when there are a number of separate organizations involved. Note that this activity is to be addressed at a high level. The identification of specific pieces of data is a separate, but related, activity, and the identification of required properties is another activity.

As development progresses, the system description is likely to be refined. This may enable refinement of the data safety system context, supporting the data safety management planning.

In the case of data flows between different processing units, it is important to define the ways that data will be combined, and the anticipated result. Figure 3.1 illustrates two ways in which similar data sources can result in very different outcomes, dependent upon the ways in which data will be combined. in the first case, low integrity data is used to check high integrity data – this is similar to the functioning of a Global Positioning System (GPS) where the addition of a data stream from an additional satellite results in a more accurate approximation of location. In the second case, data which happens to be of a high integrity is used to augment a low integrity data stream – consider the addition of an accurate fuel measurement to a set of inaccurate cargo weights, which results in an inaccurate total takeoff weight for an aircraft.

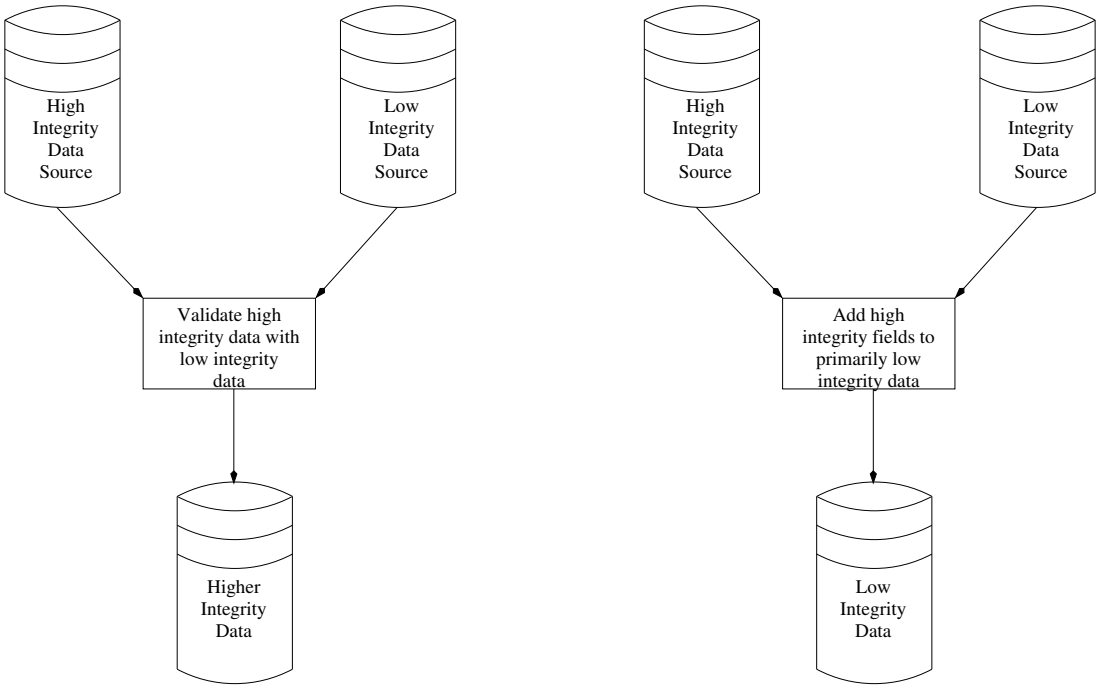


Figure 3.1: Data Streams

3.1.4 Plan the safety management

This activity involves scheduling the phases associated with the data safety management process and acquiring the necessary resources to complete them. It also involves tailoring the generic process to meet the specific needs of a particular system development.

Details of the planned safety management may be recorded in a DSMP (The DSMP may in turn be part of a higher-level planning document such as a project plan or project safety management plan). This could also be used to capture the scope of the analyses and the associated context. Together, this constitutes the first section of the DSMP. Note that if a DSMP is used, it would typically be updated with details from subsequent phases.

Planning of the data safety management requires some knowledge of the quantity and complexity of data that requires management. Therefore, the DSMP needs to be updated in subsequent phases once these details are known.

Planning of the data safety management may be done through a procurement process. Procurers may wish to know their potential supplier's understanding of data safety and their plans to implement a data safety assessment. A data safety supplier questionnaire is given in SectionG and may be used to determine the potential supplier's understanding and plans, and for auditing purposes.

3.1.5 Identify data artefacts

Data artefacts are the key pieces of data associated with the system. They may be generated, processed or consumed by the system, or used for training in its use or maintenance. They provide the foundation for the remaining phases of data risk management.

A wide variety of Data Categories have been enumerated to support the identification of data artefacts. Cross-referencing these categories against the system description can highlight the relevant artefacts.

Another way of identifying data artefacts is to consider the functions that the system performs and to establish the data that is required to support these.

A way to confirm that all relevant data artefacts have been identified is to consider the different phases of the system lifecycle. This can help prevent too narrow a focus on operational use of the system at the expense of, for example, data artefacts associated with system test and evaluation.

3.1.6 Tailoring

Including data safety within a project safety management plan is recommended for complex or highly safety-critical systems. In this case the structure of the safety management plan may be maintained, with the data safety assessment process being tailored to align to the overall safety management process.

A data maturity questionnaire (Section G) is likely to be useful, particularly when new organizational relationships are being formed. It may offer little value in situations where both organizations are familiar with each other, they have worked on data-related projects together before and there are suitable audit / review arrangements in place.

Data artefacts may be defined at a number of levels. A data artefact associated with a medical system could be described as "patient data". Alternatively, this could be split into smaller parts (e.g., "blood group").

Generally, the highest possible level consistent with the system description should be used; this prevents an excessively long list of data artefacts being developed. If necessary, where further detail is needed those data artefacts can be refined as part of an iterative process focused on key issues.

Not every data category will be relevant to every system. Furthermore, for low-risk (ODR1) systems it may be sufficient to simply consider the groupings of categories (e.g., "context", "implementation", etc.). Conversely, high-risk (ODR4) systems might need to consider every category, even if this results in a conclusion that a specific category is not relevant for the system in question.

A function-based approach to identifying data artefacts is likely to be enabled by design processes that also adopt a function-based perspective. If information from a function-based perspective is readily available, then it should be used to support the identification of data artefacts. If this information is not readily available, it is recommended that it be generated for medium and high-risk systems (i.e., ODR2 to ODR4, inclusive).

Considering data across the system lifecycle is a relatively simple activity, which is applicable to all systems (i.e., ODR1 to ODR4, inclusive).

3.2 Phase 2: Identify Risks

3.2.1 Overview

This phase involves identifying sources of risk and understanding the potential consequences. It should result in a comprehensive list of risks. These activities are likely to be concurrent with the development of more detailed system designs.

3.2.2 Activities

There are three complementary activities that can be used to identify risks. There is also an activity associated with updating planning documents:

- Review the general, historical perspective;
- Conduct a top-down analysis;
- Conduct a bottom-up analysis; and
- Update the planning documents.

3.2.2.1 Review the general, historical perspective

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 Part 3 Chapter section 5. These include cases which relate to a number of data properties, 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.

Some organizations carry out a “lessons learned” review at the end of projects, which can also give insight into potential risks.

3.2.2.2 Conduct a top-down approach

If the system under consideration has clearly identified functions, data-related risks can be assessed by considering each function in turn and analysing whichdata artefacts and, more particularly, which data properties the function depends on. This may be achieved by a variety of methods, including hazard and operability study (HAZOP) and structured what-if checklist (SWIFT).

A HAZOP [4] 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.

Appendix H lists a set of guidewords for a data-focused HAZOP, based upon the properties defined in section 10. 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. Some guidewords might not be relevant for some systems.

If there are a limited number of safety-related functions, this is usually the simplest approach. This approach also has the advantage that it integrates well with other function-based, top-down approaches to assessing system safety.

3.2.2.3 Conduct a bottom-up approach

This approach starts from the data artefacts and explores the effects of data errors. In this context an error is a situation where a required data property is not exhibited. This may be achieved by a variety of methods, including HAZOP and Failure mode and criticality analysis (FMECA).

3.2.2.4 Update planning documents

Once the data safety risks have been identified, the DSMP requires review to determine if it needs updating to take account of the quantity and complexity of the analysis and mitigation activities needed to address the risks. While the DSMP might be updated throughout the data safety management, updates may not be required for all projects.

3.2.3 Tailoring

The general, historical perspective review is a simple activity that does not require significant resources. It is recommended for all systems, regardless of risk level.

When conducting a bottom-up approach, it may not be appropriate to explicitly consider every possible property for every single artefact. For low-risk (ODR1) and medium-risk (ODR2 / ODR3) systems some form of tailoring may be expected. This might, for example, take the form of pre-selecting the properties that are most relevant, or limiting the layer of abstraction at which the system is considered.

Tailoring of the bottom-up approach may also be appropriate for some high-risk (ODR4) systems, but in this case an explicit argument that the tailoring has not adversely affected system safety would be expected. Furthermore, the risk identification process for high-risk (ODR4) systems is expected to be highly structured. To support this, a number of data-related HAZOP guidewords are presented in Chapter H.

The top-down and bottom-up approaches provide different perspectives on data-related risks. For low-risk (ODR1) systems it may be appropriate to consider just one of these perspectives. Both perspectives would be expected to be considered (to some degree) for medium-risk (ODR2 / ODR3) and high-risk (ODR4) systems.

3.3 Phase 3: Analyse Risks

3.3.1 Overview

This phase involves identifying sources of risk and understanding the potential consequences. It should result in a comprehensive list of risks. These activities are likely to be concurrent with the development of more detailed system designs.

3.3.2 Activities

This part of the risk management process involves developing an understanding of the consequences and likelihood of each risk. This understanding allows system (or safety) integrity levels or DSALs to be determined. There are two activities associated with this phase:

- Establish DSALs; and
- Analyse DSALs as part of system safety activities.

3.3.2.1 Establish required DSALs

The key activity in this phase is to establish the (untreated) likelihood and severity of each risk identified in the preceding phase.

To analyse risks and, more particularly, to align data safety with other risk management processes, problems stemming from the use of the term "likelihood" in situations where there may be no failure rates need to be overcome. For this reason the DSAL was developed. The DSAL metric is not a statistical measure of likelihood, or a literal numeric measure of integrity. Instead, the DSAL metric is an indicator for the level of rigour that an assurance argument requires. As such, DSALs share a common theoretical basis with concepts like item development assurance levels (IDALs) [5] and development process systematic capability [6].

DSALs are measured on a scale of DSAL0 (lowest-assurance) to DSAL4 (highest-assurance). They are typically allocated as indicated in a form similar to Table 3.2. The specific allocation of DSALs to likelihoods and severities will vary between industries and applications.

Table 3.2: Illustrative DSAL "risk" matrix

Severity	Likelihood		
	High	Medium	Low
Minor	DSAL1	DSAL0	DSAL0
Moderate	DSAL2	DSAL1	DSAL0
Significant	DSAL3	DSAL2	DSAL1
Major	DSAL4	DSAL3	DSAL2
Catastrophic	DSAL4	DSAL4	DSAL3

Definitions for severity and likelihood in Table 3.2 may be customised for a specific application. A default approach to the assessment of likelihood is presented in Section 3.3.2.1 and Table 4.1, while default definitions for severity are presented in Table 4.2.

Although this allocation of DSALs is typically used, there are some situations where a different allocation matrix may be more appropriate. Hence, it may be appropriate for a tailored allocation matrix to be used. Regardless of whether tailoring is used, the matrix should be reviewed and confirmed as being suitable for the intended application.

In principle, Table 4.1 and Table 4.2 are used together with Table 3.2 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 Paragraph 3.3.2.2, while potential approaches to customisation are described in section 6. 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 4.1 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 3.3: 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 4.2.

Table 3.4: 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 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.

The additional understanding developed during this part of the process may show that some previously identified data artefacts are not of consequence. Similarly, it is possible that this process may identify additional data artefacts or a need to refine the description of existing artefacts.

3.3.2.2 Analyse DSALs as part of system safety activities

Allocating a required DSAL is a significant part of controlling data safety risks, but it is not the only part. It is important that DSALs are considered as part of wider system safety activities, rather than being viewed as a separate item.

For medium-risk (ODR2 / ODR3) and high-risk (ODR4) systems it is likely that required integrity, or assurance, levels will be calculated from perspectives other than data safety. Possible examples include item / function development assurance levels from aerospace recommended practices (ARPs) 4754A [5] and safety integrity levels from IEC 61508 [6]. Where such an approach is used, the mapping to DSALs should be included within the DSMP.

This activity involves comparing DSALs with integrity, or assurance, levels. Assuming a typical scenario of a system processing or manipulating data flowing through it, there are three cases to consider:

1. Do system safety requirements place assurance requirements on the data? System safety requirements may be derived from several sources, including both system safety analyses and regulations. Whilst it may be possible to simply map between a relevant system safety standard and DSALs, it is more likely that the assignment of a DSAL to a specific piece of data will depend upon the precise system architecture selected. For example, it may be critical that a particular data item holds an accurate time stamp. If that time stamp is on an incoming data stream, then it must be preserved throughout any processing of that data, whereas if the times stamp is inserted at the end of data processing, then the assurance levels on the earlier phases of data processing may be relaxed. This is similar to the situation described in some software standards, where a high integrity requirement at system level may be met by two lower integrity systems plus a high integrity combiner.
2. Can the data affect the software? In particular, can the data affect the software such that the safe operation of the system is jeopardised? An ideal system would be able to handle any data fed into it safely without problems, but this is often not the case. An example might be a legacy system which has limited error checking and so may fail in unsafe ways if fed with data which is outside of the expected range. Formally: *given a system containing software written to a particular software assurance level (which may be none), what should the DSAL of the processed data be to preserve correct operation of the system?*

3. Can the software affect the data? In particular, can the system's software affect the data being processed or manipulated in such a way that data properties that are important for safety might be lost? Some examples might be systems which transform messages, losing any associated checksum protections, thereby possibly affecting the integrity of the data within the message; as a minimum, this removes a means of checking data integrity. Another example might be a system that can delay data flowing through it, (e.g., due to buffering) when timely delivery of the data is critical. Formally: *given data at a particular DSAL, what should the software assurance level of the software in the system be to preserve the DSAL of the data?*

3.3.3 Tailoring

DSALs are determined by the use of the data but are used to define properties of systems that manipulate the data (typically via derived requirements). They can therefore be derived at different levels and from different constructs. For example, in the case of simple, low-risk systems it may be appropriate to derive a single DSAL for an entire system. Alternatively, it may be appropriate to derive DSALs for sub-systems, for example, to match the level at which other integrity, or assurance, levels have been determined.

Another option is to derive DSALs from data artefacts rather than directly from systems. This approach has the advantage that treatments are often related to data artefacts; it can work well where there is a simple relationship between data artefacts and risks.

3.4 Phase 4: Evaluate and Treat Risks

3.4.1 Overview

This phase involves deciding what action (if any) should be taken for each of the risks identified in preceding phases. This decision will be influenced by the organization's risk appetite and other factors determined as part of the establish context phase. From some perspectives it may seem strange that requirements are identified at such a late phase. This is a consequence of explicitly linking data safety requirements to risks associated with data properties of data artefacts, and the use of DSALs to describe levels of rigour.

This phase also involves identifying, implementing and verifying treatments for the risks emerging from the previous phase. Verifying the treatment includes checking technical details of the chosen approach and re-assessing the post-treatment risk to determine whether it is now acceptable.

3.4.2 Activities

3.4.2.1 Evaluate Risk

In this phase each risk, including the associated DSAL is reviewed and appropriate responses are determined. This phase aims to answer the question: can we accept this risk or does some action need to be taken? This is likely to require discussion amongst a number of stakeholders. From a system safety perspective, there is nothing intrinsically special about data-related risks. Hence, it is recommended that evaluation of data-related risks be conducted alongside the evaluation of other system risks, as part of an organization's standard risk evaluation process.

3.4.2.2 Treat risk

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 treated by other means. If this is the situation then appropriate measures need to be put in place *within the data* to treated 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 treats 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 [7]; this information is likely to be applicable to other domains as well.

A range of methods and approaches can be used to treat the identified data safety risks. Since mitigation can be a complex process, requiring collaboration with all system engineering elements, a collection of high-level data-specific mitigation measures is given in Table 4.3. 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 data-related mitigation measures may prove sufficient for practitioners managing the safety ofsystems which do not have a high safety criticality.

Avoid Risk avoidance can be employed where a risk can be eliminated by using different approaches to the design and / or operation of the system. It may also be the case that very significant risks cannot be adequately treated and the only option is to avoid the untenable risk by not proceeding with the project;

Accept For low likelihood and low severity risks (e.g., those ranked as DSAL0), where the cost of further risk reduction is judged to be unacceptable, the risk may be accepted as-is and managed as such. Appropriate justification is likely to be required for acceptance of risks ranked higher than DSAL0;

Transfer Ownership of the risk's consequence can be transferred, for example to another organization (for example by taking out an insurance policy if the risk is to equipment or property) or to another part of the system that will address the risk. It is important that any such risk transfers are formally documented, understood and agreed by both parties; or

Treat The risk can be reduced. This can be achieved by reducing the severity, or the likelihood or both. Choosing this response often involves having an outline view of how the risk may be reduced. If a decision is made to treat a risk, suitable methods and approaches should be identified. A range of potential methods and approaches are included in this guidance. These are mapped against DSALs, data properties and a selection of lifecycle data categories.

Once a treatment strategy has been established and implemented, whether the expected risk reduction has been achieved needs to be determined. Equivalently, there is a need to consider whether the residual risk may now be accepted or whether another one of the responses identified above is necessary.

In addition to deciding on and documenting the appropriate mitigation for each risk, this phase also includes gaining approval for these decisions.

3.4.2.3 Treating risk

3.4.3 Tailoring

Records of the discussions that occur as part of risk evaluation should be kept. For low-risk (ODR1) systems, this may be in the form of a brief memo. For high-risk (ODR4) systems, a detailed, structured record, which is placed under formal control, may be required; in this case these discussions may be recorded as part of the system's hazard log or as part of a data safety management plan.

As outlined in section 3.3.3, DSALs can be applied at varying levels of abstraction. For small-scale or low-risk (ODR1) systems it may be appropriate to consider treatments at higher levels of system abstraction. For example, this could involve applying a single DSAL to a sub-system or even to the system in its entirety and implementing risk treatment techniques at that level. The latter approach could be appropriate where the data in the system interacts in complex ways and the associated safety risk does not warrant a detailed investigation of these interactions.

A significant amount of tailoring is implicit in the way the tables of methods and approaches are constructed. At best a method / approach may be highly recommended as a way of maintaining a required data property at a given DSAL. The tables are not exhaustive; additional or alternative methods and approaches can be used.

4 Detailed Guidance

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

4.1 Establish Context

4.1.1 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.

4.1.2 Organizational Data Risk Form

section E 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. The 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 [3] 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 ISO 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.

4.1.3 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, Appendix F 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.

4.1.4 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.

4.1.5 Abstraction, Hierarchy and Flowdown

It is useful to consider different levels of abstraction of the system and also the data used within the system. An abstraction model of the system and its data enables reasoning and communication of the assurance requirements, and supports flowdown of the assurance requirements. Moreover it allows clearer reasoning about its safety. Many design approaches and tools support hierarchical views and these can be used as needed. When a hierarchical design is constructed it is possible to apply requirements and DSALs at the

top level and flow these down, decomposing as needed to lower levels of design or abstraction. Tools can support tracing of requirements through design hierarchies.

4.1.6 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. section G includes a questionnaire to help ensure that the supplier has suitable processes in place to manage data safety-related issues.

4.1.7 Lifecycle Data Categories

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

section 10 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 according to the system lifecycle.

4.2 Identify Risks

4.2.1 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 Vol. 3 Chapter 5. These include cases which relate to a number of data properties, 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.

4.2.2 HAZOP Guidewords

A HAZOP [4] 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.

Appendix H lists a set of guidewords for a data-focused HAZOP, based upon the properties defined in section 10. 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. Some guidewords might not be relevant for some systems.

4.3 Analyse Risks

4.3.1 Establishing DSALs

Paragraph 3.3.2.1 presented a method for assigning DSALs, by combining severity and likelihood through a risk matrix, such as Table 3.2. 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 4.1 and Table 4.2 are used together with Table 3.2 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 Paragraph 3.3.2.2, while potential approaches to customisation are described in section 6. 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 4.1 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 4.1: 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 4.2.

Table 4.2: 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 IDALs and 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.

4.3.2 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 [7]; this information is likely to be applicable to other domains as well.

4.4 Evaluate and treat Risks

4.4.1 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.

4.4.2 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 managing the safety ofsystems which do not have a high safety criticality.

For practitioners managing the safety of systems 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.

4.4.2.1 High-level mitigation measures

Table 4.3 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 being managed.

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

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

Table 4.3: 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, 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[1] 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
M.14	Identification of Missing Data	R	R	HR	HR	Analyses or reviews that establish what might have been left out
M.15	Identification of Hidden or Obscured Data	R	R	HR	HR	Analyses or reviews that establish what might be hidden or obscured

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

4.4.2.2 Detailed methods and approaches

The following tables detail methods and approaches which may be used by practitioners conducting safety management. The tables map the methods and approaches to lifecycle data categories. These lifecycle data categories are abbreviated using the scheme shown in Table 4.4.

The five lifecycle data categories presented in Table 4.4 are a subset of those presented in Table 10.1, which presents a comprehensive list of lifecycle data categories. The five presented in Table 4.4 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 4.4: Lifecycle data category abbreviations

Lifecycle 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 10.1 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 Table 3.2 in association with Table 4.1 and Table 4.2.
- Determine the lifecycle data category, using the abbreviations in Table 4.4.
- Determine the list of applicable data properties, using the abbreviations in Table 10.1. 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 lifecycle data category matches the lifecycle 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 4.5. This technique will apply to data artefacts where:

- The lifecycle data category is "dynamic" ("D" from Table 4.4) and
- The data artefact holds one or more of the propertiesintegrity, completeness or verifiability ("I", "C" or "V" from Table 10.1).

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 lifecycle 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 lifecycle 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 10.1 can easily be seen to appear against the first two techniques of Table 4.5, 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.

4.4.2.3 System Design

Table 4.5: Mitigation methods: system design

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data property
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.....
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...

Continued on next page

Table 4.5: Mitigation methods: system design (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data property
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.
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.....

Continued on next page

Table 4.5: Mitigation methods: system design (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data property
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...
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.

Continued on next page

Table 4.5: Mitigation methods: system design (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data property
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.
SD.25	"Known unknown" handling	..D..	-	R	HR	HR	Consideration where there is a known absence of information and the impact this may have, through its absence or when the information becomes subsequently known. For example, through use of warnings, training, restrictions of use, etc.	IC.YOARTMVL.PQU..HE.GZXK
SD.26	Defined data collection opportunities	..D.J	-	R	HR	HR	Assessing the data collection opportunities gained or lost from the design or architectural approach – at design time	IC.YOARTMVL.PQU..HE.GZXK

Continued on next page

Table 4.5: Mitigation methods: system design (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data property
SD.27	Handle unknown extras	`..D.J`	-	R	HR	HR	Establish an approach for handling unknown extras: data found to be erroneous that should not be in the system. Possible approaches might be: 1. Delete the erroneous data; 2. Accept the erroneous data, i.e. continue using it; 3. Report that the data was erroneous but take no further action; 4. Establish under what circumstances and situations the erroneous data would have had an effect; or 5. Perform a full impact analysis on the effects of the erroneous data and then act according to the results.	`IC.YOARTMVL.PQU..HE.GZXK`

4.4.2.4 Data Design

Table 4.6 addresses design aspects, including the construction of data storage structures and methods.

Table 4.6: 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. Note 3.	.......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. See Note 3.	..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. See Note 2 and Note 3.	IC.Y.................
DD.07	Configuration management	VIDPJ	HR	HR	HR	HR	A formal process that controls changes to the data and data model. See Note 1 and Note 2.	.......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. See Note 1.	..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.

Continued on next page

Table 4.6: Mitigation methods: data design (continued)

Serial	Technique	Data	DSAL				Notes	Data Property
			1	2	3	4		
DD.10	Sensitivity analysis	..D..	-	R	R	HR	Assess impact of missing data, for instance, what decisions might be made erroneously due to certain data values being omitted, late or slightly perturbed. Assess impact of missing data, for instance, what decisions might be made erroneously due to certain data values being omitted, late or slightly perturbed.	IC.YOARTMVL.PQU..HE.GZXK
DD.11	Outlier analysis in design	..D..	-	R	R	HR	Use of edge / corner cases, boundary cases, and testing well beyond normal or acceptable ranges to give insight into the system degradation behaviour. This could apply to machine learning training data or other data.	IC.YOARTMVL.PQU..HE.GZXK
DD.12	Outlier analysis in verification	V....	-	R	R	HR	Use of edge / corner cases, boundary cases, and testing well beyond normal or acceptable ranges to give insight into the system degradation behaviour. This could apply to machine learning verification data or other data.	IC.YOARTMVL.PQU..HE.GZXK

Continued on next page

Table 4.6: Mitigation methods: data design (continued)

Serial	Technique	Data	DSAL				Notes	Data Property
			1	2	3	4		
DD.13	Sensor data management	..D..	-	R	HR	HR	Independently monitor sensors and ensure regular maintenance, calibration or replacement of hardware. Sensor interfaces need to be able to detect any problems and report faults. In particularly critical applications, it may be necessary to minimise the potential for common mode failures by using disparate technologies, or at least sensors from different manufacturers.	IC.YOARTMVL.PQU..HE.GZXK
DD.14	Data definition correctness	..D..	-	R	HR	HR	Ensure definitions of data are correct by making sure interface specifications are clear and unambiguous, rejecting incorrect size or scale datasets, keeping a list of known changes / incompatibilities and documenting the changes or fixes that have to be applied to make the data usable or consistent.	IC.YOARTMVL.PQU..HE.GZXK

Note 1 One useful approach is to keep a list of known changes / incompatibilities and documenting the changes or fixes that have to be applied to make the data usable or consistent.

Note 2 For verification lifecycle data categories, comparisons, reviews and data audits may help to manage the risk of information asymmetry . In general, the issue of divergence across multiple data sources needs to be recognised and addressed in the most appropriate technical, and ideally automated, way.

Note 3 Ensure risks of known extras are handled, for example, through careful filtering, review, use of concise notations such as Goal Structuring Notation (GSN) to extract the essence, and periodic audits and cleans.

4.4.2.5 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 4.7 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 4.7: 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. See Note 1, Note 5, Note 6, Note 8 and Note 10.	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. See Note 9.	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. See Note 1, Note 5 and Note 6.	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. See Note 2.	.....V.....E.....
DI.07	Defined data lifetime(s)	VIDPJ	R	R	HR	HR	Information showing when data validity expires.	.....E.....

Continued on next page

Table 4.7: Mitigation methods: data verification (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
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). See Note 2, Note 4, Note 8.	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. See Note 7.	.............UBS..D....
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. See Note 10.	IC.Y.....V.F.........X.
DI.12	Independent check	VIDPJ	-	-	R	HR	A separate person or system is used to check the data independently. See Note 1 and note 5.	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). See Note 3 and note 6.	.......T.........H..G...

Continued on next page

Table 4.7: Mitigation methods: data verification (continued)

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
DI.14	"Known unknown" handling	`....J`	-	R	R	HR	Consider where there is a known absence of information and the impact this may have, through its absence or when the information becomes subsequently know. For example, evidence can also be substituted with other more indirect assurance e.g.established organisational track-record in the sector, or audit reports.	`IC.YOARTMVL.PQU..HE.GZXK`

Note 1 It is important to guard against choosing just some cases or self-selection and missing what matters. For example, it is important to ask questions and challenge the data selected to be sure it can be justified. Also taking a step back" to look at the bigger picture.

Note 2 It is important to keep critical data up to date and regularly refreshed, and to know how old the data is in the first place.

Note 3 One useful approach is to keep a list of known changes / incompatibilities and documenting the changes or fixes that have to be applied to make the data usable or consistent.

Note 4 So sensors need to be regularly maintained and calibrated or replaced and their interfaces need to be able to detect any problems and report faults.

Note 5 For justification lifecycle data categories, the use of dialectic argument approaches may be useful (where both positive and negative cases can be considered) to deal with the risks of feedback and gaming. It is also useful to get a second opinion, independent review or a "fresh pair of eyes" to check the justification.

Note 6 For verification lifecycle data categories, comparisons, reviews and data audits may help to manage the risk of Information Asymmetry. In general, the issue of divergence across multiple data sources needs to be recognised and addressed in the most appropriate technical, and ideally automated, way.

Note 7 Comparisons with current real-world data and checking with historical data can help guard against intentionally darkened data, fabricated and synthetic data.

Note 8 Ensure risks of known extras are handled, for example, through careful filtering, review, use of concise notations (such as GSN) to extract the essence, and periodic audits and cleans.

Note 9 Also, simply taking a step back to look at the bigger picture.

Note 10 It is important to ensure unnecessary data does not obscure the meaning of the data set through review and completeness checks. Note that over-aggressive filtering would create cases of dark data.

4.4.2.6 Data migration

Table 4.8 addresses migration of data from one system to another system.

Table 4.8: 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 4.8: 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.

4.4.2.7 Data checking

Table 4.9: 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..........

4.4.2.8 Test data

Table 4.10: 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 4.10: 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...........

4.4.2.9 Media – Paper

Table 4.11: 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....

4.4.2.10 Media – electronic

Table 4.12: Mitigation methods: data media handling – electronic storage

Serial	Technique	Data Category	DSAL 1	DSAL 2	DSAL 3	DSAL 4	Notes	Data Property
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 4.12: 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.....

4.4.2.11 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 being managed;
- 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

5 Worked Example

> *There is a forest of data and we need to create a path through.*
> ***Tom Adams***

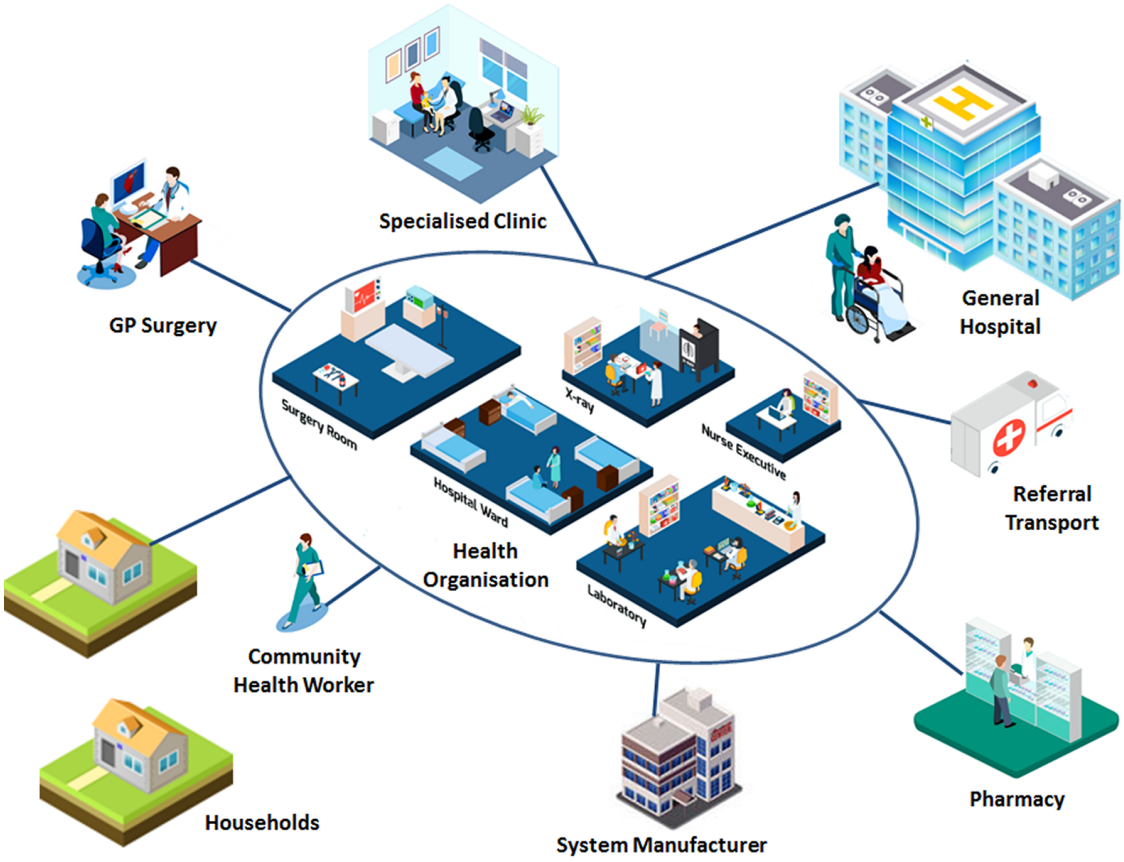
5.1 Purpose

This section provides a worked example of applying the DSG to a hypothetical system in the healthcare sector. Although some aspects of the example have been simplified, it is intended to be sufficiently realistic to allow key features of the guidance to be illustrated.

5.2 Establish Context

5.2.1 Background

A manufacturer is building a new integrated health and social care system to support holistic care for community health services. The system will support clinical workflows for aspects such as referrals, tracking clinical encounters, appointment scheduling, outcome measures through to letter and report generation. The development will follow a typical development lifecycle as a series of phases: business modelling, requirements, analysis and design, implementation, test, and deployment.



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

Fundamental to fulfilling these requirements is establishing the context within which data is used in system development, enhancement, introduction, integration or operation. This should establish the risk appetite: essentially, how much effort is to be devoted to making data risks as low as practicable. In turn, this will inform the nature and scope of assessments and treatments that are to be conducted during system development and its introduction into operational service. To meet the 'establish context' objectives set out in the guidance the following activities are recommended:

- describe the organizational context;
- describe the system context;
- plan the safety management; and
- identify data artefacts.

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



System Manufacturer

5.2.2 ODR Assessment

The manufacturer considered each of the questions within the ODR form from Chapter section E of this document. The questions in the ODR and the manufacturer's assessments were:

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

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

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

Unfortunately, accidents in the health domain are relatively frequent. There are many injuries and deaths attributed to medical errors, but these are largely tolerated by the public and grievances are usually settled by financial settlements through the courts. The manufacturer believes their contractual arrangements mean that the health organization would be liable for any claims even if they were attributed to an error in the manufacturer's system's handling of data. On balance, this question is assessed as **2b**; Score **2**.

Q3:	How much responsibility does this organization have for data safety?

The manufacturer is responsible for building the system in compliance with DCB0129 [8] and so is responsible for executing the associated safety management system to manage risk. The manufacturer, however, plans to sell the product with a condition of use that places end responsibility for patient safety on the client. On balance, this question is assessed as **3b**; Score **2**.

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

The work will be contracted under UK law and subject to the data coordination board (DCB) standards for health IT Systems. However, there is no regulator currently empowered to intervene in the delivery of healthcare systems, so the standards are not currently enforced through law. Instead, the health organization will make compliance with the standards a contractual requirement. On balance, this question is assessed as **4c**; Score **4**.

Q5:	How mature is this organization regarding data safety?

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

Q6:	How widely used is the data and who by?

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

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

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

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

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

The final score is **26**, which corresponds to **ODR2**. The manufacturer therefore concludes that there is low to medium risk that loss of properties of data in the system can contribute to or give rise to harm. The manufacturer has an internal policy for engagements based on the ODR level that dictates how the organization will plan the safety management. This policy dictates the amount of proportional effort the manufacturer needs to spend on safety data management and the level of rigour to be employed. In

this case, the policy dictates, amongst other requirements, that a separate section covering safety data management is required in its clinical risk management plan.

Next, the manufacturer identifies data artefacts that are potential sources of safety hazards. The manufacturer knows that the safety dependency of data is dictated by the context in which it is used so it also develops an understanding of when in its process lifecycle the data will be used and relied upon. The manufacturer plans to build an early prototype to show to clients to help elicit requirements definition. To support this, the manufacturer plans to create a test dataset that comprises a typical range of scenarios that the system will encounter. This form of data is identified as verification data. The system also needs to be configured to support deploying health organizations' policies. This data is infrastructure data and, for the prototyping phase, the manufacturer plans to use largely default values.

In later phases, when the system functionality is specified and the system is being built, the manufacturer plans to create a test dataset that will be key to demonstrating the correct functioning of the system and hence acceptance by the deploying health organization. This still involves the use of verification data and infrastructure data but there will be far greater dependency on these datasets than the prototyping case. The manufacturer therefore documents in the clinical risk management plan the planned use of each of the lifecycle data categories during the entire delivery lifecycle.

The procuring health organization will have a different perspective of the IT system that it will deploy into its organization. It will already have many integrated systems in live operation and as part of establishing the context for the system's deployment they will need to consider many different types of datasets:



Infrastructure data: how the system will be configured in the specific environment;

Verification data: the test datasets to be used to support certain deployments such as integration testing and training; and

Dynamic data: the data entered or fed into the system and the data presented to the user, generated in the form of reports or data passed to other systems.

The health organization decides to complete an ODR assessment so it can describe the organizational context and describe the system context. The scoring is similar to the manufacturer with the following notable differences:

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

The health organization would bear the brunt of any publicity and litigation in the event of an accident and so assesses this question as **2c**, Score **4**.

Q3:	How much responsibility does this organization have for data safety?

The health organization is responsible for deploying systems in compliance with DCB0160 [9] and is ultimately responsible for patient safety. The health organization assesses this as **3e**, Score **12**.

Q5:	How mature is this organization regarding data safety?

The health organization has only recently acquired the expertise to apply DCB0160 and is still developing its capability. Safety data management is new to the organization and it anticipates some resistance from senior management to the expenditure incurred in rolling out a DCB0160 compliant safety management system. This question is assessed as **5d**, Score **7**.

The resulting score for the health organization is **43**. This is **ODR3**: medium to high risk. The health organization aims to plan the safety management through a clinical risk management plan. This plan defines the organization and system context in more detail and lays out the planned activities for identifying, evaluating and treating data safety related risks.

As with the manufacturer, the health organization needs to identify data artefacts that are potential sources of safety hazards and understand the context of their use in the procurement / deployment lifecycles of the health organization. Post acceptance, the procuring health organization plans to run a series of user training sessions for clinicians. Once users are trained, the system will be integrated into live operations. The health organization identifies the infrastructure, verification and dynamic lifecycle data categories to be used during these phases. The health organization also realises that the system will form part of a data supply chain, because a number of external organizations and departments within their own organization engage in the procurement and use of safety-related data. For example, it will receive referral data from a number of other general practitioner (GP) systems, it will receive outcome measures from hospitals, and it will receive clinical data acquired from remote workers visiting patients in the community and from the patients themselves using the system's online portal. The system also produces data for other external systems such as electronic prescriptions for pharmacies.

The health organization sees that by using the new system it will become a commissioning user, as it will require and be a consumer of data from a variety of sources: GP systems, hospital systems, systems used by remote workers in the community, and the system's portal capturing data entered by the patients themselves. All of these act as data provisioners. Those health care professionals patients gathering patient data through physical inspections and measurement are the data acquirers.

The health organization defines in its clinical risk management plan the data supply chain relevant to the system including the roles and interfaces involved. This will therefore show where there are dependencies on dynamic data used and produced by the system.

Questions the health organization will need to address when establishing the context are:

- have all the dependent interfaces been identified?
- have the roles of commissioning user / data provider / data acquirer been established and acknowledged?
- what service level agreements or contracts exist for the delivery of the data?
- what level of assurance do data providers / data acquirers provide for their data?

5.3 Risk Identification

The manufacturer aims to carry out the following activities to meet the guidance objectives for the risk Identification phase:



- review the general, historical perspective;
- conduct a top-down approach;
- conduct a bottom-up approach; and
- update planning documents.

Before embarking on any hazard analysis, the manufacturer ensures that stakeholders review the general, historical perspective. This takes the form of a refresh briefing to raise awareness of issues that are specific to data such as ageing, biasing and defaults.

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

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

To analyse the risks in more detail, the manufacturer uses a top-down approach and a bottom-up approach. In the first approach it considers each of the system functions (such as clinical screens) and analyses where there is a dependency on data and what properties need to be preserved. In the second, as much of the functionality is driven by data flows in and out of the system, the manufacturer also looks at specific data flows and assesses the impact of loss of properties for the data in those flows.

On completion of the risk identification phase the manufacturer updates planning documents such as the clinical risk management plan, to reflect the outcome of the analysis.

The health organization will likewise need to conduct risk identification relevant to their deployment context. Hazards arising from data sources that are to be delivered into the new system from existing systems need to be assessed for data risks. As with the manufacturer, a briefing to stakeholders to review the general, historical perspective is first conducted to cover generic data safety issues but also to highlight lessons learned from previous accidents and incidents that have occurred in the health organization itself.



As with the manufacturer, both a top-down approach and a bottom-up approach are adopted. From the health organization's perspective, one key focus for hazard identification is in the use of dynamic data, i.e., data that will be delivered into the new system from existing system data sources and the data presented to the user. For the interactions identified in the supply chain, the health organization needs to consider the risks associated with loss of properties of the data it will receive. Questions the health organization will need to consider and address more formally in the clinical risk management plan are:

- which datasets or data items being received from other systems have data properties (such as timeliness, completeness, consistency, fidelity / representation etc.) that are significant to patient safety?
- what data presented to the user has data properties (such as availability, format, resolution, etc.) that are significant to patient safety?
- what existing barriers or mitigations (physical, technical, procedural) exist to reduce the risk of loss of data properties? and
- will any existing barriers be lost as a consequence of the new system?

On completion of the Risk Identification phase the health organization will update planning documents such as the clinical risk management plan to reflect the outcome of the analysis.

5.4 Risk Analysis

In this phase identified hazards are assessed to determine their likelihood and severity. To meet the guidance objectives the following activities are carried out:

- establish DSALs; and
- analyse DSALs as part of system safety activities.



System Manufacturer

The manufacturer will establish DSALs by considering cases where the use of specific categories of data could give rise to hazards. In the prototyping phase, the manufacturer sees no use of the data that can give rise to credible clinical risk and assesses the DSAL for that dataset as **DSAL0**.

In the second phase of the development lifecycle, where verification data and infrastructure data is being used to demonstrate the correct functioning of the system, the manufacturer considers that loss of any of the data properties of integrity, completeness, consistency, continuity, format, accuracy, resolution, timeliness, availability, fidelity / representation, sequencing, intended destination / usage, Goldilocks, and explainability of this data could give rise to hazards.

For example, if the verification dataset selected is not representative of the eventual diversity experienced in practice (loss of fidelity / representation), then it is possible that the system may contain latent software errors that could give rise to harm. However, the manufacturer acknowledges that the system will be subject to further testing and trials in the clinical setting and so there will be other opportunities to detect errors in the system. Overall:

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

The manufacturer therefore assesses these lifecycle data categories as **DSAL1** in this particular context of use.

The manufacturer takes care to analyse DSALs as part of system safety activities by documenting these assessments along with other hardware and software safety considerations, for example those arising from DCB0129, in the clinical risk management plan.



From the health organization's perspective, the main focus for risk assessment to establish DSALs is in the use of dynamic data. For the interactions identified in the supply chain the health organization needs to consider the risks associated with loss of properties of the data it will receive and present to the user. Questions the health organization will need to consider and address more formally in the clinical risk management plan are as follows:

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

Concerns were raised about the vulnerability of the system to inadvertent data overload from the large number of potential data sources, and possibly from malicious activity. This led to the addition of the Goldilocks property to the list of data properties applicable to the live system.

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

- The health organizations assesses the likelihood of this loss of property as **High**; and
- The impact of such errors, although not realistically likely to lead to death, could result in delays to treatment that could result in serious injury and hence **Moderate** impact.

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

The health organization takes care to analyse DSALs as part of system safety activities by documenting these assessments along with other hardware and software safety considerations (e.g., arising from DCB0160 requirements) in the clinical risk management plan.

5.5 Risk Evaluation and Treatment

The manufacturer carries out the following activities to meet the guidance objectives:



- review each risk and either avoid, accept, transfer, treat;
- establish methods for relevant risks; and
- implement and verify treatment methods.

The manufacturer decides to review each risk and either avoid, accept, transfer, or treat the risk. It decides to accept the **DSAL0** risk, but as it has determined there is some, albeit low, risk (**DSAL1**) associated with

its use of data at a specific point of its lifecycle, it decides to treat that risk. The manufacturer evaluates this risk and considers that the risks should be reduced further by taking some reasonably practicable steps.

Having decided that further risk reduction is necessary, the manufacturer needs to establish treatment methods for relevant risks that are appropriate for DSAL1 data, and in doing so demonstrate that reasonably practicable steps have been taken to reduce the risk. The manufacturer therefore refers to the tables in the DSG document. The manufacturer then documents in its clinical risk management plan:

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

For **DSAL1** verification data, the tables show that the following are recommended (R) or highly recommended (HR) where loss of data properties integrity, completeness, consistency, continuity, format, accuracy, resolution, timeliness, availability, fidelity / representation, sequencing, intended destination / usage, Goldilocks, and explainability can give rise to a number of hazards. The extracts below indicate the mitigation techniques identified from the tables which should form the basis of the data safety requirements.

Table 5.1: Worked example: filtered techniques tables

Ref	Technique	R / HR	System Design (extracted from Table 4.5)
SD.11	Logging facilities	R	Data processing events are logged to allow support staff to monitor the health of the system and provide diagnostic information.
SD.17	Credibility / reasonability checks	R	Dedicated processing 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 - NaN).
SD.20	Syntax checks	R	Semantic checking of data values and sequences based on defined rule sets.

Ref	Technique	R / HR	Data Design (extracted from Table 4.6)
DD.01	Governance model	R	A governance model is established that defines, e.g., data ownership, processing roles and responsibilities, processing authorizations and permissions.
DD.03	Data flow diagram	HR	To describe the data flow in a diagrammatic form.
DD.04	Data model	HR	To articulate how data is organized.
DD.05	Client sign-off	R	Agreement from the client that the system architecture and design is appropriate for the data considered
DD.08	Data dictionary	HR	A collection of descriptions of the data objects or data items in a data model for the benefit of data users.

Ref	Technique	R / HR	Data Implementation (extracted from Table 4.7)
DI.01	Review / inspection	HR	Manual review / inspection of data possibly involving data visualization tools.
DI.03	Ground-truth check	R	Inspection against physical measurements (e.g., lengths, positions, heights) taken in the real world.
DI.04	Auditing	R	A period of comprehensive internal and external testing of the data quality process.
DI.09	Authorization	R	A security model is established to control who is authorized to create, view, edit, delete the data.
DI.11	Defined confidence / Trust Levels	R	Criteria are established to provide an objective measurement of the confidence or trust in a given dataset.

Ref	Technique	R / HR	Data Migration (extracted from Table 4.8)
No relevant techniques			

Ref	Technique	R / HR	Data Checking (extracted from Table 4.9)
No relevant techniques			

Ref	Technique	R / HR	Test Data (extracted from Table 4.10)
TD.01	Using informal / ad-hoc means	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.
TD.05	Using manual means	R	Simple test data can be produced by manual means, although this may be prone to human error.
TD.08	Using initial runs of new system	R	This method is often used where the system is breaking new ground and there is no prototype or legacy system to produce test data. Initial operations may differ from eventual usage, so test data must evolve.
TD.09	Derived from real data	R	Where real data is available this is usually a good basis for generating test data (e.g., by modification to increase the test space coverage).
TD.11	Produced by client	R	Ideally the client is involved in producing or at least checking the test data.
TD.12	Client sign-off	R	Where possible, the client should formally agree and sign off the test data as appropriate.
TD.13	Error seeding	R	This is where errors are deliberately inserted into the dataset to demonstrate the effectiveness of data validation.
TD.14	Data reuse	R	Reusing data for one project that was created and thoroughly assured for another project. This can be effective but the read-across should be established.

Ref	Technique	R / HR	Test Data
TD.15	Feedback testing	R	To check output data by comparing it with the input source.

Ref	Technique	R / HR	Media – paper (extracted from Table 4.11)
MP.01	Photographic copies	R	Photocopy and store separately.
MP.02	Scan to electronic format	R	Retain both paper and electronic copies.
MP.10	Indexing / cataloguing	R	To support efficient availability.

Ref	Technique	R / HR	Media – electronic (extracted from Table 4.12)
ME.01	Regular refresh / rewrite	R	Of magnetic media or flash memory.
ME.02	Suitable physical environment	R	Store media in a clean, low-humidity environment at a steady temperature, cool but not cold.
ME.03	Copies at different locations	R	Physically separate to cover natural disasters, accidental or malicious damage.
ME.04	Backups / duplication	R	Backups are essential. Frequency of backup depends on rate of change. The number of generations to keep relates to the impact of data loss.
ME.05	Sample restores	R	Sample restores should be performed at intervals to ensure that the backups are readable and retrievable.

From these tables the manufacturer decides on a series of activities to implement the recommendations that are applicable to its particular endeavour. These activities are expressed as a series of requirements that can be placed on the manufacturer's delivery organization and tracked through to completion.

Table 5.2: Worked example: derived data safety requirements

Ref	Requirement	Guidance Reference
R1	The verification data shall be carefully controlled in the manufacturer's configuration management system. There shall be a configuration management plan that shall define who has responsibility for the data and who is authorized to create and amend it.	DD.01, DI.09
R2	The verification data shall be held on an industry standard file share that is regularly backed up with copies moved periodically to off-site storage. The Backup / Recovery plans shall include periodic sampling of restores.	ME.01, ME.02, ME.03, ME.04, ME.05

Continued on next page

Table 5.2: Worked example: derived (continued)

Ref	Requirement	Guidance Reference
R3	The data shall be modelled as a series of patient "journeys" that cover the entire lifecycle of data from first encounter through to archival and deletion of data. The complete set of journeys shall be chosen to exercise all the functionality of the system. The modelling shall include a data dictionary, data flow diagrams and a data model.	DD.03, DD.04, DD.08
R4	To model data from external systems, the manufacturer shall use manual data entry and spreadsheet based records to hold the data.	TD.01, TD.05
R5	The manufacturer has a set of clinical standing data that was used for another system and derived from real data. It includes encounter codes, clinical terms, consultant names, surgery and hospital addresses etc. and can be reused for this system. The manufacturer's Clinical Safety Officer has reviewed the data and agreed its suitability for reuse.	TD.09, TD.14
R6	Some of the verification datasets shall include errors deliberately inserted to check the effectiveness of data validation.	TD.13
R7	The controlled verification dataset shall be subject to review and analysis against defined confidence / trust criteria. Scripts shall be written to check for syntax and semantic consistency of the data and provide a basic credibility check. The scripts themselves shall be validated and verified before use.	SD.17, SD.20, DI.01, DI.03, DI.11
R8	The project shall be subject to an internal delivery quality assurance audit.	DI.04
R9	Data loaded from external system into the system and displayed to the user shall be crosschecked against the original source data, using manual spot-checks.	TD.15
R10	The level of rigour employed in verifying all the above requirements shall be commensurate with the DSAL criticality and so an ISO9001 compliant quality management system shall be adopted.	All
R11	Data processing events shall be logged to allow monitoring of the health of the system, provide diagnostic information, allow audit trails and support the production of explanations.	SD.11
R12	Each contracted client will be asked to ensure and signoff that test, configuration and adaptation data is appropriate (eg. in terms coverage and representative of real-life values) for their particular circumstance.	DD.05, TD.11, TD.12

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

Table 5.3: Worked example: rejected data safety requirements

Ref	Guidance Reference	Justification	Action
E1	TD.08	The data will be used before any initial run of the system.	Review when data from initial runs is available.
E2	MP.01, MP.02, MP.10	There are no paper based resources for this system.	No further action.

Having determined the requirements arising from the data safety analysis the manufacturer ensures these

are included along with other system requirements as part of the overall delivery and operation of the system. It then remains for the manufacturer to implement and verify methods, that is, as well as defining requirements the clinical risk management plan needs to ensure activities are in place to verify and evidence that treatments have actually been implemented.

Likewise, the health organization has identified DSAL2 data and in deciding to treat the risk, it aims to ensure risks are reduced as low as reasonably practicable.



This page is intentionally blank

6 DSAL Customisation

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

6.1 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 4.1, so it is highly likely that customisation will be required.

6.2 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 4.1 illustrated in Table 6.1

Table 6.1: 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 6.2.

Table 6.2: Likelihood assessment

Low	Medium	High
0–2	3–6	>6

6.3 Weighted characteristics

The method presented in section 6.2 may be enhanced to provide a more holistic overall assessment of the likelihood based on all characteristics. Note that the scheme in section 6.2 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 6.3.

Table 6.3: 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.

7 Tool Confidence and Tool Qualification of Data Processing Tools

> *Don't fix the blame, fix the problem*
> ***Keith Pennington***

7.1 Introduction

A tool is a software that is required to build the product, but itself is not part of the product. For example, a compiler or a test tool.

7.2 What the Standards say

Main safety standards, like IEC 61508, ISO 26262, DO-178, EN-50128 require an assessment of "tool confidence" or the application of "tool qualification" to ensure that a lack of integrity in the tools does not impact the safety of the product.

These safety standards use an approach which may be broadly defined as a three step process:

1. Analyse the risks of the tool, so called "tool classification"
2. Make the tools safe, based on the tool classification, so called "tool qualification", usually by testing the tool. Note if a tool is classified as non-critical, for example, Tool Confidence Level 1 in ISO 26262, it does not need to be qualified.
3. Use the tools as classified and qualified, such as by following a so called "tool safety manual"

The approach to classification differs in different standards and also the proposed qualification methods differ slightly, but all share the same principle: use the tool carefully (safety manual) or qualify that the tool is doing what it should do. More details can be found in the safety standards.

7.3 Data tools

For safety of data the same should be applied to all used tools – they should be classified, eventually qualified and used according to the safety manuals.

A specific approach is not proposed here for data tools, but a recommendation to use the standards with which a product has to be compliant. For example, if data is to be used in the Automotive sector, then consideration should be given to chapters 8–11 "Confidence in the use of software tools" of ISO 26262.

The following generic recommendations are intended to manage tool confidence according to the standards applicable to the domain:

- The tool scope is wider: not only do tools for building the product have to be considered, but also the tools that create the data which impact the system, such as the maps of a self-driving car, or the training data which is used to train AI systems.

- Consider all tools that are working with the data:
  - Creation tools that create the data, for example, tools processing sensor data
  - Data storing tools, that store the data, perhaps into a database
  - Data manipulation tools, that transform data, such as by translation to a different geodetic system, merging of different data sources or the creation of summaries
  - Data visualization tools that show the data – viewers that display the results
- Consider all outputs of the tools. For example an AI training tool usually creates two outputs:
  1. The trained network and
  2. The achieved accuracy.

  Each can be impacted by tool errors and therefore both features of the tool have to be classified/qualified and used safely.
- When working with generic tools like "Excel", "SQL", or "Python", consider every new application as a new use of the tool – in other words, classification of a tool can never be truly generic, but must be reviewed for each specific use. An example of failing to carry out this reassessment is the Covid Summary Table (see Part 3 Section 5.11), which was written / created in Excel using some tables/schemes into which the data was imported and analysed.
- The classification and qualification do not depend on the DSAL, however the DSAL should be considered when testing the tool or when classifying the risks. For example if a tool is to be qualified by testing, it should be done with the same rigour that is required for the DSAL.
- When looking to the risks of tools, it is not sufficient to consider only the functional risks like "wrong data created" (such as the potential that 1+1 could be computed to be 0 instead of 2). In addition, the data-specific risks specified in this document (Subsection 14 Ways that Data Can Cause Problems) should be considered for impact from the tools. Data properties that can be impacted from tools should also be considered. Therefore the following aspects of tools to data should be analysed during tool classification:
  - Each issue listed in Subsection 14; and
  - Each Property listed in Vol. 1 Table 2.1

For all above aspects the impact of the tool has to be considered, for example "Interpretation", the risk of "Wrong Interpretation" can be expressed in the question: "Can the tool change the data, such that it will be wrongly interpreted?". If this is not possible / makes no sense, then the error "Wrong Interpretation", can be classified as "no impact". Otherwise the risk is valid and has either to be mitigated by a safety guideline in the safety manual, or it should be excluded by systematic qualification tests which show that the error "Wrong Interpretation" cannot occur.

Part II

Additional Data Safety Considerations

8 AI and Machine Learning

> *The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency.*
> ***Bill Gates***

8.1 Machine Learning Training Data

A complete discussion about the assurance of ML /AI algorithms is beyond the scope of this document. However, as such algorithms are inherently data-driven, the data safety guidance provides the following notes on good practice for datasets used for ML / AI applications.

The data for real-world ML scenarios may be effectively infinite (for example, with self-driving cars, where the environment is largely unbounded) and can only be approximated by a finite number of scenarios. Therefore it is crucial to have an extension process in place to extend the system when previously unknown situations are experienced. ISO 21448 [10] captures this process by introducing the term safety of the intended functionality (SOTIF) and requires the systematic handling of unknowns. Although specifically aimed at road vehicles, SOTIF may be of interest to other domains.

For more general guidance on managing the safety of autonomous systems, please refer to the safety assurance objectives guidance produced by the SCSC Safety of Autonomous Systems Working Group (SASWG) [11].

8.2 Lifecycle Data Categories

Three lifecycle data categories are discussed: ML training data (the data used to train the model), ML test data (used to provide a measure the trained model's accuracy), and ML validation data (used to independently verify the trained model).

8.3 Real-world Data

It is important to obtain sufficient quantities of data for the training data. It is generally assumed that this is sufficiently representative of situations that could be encountered to allow the algorithm to generalise the appropriate behaviour for all encountered situations. However, there are problems with real-life training data. Firstly, it is possible that rare, hazardous cases are never encountered during the capture of a dataset (as those events have been engineered to be statistically unlikely) and are too dangerous, or expensive, to generate deliberately. Poor representation of rare cases in training data will result in a model that is insufficiently prepared to handle those important scenarios. Poor representation (or weighting) of rare cases in ML test data or ML validation data can result in error heuristics that do not provide adequate insight into the system's performance in rare situations. A second issue with real-world data is that amassing, understanding, verifying and validating it at scale is resource intensive. Hence it is not just a case of more testing, a more cost-effective approach is needed. Some possible techniques are: artificial ML test dataset generation, data fault injection and simulation. It may well be more practicable to use simulated training data during the training phase to give more control and avoid the cost and effort of real-world testing. The issue then is how realistic the simulated data (and the associated simulation) can be. Note that artificial

datasets can also be used to increase the frequency of rare events, thus biasing the algorithm to managing those rare events effectively, but doing so, could impair the "sunny day" performance of the system.

8.4 Sensor Data

In many cases sensors will be on-board and embedded (e.g. on an automated vehicle) and will be subject to many real-world influences. Sensor data may be changed or degraded by both internal influences, e.g. system environment temperature, and external factors e.g. humidity, ambient light, weather conditions such as snow or fog, and rarely encountered extremes such as lightning or flood. Sensors may also suffer degradation due to ageing and specific faults due to their positioning, construction or configuration. Sensors may need to be diverse to increase resilience and give an overall more reliable output. Realistic data variations, informed by an understanding of likely sensor failure modes, need to be incorporated in the training dataset (c.f. the AoA sensor in the Boeing 737 MAX 8, discussed in Part 3 Section 5.12). Sensor fusion can effectively create "composite data", e.g. from merging of lidar, radar and camera data; this will have different (and possibly hard to predict) properties compared to the individual sensors.

8.5 Data Coverage

Some notion of data coverage, that is, how well the training data covers the intended operational domain, is needed: this could be based on use of all known datasets, all known hazardous situations or some other measure. White box techniques involved in exercising all connected sensors or coverage of nodes in a neural network may have some value. Regardless, it will be essential that the training data covers all required safety test scenarios.

8.6 Dataset Assurance

There is a need to have assurance about the datasets themselves, including their configuration management. Assessment of the quality and suitability of the training datasets should be included in the safety argument.

The University of York has developed a method for the assurance of machine learning for use in autonomous systems (AMLAS) [12]. The University's approach is intended to assist in the development of a compelling argument about a machine learning model, to feed into a system safety case.

8.7 Data Diversity

The degree of data diversity is important. Common data items across datasets need to be avoided, especially across training and ML validation data. Common data may include data derived from the same sources, environments, simulators, tools, techniques or mathematical models.

8.8 Data Properties

Bias, sufficiency and accuracy are all concerns with the data. fidelity / representation may also be a problem as modelling real-life sensor data is difficult. Errors should be introduced for training: this may be by

analytical means, by random, automated sampling of data to determine the degree of similarity between samples, or by deliberate insertion of known edge cases. It is suggested that all the data properties mentioned in this guidance document are considered.

8.9 Hallucinations

The human brain can see and interpret patterns in unexpected ways, for example seeing a shadow that suggests a monster is waiting to pounce, or seeing an animal in a cloud formation[1]. AI tools can similarly interpret data in novel and unplanned ways, perhaps seeing patterns that a human would not perceive, or seeing shapes in noise. However with AI tools, hallucinations can appear in two different ways – in the interpretation of the question or instruction to perform, as well as in the interpretation of the source data. The problem is exacerbated in the case of AI "chatbots", where the output presentation in natural language can make the reply seem authoritative and result in it being trusted, as though the AI chatbot is an expert in the field under consideration [13].

The complete eradication of hallucinations may not be possible, so it can be dangerous to rely fully on the output of an AI without verification. Biases in the training data are regarded as a significant cause of hallucinations, so selection of training data should include consideration of its quality – whether it is diverse, balanced and well structured [14]. In addition, evaluation of the resultant behaviour should be assessed by humans, to achieve "reinforcement learning with human feedback" [15]. The results of the assessments can then be fed back into the model, thus fine-tuning it for its specific application.

Note that fine-tuning will generally only be possible for a specific application and / or a specific environment. As with all safety analyses, it is critical that these be clearly defined.

We asked two different AIs to comment on the risks that hallucinations present to data safety. Their replies were as presented in the following two boxes:

Response of the AI

AI hallucinations can have significant implications for data safety and overall trust in AI systems. Here are some key concerns:

Misinformation and Disinformation: If an AI model generates incorrect or misleading information (hallucinates), and this information is used without verification, it can lead to the spread of misinformation. In sensitive areas like healthcare, finance, or law, this could have serious consequences.

Decision-Making Risks: In scenarios where AI is used to aid decision-making, hallucinations can lead to incorrect or risky decisions. For instance, if a financial AI advises on investments based on hallucinated data, it could lead to substantial financial losses.

Trust and Reliability Issues: Frequent hallucinations can undermine the trust in AI systems. Users and organizations may become reluctant to rely on AI for critical tasks if there's a significant risk of the AI generating false information.

Security Vulnerabilities: AI hallucinations can be exploited in cybersecurity. For example, if an AI system responsible for network security misinterprets data or hallucinates a threat, it could either overlook actual threats or allocate resources to non-existent threats.

[1] These are not strictly hallucinations in the medical sense, but serve to illustrate the issue with AI interpretation of data.

Privacy Concerns: There's a risk that an AI model might inadvertently reveal or imply private information in its outputs, especially if it has been trained on sensitive datasets. While not a hallucination in the traditional sense, this is related to the broader issue of how AI handles and interprets data.

Ethical and Legal Implications: AI hallucinations can lead to ethical and legal challenges, especially if they result in harm or damage. Determining liability in cases where AI-generated misinformation leads to real-world consequences is complex.

Bias and Fairness: Hallucinations can also be a reflection of biases in the training data. An AI might generate outputs that are unfairly biased against certain groups or individuals, leading to ethical concerns and potential legal issues.

To mitigate these risks, it's crucial to have robust validation and verification processes, use AI in conjunction with human oversight, continually update and improve AI models with diverse and accurate data, and maintain transparency about the limitations and capabilities of AI systems.

Response of the AI

The implications of hallucinations for data safety are complex and far-reaching, spanning aspects of privacy, security, and misinformation. Here's a breakdown of some key concerns:

Privacy: False generation of personal data: AI hallucinations can fabricate personal information (names, addresses, financial details) that doesn't actually exist, potentially violating privacy rights and harming individuals. Exposure of sensitive data: If trained on sensitive data, hallucinations might leak this information unintentionally, even if not explicitly instructed to do so. This could include medical records, financial data, or other confidential information. Security:

Creating vulnerabilities: Hallucinations might invent security flaws or exploits that don't actually exist, leading to wasted resources and potentially even vulnerabilities being introduced into real systems. Misleading security assessments: If used for security testing, hallucinations could create false positives or negatives, providing inaccurate assessments of a system's security posture. Misinformation:

Spreading false information: AI-generated hallucinations can easily spread misinformation, as they appear convincing and authoritative even if factually incorrect. This can have significant implications for public discourse, trust in institutions, and even manipulation. Evolving deepfakes: Hallucinations could be used to create even more sophisticated deepfakes, making it harder to distinguish between real and fake content, further eroding trust and potentially influencing elections or other high-stakes situations. Additional considerations:

Impact on trust: As reliance on AI grows, hallucinations can erode trust in its outputs, making it harder to distinguish reliable information from fabricated content. Regulation and accountability: Establishing clear guidelines and accountability mechanisms for AI systems prone to hallucinations is crucial to mitigate their risks. Overall, addressing the implications of hallucinations for data safety requires a multi-pronged approach:

Developing robust and transparent AI systems: Techniques like adversarial training and explainable AI can help reduce hallucinations and improve transparency. Implementing data governance and privacy measures: Strong data protection frameworks and responsible data handling practices are essential to minimize the risk of sensitive data exposure. Promoting media literacy and critical thinking: Educating users to critically evaluate information, regardless of its source, is crucial in the age of

AI-generated content. Establishing ethical guidelines and regulations: Clear ethical frameworks and regulations for AI development and deployment are necessary to ensure responsible use and mitigate potential risks. By acknowledging the dangers of hallucinations and taking proactive steps to address them, we can minimize their impact on data safety and build trust in AI-powered technologies.

8.10 Dataset Management

Datasets will be large and complex and ensuring quality and consistency will be difficult. Measures will have to be taken to control, manage and archive large amounts of data. Datasets should be subject to formal change control and monitoring. Tools will be needed to manage the data.

8.11 Optimization

What is an optimal size of a training or ML validation dataset? A dataset may contain many repeated or redundant values. If these do not add to the training outcome it is possible that some values may be removed. Also if there are ways in which the dataset could be compressed or optimized then these should be considered to assist management and curation.

8.12 Compliance

A "safety by compliance" strategy for such data is difficult to implement as there are currently few existing standards and little guidance for data in this area. It is therefore difficult to quantify data sufficiency and as a result, no generally agreed "ALARP" principle for data.

It may be possible to demonstrate Data Safety through statistical methods, such as when millions of representative miles have been driven. However such approaches have generally been found insufficiently robust to support safety-critical systems in other domains such as aviation, and so must be treated with caution. The demonstration of compliance for the data is likely to require a compliance argument for the tools that produced the data, the preservation of data properties as described within this document, and an update process such as that specified in ISO 21448 (SOTIF) [10].

8.13 References

This appendix provides only a brief overview of the issues around the application of data safety techniques to machine learning. It is based upon information in the following documents, which should be consulted for further details:

- Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges [16]
- Quantifying Data Set Properties for Systematic Artificial Neural Network Classifier Verification [17]
- Safety Critical Integrity Assurance in Large Data Sets [18]

This page is intentionally blank

9 AI and autonomy

> *When building a god, we must be very, very careful.*
> ***Christopher DiCarlo***

9.1 Introduction

This section describes a range of harms that can arise from AI/ML systems. These are potentially serious because the outputs of AI/ML systems can be presented in à very credible way, leading people to make bad decisions based on unreliable outputs. There is a database of AI incidents at `https://airisk.mit.edu/ai-incident-tracker`.

9.2 Job Displacement

AI and automation can lead to the displacement of jobs, as machines can perform some tasks more efficiently than humans. This can have significant socioeconomic implications, including increased unemployment and wage stagnation in affected industries. The International Monetary Fund (IMF), in January 2024, gave an estimate that AI will impact 40% of jobs globally, and 60% of highly skilled work, ([19]). The report balances the already known complementary nature of AI to human work against the likely detrimental aspects. Unusually for technological advances, AI is expected to impact high-skilled jobs to a greater extent than manual skills. The report states, "advanced economies face greater risks from AI – but also more opportunities to leverage its benefits". Labour requirements could significantly reduce, wages could be lower and of course jobs will disappear. Managing director of the IMF, Kristalina Georgieva, said that "AI will likely worsen overall inequality , a troubling trend that policymakers must proactively address to prevent the technology from further stoking social tensions." ([20]). How economists are changing the standard growth view of economics and how engineering and other businesses can engage with such developing problems is discussed in an article in the February 2024 edition of the SCSC Newsletter ([21]). The data used to determine the impact of an AI system, especially when it affects currently underdeveloped nations, must ensure that the social tension Georgieva refers to does not reach a breaking point. Like well digging, rail construction and careful dam placement, AI technology could raise many millions out of poverty if it is applied to benefit life on earth.

9.3 Bias and Discrimination

AI systems can inherit biases present in their training data, leading to discriminatory practices. For instance, facial recognition software has been shown to have higher error rates for people of certain racial and ethnic groups. At a very basic level, slight imperfections in coatings for car registration plates, easily ignored by humans, has made some number plates unrecognisable to automatic number plate recognition cameras at car parks. This can be easily missed in training the system since it would have been difficult to anticipate every possible imperfect state of a number plate. In this case advances are made from experience but there are many situations in which day one of a system's deployment has to be near perfect. Training data sets should be vetted by committees, as safety usually is and should be always. That technique is more likely to bring out overlooked areas, obvious to someone on the committee but totally missed by others. Testing of developed systems should also cover the full range of possibilities, whether or not those involved in the development provide assurances, as assurances can hide failings.

9.4 Errors of Omission

Because outputs of AI systems can be highly convincing and credible, it is very hard to see what has been left out. This can be in the summary outputs that are produced or in the input data used to train systems such as the absence of some types of road user in autonomous vehicle training data. Data which is missing from AI training data will not be taken into account in the AI outputs.

9.5 Loss of Human Skills

With AI taking over tasks such as navigation or memory-dependent activities, there is a risk that humans may lose certain skills that are not used regularly, potentially reducing cognitive abilities over time. However, the opposite is also true because, driven by the threat of automation, many people could turn to artisan skills to create unique or less ubiquitous products providing an element of appealing individuality. This could turn supermarket checkout staff into, for instance, entrepreneurial upholsterers or cosmetic developers when and where all checkouts have become automated. Regardless of whether or not skills are lost, it is certain that in order to verify and validate data intensive artificial intelligence systems someone will be required to maintain those basic skills to provide the assurance. To make a simple analogy, one cannot validate a French grammar correcting AI system if one does not speak French fluently.

9.6 Security Risks

AI can be used to develop sophisticated cyber-attacks, and AI systems themselves can be vulnerable to such attacks. The integration of AI into critical infrastructure heightens the potential impact of these risks. The issue of autonomous vehicles in military applications has been discussed in an SCSC newsletter, ([22]), which quotes "Formal Verification of Ethical Choices in an Autonomous System" ([23]): "All participants in society are required to follow specific regulations and laws. An autonomous system cannot be an exception". Given this as an accepted principle, it is then incumbent on engineers to design systems which follow ethical regulations. This presents difficulties and it demands that the most sophisticated sensors be employed on killer robots when decisions will need to be made as to whether or not to engage in an action which may result in death or excessive damage. Such examples are, firing a missile at a market square or bridge where known enemies are observed. At what point does the death of others not engaged in wrong doing become acceptable, known as collateral damage. It is important to stress that the data used to train systems covers all possibilities, just as in human vs. human conflict. If the machine is not programmed to make the decision, the machine must request clarification from human controllers. The issue is that it will almost certainly prove impossible to develop military killer systems that can be programmed with sufficient data to make every decision required in all circumstances without needing to consult humans, though direct accurate hits on isolated targets with little risk to the combatants of the targeting force is likely to remain an acceptable usage. But it would be sad if the governments of the world end up pitting machines they don't understand against each other with all the horrendous possible consequences.

9.7 Privacy Erosion

AI's ability to analyse vast quantities of personal data can lead to erosion of privacy. For instance, AI can be used to make highly accurate predictions about individuals' behaviours, preferences, and even future actions. To a limited extent this is already operational and desired by police forces wanting to identify faces and even the gait of wanted persons. The usual claim is that "If you are not doing anything wrong you have

nothing to fear". But many democracies are turning against the principle of eroding privacy, considering some policing techniques AI makes available are too intrusive. Even when the nation is particularly controlling, the creation of the network of cameras across a nation does provide the environment for a takeover by controlling elements. The Metropolitan Police of London are using live facial recognition (LFR) cameras ([24]). Their policy is "LFR cameras are focused on a specific area; when people pass through that area their images are streamed directly to the live facial recognition system. This system contains a watchlist: a list of offenders wanted by the police and/or the courts, or those who pose a risk of harm to themselves or others." They currently assert that "LFR is not a ubiquitous tool that uses lots of closed circuit television (CCTV) cameras from across London to track every person's movements". But a change of national or local government could change that. The risk is that data may be abused or false.

9.8 Control, Autonomy, Ethical and Moral Considerations

As AI systems become more autonomous, there is a risk that they may act in unforeseen ways that are not aligned with human intentions or may be manipulated to act against human interests. Also there are significant ethical questions around AI, including the morality of decisions made by AI systems, particularly in life-and-death situations such as in autonomous vehicles or military applications. Although mainly concerned with military questions "The Troubling Aspects of Autonomy" ([22]) discusses the issues of control that arise for dangerous life-threatening situations and is pertinent to non-military systems.

9.9 Economic Inequality

The benefits of AI may accrue disproportionately to those who own the technology, potentially exacerbating economic inequality. Companies and nations that can invest in AI could gain significant economic advantages, leaving others behind. The likely beneficiaries are the developed economies. 40% globally will be affected according to an IMF discussion paper and 60% in the advanced technology world. In the fast world of competitive developments in AI which has now begun, it is difficult to believe that the potential pitfalls are ever fully examined when, just as in more conventional engineering systems, the bottom line remains money gained through business success. Only regulation by governments and other authorities can be expected to be effective in limiting damage outside the bounds of AI systems. Some are now advocating a new form of economics known as Doughnut Economics as advocated by Professor Kate Raworth, senior associate at Oxford University's Environmental Change Institute in her book on the subject ([25]). A brief discussion of the some of the issues of such economics can be found in ([21]). There could be ways to support business in developing countries if AI tools are made available on smart phones at low cost to, for instance, African women's business collectives. Such a system may just have emerged with China's DeepSeek tool. Data input and received from such applications will need to be safe and accurate if it is to benefit the less well-off of the world. Businesses must have good data to when using AI that provides reasonably accurate appraisals of outcomes from various actions. Ultimately, nobody should be left falling short of the social foundation that the doughnut economy declares the future 'bottom line' to be. In the appendix to Raworth's book, a table is given showing the issues that need addressing and declaring the sources. Such sources should be consulted for trusted data when new projects are developed.

9.10 Dependency

Over-reliance on AI can lead to a lack of preparedness when systems fail. If critical infrastructure or services are AI-dependent, outages or malfunctions could have severe consequences. Simple growth of single point

failure understanding applied to AI should mitigate a lot of the risk of loss of urgent data, such as fly-by-wire aircraft having three computers with a voting sub-system in overall control.

9.11 Manipulation and Fake Content

AI can be used to create deep-fakes and synthetic media, which can be used to manipulate public opinion, perpetrate fraud, or spread misinformation. The Partnership on AI (PAI), ([26]), has pursued the concept of there being a spectrum of harm from manipulated media. How this will all resolve is as yet unknown; the problem is in its relative infancy. With partners, a working definition arrived at is "any image or video with content or context edited along the 'cheap fake' to 'deepfake' spectrum with the potential to mislead and cause harm" ([27]). A diagram and explanation of the spectrum can be found at Data and Society's website ([28]). Twelve principles for labelling manipulated data can be found at the PAI on AI ([29]) . As to whether any of this is enough, only time will tell. Much is seen in the news of demands for content providers to do better at eliminating bad data, be it manipulated or such things as sites encouraging suicide. However, it is worth noting that The Future of Humanity Institute (FHI), a research group within the Oxford Martin School, has joined the PAI and that organisation has as members many of the known big players in the data intensive media field: Amazon, Apple, Google/DeepMind, Facebook, IBM and Microsoft. There goal is to formulate socially beneficial best practices for AI development. How much their belief in their achievements coincides with politicians requirements we must wait and see. The PAIhas produced a document "PAI's responsible Practices for Synthetic Media", ([30]), which argues the best way to overcome fake media is by the organisations not trying to deceive anybody to have codes of practice. They seek to advance ethical and responsible behaviour. Synthetic media, also known as generative media is defined as "..visual, auditory or multimodal content that has been generated or modified, commonly by artificial intelligence."

Legitimate use of such media may be, for example, entertainment, art, satire, education, and research. Furthermore, techniques can be used legitimately or harmfully; there is no barrier. One recommendation for technology and infrastructure creators is "aim to disclose in a manner that mitigates speculation about content, strives towards resilience to manipulation or forgery, is accurately applied and also when necessary, communicates uncertainty without furthering speculation". The guide goes on to recommend practices for 'Creators' and 'Distributors and Publishers'. Transparency and disclosure are the main thrusts. It therefore appears that the legal definition for consumer risk will mirror goods purchase: *caveat emptor*, let the buyer beware, but with safeguards provided by companies wishing to take ethical stances.

9.12 Existential Risk

In the long term, some theorists and scientists worry that the creation of super-intelligent AI could pose an existential risk to humanity if it becomes uncontrollable or its goals are misaligned with human values. This particular aspect of data safety with AI systems shows how important it is that the data used in AI systems reflects humanity's values. The doughnut economic model uses, as the social foundation for humanity, twelve of the seventeen goals of the United Nations. Any AI system that is not trained, where relevant, using data that will ultimately achieve those goals, is almost certainly not an ethical system designed to align with the established values of the global humanitarian consciousness. Also, many systems capable of destroying that have been granted autonomy, such as weapons, need to have been trained using the relevant data reflecting relevant guidelines (in the military case, on conflict rules). Issues related to the military problems are discussed in [22].

9.13 An AI's view of the risks

We asked an AI "What sort of mistakes due to AI can be hard to detect?." The AI replied as follows:

Response of the AI

Artificial intelligence (AI) systems, while powerful and increasingly integral to many aspects of modern life, are not infallible. There are several types of mistakes or errors in AI systems that can be particularly hard to detect:

Subtle Data Bias: AI algorithms are often only as good as the data they are trained on. If this data contains biases, the AI can perpetuate or even amplify these biases. These biases can be subtle and deeply ingrained in the data, making them difficult to detect without thorough analysis.

Overfitting: This occurs when an AI model is too closely tailored to the specifics of its training data and fails to generalize well to new, unseen data. Detecting overfitting can be challenging because the model may perform exceptionally well on training data but poorly in real-world applications.

Black Box Models: Many advanced AI models, like deep neural networks, are often considered "black boxes" because their decision-making processes are not transparent or interpretable. This lack of transparency makes it difficult to understand why the AI made a certain decision and, consequently, to identify when it makes mistakes.

Contextual Misinterpretation: AI may not fully understand or consider the context in which data exists, leading to misinterpretations. These errors can be hard to detect because the AI's output may be logically consistent within the data it has, but contextually inappropriate.

Feedback Loops: In systems where AI output influences future input data, feedback loops can occur. For instance, in content recommendation algorithms, the system can get trapped in a loop of reinforcing its own biases, making it hard to detect errors in content filtering or recommendation.

Anomalous Data or Edge Cases: AI systems might struggle with data or situations that are significantly different from what they were trained on, known as edge cases. These anomalies can lead to errors that are hard to anticipate and detect because they occur outside the 'normal' operating parameters.

Errors in Complex Interactions: In systems where multiple AI models interact, errors can arise from the complex interplay between these systems. These errors can be particularly challenging to detect and diagnose because they may not arise from any single model but from the interaction between models.

Evolution of Language and Social Norms: AI systems trained on language data can become outdated as language and social norms evolve. Detecting when an AI system is no longer accurately interpreting or generating language according to contemporary standards can be difficult.

Misalignment with Human Values: AI might make decisions that are technically correct but misaligned with human ethics or values. These types of mistakes can be particularly hard to detect because they require nuanced understanding of human ethics, culture, and values.

To mitigate these issues, continuous monitoring, testing, diverse data sets, and incorporating explainability in AI systems are crucial. Additionally, involving human oversight and interdisciplinary expertise can help in detecting and correcting such hard-to-detect mistakes.

This page is intentionally blank

10 Lifecycle Data Categories

> *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 10.1 provides information, in the form of explanations and lists of typical containers, for the identified lifecycle data categories.

Table 10.1: 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 10.1: 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 10.1: 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 10.1: 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 10.1: 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

11 Data Migrating, Porting, Importing and Exporting

> *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 chapter highlights some of the issues that arise, links to the data safety properties and suggests some mitigations.

11.1 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 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.

11.2 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 11.1):

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.

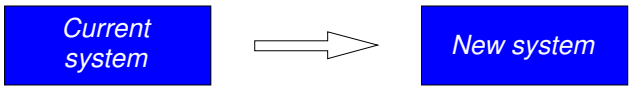


Figure 11.1: Simple migration path

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.

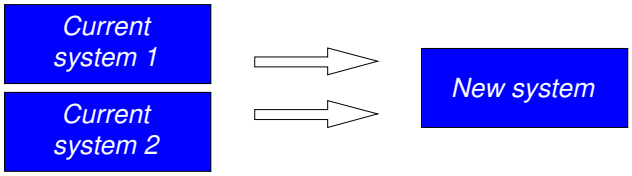


Figure 11.2: Many-to-one migration path

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

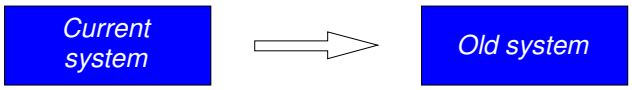


Figure 11.3: Reversion migration path

11.3 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.

Table 11.1: Safety issues due to migration

No.	Migration Case	What are safety issues in each case?	Mapping to properties/loss of properties	Possible Mitigations (Table 4.8)
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 11.1: 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 4.8)
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

12 Dark Data and Dazzle Data

> *…we can be blind to missing data … that can lead us to conclusions and actions that are mistaken, dangerous, or even disastrous.*
> ***David Hand***

12.1 Introduction

Dark and dazzle data may be viewed as two sides of the same coin. In the former data is hidden, whilst in the latter the important information is hidden by an abundance of other data.

12.2 Dark Data

12.2.1 Introduction to Dark Data

This section outlines some of the safety implications arising from the issues identified by Prof. David Hand in his work on "dark data" presented in his book: "Dark Data: Why What You Don't Know Matters" [31] and website [32]. Dark data relates to data that is not available, but nevertheless is important, and indeed, in some cases, more important than the data that is available.

Unintentionally Donald Rumsfeld popularised the concept of dark data with his famous speech:

> *"…there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know…"*

Figure 12.1 neatly shows the dark varieties in grey:

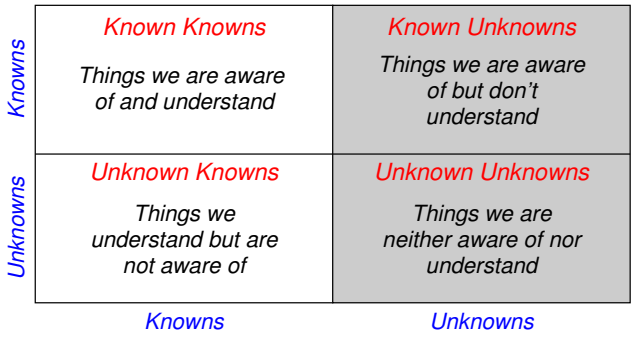


Figure 12.1: Varieties of dark data

12.2.2 Dark Data Example

A meeting was held prior to the fatal Challenger disaster in January 1986, to decide whether the flight should go ahead due to the prevailing cold weather and concerns over the performance of O-ring seals at low temperatures. The decision to launch was made on the basis of the following data points that show the temperature at which previous O-ring problems had occurred.

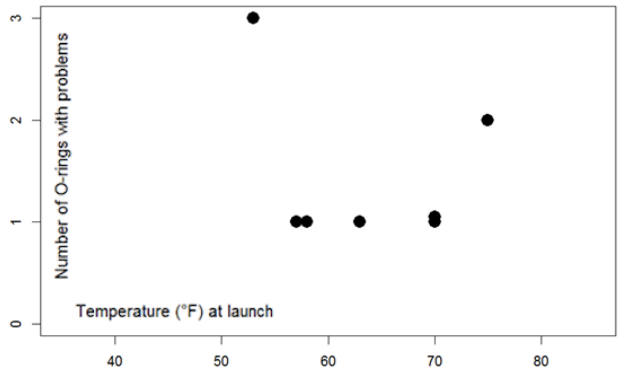


Figure 12.2: O-ring failures by temperature

On this basis, it was concluded that: "there is nothing irregular in the distribution of O-ring 'distress' over the spectrum of joint temperatures at launch between 53 degrees Fahrenheit and 75 degrees Fahrenheit." However, there was dark data in this dataset – the data points that show the temperatures for successful launches when there were no O-ring problems were left out. These are shown as the red points in Figure 12.3.

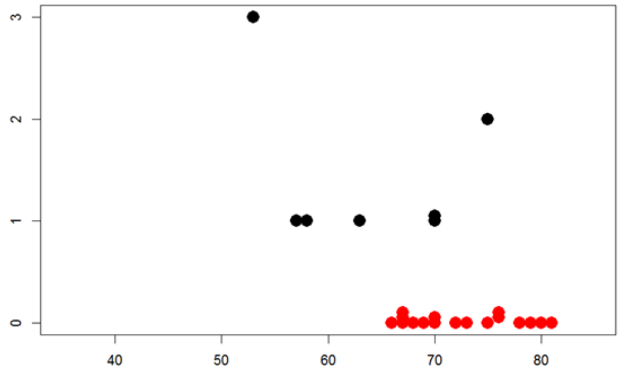


Figure 12.3: O-ring performance by temperature

When the Dark Data is included, there is an obvious correlation between temperature and O-ring failures and it is possible the decision to launch may have been different if this additional data had been presented.

12.2.3 Dark Data Varieties and Safety Examples

In his book, David introduces a taxonomy of 15 varieties of dark data. This section discuses each of these from a data safety perspective.

12.2.3.1 Data We Know Are Missing: "Known Unknowns"

This case is very common in safety justifications where assurance information may be withheld for commercial reasons or does not exist, but we know, or are informed, that it isn't available. Common examples include:

- Assurance information for COTS components is unavailable
- Information about legacy systems was never produced or is now lost

- ML training data restricted to a particular context of use

If this is the case, it can be mitigated in several ways, including use of warnings, training, restrictions of use, etc. Evidence can also be substituted with other more indirect assurance e.g. established organisational track-record in the sector, or audit reports.

12.2.3.2 Data We Don't Know Are Missing: "Unknown unknowns"

This is the most serious and far-reaching case. Occurrence is hopefully less common than case 12.2.3.1, but it is important to acknowledge that it does happen. Some examples are:

- The Covid-19 Track and Trace data loss (Part 3 Section 5.11), where the organisation handling the data was unaware that rows were missing from a spreadsheet for some time is illustrated in Figure 12.4
- Somebody knows a problem with a system but does not tell
- Machine learning training data missing edge / corner cases
- Key safety requirements missed
- Change of use never anticipated
- Data loss which may be discovered after some period (or indeed never)
- Test cases never thought of, so never created or executed

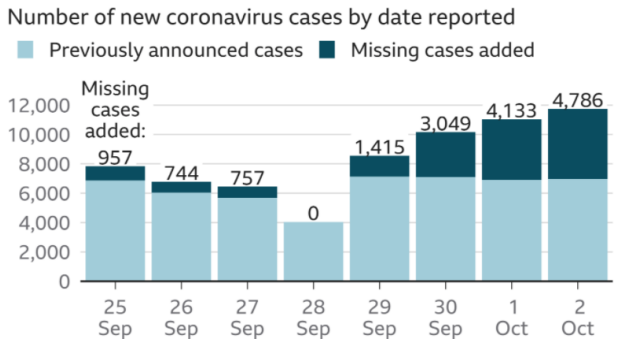


Figure 12.4: Covid-19 track-and-trace Data Loss

In many cases the loss may be discovered after some time, and it is incumbent on the organisation involved to analyse the impact of the missing data over the time period, including subsequent decisions and actions. Its effects should not be underestimated. This case can fundamentally change the safety picture and is probably in the highest safety risk category.

Data that was missing can subsequently be found. The following options show possible approaches to handling this "rediscovered data":

i) Apply the missing data

ii) Ignore the missing data

iii) Mention that the data was missing but not take it into account, and

iv) Perform an impact analysis on the missing data and then act according to the results

12.2.3.3 Choosing Just Some Cases

This is where something or somebody has been selective. Examples might be:

- Selection of test runs that succeeded (ignoring failed runs and their diagnostics)
- Selective sampling from sensors, or where the sampling intervals are chosen badly
- Incorrect filtering of the data, leaving out more cases than intended
- Using data from completed forms, finished tasks or only including data which, intentionally or not, meets some criteria which excludes the important data. This can lead to "Survivorship Bias", [Wikipedia ref: https://en.wikipedia.org/wiki/Survivorship_bias] of which the following article has a good example from WW2 aircraft where examining the bullet holes on aircraft returning from missions could have led to the wrong conclusion: https://www.dgsiegel.net/talks/the-bullet-hole-misconception

Note that with complex or informal criteria the effects could be as case 12.2.3.2, i.e. you don't know what has been left out.

Mitigations include use of peer review, independent teams, and audits. It is important to ask questions and challenge the data selected to be sure it can be justified.

12.2.3.4 Self-Selection

This is considered to be similar to case 12.2.3.3, but could be even more informal or ambiguous. Again mitigations include: use of peer review, independent teams, and audits.

12.2.3.5 Missing What Matters

This was considered similar to case 12.2.3.2 in impact terms, and might well be considered the "Elephant in the Room". Examples might be:

- Measuring the wrong things, e.g. poor safety metrics / indicators
- Too much data to deal with or analyse, so some is ignored
- Too much filtering or processing, so losing information along the way
- Being too close to the data, i.e. the "wood for the trees". This is when the detail masks the overall issue with data, e.g. a slow trend or bias masked by peaks.

Mitigations include independence as 12.2.3.3 and 12.2.3.4 but also "taking a step back" to look at the bigger picture.

12.2.3.6 Data Which Might Have Been

This is where it is impossible to obtain the relevant data as a consequence of how the system / scenario is constructed. This was considered an interesting and important case.

This could for example, be due to an inappropriate system architecture e.g. single data channel when a multiple channel approach should have been used, such as in the Boeing 737 MAX 8 accidents (see section

H.2). It is therefore important that Mitigations are in place at design time, as they are often very difficult to retro-fit.

Mitigations include assessing the data collection opportunities gained or lost from the design or architectural approach – at design time.

12.2.3.7 Changes with Time

This is a common problem in a safety context. Data in safety systems often becomes obsolete or out of date and may still be mistakenly used. Some examples are:

- System configuration data not kept up to date as software or hardware changes
- Medical drug interaction databases
- Software patches requiring updates to configuration or system data, which is not always done

Mitigations include keeping critical data up to date and regularly refreshed, and knowing how old the data is in the first place.

12.2.3.8 Definitions of Data

This was considered a common case for systems with databases or those exchanging data with external systems, e.g.

- Taxonomies e.g. in machine learning categorisation schemes
- Data schemas in medical record systems
- Interface specifications and data definitions, which often evolve over time. These can render old data obsolete / subject to misinterpretation, and often needing migration / translation or re-interpretation.

Mitigations include keeping a list of known changes / incompatibilities and documenting the changes or fixes that have to be applied to make the data usable or consistent.

12.2.3.9 Summaries of Data

This is often seen in data about safety systems and projects:

- Safety metrics or indicators where data is aggregated to create a composite value
- Data fusion across multiple sensors Summaries can be misleading or cause "boundary reactions", e.g. Red-Amber boundary.

Mitigations include thinking about what missing data could cause (i.e. performing sensitivity analysis) – for instance, what decisions might be made erroneously due to certain data values being omitted, late or slightly perturbed.

12.2.3.10 Measurement Error and Uncertainty

Sensors can degrade and fail over time, especially in harsh environments such as automotive, marine or aviation:

- Sensors may feed faulty or biased data into systems
- Sampling techniques can cause some artefacts in themselves
- Interval polling interval can lead to misleading readings
- Data fusion can be impacted if multiple sensor values are combined

Hence sensors need to be regularly maintained and calibrated or replaced, and their interfaces need to be able to detect any problems and report faults.

12.2.3.11 Feedback and Gaming

This can often happen in safety justifications or test case production:

- Early production of a safety argument could lead to only those artefacts that support the claim being generated, e.g. requirements-based testing vs. stress testing
- Confirmation bias in safety justifications and can cause over-optimism

Mitigations include the use of dialectic argument approaches (where both positive and negative cases can be considered). It is also useful to get a second opinion, independent review or a 'fresh pair of eyes' to check the justification.

12.2.3.12 information Asymmetry

This is common where there are multiple stores or sources of the same data:

- Multiple / backup databases where they are not kept in sync
- Image recognition systems that learn, but fail to share data
- Issues of divergence of data across multiple sources.

Comparisons, reviews and data audits may help in these cases. In general the issue of divergence across multiple data sources needs to be recognised and addressed in the most appropriate technical, and ideally automated, way.

12.2.3.13 Intentionally Darkened Data

This can and does happen, e.g.

- Defence, security and government sectors where data is purposefully hidden or destroyed
- Records intentionally deleted after an accident to cover up what really happened

This can be mitigated using appropriate technical measures (e.g. blockchain, digital signatures, off-site backups), and also robust procedures with oversight.

12.2.3.14 Fabricated and Synthetic Data

Fabrication (i.e. making up the data) is surprisingly common, e.g. in medical, policing and maritime sectors. Sometimes it is created with the best of intentions as it was not produced or collected at the right time and is now required (e.g. for an audit), but sometimes it is created to mask a problem. Synthetic data is often used where there are difficulties in producing enough real data with the right characteristics:

- Data is retrospectively entered / patched to make a "clean" record
- Synthetic autonomous vehicle training databases can have issues with artificial data if not realistic

Possible mitigations include comparisons with current real-world data and checking with historical data.

12.2.3.15 Extrapolating Beyond Your Data

Systems, especially machine learning systems (see Appendix J) have to cope with values outside of their training data, but the outcomes may be unexpected:

- Bizarre results from recognition / detection systems may result

Mitigations include use of machine learning training data containing edge / corner cases, boundary cases, and testing well beyond normal or acceptable ranges to give insight into the system degradation behaviour.

12.2.4 Summary of Dark Data

Dark data is a very useful way of thinking about problems and solutions. It is important to always think of the bigger picture, considering what information is not known:

- What might have been left out, intentionally or otherwise?
- What might be missing due to the way things are done (or the way those things are being executed)?
- Could missing data items cause any safety problems?
- What is outside of the defined operational domain?
- Can you present your results / outputs in a way that shows the dark data?

12.2.5 Further Reading

A number of resources are available for further information on how dark relates to data safety:

- The October 2020 edition of the SCSC Newsletter [33] contains articles both by David Hand and on dark data safety.

- A presentation on dark data given by David Hand to the DSIWG [34];
- A presentation by Mike Parsons given at the University of York [35].

12.3 Dazzle Data

12.3.1 Introduction to Dazzle Data

"Dazzle data" refers to data hiding by spurious, superfluous or unwanted data; this obscuring data confuses and can reduce the ability to see the whole picture. The data which is obscured becomes 'dark' because you can't see it or use it. Examples of dazzle data might be electrical noise masking a weak signal or camouflage hiding military assets. Dazzle data may be data that you don't want, arrives more frequently than expected or data that arrives at an unexpected or inconvenient time.

At first sight it might be thought that dazzle data hiding the wanted data is just an annoyance rather than a problem; however, this can lead to prevention of use of the real data. It can hide the true picture as in "wood for the trees" and in the worst case create serious problems. Examples include false positive alerts that cause real alerts to be ignored as in the classic case of "Crying Wolf". Another example is a Denial of Service (DoS) attack, when servers can be blocked with useless requests that cannot be distinguished from the real requests. Data which is truly missed (e.g. masked by extensive electrical noise, or hidden by fabricated data to mask a fraud) could lead to potentially hazardous results.

Varieties of dazzle data may be identified in the same grid manner as Dark Data, as shown in Figure 12.5, where the yellow areas indicate potential dazzle data.

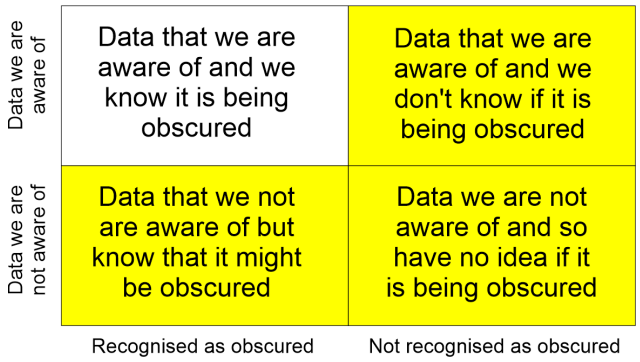


Figure 12.5: Dazzle data varieties

12.3.2 Dazzle Data Examples

12.3.2.1 Data We Are Aware Of – Recognised as Obscured

Repeated false positive alerts (e.g. from faulty sensors or alarms). These tend to be annoying irritants rather than safety issues, however they can lead to alarms being permanently switched off, masking true problems. An example from everyday life might be removing the battery from a badly placed smoke alarm in a kitchen which repeatedly sounds when the toaster is used.

12.3.2.2 Data We Are Aware Of – Not Recognised as Obscured

Electromagnetic interference (EMI) affecting transmitted data is a good example of this case; we know it exists as a possibility but don't know how it might affect data in a system or a communications path, and this could be in different ways at different times or days.

12.3.2.3 Data We Are Not Aware Of – Recognised as Obscured

In many cases we do not know what changes or corruptions may affect the data, but we have strong correction mechanisms to cope with the problems regardless. A good example is data stores on a spacecraft which have redundancy and error correcting codes built-in and so can deal with the majority of corruptions caused by cosmic rays.

12.3.2.4 Data We Are Not Aware Of – Not Recognised as Obscured

An example of this is where data is mistakenly added to a machine learning training dataset resulting in bizarre and possibly hazardous behaviours of the system using it (e.g. an autonomous road vehicle). It is paramount that data used for such image recognition systems filters and excludes potentially dangerous extraneous data arising from unanticipated situations

12.3.3 Dazzle Data Varieties and Safety Examples

In his book, David Hand introduces a taxonomy of fifteen varieties of Dark Data. This section discuses each form of Dazzle Data and considers it from a systems and safety assurance perspective. Some of the varieties are the same as Dark Data, but some are very different.

12.3.3.1 Data We Know are Superfluous or Unneeded: "Known Extras"

This case is very common in systems where an interface may be flooded with unsolicited messages. It is also common in safety justifications where assurance information may be padded out with extra, irrelevant information that can mask the underlying safety argument (which may or may not be intentional). It can be a serious problem in large, complex safety documents or detailed safety analyses. Luckily, in this case, the data is recognised as superfluous and can be ignored. Common examples include:

- Mailboxes filled with spam mail, preventing important messages being read
- Repeated stall warnings when the aircraft is under control

- Assurance information in a safety case is included for components which are not part of the solution in use at that site or situation
- Assurance information about old or legacy components is included but these are now not in service
- Machine learning training data contains many cases which are duplicates or very closely related, so adding nothing to the learned behaviour but making the dataset hard to deal with

This case can be mitigated in several ways, including careful filtering, review, use of concise notations (such as Goal Structuring Notation) to extract the essence, and periodic audits and cleans.

12.3.3.2 Data We Don't Know Are Unneeded or Spurious: "Unknown Extras"

This is the most serious and far-reaching case, where the additional data is not recognised as unneeded and may be processed, analysed or be left in place when it should be removed. Some examples are:

- Legacy code in software where nobody understands its function, or how it relates to the current code and so is reluctant to remove it. The Ariane 4 Inertial Reference function left in place in the Ariane 501 launch disaster might be such a case. Note that code whose function is known would be under the first case, discussed in 12.3.3.1
- Parts of the safety case or safety argument are supplied separately (e.g. by a subcontractor or 3rd party) and they are obscure. In this case it will not be known which parts are relevant to the particular situation. Some information may be irrelevant or worse, misleading.
- Machine learning training data containing too many outliers, which are not recognised as such

In many cases the extra data may be discovered after some time, and it is incumbent on the organisation involved to analyse the impact of the complete set of data which may have been used over the time period, including subsequent decisions and actions. The effect of spurious data should not be underestimated, as it may have masked, prevented or distorted the intended use of the nominal data for some time. This case can fundamentally change the safety picture and is probably in the highest safety risk category.

Data found to be erroneous (i.e. it should not ever have been in the system or in the safety argument) is a problem. The following options show possible approaches to handling this data:

i) Delete the erroneous data

ii) Accept the erroneous data, i.e. continue using it

iii) Report that the data was erroneous but take no further action

iv) Establish under what circumstances and situations the erroneous data would have had an effect

v) Perform a full impact analysis on the effects of the erroneous data and then act according to the results

12.3.3.3 Data Obscuration: Missing What Matters

This is where the meaning of the overall dataset becomes obscured due to the extra unnecessary data. Examples might be:

- Measuring the wrong things due to excessive noise or too much data to deal with
- Processing involving sampling only picks bad data elements
- Being too close to the data, i.e. the "wood for the trees". This is when the extra data masks the overall issue with the data, e.g. a slow trend or bias hidden by large local variations.

Mitigations include review, statistical checks and "taking a step back" to look at the bigger picture.

12.3.3.4 Data Masking in Specific Cases

This is where the extra data specifically masks, obscures or hides particular data elements (but not all). A particularly nasty case of this is where filters are put in place to remove unwanted data values, but those filters actually remove less (or more) than they should (i.e. don't remove all unwanted cases or remove valid values as well). Some examples might be:

- The number of successful test runs vastly outweighs failed runs and so the failures are not investigated
- Incorrect filtering of the data, leaving in some cases that should have been excluded

Mitigations include use of review and completeness checks. Note that over-aggressive filtering would create cases of Dark Data.

12.3.3.5 Masquerade or Fraudulent Data

This is where data has been constructed to fool the system consuming it, hiding, overlaying or replacing the correct data. Often this will be malicious and should be filtered or rejected by the target system, but of course may not be.

- Intentional fraud
- Some security attacks

Mitigations include audit, blockchain, monitoring, intelligent profiling of data and detection of changes.

12.3.3.6 Incorrect Definitions of Data

This is considered a common case for systems exchanging data with external systems, where they may duplicate, translate or incorrectly send multiple messages. They may be time-separated, for instance if communications are lost and then regained, leading to confusion.

- Data sent in the wrong rate or units (e.g. every millisecond rather than second)
- Data given for the wrong range or duration (e.g. for a month rather than a day
- Data sent for the wrong domain or scale (e.g. national values rather than regional)
- Text messages re-sent to a mobile phone when coverage restored

Mitigations include making sure interface specifications are clear and ambiguous, rejecting incorrect size or scale datasets, keeping a list of known changes / incompatibilities and documenting the changes or fixes that have to be applied to make the data usable or consistent.

12.3.3.7 Summaries of Data

Spurious data can affect summaries in a significant way if composite values are calculated, e.g.

- Calculations of averages or other statistics affected by duplicates
- Spurious outliers can lead to over fitting when analysing data trends

Mitigations include audit and assessing what extra data could cause (i.e. performing sensitivity analysis) and establishing what decisions might be made erroneously due to certain data values being present.

12.3.3.8 Information Asymmetry

This is common where there are multiple stores or sources of the same data. If one has erroneous duplicate values (i.e. many null values used for padding) then comparisons between them may fail.

- Multiple / backup databases where they are not kept in sync
- Issues of divergence of data across multiple sources.

Comparisons, reviews and data audits may help in these cases. In general the issue of divergence across multiple data sources needs to be recognised and addressed in an automated way.

12.3.3.9 Intentionally Dazzled Data

This might happen in a secure context where key data is effectively "camouflaged" by hiding within large data items (e.g. images) or the use of large amounts of apparently routine data to mask the secure data.

- Defence, security and government sectors where data is purposefully hidden
- An organisation may provide copious amounts of safety case collateral to hide a weak safety case

12.3.3.10 Extrapolating Beyond Your Data

Machine learning systems have to cope with values outside of their training data, but the outcomes may be unexpected if the training data contains many spurious values:

- Bizarre results from recognition / detection systems due to out of range values
- Image components mislabelled leading to strange results

Mitigations include checking of the values that are used in training, analysis of outliers and repeats, deletion of duplicates and use of machine learning validation data containing edge / corner cases and boundary cases.

12.3.3.11 A Note on Sensors

Sensors can degrade and fail over time, especially in harsh environments such as automotive, marine or aviation. In such cases sensors may feed erroneous or additional data into systems (e.g. if an out of normal range situation is detected), leading to misleading processing or false alarms. The Boeing 737 MAX 8 accidents might fall into this category as the angle-of-attack sensor erroneously generated high values, and certainly the systems / processing which passed on the values created spurious data. Some examples are:

- Faulty sensors
- Multiple sensors that do not have their values properly combined
- Sampling techniques which generate additional data
- Interval polling interval can lead to misleading readings
- Data fusion can be impacted if multiple sensor values are combined

Mitigations include independent monitoring of sensors and regular maintenance, calibration or replacement of hardware. Their interfaces need to be able to detect any problems and report faults. In particularly critical applications, it may be necessary to minimise the potential for common mode failures by using disparate technologies, or at least sensors from different manufacturers.

An example may be found in https://en.wikipedia.org/wiki/Qantas_Flight_72 where:

> ...the [Central processing unit] of the ADIRU [Air data inertial reference unit] corrupted the AoA [Angle of attack] data. The exact nature of the corruption was that the ADIRU CPU erroneously re-labelled the altitude data word so that the binary data that represented 37,012 (the altitude at the time of the incident) would represent an angle of attack of 50.625 degrees. The FCPC [Flight control primary computer] then processed the erroneously high angle of attack (AoA) data, triggering the high-AoA protection mode, which sent a command to the electrical flight control system (EFCS) to pitch the nose down.
>
> The FCPC algorithm was very effective, but it could not correctly manage a scenario where there were multiple spikes in either AoA 1 or AoA 2 that were 1.2 seconds apart – i.e., if the 1.2-second period of use of the memorised value happened to end while another spike was happening.

12.3.4 Summary of Dazzle Data

Dazzle data is a very useful way of thinking about data problems and solutions. It is important to always think of the bigger picture, considering what information is superfluous and may be distorting or masking the real picture:

- What might be hidden intentionally or otherwise due to the extra values?
- Could the extra data items cause any safety problems, e.g. change a numerical calculation?
- How would a system cope if it received data it was not expecting?
- How do you characterise the nature of unexpected data so as to ensure it can be handled if it does occur?
- Can you present your results / outputs in a way that shows the extra data?

12.4 Links between dark data / Dazzle Data and

Fundamentally, dark data is data that is not available, but nevertheless is important, and indeed, in some cases, more important than the data that is available.

Similarly, Dazzle data refers to spurious, superfluous or unexpected data that masks and confuses the picture and can reduce your ability to see details, and indeed the whole picture – a bit like noise or camouflage.

The opposite attributes of dark data and dazzle data might be termed Brightness and Clarity respectively. These are potential new properties or meta-properties of data.

Table 12.1 illustrates the properties which can be affected by dark data and dazzle data issues.

Table 12.1: Data properties affected by dark data and / or dazzle data issues

Data property	Explanation	Dark Data – missing data	Dazzle data – obscuring data	Notes
Integrity (I)	The data is correct, true and unaltered	✓	✓	E.g., Corruptions to a database could change the original values so the original data is lost, or hide or mask values such as by inserting End-of-File markers
Completeness (C)	The data has nothing missing or lost	✓	-	If data is lost then it is usually unknown and hence dark. However use of parity, CRCs and digital signatures may be able to detect the loss and, in some cases, fill in the missing data.
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	✓	-	Gaps or breaks in a data stream would be dark – the missing values are unknown. Methods may be available to mitigate if detected, e.g., interpolation.
Format (O)	The data is represented in a way which is readable by those that need to use it	✓	✓	If data is not in the correct format then it may not be usable so becomes dark. If lots of data is in the wrong format it may obscure other (good) data, and hence can dazzle.
Accuracy (A)	The data has sufficient detail for its intended use	✓	-	If the data has lost detail (e.g., due to sensor sampling or representation) then this is a dark case.

Continued on next page

Table 12.1: Data properties affected by dark data and / or dazzle data issues (continued)

Data property	Explanation	Dark Data – missing data	Dazzle data – obscuring data	Notes
Resolution (R)	The smallest difference between two adjacent values that can be represented in a data storage, display or transfer system	✓	✓	If the system cannot distinguish between different data values that actually are different, then data could become hidden or lost and hence dark. If the data is masked then it may also be a case of dazzle data.
Traceability (T)	The data can be linked back to its source or derivation	✓	-	If traceability is lost then the data effectively becomes darker as its origins and derivation are lost, and may not be able to be used.
Timeliness (M)	The data is as up to date as required	✓	✓	If data is not, or thought to be not, up to date then it may be rejected and hence become dark data. But clocks can become out of sync across different systems leading to erroneous rejection so the timestamps may then be thought of as dazzling the system.
Verifiability (V)	The data can be checked and its properties demonstrated to be correct	✓	-	If data can't be checked then it may be rejected (and become dark), even if it is valid.
Availability (L)	The data is accessible and usable when an authorised entity demands access	✓	-	If data is not available as required then it is dark (whether it exists or not).
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	✓	✓	If data is not prioritised to be available when needed it is effectively dark. This could be because other data is taking precedence so could be a dazzle case.

Continued on next page

Table 12.1: Data properties affected by dark data and / or dazzle data issues (continued)

Data property	Explanation	Dark Data – missing data	Dazzle data – obscuring data	Notes
sequencing (Q)	The data is preserved in the order required	✓	✓	If data is not sequenced as needed it may be rejected or ignored and then is effectively dark. Other data which is out of order or sequence may cause the whole message, set or file to be rejected – a dazzle case.
Intended destination / usage (U)	The data is only sent to those that should have access to it	✓	-	If data is not sent to the correct destination it may be a dark data situation as it is effectively missing.
Accessibility (B)	The data is visible only to those that should see it	-	✓	If data is more widely visible or distributed than intended it could cause a dazzle situation due to e.g. bandwidth issues.
Suppression (S)	The data is intended never to be used again	-	✓	If data re-emerges after it should have been deleted it could cause all sorts of dazzle-related problems.
History (H)	The data has an audit trail of changes	✓	-	If the audit trail is lost then the data is darkened as its origins and derivation could be important.
Lifetime (E)	When does the safety-related data expire	✓	✓	If the lifetime erroneously ends too early then the data is lost – a dark data problem. Conversely if the lifetime is too long it is potentially a dazzle situation.
Disposability / deletability (D)	The data can be permanently removed when required	-	✓	If the data cannot be removed it is potentially a dazzle situation, as it may obscure the real or current data.
Goldilocks (G)	The data is just the right size — not too much and not too little	✓	✓	Too little data may be a dark data case and too much data to deal with is a dazzle case.
Analysability (Z)	The data is of a suitable size, type and representation (including any metadata) to enable it be usefully analysed	✓	✓	If the data can't be analysed it may not be able to be used causing a dark data issue. If there is too much of it or it is obscured (e.g., by noise on a comms line) then it is a dazzle case.

Continued on next page

Table 12.1: Data properties affected by dark data and / or dazzle data issues (continued)

Data property	Explanation	Dark Data – missing data	Dazzle data – obscuring data	Notes
Explainability (X)	The data can be meaningfully explained to those who need to understand it by a suitable mechanism	✓	✓	If the data can't be explained it may not be able to be used causing a dark data issue. If there is too much of it or suitable management and analysis tools are not available then it is potentially both cases.
Clarity (K)	The data is visible, unmasked, not hidden or obscured, and not camouflaged	✓	✓	Clarity covers the cases where data is invisible or not included for some reason, often because it is hidden by other data. This could be due to either.

This page is intentionally blank

13 Data Cygnology

> *Here be dragons*
> ***Anon***

13.1 Introduction

The term "Black Swan" is used to describe an unpredictable event that is beyond what is normally expected of a situation and has potentially severe consequences. Black Swan events were originally discussed in a financial context by Nassim Taleb in his 2004 book *"Fooled By Randomness"* [36] and further expanded to cover other topics in *"The Black Swan: The Impact of the Highly Improbable"* [37]. The term quickly became accepted, and Black Swan events are characterized by their extreme rarity, severe impact, and the widespread insistence they were obvious in hindsight. Taleb gave examples such as the rise of the Internet, the personal computer and financial crashes as "Black Swan" events, but the term can equally be applied to unexpected safety-related events such as the Boeing 737 MAX crashes of 2018 / 2019 [38].

While Black Swans are events that are only obvious in hindsight, other types of event may be extremely rare and have severe impact but are predictable beforehand, for example pandemics. Another example comes from the sun: a Coronal Mass Ejection (CME) of such severity can be produced that it causes widespread life-threatening electromagnetic interference on Earth. This type of event can be predicted to occur to some extent, and has happened before, for example the "Carrington Event" of 1859 [39]. These rare events which have rapid escalation, but are somewhat predictable are termed "Dragon King" events.

There are several more event types noted in the literature. The practice of studying these types of high-impact but rare variants has been termed "Cygnology", a term that appears in work by Ale, Hartford and Slater in their "Dragons, black swans and decisions" paper of 2020 [40].

While Cygnology can be applied to safety in general, it can also be applied to data safety and recent research has identified four important categorisations of data-related risks:

- Black Swan Data
- Dragon King Data
- Perfect Storm Data
- Pudding Lane Data

These types of data are discussed in more detail in the following sections along with recommended strategies for managing their associated risks.

13.2 Black Swan Data

Definition: Data that is totally unexpected by those receiving it (i.e. 'out of the blue') and has huge (detrimental) impact, but in retrospect should have been anticipated.

In 2013, a C-130 Hercules aircraft landed at Bar Yehuda airport near the Dead Sea, a saltwater lake sitting astride Israel and Jordan. The Dead Sea is the lowest place on Earth and the airfield lies -1,210 feet below sea level.

The aircraft's navigation system became unresponsive and the constellation of GPS satellites above, mysteriously winked out of existence. As it turned out, the plane's navigation electronics were not designed to operate at altitudes less than 400 feet below mean sea level. In a sense, the plane thought it was underwater.



This is an example of what is termed a "Black Swan Data" event; the stream of altitude data going to the navigation systems surprisingly turned significantly negative and could have had a catastrophic result; a situation not considered possible. Yet, with hindsight, and knowledge of earth's geography, we can rationalise this and question why the assumption of positive altitude data was ever thought to be true in the first place.

13.2.1 Managing Black Swan Data Risks

The following strategies are recommended for managing Black Swan Data Risks:

- Be aware of changes of usage, environmental conditions or failures that may produce data that is not expected. The Ariane 501 launch disaster may be thought of in these terms, as the horizontal bias values that caused the exception leading to the complete loss of mission were not anticipated;
- Continually test the established thinking; challenge the oft quoted "that's just how we do things around here";
- What assumptions have been made about the data? Are these really valid and do they continue to be valid in a changing and dynamic world?
- What happens if different inputs (e.g. new sensors or information flows) are added to your system?
- Think outside the box; it is the most wild and implausible (compared to current perceptions) ideas that will trip a Black Swan Data event – what could possibly be 'bad data' and how might it be handled?

13.3 Dragon King Data

Definition: Data whose effect might have been foreseen, but leads to an unexpected and explosive escalation with major impact ('things get rapidly out of control').

In risk management a Dragon King is defined as an event that is both extremely large in size or impact (a "king") and born of unique origins (a "dragon") relative to other events in the system.



Dragon King events are generated by, or correspond to, mechanisms such as positive feedback, tipping points, bifurcations, and phase transitions that tend to occur in non-linear and complex systems, and serve to amplify events to extreme levels. In the data domain, "Dragon King Data" represents incorrect or missing data whose consequences unexpectedly escalate into an unpredicted and often catastrophic system event.

An example might be the loss of data relating to Covid-19 test results in the UK due to a row limit in Microsoft Excel. This led to Public Health England losing 15,841 positive test results, which in turn, meant that 50,000 potentially infectious people were missed by contact tracers and not told to self-isolate. Because of the exponential way that viruses spread, this may have led to many more infections and deaths.

Another, older, example is the infamous comment given by the UK weather forecaster Michael Fish about the Great Storm of 1987:

> "Earlier on today, apparently, a woman rang the BBC and said she heard there was a hurricane on the way. Well, if you're watching, don't worry, there isn't!"

These few words may well have led to precautions not being taken across the whole of Southern England, affecting millions of people. In fact, the storm was the worst to hit South East England for three centuries, causing record damage and killing 19 people.

13.3.1 Managing Dragon King Data Risks

The following strategies are recommended for managing Dragon King Data Risks:

- Identify all the receiving systems or users of the data: what are the consequences of errors or losses in the data to those receiving it?
- Analyse the effects of data impact: are any effects subject to non-linear effects? Try to establish the consequence chains and knock-on effects;
- Do any of the data potentially impact systems that have large fan-out or spread?
- Create models (or even Digital Twins) of the system in a way that different combinations of data can be tested and simulated to reveal extreme non-linear behaviours.

13.4 Perfect Storm Data

Definition: Combinations, sets or occurrences of data (or absences of data) that would never have been thought possible to occur together, and when they do have a large and undesirable impact.



A Perfect Storm in the risk world is when everything bad happens at the same time, and it was not anticipated that this could happen. These risks get their name from "The Perfect Storm", both a 2000 American biographical disaster drama film [41] and a 1997 non-fiction book by Sebastian Junger [42]. In the data world, a Perfect Storm is when the data that the system relies upon are all wrong (or missing) at the same time, and can be layers deep – much akin to Reason's Swiss cheese model. A classic example might be to try and restore deleted data from a backup system, only to find that the backups had not been working properly due to a configuration problem. Perfect Storm Data is therefore when multiple data faults or losses conspire to create a situation that has disproportionately catastrophic outcomes.

Another example is from the world of security; on 10th Dec 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, permitted remote code execution on the vulnerable server with system-level privileges. The exploit was a combination of both the Java code that contains different logging functions and the settings of a configuration file.

Another example is the China Airlines A333 at Taipei on 14th June 2020 when all primary computers, reversers and auto brakes failed on touchdown. During landing, flight controls reconfigured from 'normal

law' to 'direct law' after all three flight control primary computers (FCPCs) became inoperative. While all aircraft primary control surfaces were still controllable, the deceleration devices including ground spoilers, thrust reversers, and autobrake were lost; the deceleration of the aircraft had to be performed by manual braking by the pilots.

13.4.1 Managing Perfect Storm Data Risks

The following strategies are recommended for managing Perfect Storm Data Risks:

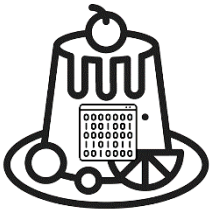
- Analyse combinations of data that are important: how might one dataset being wrong influence another?
- Look at dependencies in the data: are seemingly different datasets all derived from a common source?
- Examine as many levels of the system to establish commonalities and dependencies;
- Even if no commonalities are found, consider the unlikely cases of independent data being wrong together, and follow through their consequences.

13.5 Pudding Lane Data

Definition: Data that is unknowingly and unexpectedly critical to the whole operation and if missing or incorrect in some way, has a dramatic and negative impact. After the event it may be obvious that it was critical.

> "Lord! what sad sight it was by moone-light to see, the whole City almost on fire, that you might see it plain at Woolwich, as if you were by it".

These words from Samuel Pepys diary in 1666, give a personal account of the Fire of London that destroyed 13,200 houses, 87 parish churches, The Royal Exchange, Guildhall and St. Paul's Cathedral, all from a single fire thought to have started in a humble baker's shop in Pudding Lane.



In the same way, unexpectedly critical data can also act as a catalyst to catastrophic outcomes if the dependencies are not identified and adequately mitigated. Often this data is part of an existing system and may be long forgotten (e.g. a system configuration parameter). It may also be a hard-wired parameter in legacy source code that code maintainers change because they do not understand its original purpose.

In 2015, an Airbus A400M Atlas cargo plane on a test flight crashed at La Rinconada, Spain, less than 5 kilometres from Seville Airport, killing 4 of the 6 crew. Several reports suggested that as many as three of the aircraft's four engines failed during the A400M's departure from Seville.

The suspected cause of the failure was that the torque calibration parameter data had been accidentally wiped on three engines as the engine software was being installed, which would prevent the engine control software from operating properly. In this case, the seemingly innocuous calibration data had a critical role to play in the safety of the entire platform and so the term "Pudding Lane Data" is coined to encompass this form of data.

13.5.1 Managing Pudding Lane Data Risks

The following strategies are recommended for managing Pudding Lane Data Risks:

- Understand and know all the data. The high-profile data streams will of course be first and foremost, but is there an inventory of all the data that the system depends upon and are the implications if there is loss of their data properties understood?
- Configuration and tailoring data may seem static and uninteresting but what are the dependencies on this data? If loss of properties of this data can have catastrophic outcomes, then what processes and mitigations will be put in place to assure those properties are maintained?
- Consider the reporting and monitoring data: if this is incorrect will poor decisions be taken?

13.6 Conclusions

Looking at Data Risks through Cygnology can help to tease out some unlikely but severe failures of systems involving data. It is possible that sometimes these risks are already considered in system HAZOP sessions but discounted as too unlikely to happen. The Data Cygnology approach is recommended as a useful and visual 'Tool in the Box' when trying to identify extreme data-related risks of systems. Due to their serious impact arguably they should still be included, assessed and managed alongside more 'normal' risks.

13.7 Summary Table

Table 13.1: Summary of cygnology types

Type	Brief Explanation	Icon
Black Swan Data	Data that is totally unexpected by those receiving it (i.e. 'out of the blue') and has huge (detrimental) impact, but in retrospect should have been anticipated.	
Dragon King Data	Data whose effect might have been foreseen, but leads to an unexpected and explosive escalation with major impact ('things get rapidly out of control').	
Perfect Storm Data	Combinations, sets or occurrences of data (or absences of data) that would never have been thought possible to occur together, and when they do have a large and undesirable impact.	
Pudding Lane Data	Data that is unknowingly and unexpectedly critical to the whole operation and if missing or incorrect in some way, has a dramatic and negative impact. After the event it may be obvious that it was critical.	

This page is intentionally blank

14 Additional Ways that Data Can Cause Problems

> *When solving problems, dig at the roots instead of just hacking at the leaves.*
> ***Anthony J. D'Angelo***

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.

Table 14.1 illustrates where the likely relationships between data properties and these issues.

Table 14.1: 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	Clarity
Issue	**I**	**C**	**N**	**Y**	**O**	**A**	**R**	**T**	**M**	**V**	**L**	**F**	**P**	**Q**	**U**	**B**	**S**	**H**	**E**	**D**	**G**	**Z**	**X**	**K**
Fluidity								x		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	x
Ownership								x		x	x				x	x		x					x	
Archiving / retrieval															x	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	x
Defaulting	x	x	x	x		x						x										x		x
Sentinels	x	x		x		x								x								x		x
Aliasing	x	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	x
Masking	x	x	x	x		x				x		x										x	x	x
Incompleteness	x	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	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	x

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.

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" [43]. 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 Vol. 3 Section 5.8.

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 Section 12.2) 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).

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* [44].

15 Acronyms, Definitions and Glossary

> *What I'm doing is not really based on a definite identification or a definition of what it is. It's intended to be open to interpretation.*
> ***Reggie Watts***

15.1 Acronyms

ADIRU	air data inertial reference unit
AI	artificial intelligence
AoA	angle of attack
ARP	aerospace recommended practice
ARQ	automatic repeat-request
BIT	built in test
BITE	built in test equipment
CCTV	closed circuit television
CoD	certificate of design
COTS	commercial off-the-shelf
CPU	central processing unit
CRC	cyclic redundancy check
CSV	comma separated variable
DCB	data coordination board
DoS	Denial of Service
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
FCPC	flight control primary computer
FDAL	functional design assurance level

FHI	Future of Humanity Institute
FMECA	Failure mode and criticality analysis
GP	general practitioner
GPS	Global Positioning System
GSN	Goal Structuring Notation
HAZOP	hazard and operability study
ICD	interface control document
IDAL	item development assurance level
IMF	International Monetary Fund
ISO	International Standards Organization
LFR	live facial recognition
LLM	large language model
ML	machine learning
NaN	not a number
ODR	organizational data risk
OED	Oxford English Dictionary
PAI	Partnership on AI
SASWG	Safety of Autonomous Systems Working Group
SCSC	Safety-Critical Systems Club
SIL	safety integrity level
SMP	safety management plan
SOP	standard operating procedure
SOTIF	safety of the intended functionality
SSS	Safety Critical Systems Symposium
SWIFT	structured what-if checklist
XML	extensible markup language

15.2 Definitions and Glossary

accessibility

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

accuracy

Closeness of agreement between a measured quantity value and a true quantity value of a measurand. [45]

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 [46]

aeronautical data

- A representation of aeronautical facts, concepts or instructions in a formalized manner suitable for communication, interpretation or processing. (EU) No 73/2010 [47]
- Data used for aeronautical applications such as navigation, flight planning, flight simulators, terrain awareness, and other purposes. RTCA/DO-178C [48]

analysability

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

artefact, data

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 [49]

availability

The property of being accessible and usable upon demand by an authorized entity. ISO 27001:2013 [50]

clarity

The property that the data is visible, unmasked, not hidden or obscured, and not camouflaged

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 [50]

configuration data

- Data that configures a generic software system to a particular instance of its use. (EC) No 482/2008 [49]
- 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 [46]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 [46]

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 [6]

critical data

- Data whose loss of specific properties could contribute to a hazardous system state
- Data with an integrity level as defined in Chapter 3, Section 3.2 point 3.2.8(a) of Annex 15 to the Chicago Convention, i.e., integrity level one in one hundred million: there is a high probability when using corrupted critical data that the continued safe flight and landing of an aircraft would be severely at risk with the potential for catastrophe. (EU) No 73/2010 [47]

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 [51]

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 [52]

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 [53]

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 [54]

disposability / deletability

The property that the data can be permanently removed when required

error, data

- Discrepancy with the universe of discourse. ISO 19138:2006 [55]
- Discrepancy between a data value and the true, specified or theoretically correct value or condition. P. Ensor [56]

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 way 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 [52]

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 [51]
- 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 [47]
- 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 [57]
- The property of protecting the accuracy and completeness of assets, i.e., that which has value to the organization. ISO 27001:2013 [50]

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 [47]

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 [58]

mitigation

Steps taken to control or prevent a hazard from causing harm and reduce risk to a tolerable or acceptable level. [59]

owner, data

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

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 [51]
- 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 [47]

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 [51]
- The number of units or digits to which a measured or calculated value is expressed and used. (EU) No 73/2010 [47]

response

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

safety

- The state of being protected from or guarded against hurt or injury; freedom from danger. OED
- Freedom from conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment. [2]

safety assessment, data

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

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

safety requirement, data

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.

software lifecycle data

Data that is produced during the software lifecycle to plan, direct, explain, define, record, or provide evidence of activities (including the software product itself). This data enables the software lifecycle processes, system or equipment approval and post-approval modification of the software product. ED-153 [46]

stakeholder

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 [56]
- The difference between the time of output of a data item and the time of applicability of that data item. (EU) No 1207/2011 [57]

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 [48]

traceability

Ability to determine the origin of the data. RTCA/DO-200A [51]

treat

To apply a treatment

treatment

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 [51]
- Process of ensuring that data meets the requirements for the specified application or intended use. (EU) No 73/2010 [47]

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 [47]

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 [47]

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 [47]

16 References

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

Bibliography

[1] Safety, n., i.1.a. In *OED Online*. Oxford University Press.

[2] System safety. standard Mil-Std-882E, Department of Defense, 2012.

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

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

[5] Society of Automotive Engineers International. *SAE Aerospace Recommended Practice 4754A: Guidelines for development of civil aircraft and systems*. Society of Automotive Engineers, 2010.

[6] Functional safety of electrical/electronic/ programmable electronic safety-related systems: Software requirements. Standard BS EN 61508-3:2010, International Electrotechnical Commission, June 2010.

[7] 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.

[8] *DCB0129: Clinical Risk Management: its Application in the Manufacture of Health IT Systems*. NHS Digital, May 2018. Version 4.2.

[9] *DCB0160: Clinical Risk Management: its Application in the Deployment and Use of Health IT Systems*. NHS Digital, May 2018. Version 3.2.

[10] Road vehicles — Safety of the intended functionality. Standard ISO/PAS 21448:2019, International Standards Organisation, 2019.

[11] The Safety of Autonomous Systems Working Group [SASWG]. *Safety Assurance Objectives for Autonomous Systems*. SCSC-153A. SCSC, January 2020. ISBN 978-1654029050.

[12] University of York. Website: Assurance of Machine Learning for use in Autonomous Systems (AMLAS). URL https://www.york.ac.uk/assuring-autonomy/guidance/amlas/. Accessed: 25 January 2022.

[13] MIT Technology Review. Website: Chatbots could one day replace search engines. here's why that's a terrible idea. |. URL https://www.technologyreview.com/2022/03/29/1048439/chatbots-replace-search-engine-terrible-idea/. Accessed: 11 February 2024.

[14] IBM. Website: What are ai hallucinations? URL https://www.ibm.com/topics/ai-hallucinations. Accessed: 11 February 2024.

[15] The New York Times. Website: How chatbots and large language models, or llms, actually work. URL https://www.nytimes.com/2023/03/28/technology/ai-chatbots-chatgpt-bing-bard-llm.html. Accessed: 14 April 2023.

[16] R Ashmore, R Calinescu, and C Paterson. *Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges*. Assuring Autonomy International Programme, University of York, 2019. URL https://arxiv.org/pdf/1905.04223.pdf. Accessed: 17 January 2023.

[17] D Hond, A White, and H Asgari. Quantifying dataset properties for systematic artificial neural network classifier verification. In *Assuring Safe Autonomy, Proceedings of the 28th Safety-Critical Systems Symposium (SSS'20)* [60]. ISBN 9781713305668.

[18] GS Sutherland and A Hessami. Safety critical integrity assurance in large datasets. In *Assuring Safe Autonomy, Proceedings of the 28th Safety-Critical Systems Symposium (SSS'20)* [60]. ISBN 9781713305668.

[19] Mauro Cazziniga, Florence Jaumotte, Longji Li, Giovanni Melina, Augustus J. Panton, Carlo Pizzinelli, Emma Rockall, and Marina M. Tavares. Gen-AI; artificial intelligence and the future of work. Staff Discussion NOte SDN/2024/001, International Monetary Fund, 2024.

[20] Michelle Toh. Jobs may disappear. URL https://edition.cnn.com/2024/01/15/tech/imf-global-employment-risk-ai-intl-hnk/index.html, 2024. Accessed: 30 January 2025.

[21] Nicholas Hales. Engineering, ethics, and 'doughnut' economics. URL https://scsc.uk/scsc-193, February 2024. vol.32-2. Accessed: 30 January 2025.

[22] Nicholas Hales. Autonomy's troubling questions. URL https://scsc.uk/scsc-152, January 2019.

[23] Louise Dennis, Michael Fisher, Marija Slavkovik, and Matt Webster. Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems*, 77:1–14, 2016.

[24] Metropolitan Police. Facial recognition technology. URL https://www.met.police.uk/advice/advice-and-information/fr/facial-recognition-technology/. Accessed: 30 January 2025.

[25] Kate Raworth. *Doughnut Economics – Seven Ways to Think like a 21st Century Economist*. Penguin Books, 2017.

[26] Partnership on AI. Partnership on ai. Accessed: 30 January 2025.

[27] Britt Paris and Joan Donavan. *Deepfakes and Cheapfakes: The Manipulation of Audio and Visual Evidence*. Data & Society's Media Manipulation Research Initiative. URL https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf,accessed30January2025.

[28] Data & Society. The deepfakes / cheapfakes spectrum. URL https://datasociety.net/wp-content/uploads/2019/09/Deep_fakes_spectrum.png. Accessed: 23 January 2025.

[29] Tommy Shane. It matters how platforms label manipulated media. here are 12 principles designers should follow. URL https://partnershiponai.org/it-matters-how-platforms-label-manipulated-media-here-are-12-principles-designers-should-follow/. Accessed; 30 January 2025.

[30] Partnership on AI. PAI's responsible practices for synthetic media: A framework for collective action. URL https://partnershiponai.org/wp-content/uploads/2023/02/PAI_synthetic_media_framework.pdf. Accessed: 30 January 2025.

[31] David J Hand. *Dark Data: Why What You Don't Know Matters*. Princeton University Press, January 2020. ISBN 069118237X. Available from https://www.amazon.co.uk/dp/069118237X. Accessed: 27 January 2021.

[32] David Hand. Website: Home — Dark Data. URL https://darkdata.website/. Accessed: 27 January 2021.

[33] SCSC Systems Safety Newsletter. URL https://scsc.uk/scsc-160, October 2020. vol.28-3. Accessed: 27 January 2021.

[34] David Hand. Presentation on Dark Data given to the DSIWG. URL https://scsc.uk/file/gd/DARK_DATA_David_Hand_DSIWG_talk-863.pptx, 2020. Accessed: 27 January 2021.

[35] Mike Parsons. Presentation on Dark Data given at the University of York. URL https://scsc.uk/file/gd/Dark_Data_ABC_Slides_v2-963.pptx, December 2020. Accessed: 27 January 2021.

[36] Nassim Nicholas Taleb. *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*. Penguin, 2004. ISBN 9780141031484.

[37] Nassim Nicholas Taleb. *The Black Swan: The Impact of the Highly Improbable*. Penguin, 2007. ISBN 9780141034591.

[38] Paul Hampton and Dewi Daniels. Boeing 737 MAX Safe to Fly? In *The Safety-Critical Systems Club Newsletter*, volume 29, Number 1. SCSC, 2021. Available from https://scsc.uk/scsc-162. Accessed: 9 January 2023.

[39] Website: Carrington Event. URL https://en.wikipedia.org/wiki/Carrington_Event. Accessed: December 2022.

[40] Ben J.M. Ale, Des N.D. Hartford, and David H. Slater. Dragons, black swans and decisions. In *Environmental Research*, volume 183. ScienceDirect, 2020. Article 109127.

[41] Wikipedia. Website: The perfect storm (film). URL https://en.wikipedia.org/wiki/The_Perfect_Storm_(film). Accessed: 17 January 2023.

[42] Sebastian Junger. *The Perfect Storm*. W.W. Norton, 1997. ISBN 0-393-04016-X.

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

[44] 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.

[45] ISO/IEC. Information technology — data governance — data collaboration framework. https://www.iso.org/obp/ui#iso:std:iso-iec:25642:ed-1:v1:en:term:3.9. Accessed: 2 October 2025.

[46] Guidelines for ANS software safety assurance. Standard EUROCAE/ED-153, European Organisation for Civil Aviation Equipment, August 2009. Used for definitions only.

[47] Commission Regulation (EU) No 73/2010 of 26 January 2010 laying down requirements on the quality of aeronautical data and aeronautical information for the single European sky. URL http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:023:0006:0027:EN:PDF, January 2010. Accessed: 27 January 2021.

[48] Software Considerations in Airborne Systems and Equipment Certification. Standard RTCA/DO-178C, EUROCAE/ED-12C, Radio Technical Commission for Aeronautics / European Organisation for Civil Aviation Equipment, January 2012.

[49] Commission Regulation (EC) No 482/2008 of 30 May 2008 establishing a software safety assurance system, as amended. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32008R0482&qid=1611781633022, May 2008. Accessed: 27 January 2021, but no longer in force.

[50] Information Technology — Security Techniques — Information Security Management Systems — Requirements. Standard BS ISO/IEC 27001:2013, International Standards Organisation, 2013.

[51] Standards for processing aeronautical data. Standard RTCA/DO-200A, EUROCAE/ED-76, Radio Technical Commission for Aeronautics / European Organisation for Civil Aviation Equipment, September 1998.

[52] Information Technology. Vocabulary. Part 1: Fundamental terms. Standard ISO/IEC 2382-1:1993, International Standards Organisation, 1993.

[53] Functional safety of electrical/electronic/ programmable electronic safety related systems: Definitions and abbreviations. Standard BS EN 61508-4:2010, International Electrotechnical Commission, June 2010.

[54] Geographic Information. Data Product Specifications. Standard BS EN ISO 19131:2008, International Standards Organisation, 2008.

[55] Geographic Information. Data Quality Measures. Standard ISO/TS 19138:2006, International Standards Organisation, 2006.

[56] P Ensor. *Safety Analysis of Navigational Data*. University of York, September 2009.

[57] Commission Implementing Regulation (EU) No 1207/2011 of 22 November 2011 laying down requirements for the performance and the interoperability of surveillance for the single european sky. URL http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:305:0035:0052:EN:PDF, November 2011. Accessed: 27 January 2021.

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

[59] Use of safety management systems by ATM service providers. Standard ESARR3, EUROCONTROL, 2011.

[60] *Assuring Safe Autonomy, Proceedings of the 28th Safety-Critical Systems Symposium (SSS'20)*, SCSC-154. SCSC, 2020. ISBN 9781713305668.

Part III

Appendices

A Scenarios

> *Scenarios help solve problems differently because they can illuminate new possibilities and ignite hope.*
> ***Roger Spitz***

Table A.1 identifies several scenarios under which it might be useful to recognise data as safety significant.

The levels of necessary assurance given in Section 3.3 depend on the relevant scenario, and these scenarios are also used in subsequent chapters to indicate the relevance of different tools and techniques. The scenarios may be applicable one-off, regular, scheduled, or on update.

Table A.1: Data Safety Scenarios

No.	Name	Description
Scenarios with Systems Using Data		
S-0	Assure	A safety assurance position is to be created for a system that uses or produces data
S-1	About	Development/Engineering data about the system that uses data is produced or acquired. This could include requirements, design, implementation, deployment or performance monitoring data.
S-2	Investigate	A safety incident has occurred with the system related to data and the cause of this needs to be understood and mitigations proposed.
S-3	New Use	The system using data has changed use and an impact analysis needs to be undertaken.
S-4	Change	AA new or replacement operational system using data is created, procured or made operational.
S-5	V&V	A system is verified or validated using data (e.g. test cases).
S-6	Analysis	Behaviours or results from a system are analysed producing reports or conclusions
Scenarios for Tools and Data Chains		
A-1	Train ML	An ML model is developed using training data.
A-2	Train ML	An ML model is validated using a data set.
Data Scenarios		
T-1	Train ML	A data processing chain is constructed or configured from tools or processing elements.
T-2	Train ML	A tool which can transform or change the data (as part of a chain) is constructed or configured.
Data Scenarios		
D-1	Create	A new data set for a system is to be created from scratch. This may use a variety of methods including simulation, other system, prototypes, or manual means.
D-1a	Create: Training	A new training data set for an ML system is to be created from scratch. This may use a variety of methods including simulation, other system, prototypes, or manual means.
D-1b	Create: Validation	A new validation data set for an ML system is to be created from scratch. This may use a variety of methods including simulation, other system, prototypes, or manual means.
D-1c	Create: Measure	New data is created by measuring the environment or system. E.g. acquiring or observing from sensors.

Continued on next page

Table A.1: Data Safety Scenarios (continued)

No.	Name	Description
D-1e	Create: Mash-up	Data is to be re-purposed from various sources, primarily not intended for this usage, and may be re-interpreted in its use
D-2	Migrate	Data is to be migrated from one system to another, or one version of software to another. This may involve re-formatting, transformations, mappings, translations or other changes.
D-3	Restructure	A data set for a system is to be reorganised, reformatted or reworked.
D-4	Increment	A data set is incrementally added to (may be over some time).
D-5	Degrade	The data for a system is reduced in quantity or quality, and the impact of this needs analysis.
D-6	Cleanse	A data set is to be cleaned to remove old, obsolete, incorrect or other unwanted data.
D-7	Audit	A data set is checked for certain aspects on a sampling basis.
D-8	Select	A data set is to be filtered or selected to produce a new subset of data.
D-9	Find	Missing or obscured parts of the data set are to be identified and made available.
D-10	Control	A data set is to be formally managed and change controlled.
D-11	Archive	A data set is to be copied to long-term storage and removed from current use.
D-12	Delete	A data set or part of a data set is to be deleted.
D-13	Secure	A data set is to be checked for security-sensitive or insecure content, which may be flagged or removed.
D-14	Safe	A data set is to be checked for dangerous or harmful content which may be removed.
D-15	Anonymise	A data set has all parts which could be used to identify a person, location or other sensitive aspect, removed or obscured.
D-16	Split	A data set is to be decomposed into smaller data sets, which could be based on the data itself (e.g. by date), by its storage location, or other aspect.
D-17	Distribute	A data set is to be distributed across several systems; this could involve duplication or splitting.
Development/Engineering Data		
E-1	Specify	A requirements specification (data) of a system is created
E-2	Design	A design/architecture data is created documenting the architecture of a system
E-3	Implement	The implementation of a system / software is created
Tool/Transforming Data		
T-1	Input	Inputs data for a tool/transformer can be used
T-2	Output	Output data from a tool/transformer that has been produced
T-3	Constructive	Constructive data is data that is used to construct the system, e.g. a requirement or source code
T-4	Verification	Verification data is data which is produced when verifying the system, e.g. a test report
T-5	Persistent	Persistent data is data that is stored permanently, e.g. in a configuration management system
T-6	Temporarily	Temporarily data is data which is not stored, e.g. is only temporarily available, e.g. the values of a memory (RAM)

Continued on next page

Table A.1: Data Safety Scenarios (continued)

No.	Name	Description
T-7	AI Model	A set of data (“weights”) that relates the inputs to the outputs

This page is intentionally blank

B Data Safety Management Plan

> *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.

   Behaviour Monitoring Identification of measures to monitor the behaviour of the SCS at appropriate intervals throughout its life.

   Change of Use Identification of procedures to monitor changes of use and for an assurance impact analysis to be undertaken where necessary.

   Evolution and Incremental Change Identification of procedures to manage evolution and incremental change of the system.

   Lifetime Identification of procedures to manage longevity and obsolescence of the data SCS.

Lessons Learned Identification of procedures to identify and review lessons learned from successful operation, failures and incidents.

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?).

C Interactions Report

> *For good ideas and true innovation, you need human interaction, conflict, argument, debate.*
> ***Margaret Heffernan***

This section gives a suggested interactions report table of contents.

C.1 Introduction

Scope and context briefly sets the scene, describes the project;

Scenario ;

Stakeholders ;

Concept of operations, etc (may be done by reference);

Assumptions ;

References ;

Abbreviations and acronyms ,

C.2 Interactions Description

Description of system boundaries and interfaces ;

Producers and consumers of the data ;

A list of known data interactions with other entities (may be systems, people, etc). Interactions involving data flows (e.g. messaging, networking) are of primary interest here, and control interactions (e.g based on configuration data) are also considered. It should include both external and internal data interactions to the system.

Other possible interactions

C.3 Interactions Analysis

The results of the failure analysis of the interactions, which may use such techniques as HAZOP, data FMECA, STAMP, FRAM, etc.

C.4 Summary of Findings

A description of the significant outcomes of the analysis and the residual risks.

C.5 Conclusions

Includes actions and recommendations to manage the risks.

This page is intentionally blank

D Unintended Behaviours Report

> *It seems to be a law of design that for every advantage introduced through redesign, there is an accompanying unintended disadvantage.*
> ***Henry Petroski***

This section gives a suggested unintended behaviours report table of contents.

Scope and context briefly sets the scene, describes the project;

Scenario ;

Stakeholders ;

Concept of operations, etc (may be done by reference);

Assumptions ;

References ;

Abbreviations and acronyms ,

D.1 Identification of Unintended Behaviours

Describes the steps taken to attempt to identify unintended and unexpected behaviour due to data. Techniques for management of unknown risks and system resilience might be useful.

D.2 Unintended Behaviour Analysis

The results of the unintended behaviour analysis of the data, which may use such techniques as HAZOP, data FMECA, STAMP, FRAM, brainstorming, and so on.

D.3 Summary of Findings

A description of the significant outcomes of the analysis and the residual risks, and an assessment of the coverage of unintended behaviours.

D.4 Conclusions

Includes actions and recommendations to manage the risks.

This page is intentionally blank

E Organizational Data Risk

> *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	A use of the data could lead to an accident resulting in death for several people. The accident could be caused by data or lack of 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:	

F Data Safety Culture Questionnaire

> *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. This form should be distributed to all those who are involved with data on the programme / project

Data Safety Culture Questionnaire Form			
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 believe we have a good position regarding data safety.	❑	❑	❑	❑	❑	❑
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 think the programme is adequately resourced to manage data safety risks.	❑	❑	❑	❑	❑	❑
2c	I think the programme meets its obligations regarding data safety.	❑	❑	❑	❑	❑	❑
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:							

F.1 Interpretation of Results

- If all responses are mainly in the "Agree/Strongly Agree" areas then the data safety culture on the programme is strong and data safety risks are likely to be being managed well. There are no further actions needed except that new team members should be be asked to to complete the survey after induction.
- If there are some "Maybes" then further work to improve the data safety culture may be necessary. This could involve specific data safety training for team members and awareness training for

management. It is suggested that the latest version of this guidance document is made part of the programme.

- If the responses are mainly “Disagree” or below then the data safety culture is poor and data safety is not likely to be being managed on the programme. Immediate remediation is required. This could involve appointment of a data safety champion, training, consultancy from a data safety expert, higher management involvement, safety audits, etc. It is suggested that the questionnaire is repeated after actions have been taken to measure if there has been any changes. If there is particular trend in one question area then specific actions to improve that area is recommended.
- The survey should be given to all staff on a programme periodically, e.g. every 12 months.

This page is intentionally blank

G Supplier Data Maturity

> *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?

H HAZOP Guidewords

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

A HAZOP [4] 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 H.1 lists a set of guidewords for a data-focused HAZOP, based on the properties defined in Vol. 1 Section 2.1. 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.

Table H.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, 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, 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	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, insufficient	Accurate, true, correct, undistorted, unbiased, faultless.

Continued on next page

Table H.1: HAZOP guidewords: detailed definitions (continued)

Property	Description	HAZOP Data Guidewords	HAZOP Data Considerations
Resolution	The smallest difference between two adjacent values that can be represented in a data storage, display or transfer system	Loss, partial loss, incorrect, insufficient	Exact, untruncated, retention of detail, clarity, determination, distinguishable, clear, within range, distinct, separated, discernible, discriminatable, unconfused, divisible, unaliased, granularity, precision.
Traceability	The data can be linked back to its source or derivation	Loss, partial loss, incorrect, 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, 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, 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, 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, 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, too early, too late, loss of sequence	Directed, delivered, copied, sent, transmitted, correct recipient, unintercepted, unseen, integral, received, acknowledged, forwarded, filtered.

Continued on next page

Table H.1: HAZOP guidewords: detailed definitions (continued)

Property	Description	HAZOP Data Guidewords	HAZOP Data Considerations
Accessibility	The data is visible only to those that should see it	Loss, partial loss, incorrect, 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.
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, 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
Clarity	The Data is visible, unmasked, not hidden or obscured, and not camouflaged	Dark, Invisible, Obfuscated, Murky, Indistinct, Noisy	“Hiding in plain sight”, “Wood for Trees”, “Cry Wolf”, “Elephant in the Room”

This page is intentionally blank

Index

Accessibility Property, 114, 123, 124, 167
Accuracy Property, 51, 63, 65, 74, 77, 78, 112, 123, 132, 133, 165
Adaptation Data, 15, 27, **90**, 129
ADIRU, 111
AI, 77–81, 83–87
  Bias, 83
  Control, Autonomy, Ethical and Moral Considerations, 85
  Dependency, 85
  Discrimination, 83
  Economic inequality, 85
  Existential risk, 86
  Fake content, 86
  Job displacement, 83
  Manipulation, 86
  Privacy erosion, 84
  Security risks, 84
Analysability Property, 114, 123, 167
Artefact
  Data, 132, 133
Artefact, Data, 9, 13–15, 20, 21, 26, 31, 35, 36
Asset Data, **91**
Assurance Level, 124, 133
  Data, 17, **17**, 18–24, 29–31, 33–55, 63–65, 68, 69, 71, 145, 146, 151
  Development, 17, 20
  Software, 20–22, 31, 32
Automation, 83, **83**, 84, 86
Autonomy
  Loss of human skills, 84
Availability Property, 53, 63, 65, 113, 123, 166

Behavioural Data, **90**
Boeing 737, 102, 111

Category
  Approach
    Data Design, 35
    Data Implementation, 35
    Data Migration, 35
    Data Testing, 35
    Media – Electronic, 35
    Media – Paper, 35
    System Design, 35
    Test, 35
  Data, 14, 15, 27, 35–37, 42–55, 60, 63, **89**, 146
  Verification, 77
  Lifecycle Data, **27**, 35
Clarity Property, 115, 123, 167
Completeness
  Checks, 109, 124
  Hazard Identification, 15, 27, 165
  Property, 15, 27, 36, 63, 65, 97, 98, 112, 123, 126, 165
Completeness Property, 132, 133
Consistency
  Property, 63, 65, 81, 97, 98, 112, 123, 165
  Self, 130
  Semantic, 68
  With Inputs, 134
  With Requirements, 130
Continuity Property, 63, 65, 112, 123, 165
COTS, 100
Covid-19, 125
Cygnology
  Dragon King Data, 118
  Perfect Storm Data, 119
  Pudding Lane Data, 120

Dark Data, 99–106, 106, 107, 109, 112–115, 126
  Comparison with Dazzle Data, 112
Data
  Artefact, 9, 13–15, 20, 21, 26, 31, 35, 36
  Category, 14
  Dark, *see* Dark Data
  Dazzle, *see* Dazzle Data, 106, **106**, 107–113
  Definition, 103
  Entry, 68, 132
  Owner, 25, 34, 42, 65, 123, 145, 156
Data Cygnology, 117
  Black Swan Data, 117
  Summary, 121
Data Migration, 95
  Many to one, 96
  Reversion, 96
  Simple, 96
Data Property, 16, 20–24, 27, 29, 31, 32, 35–55, 59, 62–64, 78, 79, 81, 124, 166
  Accessibility, 114, 123, 167
  Accuracy, 51, 63, 65, 74, 77, 78, 112, 123, 132, 133, 165
  Analysability, 114, 123, 167
  Availability, 53, 63, 65, 113, 123, 166

Clarity, 115, 123, 167
Completeness, 15, 27, 36, 63, 65, 97, 98, 112, 123, 132, 133, 165
Consistency, 123
Continuity, 63, 65, 112, 123, 165
Disposability, 123
Explainability, 123
Fidelity, 123
Format, 123
Goldilocks, 123
History, 123
Integrity, 123, 130, 133
Intended Destination, 123
Lifetime, 123
Priority, 123
Resolution, 123
Sequencing, 123
Suppression, 123
Timeliness, 123
Traceability, 123
Verifiability, 123
Denial of Service, 106
Design and Development Data, **90**
Discursive Text, ii
Disposability Property, 114, 123, 167
Dynamic Data, 35, 36, 60, 61, 64, **92**

End of Life Data, **91**
Evolution
Data, **91**
Explainability Property, 63, 65, 87, 115, 123, 167

Fidelity Property, 63, 65, 78, 97, 113, 123, 166
FMECA, 16
Format Property, 63, 65, 97, 112, 123, 165, 167

Goldilocks Property, 63–65, 114, 123, 167

Hallucinations, **79**
HAZOP, 15, 16, 27, 165
History Property, 97, 114, 123, 167

Informative Text, ii, b
Infrastructure Data, 35, 60–63, **90**
Instructional Data, **91**
Integrity, 12, 17, 19–21, 30, 55, 73, 98
Property, 15, 16, 27, 36–38, 63, 65, 97, 98, 112, 123, 124, 165
Integrity Property, 130, 133, 165
Intended Destination Property, 63, 65, 114, 123, 166
Interface
Control Document, 11, 34, 89
Flooding, 107
Organizational, 25, 156
organizational, 10
Sensor, 104, 111
Specification, 103, 110
System, 9, 26, 89, 92, 157, 158
User, 164
Interface Data, **89**
Investigation
Incident / Accident, 93
Investigation Data, **93**

Justification
Safety, 104
Justification Data, 35, **92**
Justification, Safety, 23, 100, 104, 107

Known Knowns, 99
Known Unknowns, 99

Lifecycle, 61, 77
Data, 23, 25, 34, 65, 68, 146
Delivery, 60
Development, 63
Process, 60
Software, 133, 134
System, 13, 14, 27, 57, 124
Lifetime Property, 114, 123, 124, 167

Machine Learning Data, 77, **77**, **90**, 101, 105, 107, 108, 110, 126
Metropolitan Police, 85
Mitigation, 10, 16, 22, 23, 31–33, **33**, 34, 37–55, 63, 65, 91, 101–105, 108–111, 125

Normative Text, ii, b

Performance Data, 35, **91**
Performance, System, 77, 78
Predictive Data, **89**
Priority Property, 113, 123, 166
Property
Data, *see* Data Property, i, 165
Safety, 22, 31

Qantas, 111

Reference or Lookup Data, **89**
Release Data, **91**
Requirement Data, **89**

Resolution Property, 63, 65, 113, 123, 133, 166
Response, 11, 18, 21, 23, 26, 29, 151, 152
Rumsfeld, Donald, 99

Safety Assessment
    Data, 6, 13, 16, 28
    System, 16, 28, 165
Safety Requirement, 101
    Data, 21, 65
        Derived, 67, 68
        Rejected, 68
    System, 18, 29
Scenario, **141**
Scope, Assumption and Context Data, **89**
Sequencing Property, 63, 65, 114, 123, 166
Software
    Legacy, 108
Software Changes, 103
Software Data, **90**
Staffing Data, **91**
Stakeholder, 9–11, 21, 25, 26, 62
Standards and Regulatory Data, **92**
Stored Data, **92**
Suppression Property, 114, 123, 167
SWIFT, 15
System Safety Assessment, *see* Safety Assessment, System, *see* Safety Assessment, System, 165

Tailoring, 9, **9**, 10, 13, 16, 18, 21, 23, 24, 51, 151
Timeliness Property, 15, 27, 63, 65, 98, 113, 123, 124, 133, 157, 166
Tools, 73
Traceability Property, 97, 113, 123, 133, 166
Training
    AI, 73, 74, 77–81, 83, 87, 101, 105, 107, 108, 110, 124
    Personnel, 11, 13, 26, 60, 61, 91, 101, 156, 160, 163
Treatment
    Medical, 58, 63, 64
    Risk, 21–23, 31, 33, 59, 64, 65, 69
Trustworthiness, **93**
Twinning Data, **92**

Unemployment, 83
Unknown Knowns, 99
Unknown Unknowns, 99, 101

Verifiability Property, 36, 113, 123, 166
Verification Data, 35, 60–63, **90**

Data is here. Data is Growing. Data is Causing Harm

This guidance has been developed by the Safety-Critical Systems Club Data Safety Initiative Working Group (DSIWG) to assist with the management of data, as distinct from hardware and software, 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 a major rewrite of the version 3 guidance, to improve the structure, clarify the workflow, improve the consistency of the normative content and to move much of the appendix material of version 3 into the body of the text. Although the fundamental process has not changed, conformance to previous versions does not guarantee conformance to this version.

SCSC

For Everyone Working in System Safety