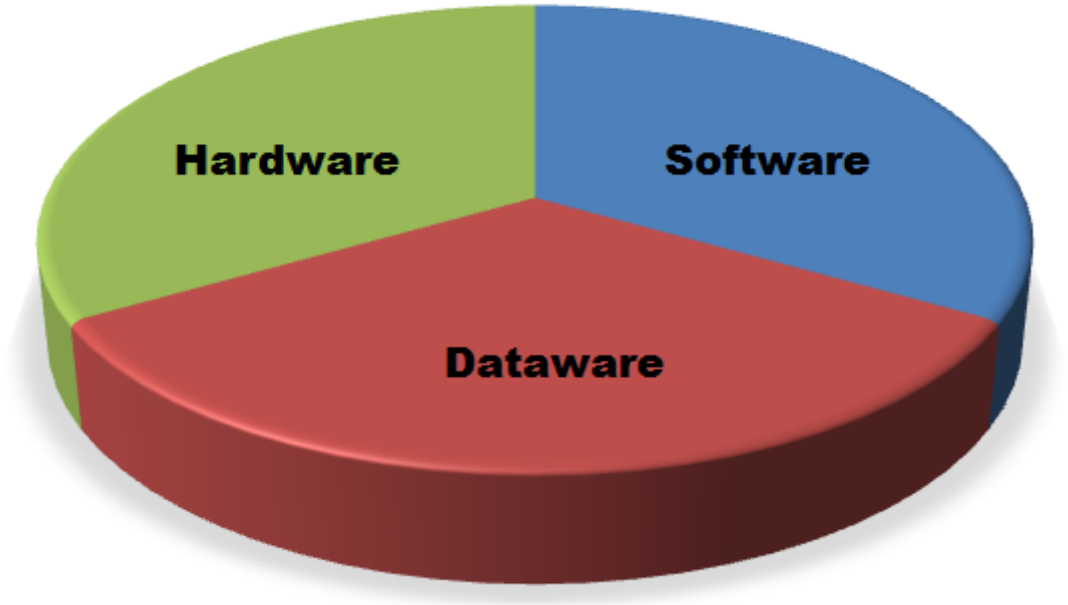
Data Safety *by the SCSC Data Safety Initiative Working Group [DSIWG]*



"Data is a precious thing and will last longer than the systems themselves."

Tim Berners-Lee

"In any collection of data, the figure most obviously correct, beyond all need of checking, is the mistake."

Finagle's Third Law

"We're entering a new world in which data may be more important than software."

Tim O'Reilly

Change History

Version	By	Status	Date
1.0	The DSIWG Team	First draft for external review	31/01/2014

1. Executive Summary

It is increasingly clear that data, not just systems and software, can present a safety problem; consequently the UK Safety Critical Systems Club has promoted efforts to look at the issue of data in safety systems. This started with a seminar 'How to stop data causing harm' held in December 2012 and has progressed with regular meetings of a working group during 2013 and into 2014. The group, comprising industry, academic, government and independent consultants has produced this guidance document.

The objective of the guidance is to describe the data safety problem and initially, for key data types and lifecycles, provide methods for defining the level of risk and recommend strategies and methods for safety risk reduction.

At first it may seem strange that this document is required - after all there are a plethora of existing safety standards and guides covering most sectors that need to consider it. Surely this problem has been addressed? Strange as it may seem, in most sectors, data as separate entity has hardly

been considered, possibly with the exception of the aviation domain. Sometimes certain types of safety data have been identified, but then little guidance on how to manage the risks is provided. In many standards data is treated in a similar way to software, although it is obviously very different. There are also several sectors which now produce and manipulate safety-related data in vast quantities which are not thought to be covered by any existing safety standards, for instance the Police or Government sectors. They need some guidance for data in the absence of anything else.

So this document is trying to "plug the gap", and to do so in a general way that is universally applicable across different sectors. In doing this it identifies a much wider spectrum of safety data that exists in many forms within systems: from specification and requirements data to maintenance and disposal data and everything in between including operational or application data. All of these types of data can have safety implications.

A summary lifecycle is proposed for managing data in safety systems, together with the roles and documents required. This includes steps for assessing the risk, developing a plan for the management of the risk and the execution of the management plan.

A complete list of data types (and how they fit into development, service and data acquisition lifecycles) together with the properties which need to be maintained for safety is presented.

A risk model is developed and this is linked into existing risk management standards.

The different cases of safety data in highly regulated sectors (Nuclear, Defence) is contrasted with that of unregulated ones (Criminal Justice, Government).

Data is considered to have required integrity levels ("DILs") in a similar way to software or systems have SILs, and methods are given for assessing and using the DIL. It also covers data in an organisational context, looking at the risk safety data presents to a particular organisation (for instance a company contracted to build a system or a service organisation using the system) and provides an assessment form to calculate an Organisations Data Risk (ODR) level.

Appropriate data handling and manipulation methods are presented with their applicability by DIL. These cover areas such as System Design, Data Assurance and Test Data Generation.

To maintain relevance to the real world a list of safety incidents and accidents in which data played a role is included as an appendix.

This document is aimed at all those who have an interest or a responsibility for safety data within systems: Managers, Developers, Safety Engineers, Assurers (including ISAs), Regulators and Operators.

This document is a starting position in this complex and evolving area. The aim is to develop this guidance further and integrate it into existing safety standards over the coming years.

2. Background

System safety depends not only on the hardware and software comprising a system, but also on the data within it and in its environment. This data takes many forms:

- It might be for the application, e.g. patient medical records in a hospital;
- About the system itself, e.g. configuration data for a satellite navigation system; or
- About users of the system, e.g. operator competence data in a nuclear power plant.

Incorrect data could clearly have serious safety consequences for the people or other systems using it. Data could become hazardous (that is, lead to harm) in many ways, for instance it could be corrupted, lost, not in the right place, or out-of-date. In many cases the implications of the data being wrong are the same as for serious system failure.

Current safety standards and regulations focus strongly on systems, hardware and software development with data aspects poorly covered. Yet, increasingly, data is at the heart of the system, and, therefore, at least as important as the hardware and software used to generate, store and manipulate it. A new approach is needed, together with standards and guidelines, to ensure that safety data is created, maintained, used and protected appropriately.

2.1. History

This task was taken on by a Working Group of the UK Safety Critical Systems Club. A series of collaborative meetings were held during 2013 and early 2014. This document presents the initial position of the group.

2.2. Aims

The first meeting of the working group agreed the following vision:

To have clear guidance on how data (as distinct from software and hardware) should be managed in a safety-related context, which will reflect emerging best practice.

and the following objectives for 2013:

1. Produce cross-sector guidance by end of the year including a clear statement on handling of data as a separate component within safety related systems;
2. Produce a high level strategic plan by the end of the year for fuller adoption, e.g. into existing standards or a new standard;
3. Influence standard updates currently in progress where we can;
4. Actively promote and disseminate objectives and outcomes of the initiative to the wider community, professional bodies, etc.

2.3. Current situation

In most sectors there are extant formal safety standards and/or guides for development of safety-related systems, with supporting materials to assist in their application. This approach is well-established and focuses on demonstrably assured techniques for both design and verification, e.g. detailed requirements-based testing.

2.4. The change in safety systems

"The world is one big data problem."

Andrew McAfee

Traditionally safety components have been engineered to specific standards and then configured by data to perform a bespoke part of a specific safety system. Today safety systems are becoming more data-intensive and/or data-centric, sometimes using large and complex data sets, which may be stored in commercially available databases. This type of system is different to bespoke safety systems, in that the components are often generic COTS (commercial off-the-shelf) products, not necessarily developed to any formal Assurance or Integrity level, and the criticality is inherently with the data rather than in a directly controlling function. These data-intensive safety systems are often used as decision support or advisory systems, where there is usually a trained and experienced operator, who may be able to detect and correct data problems. However, data is now so complex and of such a large volume it is becoming increasingly unlikely that a user would spot the data errors, and it would be unreasonable to expect them to do so. An example is in the medical field, where the amount and sophistication of medical data stored for a patient is increasing to form a 'data explosion': there is no way a busy clinician could be expected to spot subtle and yet critical data errors, for instance if diagnostic test results were swapped between two similar patients.

There are industry trends which make this initiative very timely: the push for 'Big Data' systems means that safety-related data will be used in more and varied ways, and as part of very large aggregated databases, often via the Internet. Increasing use of Systems of Systems technology (SoS) means that data systems are becoming more and more connected using data from a variety of sources. This means that mapping and translation of data between diverse systems becomes an issue, as well as dealing with data conflicts across multiple systems. The usage of the data is hard to predict in some of these applications; there is a danger that the safety aspects of the data will become lost due to the nature and scale of these systems.

There is also a big push towards much more distributed access to data, e.g. via mobile devices, which means that it much harder to establish the overall data integrity picture. It should be noted that many organisations are becoming highly data dependent: if their data is incorrect (or unavailable) then the organisation cannot function effectively. This is particularly the case for safety-related data.

2.5. Literature

"In God we trust; all others must bring data."

W. Edwards Deming

The data is of fundamental importance. Several authors have proposed over the years that data itself should be subject to hazard and risk assessment and intuitively this appears correct: instead of focussing primarily on the software or hardware, we should concentrate on the data and its usage. One of the few (de facto) standards with specific guidance on data integrity is RTCA/DO-200A (EUROCAE Document ED-76)[1], which is concerned with aeronautical data. This identifies a number of aspects of data quality including Accuracy, Resolution, Confidence, Traceability, Timeliness, Completeness, and Format of the data. It also proposes a concept of Data Processing Assurance Level related to the scheme proposed

below. Work has also been conducted in the nautical navigational field which is highly relevant. See [2], [3] & [4].

2.6. The data-centric picture

There are several properties of data which, if not maintained, could potentially give risk to harm in safety systems. Properties such as Data Integrity and Availability are well established and familiar but it emerges that there is a much wider set of properties that apply to data where loss of fidelity of that property could have safety significance. For example, a small loss of precision of a data item may or may not be significant for a particular system. For example, loss of a 10th of a cm off information about a runway length is unlikely to be significant but loss of a 10th of a cm precision from the position of an atoll on a hydrographic map could be. The key point therefore is that all properties of data (the data-centric picture) should be considered and dependencies on the fidelity of those properties carefully understood. See Section 2.14 for a list of typical properties for consideration.

2.7. Types of data considered in this version

The full set of data types which can have safety implications is large: to date some 19 types have been identified and are documented later in this section. This full set is clearly too much to consider in this first version of the document. Hence for this version the following 5 types have been used to drive the work; the remaining types will be integrated into the document over the coming years. The types below have been chosen as those which give a "quick win", and are easily identifiable within a safety system.

For this phase of the work we have restricted ourselves to the following:

1. Verification (data used to test and analyse the system)
2. Configuration (data used to configure, tailor or instantiate the system)
3. Application (data used in the system during operations)
4. Operational (data collected or produced about the system during trials, pre-operational phases and live operations)
5. Justification (data used to justify the safety position of the system)

2.8. A way forward

Data has to be seen as a “first-class citizen” in system and software design, i.e. data has its own risks, and it can become an asset or liability for the organisation. There is a need to justify the safe handling of the data in the safety case in the same way as the hardware and software comprising a system. If we cannot be sure that the data (in its many forms) is generated, stored, manipulated, distributed and destroyed safely, then all of the other aspects of the safety case are, at best, of indirect help.

2.9. Vision

This work aims to provide guidance for eventual incorporation into the panoply of safety standards. This may take some time, so the initial work is focussed on production of this guidance note. Deciding how, or whether, this work is incorporated into each type of standard is an activity planned for 2014.

2.10. Guidance for use of this document

The SCSC Data Safety Initiative Working Group has developed this guidance to advise a wide audience of the issues related to the use of data in systems where safety could be compromised. It is not written solely for safety professionals, but for anyone with an interest in the increasing use, and reliance, of data in all manner of systems.

The following shows the structure of the document:

Section 1 an executive summary for senior management to understand the purpose and applicability of the document.

Section 2 (this section) provides background information for the document: it describes the data safety issue problem at large and the vision for how the guidance will be used to address it. It also defines the intended readership for the guidance and the scope and applicability for this version of the document.

Section 3 suggests a risk model taxonomy for specifying data criticality levels, or Data Integrity Levels as they have been defined in this guidance.

Section 4 offers guidance related to determining models for the data lifecycles such as data aquisition, system development and system operation.

Section 5 examines data governance issues including data ownership, roles and responsibilities and organisational issues.

Section 6 suggests methods and approaches which are intended to support the safe use of data, including procedural / process assurance methods and mitigation means.

Section 7 discusses qualification of tools that could be valuable in assuring safety through the data lifecycle.

Section 8 discusses the current ongoing development and plans for future work.

Annex A contains a selection of 'War Stories'; real world examples of data safety issues.

Annex B presents a template for a possible Data Management Plan.

Annex C presents the Organisational Data Risk (PDR) questionnaire.

Annex D presents the definitions, acronyms and glossary that have been agreed by the cross-industry Data Safety Initiative Working Group.

Annex E provides a list of references used in the guidance.

Annex F provides the list of people of have contributed to the initiative.

Annex G presents acknowledgments and thanks to those people and organisations without whose help the work would not be possible.

Existing 'sector-specific' standards and documentation related to data safety have been reviewed by the document authors in order to determine best practice guidance. It is important to note that this document in no way replaces or supersedes any standard, nor does it (currently) provide an acceptable means of compliance to any particular standard. Some industry sectors have regulatory and/or legislative requirements on data safety or integrity; this document does not constitute an agreed means of compliance to such requirements.

Guidance is provided to support understanding of data safety issues, to identify the data critical elements of a system and to assist in the application of models, tools and techniques to provide assurance that data is given appropriate priority in system safety analysis.

The document has been structured to guide the user through an intuitive 'data safety' lifecycle comprising the three main steps of: - assessing the risk - developing a plan for the management of the risk and - the execution of the management plan.

The following diagram shows these steps and where specific parts of the guidance can be used to inform that particular activity. This therefore provides a roadmap for how parts of the guidance can be incorporated into the relevant safety management processes where these already exists of when they are elaborated for the particular engagement.

The lifecycle is triggered when the safety related nature of data becomes apparent in the particular system the organisation is involved in. This might be in the early stages of a bid or for an existing operational system where data safety risks have not previously been considered. Once there is an appreciation that there may be some safety dependency further analysis is required to determine the risk. This is normally an informal step.

The next main step is to assess the organisational data risk and secure management commitment to further work. This step is intended to be a relatively quick and cost effective method to understand the overall scale of the risk as far as the relevant organisation is concerned. This step is informed by completion of the **Organisational Data Risk (ODR) Assessment Form**: This is a short and readily digestible assessment of risks that all stakeholders can understand and in particular top management who will inevitably be responsible for committing resources to further effort in assessing and managing the risk. The ODR Assessment form is given in Annex C.

If more detail is required to inform the decision making at this stage, then a **Dataware Framework Assessment Report** can be completed. This is a more structured and detailed analysis of the safety risks as it applies to the organisation. The report provides a framework to better explore tangible hazards in the use of data. This hazard analysis is then used to inform a better understanding of key areas of concern (e.g. specific data exchanges or processing in the scope of supply or operational system) and to help set the direction for further in-depth analysis to develop and implement appropriate mitigation strategies.

The next key step is to develop the plan for how the identified data risks are to be managed. This may be a standalone document or incorporated into the Safety Management Plan. This step is informed by two other activities:

1. **Develop lifecycle models and data type applicability**: this process models the organisations data lifecycle (eg. development, operational etc) and defines when in the data lifecycle(s) additional management of data will be required for specific types of data. A number of reference lifecycle models are given in Section 4 to illustrate typical models.
2. **Determine Data Integrity Level (DIL) for relevant data sets**: this process involves the detailed analysis of the relevant data types and sets and an assessment of their required Data Integrity Level. This for example, will show the level of dependency (on say particular data properties) the organisation has for safe operation. The process for assessing the DIL is given in Section 3.6.

A key feature of the plan will be the definition of which of the recommended or highly recommended data assurance techniques are to be applied, including justification where specific techniques are not to be adopted. A template for the SDMP is given in Annex B.

The final step is the execution of the Safety Data Management Plan. In this step it would be expected that amongst other activities, the planned assurance techniques would be implemented, verified that they are effective and continue to be so for the lifetime of the system.

2.11. Scope & Applicability

2.11.1. Coverage

This document is intended to cover any engineered system that uses or manipulates data that may have an effect on safety.

2.11.2. Intended audience

This document is aimed at the following audiences:

- Managers of businesses or organisations dependent on safety data
- Developers of systems which produce, store, process or distribute data that has safety aspects
- Safety Engineers
- Assurers of safety data (including ISAs)
- Regulators of sectors where data has safety implications
- Users or operators of systems involving safety-related data

2.11.3. Applicability of Sections

While the guidance is intended to be accessible to all, some sections will have greater applicability than others depending on the role of the reader. The following table therefore provides guidance on those sections that will have most significance to the given role of the reader.

Group	Applicable Sections	Relevance	Rationale	Group Risks
Managers of Businesses or Organisations	1, 2.10, 2.13.4, Annex A, Annex C	Relevant to managers of teams / departments / organisations that deal with safety data	Managers need to understand the risks from data so that systems can be procured and operated to the appropriate DIL and safety standards requirements. This will involve time, cost and staff considerations.	If a manager ignores known data risks in procurement, construction or operation they could be held largely responsible for data-related accidents
Developers / Implementers (i.e. any user of an existing safety standard such as IEC 61508 [5], DO-178C [6], etc)	1 - 6, Annex A, Annex C and Annex D	Relevant to designers, developers and implementors of systems configured by, manipulating or producing safety-related data	Developers need to understand the risks due to data and need to use appropriate techniques and methods to manage the risks, including SILs and DALs specified by the safety standard (if any) and DIL requirements in this document	If a developer or designer ignores known data risks that could be addressed by design they could be held responsible for data-related accidents
Safety Engineers	1, 2, 3, 4, Annex A, Annex C and Annex D	Relevant to all safety engineers who are assessing the safety risk of a system which includes data	Safety engineers need to understand the potential for incidents or accidents as a result of data. Data should be considered alongside Hardware, Software, Procedures and Human Factors when performing a safety analysis of a system.	If a safety engineer ignores the potential for data to cause harm within a system they could be held partly responsible for data-related accidents
Assurers of Safety Data	1, 2, Annex A and Annex D	Relevant to auditors / reviewers / ISAs who review or audit systems involving safety data	Assurers, auditors and ISAs need to verify that the guidelines of this document have been considered together with any DIL or other requirements	If an assurer ignores known data risks during review or audit they could be held partly responsible for data-related accidents
Regulators	1, 2, Annex A and Annex D	Relevant to regulators / approvers / certifiers who sign off or otherwise approve systems involving safety data	Regulators need to check that the risk management regime appropriate for their sector has been considered (especially that the data risks have been considered properly) together with any standards requirements.	If a regulator ignores known data risks during review or certification they could be held partly responsible for data-related accidents
Users or Operators	Whilst not directly involved in the data management process, users may find Annex A of interest	Relevant to all users or operators of the system who enter, configure or use safety data	Users need to know which bits of data are critical (i.e. what DIL) so that appropriate methods are employed to manage associated risks (by process and procedures outside of the system).	If a user or operator ignores known data risks that could be mitigated by process or procedure they could be held partly responsible for data-related accidents

2.11.4. Indirect and direct systems

This guidance covers data in both direct and indirect situations (i.e. safety-critical control systems that directly effect equipment, and safety advisory systems that provide data to users or operators for their further use). The data in these systems may be used or stored in many forms: configuration parameters, external databases, internal file stores, etc. All data used or stored in a digital form and which has safety aspects should be considered.

2.11.5. Full List of types of safety-related data

The table below gives the current view of the types of safety-related data that contribute to, are used by, produced or affected by, safety systems. They are presented in a typical lifecycle order starting with data that is used to establish the required behaviours and specification of the system, through data about operational performance, updates and evolutions to an established system, to data about disposal and decommissioning of the system.

No.	Data Type	Description	Explanation	Typical Containers	In Guidance v1.0?	Examples (Sector: Use)
1.	**Prediction**	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, etc.	N	Air Traffic Navigational Aids: Terrain information used in instrument landing system performance simulation
2.	**Assumption**	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	CONOPS, Safety Case Report part 1	N	TBD
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, ICDs, User Requirements documents, Safety Case Report part 1.	N	TBD
4.	**Interface**	Data used to enable interfaces between systems: for operations, initialisation or export from the system	Data that exists to enable exchange between systems. Covers start-of-life operations (data import or migration), end-of-life operations and ongoing operational exchange of data between systems.	Protocols, Interface Spec, Schemas, ICDs, Transition Plans, ETL tool specs, Cleansing and Filtering rules	N	Air Traffic Management: ASTERIX format for exchange of aircraft surveillance information (http://www.eurocontrol.int/services/asterix)
5.	**Design & Development**	Data produced during development and implementation	This is data encompassing the design & 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 and design, Test scripts, Code inspection reports, etc. Safety Case report part 2.	N	TBD
6.	**Verification**	Data used to test and analyse the system	This is data comprising the test values and test data sets used to verify the system. It may include real data, modified real data or synthetic data. It includes data used to drive stubs, and any data files used by simulators or emulators.	Test data sets, Stub data, Emulator and Simulator files	Y	Simulator platform definition files.
7.	**Configuration**	Data used to configure, tailor or instantiate the system	Data used to set up and configure the system to perform a particular function, for a particular installation, product configuration, behaviour or specific usage	Configuration files, Initialisation files, Hardware pin settings, Network addresses, Passwords, etc.	Y	Air Traffic Management: Runway Visual Range system mapping of runway light percentage settings to actual brightness of the lights. ... "Run-Time Blueprint Data" defining virtual communication channels between applications. ... Specification of aircraft characteristics for apps useable on/with different platforms.

8.	**Application**	Data used in the system during operations	This is the data processed or produced by the system which has end-user meaning. It may be displayed and used within the system or may be for transfer or distribution to other systems or downstream users. It is data that has some real “application” meaning, i.e. is not to do with the system internals.	May 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).	Y	Air Traffic Management: Aircraft identification, position and intention information at an air traffic control working position. ... Inertial and GPS Navigation, Target tracks, Maps.
9.	**Instructional**	Data used to warn, train or instruct users about the system	This is data that explains to users the risks of the systems and gives any mitigations that may be required to be implemented by users, e.g. by process, procedure, workarounds, limitations of use	Manuals, SOPs, On-line help, Training courses, etc. Safety case part 3	N	TBD
10.	**Release**	Data used to ensure safe operations per release instance	Explanation of particular features or limitations of a release or instance. May include specific time-limited workarounds and caveats for a release.	Release notes, Certificates of Design, Transfer documents, Safety case part 2 or part 3	N	TBD
11.	**Operational**	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 may be the results of various phases of introduction and may include trend analysis to look for long-term problems.	Field data, Support calls, Bug reports, NCRs, DRACAS data	Y	Service record of defects and failures over time. Equipment lifetime management data.
12.	**Evolution**	Data about changes after deployment	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, CM system outputs	N	TBD
13.	**System**	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 may cover hardware, software and data.	Inventory, asset and maintenance database systems	N	ED-153 Software Safety Folder
14.	**Justification**	Data used to justify the safety position of the system	Data used to justify, explain and make the case for starting or continuing live operations and why they are safe enough. Often passed to external bodies (regulators, HSE, ISAs) for their review.	Safety Case report, Certification case, Regulatory documents, COTS Justification file, Design Justification file	Y	Object Management Group: Structured Assurance Case Metamodel (http://www.omg.org/spec/SACM/). ... RTCA/DO-178C [6] PSAC & PHAC and compliance matrices. ... DEF-STAN 00-56 [7] Safety Case.
15.	**Staffing & Training**	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 may include training records, competency assessments, permits to work, etc.	HR records, training certificates, card systems	N	TBD
16.	**End of Life**	Data about how to stop, remove, replace or dispose of the system	This is data covering all activities related to taking the system out of service or mothballing / storage / dormant phases	Transition, Disposal and decommissioning plans	N	TBD

17.	**Investigation**	Data to support accident or incident investigations (i.e. potential evidence)	This is data collected or produced during an accident investigation which may be used in investigation reports, lessons learnt or prosecutions. This can be source data (e.g. photographs of crash site) or may be derived (accident simulations, analyses, etc)	Accident Investigation reports and supporting documents	N	TBD
18.	**Standards and Regulatory**	Data that governs the approaches, processes and procedures used to develop safety systems.	This is 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, regulatory requirements and laws	N	TBD
19.	**Reference or Lookup**	Data used across multiple systems with generic usage	Data comprising generic reference information sets used by multiple systems. Typically updated infrequently, and not specific to this system.	Dictionaries, materials information, sector data reference sets, encyclopaedias, etc	N	The Aeronautical Information Publication

2.11.6. Relationship between safety and security regulation of data

Whether at an enterprise, business or project level, both safety and security regulations must be satisfied. Challenges often exist as a result of a natural tension between compliance strategies for these regulations. This tension can, for instance, be related to the need to provide access to data (for safety reasons) and the need to constrain it (for security reasons).

Compliance is often managed by different organisational hierarchies, different compliance regimes, different sets of external stakeholders, different sets of career backgrounds, different approaches to (and definitions of) risk, different mindsets and cultures, even within a single organisation. The results can be tension, contradiction, mutual misunderstanding, operational cost, and drivers that reduce operational effectiveness.

Safety and security risks should be managed effectively within an organisation’s operations, taking a business case approach in the interests of ensuring an effective balance between organisational risk, cost and effectiveness. This approach recognises that a number of controls are related to both safety and security, including:

- Physical
- Procedural
- Personnel
- Technological

The need to manage the risk associated with data that is passed between organisations represents a further level of challenge and complexity, especially as the confidence that an organisation may have in data may be difficult to determine, and as the significance of the risk or benefit associated with the data being handled by systems and particular organisations may manifest itself elsewhere. Approaches to the integration of both safety and security considerations require that the tensions between these are managed both within and across organisational boundaries. The following principles support organisations in determining the appropriate compliance and risk management strategy for data safety and security:

1. “Buyer beware” – an organisation should ensure that it makes an outline assessment of the risk associated with data that is received and/or handled, and that it adopts an appropriate and proportionate management strategy.
2. An organisation should consider, and manage, any interdependence between safety and security compliance and risk management implications when determining compliance strategies.
3. Overall business implications should be considered in the implementation of appropriate measures for the risk or compliance management of data.
4. Organisations should consider, and manage, interdependence between compliance strategies and related controls when making changes.

The consideration and implementation of data safety and data security strategies should take place within an overall business risk management context, as appropriate to the particular enterprise and/or project under consideration. In practical terms, the interdependence between safety and security requirements needs to be recognised at the start of a complex project, and the security implications of a system need to be identified early, with safety and security issues being resolved with visibility from both sides.

2.12. Relationship to existing safety standards and guidance

This document is intended to augment existing software, systems and hardware development standards and guides; it is not intended to replace them. It is intended that existing sector-specific guidance will incorporate the ideas from this document over time.

2.13. Rationale

2.13.1. Safety-related systems, and the benefits of additional guidance relating to Safety-related data

2.13.1.1. Highly safety-regulated sectors

In highly regulated sectors such as defence and civil nuclear power, sector-specific safety regulations and standards typically identify what is meant by a 'safety-related system', and such systems are subject to rigorous development and operational processes. Safety-related data is sometimes, but not always, explicitly identified as such and is typically managed within the context of hardware and software-based systems that are subject to a development lifecycle, which includes interface and data specifications.

Standards recognise the need for the suppliers to identify functional and physical boundaries for their scope of supply, and in doing so they set the boundaries for, and context of, the safety assessment for data. The standards also recognise the need to make assumptions taking into account expected system use within its environment. Such assumptions are very important and may be wide ranging, for instance, that engineered interfaces will be compatible and interact correctly. Other assumptions may concern the training and expertise of system users. They may also include assumptions regarding the completeness or validity of manually entered data or data obtained from other external sources.

Hence data-related safety is typically addressed within the context of clearly identified responsibilities and within system boundaries, and systems are engineered with appropriate rigour to ensure the correct management of safety-related data within these defined boundaries. This is more manageable in sectors and applications where interacting systems fall under the influence or control of a common Authority, and in such cases industry-specific approaches have been developed to ensure system compatibility and interoperability. An example of this is the UK MoD's identification of a standard 'Generic Vehicle Architecture' (GVA), 'Generic Base Architecture' (GBA) and 'Generic Soldier Architecture' (GSA).

Specific data communication approaches have been identified to support the effective interaction of engineered systems, including taking into account properties of data and treating these accordingly in view of its ultimate application.

Whilst this can ensure that interfaces are compatible, the challenges related to safety related data are broader. For instance, ultimately data will be derived from a given source with a given degree of confidence, and may be acted upon in combination with other information, some time later, in support of an action or decision. The overall safety of a System may depend on the properties of data that may have been passed through many subsystems being known, in particular its accuracy and currency/validity. Even if processing and compatibility is assured, some data that is used by a system in determining an output may be stored in databases, which must also be correctly labelled and stored, other data may be entered manually. Such data may, for instance, update the values of parameters or set a mode of operation, and may have a bearing on the overall safety of a safety-related system. For that reason, this form of data is also given consideration in this guidance.

The safe operation of interacting systems largely depends upon the validity of data that is input to them, and the correct and appropriate application of data that is output from them. It is sometimes impossible to know in advance exactly how and where data used or produced by one system may be used in safety-related decision-making within a larger, distributed system.

Distributed and/or reconfigurable communication systems themselves form a part of a larger safety-related system when data and information that they transmit is used in a safety-related application. It may be also that safety-related data may be received from, or passed to, interacting systems that are yet to be specified and developed, for applications that are not yet identified. The 'owners' of clearly bounded and safety-justified subsystems therefore often need to identify assumptions regarding the validity and applicability of data in support of their safety justifications, and in support of overall System safety.

2.13.1.2. Other sectors

Other sectors also rely on data and in these cases regulators provide little or no guidance relating to the management of any associated safety implications, and there can be less standardisation. Examples are:

- In commercial shipping, where undue reliance may be placed upon the accuracy and validity of data derived from navigational and hydrographic charts (despite the charts including disclaimers regarding the level of reliance that should be placed upon them). Electronic charts are increasingly the 'norm' and these are implemented in Electronic Chart Display and Information Systems (ECDIS), which in turn are commonly used to calculate navigational passages for masters to follow. The configuration of such systems, their correct functionality, and the correctness of the navigational and hydrographic chart data all influence safety, and carry potential liability. Issuing warnings and caveats in such cases does not remove the obligation for suppliers to consider and address safety insofar as this is possible;
- In the medical sector, where the accuracy and completeness of medical records is imperative to ensuring correct and timely medical treatment. Examples could be medical records that present incorrect identities, corrupt diagnostic test results or medical histories;
- In Government sectors such as the Department of Justice, safety-related decisions are increasingly made that depend upon the validity of data that is held in databases and/or is passed between various electronic and procedural systems. In such cases, although data is not regulated or assessed for safety, it is essential to ensure the correctness and completeness of information relating to the identity, location and criminal record of individuals. In such cases, the safety of the public could be jeopardized by actions based on incorrect data. Examples could be systems that incorrectly identify individuals as suspects of serious crimes, or fail to correctly identify those who present real threats.

Whatever the regulatory environment, there is a trend towards combining data-based systems in support of decision-making, including where it may

have direct safety implications. There may therefore be opportunities for the 'owner' of a subsystem to consider and address emergent safety implications, and the potential liabilities if data derived from their systems is found to have been a contributing factor to an accident.

Similarly, overall system integrators need to be aware of the implications of combining and acting upon data that may be derived from a range of diverse sources.

A sub-system supplier's contribution and influence in providing for overall system safety is important, and this guidance proposes an organisational risk assessment at the outset of a new project that informs the organisation as to the potential significance of data from safety and liability perspectives

2.13.2. The rationale for generally applicable guidance

It is acknowledged that data has been a contributing factor in several incidents to date (examples are given in Annex A). However there is no widely recognised guidance as to how the safety of data (in whatever form) can be considered.

This guidance will be useful in framing the substantiation of data assumptions and constructing appropriate safety arguments and evidence, as well as in influencing design. It will be of most value when organisations consider how best to address new or emerging requirements for the development of systems that handle, or may produce, safety-related data.

The guidance must be applied within the context of industry-specific requirements and is consistent with the application of existing cross-sector standards.

2.13.3. Cross-sector standards and guidelines

Even within highly safety-regulated sectors, the issue of how the integrity of data relates to the safety of systems, and how this might be managed within, and across, specific system components is not well supported by existing guidance. Data that is passed between subsystems for which different organisations are responsible can start off being valid and accurate, but the passage of time may reduce its validity. These issues are better understood in some sectors than others (and appropriate assessment and management approaches are embodied in standards such as Def Stan 00-56 in Defence), and elements of experience and good practice are transferable.

Where cross-sector standards (such as IEC 61508) have been identified for the design of safety-related systems, the standards necessarily recognise that what is to be assessed must first be defined within clear boundaries, and that users of the standards can only influence design and operation within their applicable sphere of influence. IEC 61508 (IEC 61508-4, Edition 1.0 1998-12) states that:

- A safety-related system comprises everything (hardware, software and human elements) necessary to carry out one or more safety functions, where failure of the safety function would give rise to a significant increase in the risk to the safety of persons and/or the environment;
- A safety-related system can comprise stand-alone equipment dedicated to perform a particular safety function (such as a fire detection and suppression system) or can be integrated into other plant or equipment (such as motor speed control in a machine tool);
- A safety-related system is a designated system that both:
  - implements the required safety functions necessary to achieve or maintain a safe state for the Equipment Under Control (EUC); and
  - is intended to achieve, on its own or with other electrical/ electronic/ programmable electronic (E/E/PE) safety-related systems, other technology safety-related systems or external risk reduction facilities, the necessary safety integrity for the required safety functions

It follows, and IEC 61508 recognises, that where risk needs to be reduced in an overall 'system' then additional risk reduction needs to be implemented in other parts of the overall 'system'. An overall view of system safety is required, where the hazards and risks associated with data are assessed and managed accordingly.

2.13.4. Applicability of guidance

Different sectors have their own regulations, experience, constraints and abilities to influence how safety-related data is identified and treated. Whilst experience and approaches differ, there is the opportunity to draw broadly upon experience and lessons learnt in order to benefit all sectors.

This guidance is intended to assist those who are engaged with different elements or lifecycle phases, recognising that they have different spheres of influence. It takes a broad view of safety, considering an overall System perspective that may include people, systems and processes, and recognises that there are significant differences between industry sectors in the way that safety-related data is recognised and managed.

Accordingly, guidance is generally applicable, and it proposes measures that could be considered for application across any safety related data application, in any sector. It is applicable to the development of systems that recognise and manage safety data-related risk effectively.

Specifically, this guidance will support developers of engineered systems that may, in practice, become just one sub-system component of an overall System, providing specific guidance to assist the effective management of safety related data in view of its likely safety significance.

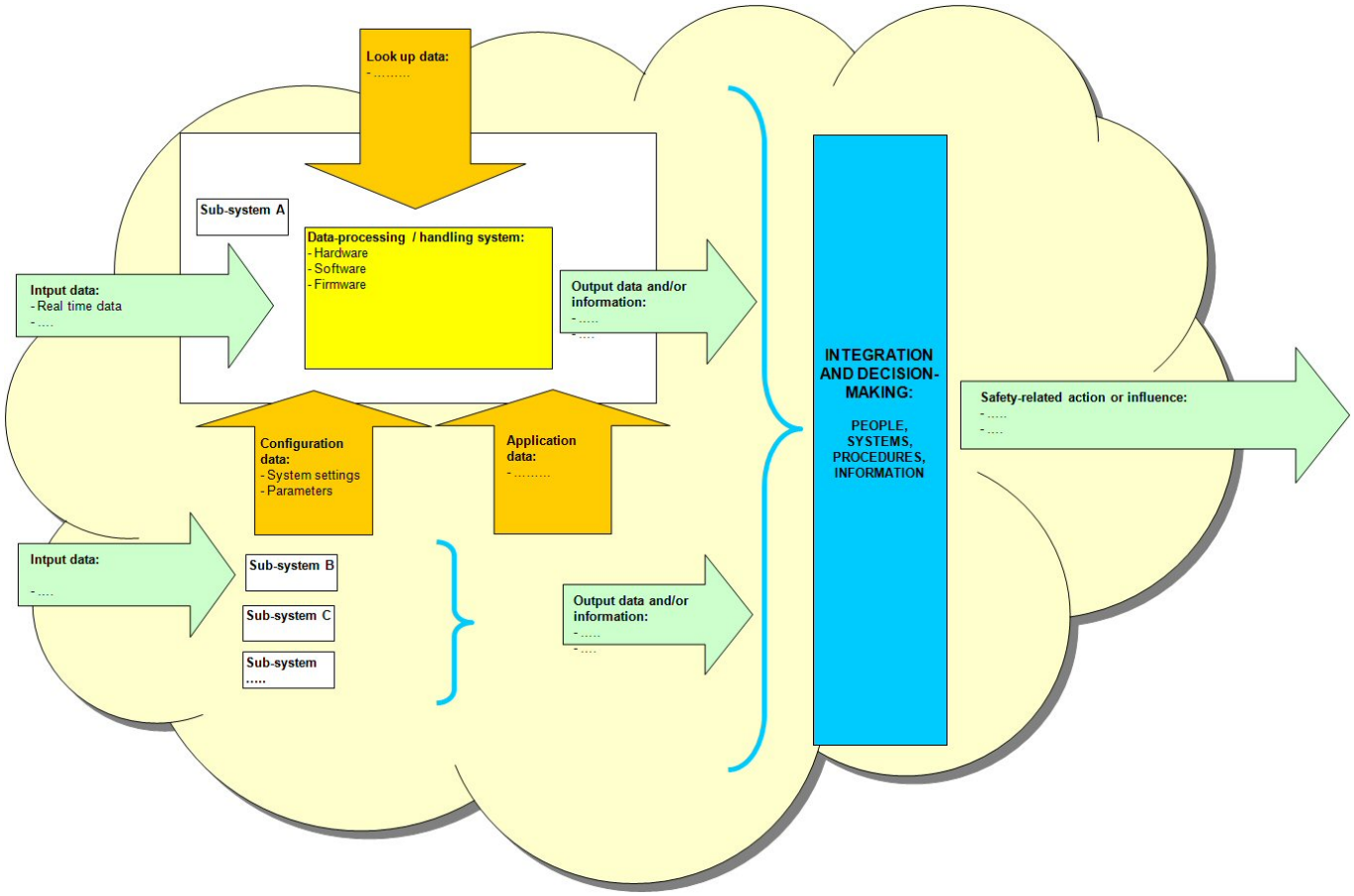
The guidance is also aimed at influencing the development of future guidelines and standards; these may recognise a number of principles that are generally relevant, such as the need to ensure that the safety-related implications of data are being appropriately managed by those who are best placed to influence safety during the development and operational phases.

2.14. Safety-related data and properties

The correct operation of safety-related systems is typically highly dependent upon such systems being configured correctly with configuration data, processing valid input data in accordance with their system specifications, and recognising and responding appropriately when deficiencies in data are present.

A number of types of data have been identified by this guidance, each of which are important the development or operation of safety-related systems. Examples would be data that is stored within systems and looked up when required, data that is manually entered to configure a system or function for a particular task, or data received electronically from another system in real time.

The figure below represents just one simple illustration of safety-related data in the context of the safety of an operational system. Other types of data are also important. For instance, in design and development, data sets are used to test that systems are functioning as required, and data is also used in the development and application of test tools that are used in the development and testing of safety-related systems.



2.14.1. Properties

The list of data properties below has proved useful in discussions, but may need modification or additions. Data (and its properties) as an abstract entity is difficult to describe and one person's interpretation may be different to another. It is suggested that the list of properties below is tested against various "real-world" situations involving data and it is refined further. For this work it is the loss or negation of the property which is of interest, e.g. if Integrity is lost, could this present a hazard? For all the properties below it is thought that this could be the case in certain situations, and therefore it is a useful property to include here.

It is noted that James Inge's work [2] produced a useful taxonomy of data types, and went on to look at faults in data. He concluded that a rigid taxonomy of data types was unhelpful due to various properties or characteristics of the data which vary independently. Nevertheless, such lists are proving useful in helping to articulate, elaborate and explore the safety dependencies that exist with data properties.

Property	Description
Integrity	the data is correct, true and unaltered
Completeness	the data has nothing missing or lost
Consistency	the data adheres to a common world view, e.g. units
Format	the data is represented in a way which is readable by those that need to use it

Accuracy	the data has sufficient detail for its intended use
Resolution	the data is stored in such a way as to preserve its accuracy
Traceability	the data can be linked back to its source or derivation
Timeliness	the data is up to date
Verifiability	the data can be checked and its properties demonstrated to be correct
Availability	the data is available when needed to be used
Fidelity / Representation	how well the data maps to the real world entity it is trying to model
Priority	the data (items) are presented / transmitted / made available in the order required
Sequencing	the data (items) are preserved in the order required
Intended Destination/Usage	the data items are only sent to those that should have them
Accessibility	the data items are only visible to those that should see them
Non-accessibility / Non-availability	the data items are intended never to be used again
History	the data has an audit trail of changes
Lifetime	when does the safety data expire
Disposability / Deletability	the data can be permanently removed when required

2.15. Common Data Safety Issues

There are some issues related to data which are different to any other element in a system (hardware and software), and there are others which, because of increasing prevalence, are now presenting a greater risk. Four specific issues are worth highlighting: reuse of data in a new system or setting, aggregation of multiple data sets into one set, ageing of data and archiving / retrieval.

In each of the following cases the properties detailed above need to be considered. If any are affected then analysis and justification may be required.

2.15.1. Reuse

Reuse in this context is use of the same data in a different system or system context. Data in safety systems can be reused similarly to software, but there are aspects to be addressed:

- Similarity of usages - how close to the original use is the new one?
- Compliance with new requirements
- Data Integrity Level (DIL), i.e. the 'amount of integrity' required, should be less than or equal to that of the previous system implementation
- Boundaries/ranges or coverage data is similar
- Similar properties with justification if not close enough

2.15.2. Aggregation

Aggregation here means concentration of data from several disparate sources into a new data set

- Combinations of data sets can result in loss or corruption of data and therefore needs to be demonstrated to be safe. Just what has been lost?
- Translation and mappings need to be checked
- Unintentional loss of context and history may occur
- Unintentional loss of data properties due to omissions or changes

2.15.3. Ageing

All safety data has a lifetime and this needs to be managed

- Need to explicitly address the data lifetime in the management regime
- Needs to be specified in the data requirement
- Need to consider: purging, deletion, flagging/alerting, stale data, fitness for purpose
- Need to keep the data as a true representation of what it is modelling (e.g. external world or usage may change)

2.15.4. Storage, Archiving & Legacy Data

Safety data needs to be available when required

- Need to think about accessibility of data over complete system lifetime
- Properties of the data have to be preserved via different media
- Have to store in a form that can be retrieved at some point in the future

3. Risk Model

3.1. Introduction

There are a number of competing definitions for "risk" which makes finding a general purpose definition difficult. However, in order to set business objectives which achieve an acceptable level of risk, it is important to define metrics and rules which allow developers and engineers to elicit verifiable requirements and measures to assure the performance of a system against the desired level of risk. "Risk" is an attribute of hazards, which is typically defined as a function of the likelihood of an outcome and the magnitude of the consequences, ("likelihood" here is being used in its colloquial sense, covering qualitative notions such as "common" and "rare" and quantitative notions such as the probability of occurrences per hour/day etc.)

Different organisations implement different types of risk management plan, but the high-level steps in ISO 31000 [8] are indicative of most processes. The list below contains the risk management processes contained with ISO 31000:

- Establishing the Context (identify risk appetite and scope of assessment)
- Risk Identification (identify sources of risk)
- Risk Analysis (assign categories or levels of risk to sources)
- Risk Evaluation (compare risk environment with risk appetite)
- Risk Treatment (resolve/circumvent/avoid risks where necessary)

ISO31000 recommends that there are two other parallel activities which monitor and review the risk assessment process and communicate and consult with the stakeholders about the risk assessment process.

The ODR assessment is designed to inform the process of establishing the context of the risk assessment (see the accomanying guidance notes in Annex C). In order to support the remaining steps and align data safety with other risk management processes, one must first overcome the problems stemming from the fact that "likelihood" becomes difficult to apply in contexts where there are no failure rates. For this reason the DIL (Data Integrity Level) was developed. The DIL is a heuristic metric which allows developers to classify data sets (existing and future) in terms of the risk they pose. From here the focus becomes not on a statistical measure of likelihood, but on the way in which an assurance argument is built up for the data meeting the requirements of the system in question.

3.2. Risk Assessment Context

3.2.1. Organisational Data Risk and ISO 31000

The organisational data risk (ODR) assessment form was generated to capture a high-level perspective on the risk posed to an organisation by data safety issues within a specific project. How it interfaces to an organisation's risk management policies is the responsibility of the implementing organisation. However, what follows is a discussion for the interested reader on how ODR assessments connect with the ISO 31000 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, ensuring that there is no overrun of the assessment's boundaries and that things which are out of scope are explicitly communicated to all stakeholders. In addition, it is the role of this activity to produce the risk criteria that a system will be judged on. The ODR assessment links directly to the sub-tasks identified by ISO31000 for establishing the risk assessment context and introduces aspects to guide the assessor into focussing on data-specific risks.

Questions 2, 3 and 4 of the ODR align directly with establishing the external context of the risk assessment (activity 5.3.2 from ISO 31000). They guide the assessor into judging the risk tolerance of external stakeholders, the level of risk that is allocated to the organisation and the regulatory environment within the project will operate.

Question 5 is concerned with establishing the internal context of the risk assessment (activity 5.3.3), inviting the assessor to comment on the maturity of the organisation in terms of their attitude to not simply risk, but specifically 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 5.3.5 : "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". This is expanded upon by questions 1, 7 and 8 which go into data-driven specifics about 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 5.3.5) which handles how to combine these different aspects of risk into a single, high-level estimate of the risks associated with a given data-driven project.

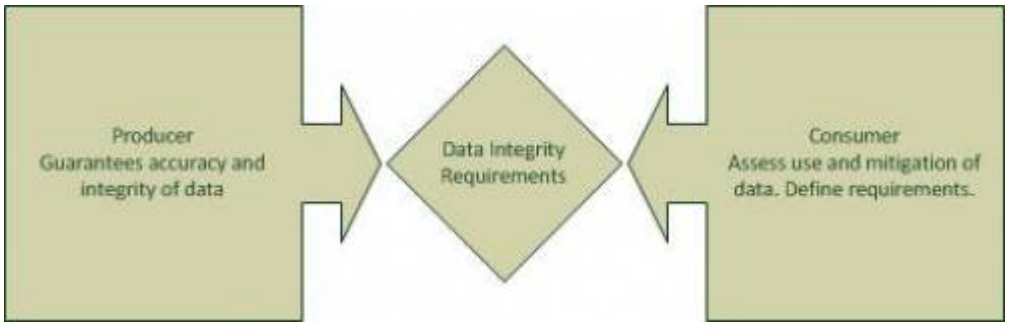
It is of note that whilst the completion of an ODR fits within the context establishment activity it also augments the ongoing activity "Communication and Consultation" activity both by providing a standardised format for capturing the relevant information and by providing a standardised reporting template in order to secure endorsement for the plan.

3.2.2. System Definition

The system under consideration should be understood and documented, including interfaces and data aspects. The process of documenting the system of interest not only furthers the understanding of the reviewer so they can make sensible judgements about the system, but also both formally declares assumptions that the reviewer is making whilst assessing the system and clearly defines the limits of the assessment. In addition, different levels of risk may be associated with composites or groups of data where it is easier to manage (or where independence cannot be demonstrated or maintained), so the partitioning of datasets should be considered and modelled at this stage.

3.2.3. Usage Scenarios

If data is incorrect it can become dangerous when used – either by making a computer system perform incorrect actions, or by misleading human users into making incorrect decisions. Because the danger can only be determined when the usage of the data is understood, risk assessment must involve the consumer of the data rather than the producer.



The **consumer** assesses the use and possible mitigations for the data, and uses this information to define **data integrity requirements** – how accurate and reliable the data must be.

The **producer** investigates how the data is collected and what errors might occur, and guarantees that the data meets the data integrity requirements.

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

3.3. Risk Identification

ISO 31000 uses the term "Risk Identification" to describe the identification of sources of risk. Here, the authors use "hazard" as a generic source of risk of accidental harm. If a hazard is missed then no action will be taken to manage it, so hazard identification must be rigorous and systematic. Two possible approaches are outlined below:

3.3.1. Function Based (top-down)

If the consumer is a system which has clearly identified safety functions, then data can be assessed by considering each function in turn and analysing what data the function depends on. If there are a limited number of safety functions, this is usually the simplest approach.

3.3.2. Data Based (bottom-up)

A more general approach is to consider each different category of data and explore the effect of possible data errors. The errors which are most likely to have safety impacts are those which effect the data properties as described in Section 2.14.3.

Data-based analysis has the advantage that it may identify safety implications that are currently unknown, but it can require considerable effort for systems with large data sets.

3.4. Risk Analysis

From a data perspective, analysing risk involves attributing a level of risk to each dataset. As data does not intrinsically have a failure rate, the use of a general classification system is recommended to simplify making judgements about how much integrity assurance is required from a given item. The table below is an example classification system, allocating Data Integrity Levels (or DILs) to data items, using a function of likelihood/consequences. Implementors of such a system are encouraged to tailor this classification system to their own needs. To make the analysis applicable to data, the "consequences" are subdivided further into five categories which impact on the level of risk:

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

		Likelihood of Data Causing Accident		
		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.
	Detection	Low or no chance of anything else detecting an error.	Some other people/systems are involved in checking the data.	Many other people/systems are involved in checking the data.
	Prevention	Difficult or impossible to guard/barrier against errors.	Possible to guard/barrier against errors.	Easy to guard/barrier against error.
	Correction	Difficult or impossible to correct or workaround errors.	Possible to correct or workaround errors.	Easy to correct or workaround errors.
Severity or Impact of data related accident				
Negligible	Negligible harm. Negligible environmental impact.	DIL0	DIL0	DIL0
Minor	Minor injury or temporary discomfort for 1 or 2 people. Minor environmental impact.	DIL1	DIL0	DIL0
Moderate	An accident resulting in minor injuries affecting several people or one serious injury. Some environmental impact.	DIL2	DIL1	DIL1
Major	A serious accident resulting in serious injuries affecting a number of people, or a single death. Major environmental impact.	DIL3	DIL3	DIL2
Catastrophic	An accident resulting in death for several people. The accident could affect the general public or have wide and catastrophic environmental impact.	DIL4*	DIL4	DIL3

*this level is not recommended as there is too much dependence.

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

3.5. Risk Evaluation

Evaluating the risk involves comparing the risks associated with the data in the system with the organisation's accepted level of risk. This evaluation may choose to either accept or treat the risk to bring the project risk back in line with the organisation's risk appetite.

This step is entirely organisation/project specific and should be in line with local risk management policies.

3.6. Risk Treatment

How an identified risk is treated is obviously implementation and system specific. However, Section 4 provides guidance on common strategies for sourcing/generating new data which will satisfy DILs and Section 5 is a compilation of suggested methods and approaches for treating data-related risks.

4. Lifecycles

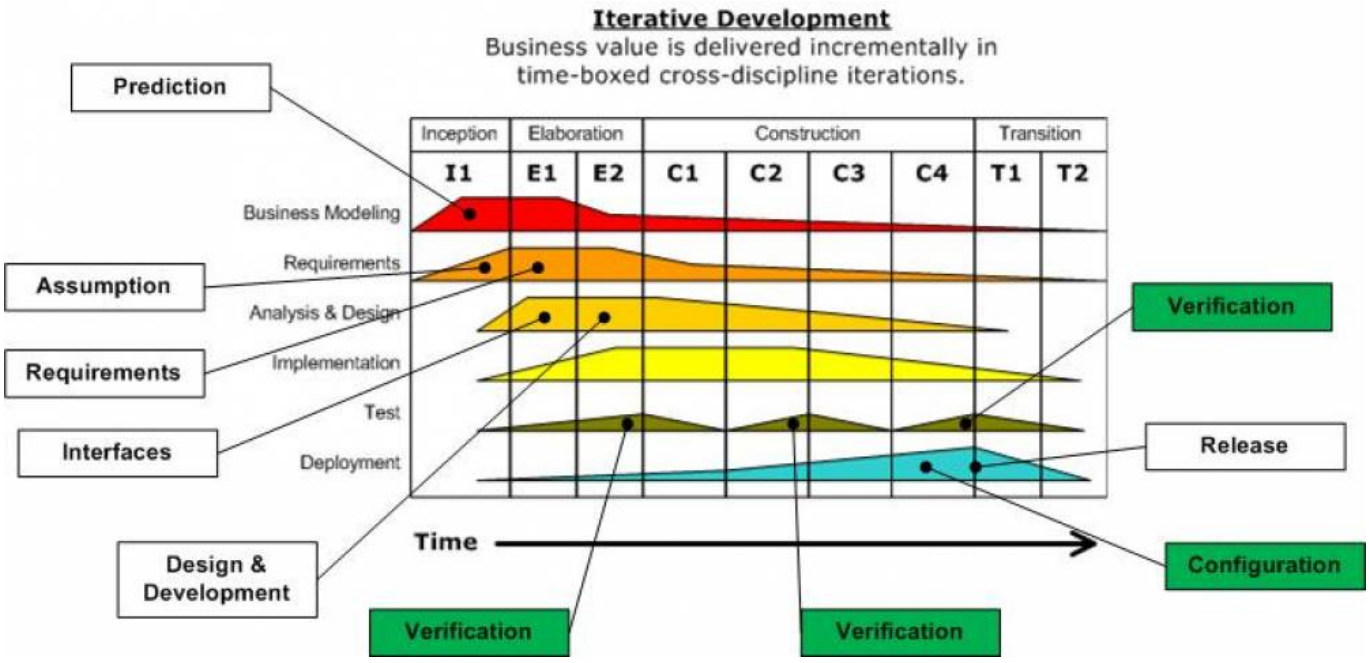
Like other components of a safety system, the safety dependency of data is dictated by the context in which it is used and the causal links that become established where loss of fidelity of one or more of its properties can contribute directly or indirectly to outcomes of the overall system that lead to harm. For example, a given data set (say configuration data) could be used in a number of separate contexts such as:

- protoyping a system to demonstate solution feasiblity of a safety system
- development testing of a safety system
- live operational use of a safety system

In these cases, the data set is the same but the context of its use changes the safety significance and therefore the level of assurance that it may require. Therefore, not only is the type of data under consideration important but also *when* in the organisation's particular process lifecycle the data will be used and depended on. It therefore follows that the assessed integrity level of a data set is also predicated on where and when in the lifecycle the data set will be applied. It is recommended these considerations are addressed in the Data Management Plan by modelling the organisation's lifecycles and explicitly documenting where a specific data set will be used and therefore subject to further assurance techniques. To illustrate this concept, a number of generic model lifecycles are discussed below and are expected to approximate to most real life scenarios. Note that these are not intended to be prescriptive or mandating any particular model but rather are being used to illustrate how the Data Management Plan could articulate these lifecycle considerations.

4.1. Development

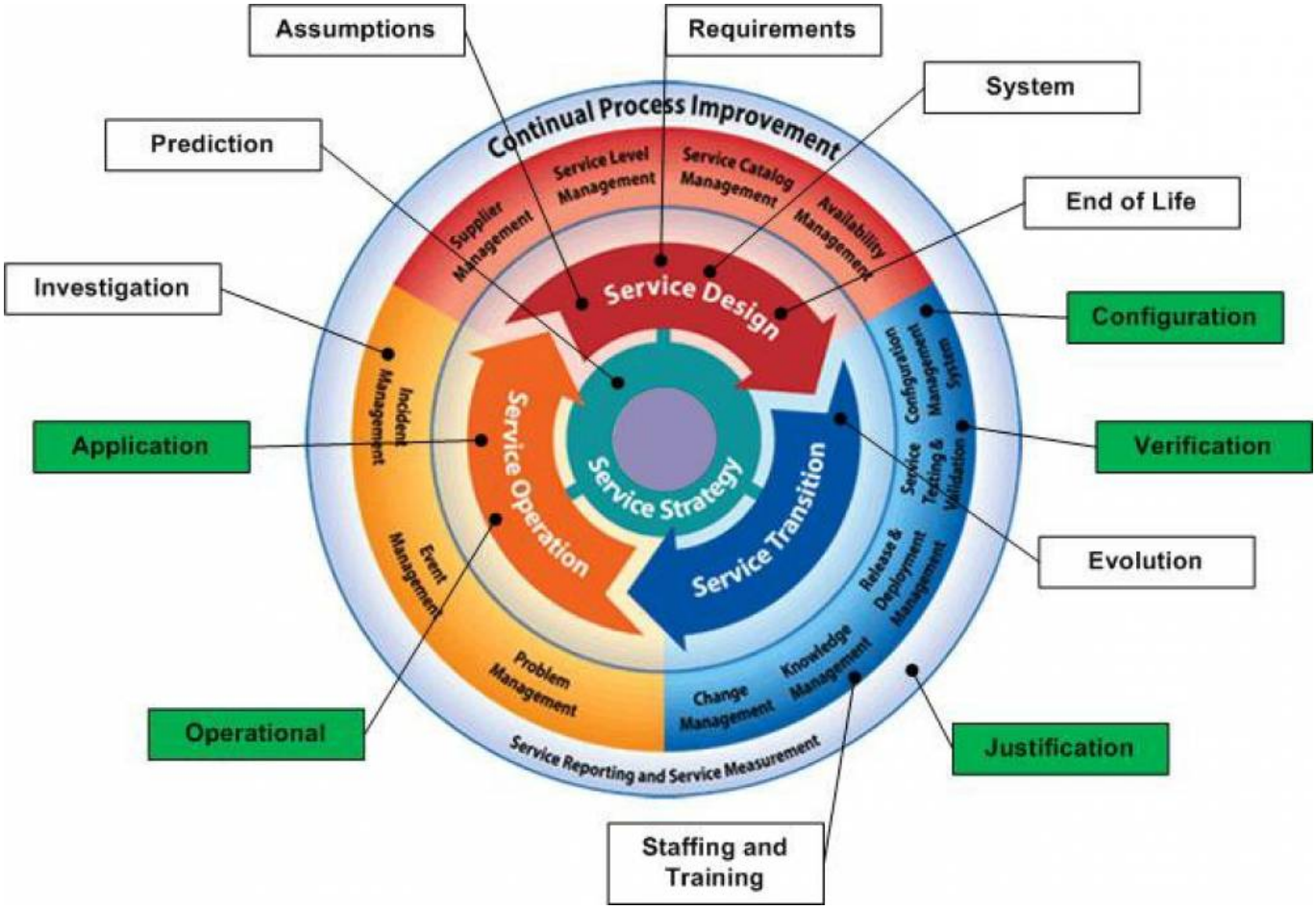
The following diagram represents a typical development lifecycle using an iterative development approach. In this model there are key phases as the software transitions from concept through to testable executable code. The process is iterative in that several cycles of functional elaboration, design,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



The model itself may vary depending on the specific needs of the project but the diagram illustrates that different data types become significant at different points of the process. It is therefore important that these considerations are explored and documented in the Data Management Plan.

4.2. Operational

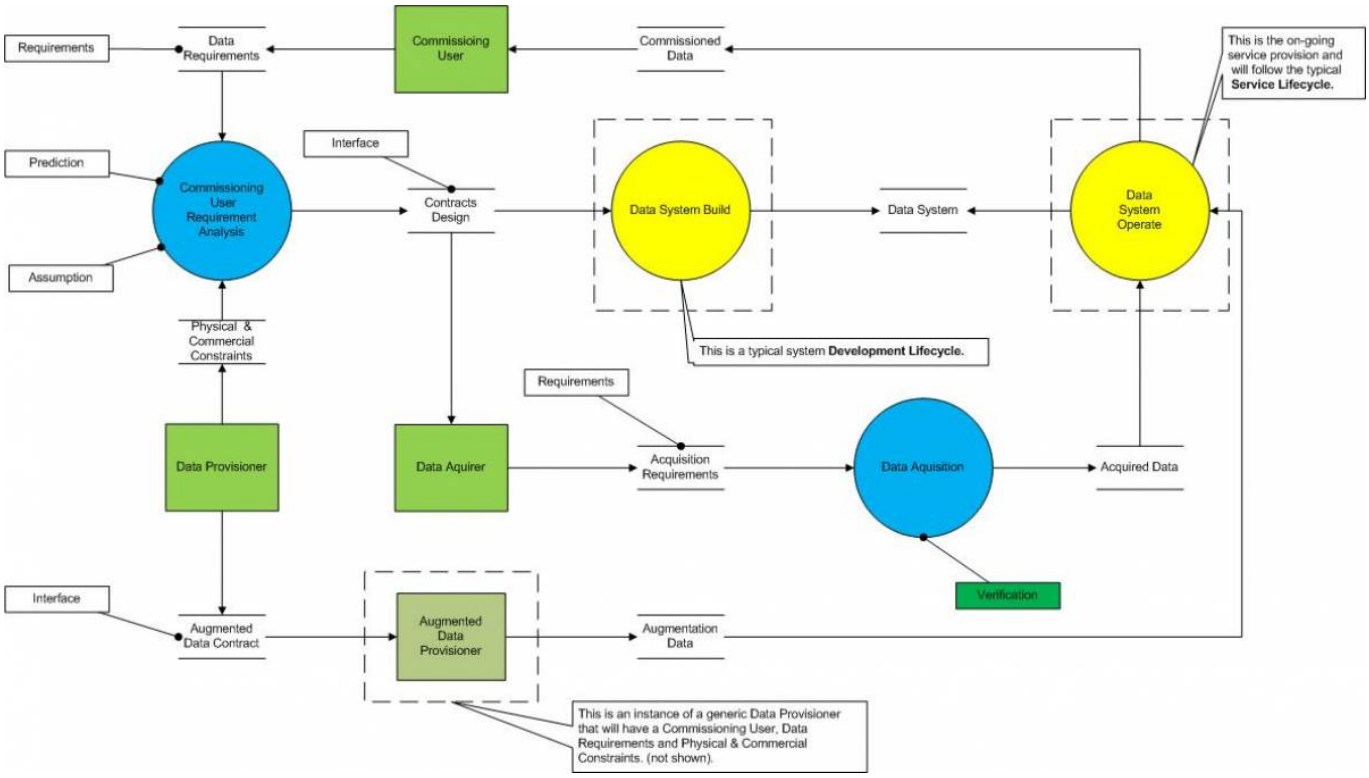
Once a system has been developed it will move into a operational style of lifecycle or indeed, if the data safety has not previously been considered for an enterprise, then the system will already be in operational use. These operational lifecycles tend to be cyclical in nature and the following diagram illustrates a typical model.



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

4.3. Data supply chains

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



This model represents the interactions between 3 key organisations:

- The Commissioning User: the organisation that has the need for the data
- The Data Provisioner: the organisation that will fulfil that need for data
- The Data Aquirer: the organisation employed by the Data Provisioner to carry out physical collection of data

The Commissioning User Requirement Analysis is the key process step where the Commissioning Users expectations for data (including fidelity levels for associated properties) are agreed with the Data Provisioner. The requirements may be adjusted because of physical constraints (eg. loss of precision because of physical measuring device constraints) and me include additional requirements to augment the captured data with additional information. (eg. airport codes added to a measurement of a given runway length).

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

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

Again, it should be stressed that these models are not intended to be prescriptive but rather there to illustrative how authors of the Data Management Plan might address the question of *when* in the process data assurance techniques should be applied for a given data type.

4.3.1. Data safety roles

It is acknowledged that, similar to software engineering, the issue of data integrity is currently largely dependent on the calibre and experience of those who exercise judgement at various stages of data development and assurance. A formal recognition of the roles involved and the kind of skill and competences required is not only an indication of where the industry is going but also recognition that people assurance is a key aspect of getting the data right.

Further elaboration of the key *safety data* roles and responsibilities is required and will be considered in future versions of this guidance. However, the current vision is that the roles will be derived from EN50128:2011 [9] Software Role Specifications-Annex B. Such a mapping will result in a number of responsibilities such as "Safety Data Requirements Manager", "Safety Data Tester", "Safety Data Reviewer" amongst others, but except for perhaps the largest of programmes, the expectation is that an individual may carry out more than one of the roles in conjunction with other duties.

What has been established however, is that it is important that roles are defined and documented in the Data Management Plan/Safety Management Plan and specific duties enumerated in the individuals job/role description or terms of reference.

5. Suggested Methods and Approaches

The following table provides a wide spectrum of methods and techniques that represent best practice in mitigating the risks associated with data. The table shows where a particular method/technique is applicable to a given lifecycle data type, and for each Data Integrity Level, whether the method/technique is:

- not applicable/not advisable (-)
- Recommend (R)
- Highly Recommended (HR)

The lifecycle data types are those under consideration for this version of the guidance as follows:

- Verification (V)
- Configuration (C)
- Application (A)
- Operational (O)
- Justification (J)

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

The Safety Management Plan or the Data Management Plan (See Appendix E) are typical vehicles to document:

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

The overall safety justification for the given project/service/operational context must then provide evidence of compliance against the plan.

Methods and Approaches – System Design	**Lifecycle Data Types**					**Data Integrity Level**				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Built-in-Test (BIT/BITE)	-	-	Y	-	-	-	R	HR	HR	the application applies tests to the configuration data or data it is processing
Cyclic / Continuous BIT	-	-	Y	-	-	-	-	R	HR	application applies tests to the data it is processing continuously (eg. for a live data stream) or periodically according to a periodicity strategy (eg. evert nth message, every hour etc)
Backward recovery	-	-	Y	-	-	R	R	HR	HR	If a fault in data has been detected, the system resets to an earlier internal data set, which has been proven consistent
Parity Checks	-	-	Y	-	-	R	R	HR	HR	Within data, eg. Hamming codes, Reed-Solomon
Automatic Error Correction	-	-	Y	-	-	R	R	R	R	Detected errors are corrected automatically
Checksums / CRCs / Hashes	-	-	Y	-	-	-	R	HR	HR	digests of datasets are produced, included with the datset and checked to provide confidence that the data is unaltered
Digital Signatures	-	-	Y	-	-	-	R	HR	HR	for non-repudiation and integrity of data
Sequence Numbers	-	-	Y	-	-	R	R	HR	HR	data bears sequence numbers so the integrity of a data stream can be checked (eg. data items not monotonically increasing, duplicate detection)
Auditing Facilities	-	-	Y	Y	-	-	R	HR	HR	changes to data properties are audited so the before and after values are recorded and also other relate information such as the author (eg. person, function or system) and the time of the change
Logging Facilities	-	-	Y	Y	-	R	R	HR	HR	data processing events are logged to allow support staff to monitor the health of the system and provide diagnostic information if problems are detected
Encapsulation	-	-	Y	-	-	R	R	HR	HR	the hiding of data so that it is only accessible through well defined interfaces
Multiple Stores	-	-	Y	-	-	-	-	R	HR	the same instance of a data set or data items is stored in multiple locations
Homogeneous Redundancy	-	-	Y	-	-	-	-	R	HR	data is processed using homogeneous redundant channels; detected faults in data of one channel will cause processing to switch to another channel
Heterogeneous Redundancy	-	-	Y	-	-	-	-	R	HR	data is processed using heterogeneous redundant channels (same functionality but different implementations) detected faults in data of one channel will cause processing to switch to another channel
N-Version Programming	-	-	Y	-	-	-	-	R	R	data is processed using heterogeneous redundant channels (same functionality but different implementations) with both channels active and a form of voting across channels to determine data output or control behaviour
Data Integrity Sampling	-	-	Y	-	-	HR	HR	R	R	the integrity of subsets of data is periodically checked, in accordance with a given selection criteria (eg. random, critical records etc), frequency and volume of data to check.

Sanity/Reasonability Checks	-	-	Y	-	-	R	R	HR	HR	dedicated processing implemented to check that data is within reasonable tolerances and/or logically/semantically consistent with what the data represents. For example, range checks, date checks, record counts, record sizes etc.
Data Correlation	-	-	Y	-	-	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, or prime source) with other sources
Data Partitioning	-	-	Y	-	-	R	R	HR	HR	To separate data that is managed differently, creating independence between data so that whole data set do not require validation after a change
Syntax Checks	Y	Y	Y	-	-	R	R	HR	HR	Semantic checking of data values and sequences based on defined rule sets
Database Management System (DBMS)	-	-	Y	-	-	HR	HR	R	R	Use of established 3rd Party products for storage and management of data
Feedback testing	-	-	Y	-	-	HR	HR	R	R	To check output data by comparing it with the input source
Information Redundancy	-	-	Y	-	-	HR	HR	R	R	Additional redundant information is supplied from diverse sources. The validity of the data coming from the diverse sources can be checked against each other
Reverse Translation	-	-	Y	-	-	-	R	HR	HR	To verify that the data output of a process is correct, by attempting to create the source data from the output data and comparing this with the source used to create the output data
Meta-data	-	-	Y	Y	-	-	R	HR	HR	Auditable data are sent with the data that is about the data, e.g. source, issue state, expiry date

Methods and Approaches – Data Assurance	**Lifecycle Data Types**					**Data Integrity Level**				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Review / Inspection	Y	Y	Y	Y	Y	HR	HR	HR	HR	manual review/inspection data possibly involving data visualisation tools
Simple Sanity Check	Y	Y	Y	Y	Y	R	R	-	-	Informal check of data, eg. total record counts, filesizes within expectations
Statistics-Based Sampling	Y	Y	Y	Y	Y	-	R	HR	HR	more appropriate for realtime large and/or volume data, not necessary if inspection gives full coverage but gives extra defence in depth. Could be manual selection, a form of random selection or comparison against statistical norms.
Ground-Truth Check	Y	Y	Y	Y	Y	R	R	HR	HR	inspection against physical measurements (eg. lengths, positions, heights) taken in the real world
Auditing	Y	Y	Y	Y	Y	R	R	HR	HR	A period of comprehensive internal and external testing of the data quality process is performed where Data is verified according to its intended use and definition.
Tracing	Y	Y	Y	Y	Y	-	R	HR	HR	ability to trace data from source across multiple participants in the data supply chain
Defined Verification Frequency	Y	Y	Y	Y	Y	-	R	HR	HR	data should contain an indicator on how often it should be revalidated against other (eg. real world) source
Defined Data Lifetime(s)	Y	Y	Y	Y	Y	R	R	R	R	When does data validity expire
Data Quality Measurement	Y	Y	Y	Y	Y	-	R	HR	HR	criteria are established to provide an objective measurement of the quality of a given dataset
Data Quality Trend Analysis	Y	Y	Y	Y	Y	-	-	R	HR	checking that a dataset is consistent with a model of the expected data behaviour. Eg. vibration data increases over time
Authorisation	Y	Y	Y	Y	Y	R	R	HR	HR	a security model is established to control who is authorised to create, view, edit, delete the data
Authentication	Y	Y	Y	Y	Y	R	R	HR	HR	data is authenticated to validate the provenance of the data
Defined Confidence / Trust Levels	Y	Y	Y	Y	Y	R	R	HR	HR	criteria are established to provide an objective measurement of the confidence or trust in a given dataset
Independent Check	Y	Y	Y	Y	Y	-	-	R	HR	a separate person or system is used to independently check the data

Methods and Approaches – Data Procedures	**Lifecycle Data Types**					**Data Integrity Level**				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Data Management Plan	Y	Y	Y	Y	Y	R	R	HR	HR	
Governance Model	Y	Y	-	-	Y	R	R	HR	HR	A governance model is established that defines aspects such as data ownership, processing roles & responsibilities (who can do what to the data), processing authorisations and permissions (what can be done to the data) etc.
Data Process Definition	Y	Y	Y	Y	Y	-	R	HR	HR	Documented and agreed process definitions for how data is handled
Data Flow Diagram	Y	Y	Y	Y	Y	HR	HR	HR	HR	To describe the data flow in a diagrammatic form
Data Model	Y	Y	Y	Y	Y	HR	HR	HR	HR	To articulate how data is organised
Data Safety Training	Y	Y	Y	Y	Y	R	R	HR	HR	For individuals
Data Safety Competence Assessment	Y	Y	-	-	Y	-	R	HR	HR	For individuals
Client Sign-Off	Y	Y	-	Y	Y	R	R	HR	HR	
Data Quality Correction Mechanisms	-	-	-	Y	-	-	R	HR	HR	A process, strategy and tooling for data that breaches a given data quality criteria
Configuration Management	Y	Y	Y	Y	Y	HR	HR	HR	HR	The recording of the production of every version of every 'significant' deliverable and of every relationship between versions of the different deliverable

Data Dictionary	Y	Y	Y	Y	Y	HR	HR	HR	HR	A data dictionary is a collection of descriptions of the data objects or items in a data model for the benefit of data users
Formal Methods	-	-	Y	-	-	-	R	R	HR	To specify data in a formal, mathematical manner
Update Comparison	Y	Y	Y	Y	Y	-	R	R	HR	Updated data is compared to its previous version. The list of changed elements can be compared with a similar list generated by the supplier

Methods and Approaches – Data Media Handling (Paper / Physical Storage)	**Lifecycle Data Types**					**Data Integrity Level**				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Photographic Copies	Y	Y	Y	Y	Y	R	R	HR	HR	
Scan to Electronic Format	Y	Y	Y	Y	Y	R	R	HR	HR	
Copies Held at Different Locations	Y	Y	Y	Y	Y	-	R	HR	HR	
Limited Access	Y	Y	Y	Y	Y	-	R	HR	HR	
Secure Storage	Y	Y	Y	Y	Y	-	R	HR	HR	
Manual Inspection	Y	Y	Y	Y	Y	-	R	HR	HR	
Suitable Physical Environment	Y	Y	Y	Y	Y	-	R	HR	HR	
Defined Handling Procedures	Y	Y	Y	Y	Y	-	R	HR	HR	
Repair / Restoration Programme	Y	Y	Y	Y	Y	-	-	R	HR	
Indexing / Cataloguing	Y	Y	Y	Y	Y	R	R	HR	HR	
Lifetime Planning	Y	Y	Y	Y	Y	-	-	R	HR	

Methods and Approaches – Data Media Handling (Electronic Storage)	**Lifecycle Data Types**					**Data Integrity Level**				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Regular Refresh/Rewrite	Y	Y	Y	Y	Y	-	R	HR	HR	Of magnetic media or flash memory
Suitable Storage/Handling	Y	Y	Y	Y	Y	-	R	HR	HR	
Suitable Physical Environment	Y	Y	Y	Y	Y	-	R	HR	HR	
Copies at Different Locations	Y	Y	Y	Y	Y	-	R	HR	HR	Physically separate to cover natural disasters
Backups/Duplication	Y	Y	Y	Y	Y	-	R	HR	HR	
Sample Restores	Y	Y	Y	Y	Y	-	R	HR	HR	
Multiple Copies	Y	Y	Y	Y	Y	-	R	HR	HR	
Copy to Latest Media Format	Y	Y	Y	Y	Y	-	R	HR	HR	
Media Physically Secured	Y	Y	Y	Y	Y	-	R	HR	HR	
Resilient / Redundant Format	Y	Y	Y	Y	Y	-	-	R	HR	
Long-Lifetime Format	Y	Y	Y	Y	Y	-	-	R	HR	
Easily Translatable / Convertible Format	Y	Y	Y	Y	Y	-	-	R	HR	
Copy to Cloud Storage	Y	Y	Y	Y	Y	-	R	HR	HR	Must specify whether a private cloud or a public cloud shall be used
Copy to Archiving Organisation	Y	Y	Y	Y	Y	-	R	HR	HR	

Methods and Approaches – Data Usage Confidence	**Lifecycle Data Types**					**Data Integrity Level**				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Limited / Pre-Operational Deployment	-	Y	Y	Y	-	-	R	HR	HR	
Client Sign-Off of Data	Y	Y	-	Y	Y	-	R	HR	HR	
Non-Critical Trialling	-	-	Y	-	-	-	R	HR	HR	
Beta Testing	Y	-	-	-	-	-	R	HR	HR	
Parallel Running	-	Y	Y	Y	-	-	R	HR	HR	
Widespread Distribution to User Community	-	Y	Y	Y	-	-	R	HR	HR	
Open Source Techniques	-	-	Y	-	-	-	R	HR	HR	

Methods and Approaches – Test Data Generation	**Lifecycle Data Types**					**Data Integrity Level**				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Using Informal / ad-hoc means	Y	-	-	-	-	R	R	-	-	
Using Testbed	Y	-	-	-	-	-	R	HR	HR	
Using Simulator	Y	-	-	-	-	-	R	HR	HR	
Using Prototype	Y	-	-	-	-	-	R	HR	HR	
Using Manual means	Y	-	-	-	-	R	R	-	-	
Using Dedicated Platform	Y	-	-	-	-	-	R	HR	HR	
Using Existing/Established System	Y	-	-	-	-	-	R	HR	HR	
Using Initial Runs of New System	Y	-	-	-	-	R	R	-	-	
Derived from Real Data	Y	-	-	-	-	R	R	HR	HR	
Statistical Profiling Post-Production	Y	-	-	-	-	-	-	R	HR	
Produced by Client	Y	-	-	-	-	-	-	R	HR	
Client Sign-Off	Y	-	-	-	-	R	R	HR	HR	

Error Seeding	Y	-	-	-	-	R	R	HR	HR	Errors are deliberately inserted into the dataset to demonstrate the effectiveness of data validation
Data Re-use	Y	-	-	-	-	R	R	HR	HR	Re-using data for one project that was created and thoroughly assured for another project
Feedback testing	Y	-	-	-	-	R	R	HR	HR	To check output data by comparing it with the input source

Methods and Approaches – Test Tools	Lifecycle Data Types					Data Integrity Level				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Informal / Office Tools	Y	-	-	-	-	R	-	-	-	
Specific Tools / Scripts	Y	-	-	-	-	R	HR	HR	R	
Formally Developed Tools / Scripts	Y	-	-	-	-	R	HR	HR	HR	
Commercial Tools	Y	-	-	-	-	-	R	HR	R	
Qualified Tools	Y	-	-	-	-	-	R	HR	HR	Formally approved according to a standard

Methods and Approaches – Test Results Analysis	Lifecycle Data Types					Data Integrity Level				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Informal / Office Tools	Y	-	-	-	-	R	R	-	-	
Specific Tools / Scripts	Y	-	-	-	-	R	HR	HR	R	
Formally Developed Tools / Scripts	Y	-	-	-	-	R	HR	HR	HR	
Dedicated Analysis Platform	Y	-	-	-	-	-	R	HR	HR	
Commercial Tools	Y	-	-	-	-	-	R	HR	HR	

Methods and Approaches – Data Migration	Lifecycle Data Types					Data Integrity Level				
Technique	**V**	**C**	**A**	**O**	**J**	**DIL 1**	**DIL 2**	**DIL 3**	**DIL 4**	**Notes**
Informal / Office Tools	-	-	Y	-	-	R	R	-	-	
Manual Load	-	-	Y	-	-	R	R	-	-	
Dedicated Translation and Loading Platform	-	Y	Y	-	-	-	R	HR	HR	For example, using mature enterprise migration COTS products
Existing/Established System Transfer	-	-	Y	-	-	-	R	HR	HR	
Client Supervision	Y	Y	Y	Y	-	-	R	HR	HR	
Client Sign-Off	Y	Y	Y	Y	Y	-	R	HR	HR	formal acceptance of the migrated datasets in the target system
Incremental switch-over	-	-	Y	-	-	-	R	HR	HR	users are incrementally switched over to the new system rather than as a 'big bang'
Parallel Load With Existing System	-	-	Y	-	-	-	R	HR	HR	parallel running of the new system alongside the existing system with data crosschecks between the two systems
Shadowing	-	-	Y	-	-	-	R	HR	HR	parallel running of the new system alongside the existing system such that only data from the existing system is used operationally and with an experienced user crosschecking between the two systems so as to validate the new one (or not, as the case may be)
End to End Import-Export Verification	Y	-	Y	-	-	-	R	HR	HR	

6. Tools

Tools in this context are considered anything that automates all or part of a process, e.g. data creation or data transformation.

Tools have different potential for adverse impact on data safety, depending on both their function and how they are to be used. For tools to be considered fit for purpose it is necessary to show that the tool meets its requirements in the context in which it is to be used. The activity to ensure a tool is fit for purpose is usually called “tool qualification”.

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

1. Use evidence of a previous certification of the tool by a trusted third party (unlikely to be available in most industry sectors);
2. Base tool qualification on the practices used when designing and developing the tool (only practical for tools developed within the organisation; the guidance of EUROCAE Document ED-215 [10], "Software Tool Qualification Considerations" may be of help);
3. Assess the potential risks presented by use of the tool and provide assurance that they are adequately managed (most practical option).

The third, risk based, approach, can be summarised as follows:

- Draft a procedure for the use of the tool to achieve the stated purpose;
- Identify threats to data safety associated with using the tool;
- Specify adequate mitigations for each identified threat;
- Augment and formally issue the tool requirements and the usage procedure to implement the specified mitigations;

- Demonstrate that the tool and its mitigations perform as expected; and
- Provide a compelling assurance argument based on the previous steps and any other evidence that will improve confidence, e.g. reputation of the supplier, configuration management of the tools and its documentation, competence of the tool user and of those who check the outputs...

7. Ongoing Development

This document is still considered a draft, but is sufficiently mature to benefit from wider review.

Comments should be sent to the editorial team via this form http://scsc.org.uk/file/gd/Updated%20Review%20Form-97.doc or lodged on the SCSC web site forum: http://scsc.org.uk/forum

Please return your comments by 7th March 2014.

Reviewers please note the guidance below:

Please help us where you can with a solution to a problem found and a clear explanation, for example:

- Delete paragraph N in section Y because ...;
- The 2nd sentence in section X.Y needs to include text “Z” because ...;
- The text in section X.Y needs moving to section Z.W as this is not appropriate to "V".
- The 2nd sentence in section X is ambiguous due to Y, suggest replace with “ZZZ...”

Please let us know if you want to be noted as a contributing reviewer.

Any queries or problems please do not hesitate to contact any of the document contributors, or:

mike.parsons@cgi.com or paul.hampton@cgi.com

Thanks!

Annex A: War Stories

> Two men were examining the output of the new computer in their department. After an hour or so of analyzing the data, one of them remarked:
>
> "Do you realize it would take 400 men at least 250 years to make a mistake this big ?”

The following 'War Stories' describe incidents in which data is considered to have been a contributory factor. A data perspective has been taken to demonstrate the need for data to be given equal footing alongside software, hardware and human factors.

Note: The analysis presented here has no legal standing whatsoever. The purpose of this section is not to discredit, contradict or undermine the existing accident analysis, the aim is simply to view these incidents from a data perspective. Where possible accident reports have been referenced with the role of data highlighted.

Lake Peigneur Drilling Accident

Summary

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

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

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

The role of data in the incident

Federal experts from the Mine Safety and Health Administration were not able to determine the cause of the accident due to confusion over whether the rig was drilling in the wrong place or that the mine's maps were inaccurate. All evidence was lost, however the incident demonstrates how catastrophic a data error, or misinterpretation of data, can be.

Sources

Wikipedia, http://en.wikipedia.org/wiki/Lake_Peigneur, last accessed 15/01/2014

Oil Rig Disasters, http://home.versatel.nl/the_sims/rig/lakepeigneur.htm, last accessed 15/01/2014

Comair Flight 5191

Summary

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

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

Role of data in the incident

"The Airport Charts used by the crew were inaccurate. The airport was under construction, and the charts were not kept current with the rapid changes that were taking place during the construction work. The chart did not accurately reflect the taxiway identifiers and the closed taxiway at the airport on August 27, 2006."

"Due to a previously unrecognised software glitch, any information the chart provider received after normal work hours on Fridays was not included in their regular updates. Furthermore, the chart provider modified the Blue Grass Airport chart after the accident to include a note that Runway 8/26 is 'daytime VMC use only', even though this information had been published since 2001."

Additionally there was a local Notice to Airmen (NOTAM) issued advising of taxiway closures due to construction work. However the crew was not provided with this information in their dispatch paperwork.

Sources

Attempted Takeoff from Wrong Runway – Comair Flight 5191 – Accident Report, National Transportation Safety Board, NTSB/AAR-07/05

Wikipedia, http://en.wikipedia.org/wiki/Comair_Flight_5191, last accessed 15/01/2014

Mars Climate Orbiter

Summary

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

The plan was that the rocket would place the spacecraft into a transfer orbit to Mars, which would be optimised along the way by a series of four

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

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

Role of data in the incident

It transpires that the orbiter's Flight Management System (FMS) software was designed to work with metric Newton seconds, whereas a FMS data-file generated by ground system software used pound-force seconds. A Newton is about 22.5% of a pound-force or a factor of 4.45. (See section §4 of reference [11].)

The cost of the mission was stated by NASA to have been $327.6 million in total ($193.1 million to develop the spacecraft, $91.7 million for launch and $42.8 million for mission operations). Your data may correctly reflect the specification, and be available when required, but is the specification consistent across interfaces? See also reference [12].

Sources

Wikipedia, http://en.wikipedia.org/wiki/Mars_Climate_Orbiter, last accessed 30/01/2014

MS Oliva

Summary

At about 0510 (UTC) on 16 March 2011, Oliva, a Maltese registered bulk carrier ran aground on the north-west coast of Nightingale Island in the Tristan Da Cunha Group. Oliva was on a loaded passage from Santos, Brazil to China. The vessel sustained severe bottom damage to almost all of her water ballast tanks that resulted in the vessel developing a 12º list to port.

On 18 March, the vessel broke up in two sections; the forward section drifted away and the aft section capsized and sank. All this resulted in wide spread pollution around the islands of Nightingale and Inaccessible because of the diesel and fuel oil that escaped from the vessel’s fuel tanks.

Role of data in the incident

“Both the second mate and chief mate were not aware that Oliva was heading towards Nightingale Island. This was because there was no indication on the plotting chart to alert them of the dangers ahead. It appeared that the bridge team was focused on following the GPS track (red course line) superimposed on the radar screen instead of monitoring the vessel’s position in relation to surrounding hazards.”

‘No Go’ areas were not marked on the chart. It appeared that the vessel did not have BA Chart 1769, which was the appropriate large scale chart covering the Tristan Islands.

Sources

‘Safety Investigation into the grounding of the bulk carrier OLIVA On Nightingale Island, Tristan Da Cunha on 16 March 2011’, Transport Malta - Marine Safety Investigation Unit, Marine Safety Investigation Report No. 14/2012

Sichem Osprey

Summary

On 10 February 2010 at 4.36 am local time, the chemical tanker SICHEM OSPREY, on her way from Panama to Ulsan (South Korea) stranded at more than 16 knots on the northeasterly part of Clipperton Island, although an OOW and a lookout AB were on the bridge and no damage was reported prior to the accident. A 100 metre fore part of the vessel had been grounded. No pollution was observed.

Role of data in the incident

Anti-collision radar alarm thresholds were not set according to the Captain's instructions. The adjustments were not reappraised by any of the Officers or the Captain. There were sizeable discrepancies between the fixes plotted on the chart and those displayed on the radar (VDR).

Sources

Stranding of the chemical tanker vessel SICHEM OSPREY on 10 February 2010 on Clipperton Island, Bureau d'enquêtes sur les évènements de mer

The Pride of Canterbury

Summary

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

Role of data in the incident

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

Although the Voyage Management System was loaded with electronic navigational charts (ENC) for the vessel's area of operation, the system had not been approved by the MCA as the owner's policy was for the VMS to be used as an aid to navigation only, with Pride of Canterbury's paper charts being utilised as the primary means for navigation. Relevant admiralty charts were supplied to the vessel for this purpose.

Although the VMS was not approved for use as the primary means of navigation, the officers on Pride of Canterbury were using it as if it was, despite the fact that many of them, including the chief officer, who was in charge at the time of the accident, were not fully trained in its use.

Sources

Report on the investigation into the grounding of Pride of Canterbury "The Downs" – off Deal, Kent 31 January 2008, Marine Accident Investigation Branch, Report No 2/2009, January 2009

LOT Flight 282

Summary

Just after takeoff from Runway 09R at London Heathrow Airport (LHR), the pilots noticed that most of the information on both of the Electronic Attitude Director Indicators (EADI) and Electronic Horizontal Situation Indicators (EHSI) had disappeared. The aircraft entered Instrument Meteorological Conditions (IMC) at about 1,500 ft aal, and the co-pilot had no option but to fly using the standby attitude indicator and standby

compass. He experienced difficulty in following radar headings. The aircraft returned to land at LHR after a flight of 27 minutes.

Role of data in the incident

The single error made by the co-pilot during the pre-flight preparation initiated the subsequent problems. This was the use of ‘E’ instead of ‘W’ when the longitude co-ordinates were entered into the FMS.

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

Sources

AAIB Bulletin 6/2008

Cedars Sinai Medical Centre - CT Scanner

Summary

A software misconfiguration in a CT scanner used for brain perfusion scanning at Cedar Sinai Medical Center in Los Angeles, California, resulted in 206 patients receiving radiation doses approximately 8 times higher than intended during an 18 month period starting in February, 2008. Some patients reported temporary hair loss and erythema. The U.S. Food and Drug Administration (FDA) has estimated that patients received doses between 3 Gy and 4 Gy.

Role of data in the incident

The problem reportedly resulted from an error made by the hospital in resetting the CT machine after it began using a new protocol for the procedure in Feb. 2008, but it wasn't detected until one of the patients reported patchy hair loss in August 2009.

"There was a misunderstanding about an embedded default setting applied by the machine," according to a statement from Cedars-Sinai. “As a result, the use of this protocol resulted in a higher than expected amount of radiation.”

Sources

Los Angeles Times, http://articles.latimes.com/2009/oct/10/local/me-cedars-sinai10, last accessed 24/01/2014

HealthImaging, http://www.healthimaging.com/topics/diagnostic-imaging/update-cedars-sinai-explains-ct-perfusion-radiation-overexposure, last accessed 24/01/2014

Fort Drum Artillery Incident

Summary

Two artillery shells were fired more than a mile off target during an Army firing exercise at Fort Drum in Northern New York in March 2002. The shells landed near a mess tent where a Battalion were having breakfast. Two soldiers were killed, 13 were injured.

Role of data in the incident

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

Sources

The New York Times, http://www.nytimes.com/2003/07/02/nyregion/officer-found-negligent-in-deaths-of-2-at-fort-drum.html, last accessed 24/01/2014

AP News Archive, http://www.apnewsarchive.com/2002/Army-Reports-on-Ft-Drum-Accident/id-539bf2ea24b8dd66009c6efee2be926c, last accessed 24/01/2014

Others

- American Airlines Flight 965 at Buga, December 1995 ~ An accident due to inconsistent data naming conventions? - reference [13]
- The Provide Comfort Blackhawk Shootdown Incident - reference [14]
- Patriot Missile versus Tornado 2003 - reference [14]
- Enduring Freedom GPS Friendly Fire Incident - reference [14]

Annex B: Data Management Plan

This section gives a suggested Safety Data Management Plan (SDMP) 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 SDMP as well as an SMP.

Safety Data Management Plan suggested contents:

1. Introduction
   - Scope & Context *(Sets the scene, describes the project, scenario, concept of operations, etc)*
   - Boundaries & Interfaces *(Describes the main interfaces and exchanges, with a scope boundary diagram)*
   - 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
   - Abbreviations and Acronyms
2. Analysis of Assigned DIL & ODR Level *(Implications of the data analyses. Note this assumes the DIL and ODR analyses have already been performed)*
   - SIL / DAL Implications *(Will the system need to be developed to a SIL/DAL or Software Level to meet the DIL?)*
   - Development Implications *(Are there any special development considerations? Derived from the SIL/DAL if there is one, otherwise what is deemed necessary for this system)*
   - Verification Implications *(Derived from the SIL/DAL if there is one, otherwise what is deemed necessary for this system)*
   - Assurance Implications *(Derived from the SIL/DAL if there is one, otherwise what is deemed necessary for this system)*
   - Process / Procedure Implications *(Derived from the SIL/DAL if there is one, otherwise what is deemed necessary for this system)*
3. Types of Safety Data in Scope *(A list of all the types 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?)*
   - 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 organisation manage the data safety risks?
   - Organisation
   - Responsibilities
   - Authorisations
   - Approvals and Signoffs
6. Justification Approach *(How will the safe usage 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. Annex: DIL Guidelines Response *(Tailored version of the tables from this document. What is considered applicable/useful and what is not)*

Annex C: Organisational Data Risk

This form is used to determine the safety risk related to data for a particular organisation and usage.

This form must be completed from the perspective of one of the organisations involved; typically this will be the organisation 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, eg. 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 determine the management regime required to mitigate the risk associated with the data.

Organisational Data Risk (ODR) Assessment Form

This form is used to determine the safety risk related to data for a particular organisation and usage.

*This form must be completed from the perspective of **one** of the organisations involved; typically this will be the organisation 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, eg. 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 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:			
Applicable Data Sets:			
Completed By:		Date Completed:	

Answer each question using the response that forms the best match for the particular scenario. Not all statements have to be satisfied and some judgment 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 uses 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 possible.	8	☐
1e	An intended use of the data could easily lead to an accident resulting in death for several people. The accident could be caused by the data with little chance of anything else detecting and mitigating data issues. The accident could affect the general public or cause catastrophic environmental impact.	16	☐

Justification:

QUESTION 2 – ORGANISATIONAL AND SOCIETAL IMPACT

What would be the impact on the organisation, 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 organisation 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 organisation. 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 organisation. Regional press interest. Client inquiry or investigation likely.	4	☐

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 organisation have for data safety?

This question considers how much legal and other responsibility and ownership the organisation has for data safety aspects within this scenario. What liabilities for consequential losses / 3rd party claims does the organisation have via the contract or other means? What is the scale of the organisation's contribution to the overall scope?

3a	The organisation is not responsible for any data safety aspects. No liabilities for accident claims related to the data lie with the organisation. Client or other party has accepted full data safety responsibility. The organisation is fully covered and indemnified by the client or a 3rd party.	1	☐
3b	The organisation 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 organisation only reviews and monitors.	2	☐
3c	The organisation is a significant part of the consortium or 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 organisation's favour. All key safety data obligations are explicitly flowed down to subcontractors.	4	☐
3d	The organisation 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 organisation provides some indemnities to others via the contract. Some significant data safety obligations are not flowed down to subcontractors.	7	☐
3e	The organisation is priming a major programme or has total data safety responsibility. Specific accident-related liabilities in the contract are large (or unlimited). The organisation 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 – ORGANISATIONAL MATURITY

How mature is this organisation regarding data safety?

This question considers the maturity of the organisation 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	☐
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 usage 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 usage. One data owner, a specialist highly trained user group. Single organisation or recipient usage only.	1	☐
6b	A number of operational data users. Simple linear supply chain. More than one data owner. Specialist user or limited public access. Small scale operation. No general web access. Few user organisations or recipients.	2	☐
6c	Regional usage. Some public or mainstream usage. A few supply chains. A few data owners. Some web access. Several user organisations or recipients.	4	☐
6d	National usage. Public or mainstream usage. Several supply chains. Several data owners. Web access. Some or varied user organisations or recipients	7	☐
6e	International usage. Extensive public or mainstream usage. Extensive web access. Many complex supply chains. Many and diverse data owners. Many and diverse user organisations 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	☐
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:			

ORGANISATIONAL DATA RISK LEVEL

Record the total score and use it to determine the ODR level based on the ranges given below. If the first 3 question's 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:	

Warning: this form only gives an initial organisational data risk level assessment. Further work is required to establish the safety data risks in detail such as determining a Data Integrity Level (DIL) for relevant data sets.

Annex D: Definitions, Acronyms & Glossary

	Definition	Source
A		
Accuracy	Closeness of agreement between a test result and the accepted reference value. NOTE A test result can be observations or measurements.	ISO 19113:2005 [15]
	a degree of conformance between the estimated or measured value and the true value	(EU) No 73/2010 [16]
Accuracy (temporal)	Correctness of the temporal references of an item (reporting of error in time measurement). Correctness of ordered events or sequences, if reported. Validity of data with respect to time.	ISO 19138:2006 [17]
Active (data)	Data that changes system functionality	??
(data) Assurance Level	the required assurance level for the aeronautical data process is identified, based on the overall system architecture through allocation of risk determined using a preliminary system safety assessment.	RTCA/DO-200A [1]
	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 [18]
Adaptation Data	Data used to customise elements of the Air Traffic Management System for their designated purpose. Adaptation data is utilised to customize elements of the CNS/ATM 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: a) Geographical Data – latitude and longitude of a radar site. b) Environmental Data – operator selectable data to provide their specific preferences. c) Airspace Data – sector-specific data. d) Procedures – operational customization 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 the 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.	ED-153 [19]
Aeronautical Data	a representation of aeronautical facts, concepts or instructions in a formalised manner suitable for communication, interpretation or processing	(EU) No 73/2010 [16]
	Data used for aeronautical applications such as navigation, flight planning, flight simulators, terrain awareness, and other purposes.	RTCA/DO-178C [6]
Application Data	Data used in the system during operations: this is the data processed or produced by the system which has end-user meaning. It may be displayed and used within the system or may be for transfer or distribution to other systems or downstream users. It is data that has some real “application” meaning, i.e. is not to do with the system internals.	SCSC Data Safety Initiative Working Group
Assumption Data	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	SCSC Data Safety Initiative Working Group
Availability	The property of being accessible and usable upon demand by an authorized entity	ISO27001:2005 [20]
B		
C		
Completeness	Completeness of the data provided	RTCA/DO-200A [1]
Configuration Data	Data that configures a generic software system to a particular instance of its use.	(EC) No 482/2008 [18]
Configuration Data	Data used to configure, tailor or instantiate the system: data used to set up and configure the system to perform a particular function, for a particular installation, product configuration, behaviour or specific usage	SCSC Data Safety Initiative Working Group
Configuration Data	Data that configures a generic software system to a particular instance of its use (for example, 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 [19]
Confidentiality	The property that information is not made available or disclosed to unauthorized individuals, entities, or processes	ISO27001:2005 [20]
(data) correctness	Completeness, self consistency, protection against alteration or corruption and consistency with the functional requirements of the data driven system	IEC 61508 Part 3 [5]
(data) coupling	The dependence of a software component on data not exclusively under the control of that software component	RTCA/DO-178C [6]
(data) Criticality	classification of data by the potential effect of erroneous data on the expected operation that is supported by that data.	RTCA/DO-200A [1]
Critical Data	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 [16]
Customisation (data)	Data used to configure a system or component	Def(Aust)5679 [21]
D		
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 [22]

Data (Aeronautical)	a representation of aeronautical facts, concepts or instructions in a formalised manner suitable for communication, interpretation or processing	(EU) No 73/2010 [16]
	Data used for aeronautical applications such as navigation, flight planning, flight simulators, terrain awareness, and other purposes.	RTCA/DO-178C [6]
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 [6]
Data Chain	An 'Aeronautical Data Chain' is a conceptual representation of the path that a set, or element of aeronautical data takes from its creation to its end use. An aeronautical data chain is a series of interrelated links wherein each link provides a function that facilitates the origination, transmission and use of aeronautical data for a specific purpose.	RTCA/DO-200A [1]
	A collection of organisational data processing functions, where data is transferred from one chain participant to another between data origination and end use.	P. Ensor [3]
	Any combination of two or more data elements, data items, data codes, and data abbreviations in a prescribed sequence to yield meaningful information; for example, "date" consists of data elements year, month, and day.	McGraw-Hill Dictionary [23]
Data Chain Participant	A key organisational/functional element within a data supply chain.	P. Ensor [3]
(data) dictionary	The detailed description of data, parameters, variables, and constants used by the system.	RTCA/DO-178C [6]
Data Driven Systems	System which relies upon configuration data or lookup tables to define the functionality of the system.	IEC 61508 Part 4 [24]
Data Intensive System	Systems which make extensive use of large amounts of data	N. Storey [25]
Design & Development Data	Data produced during development and implementation: this is data encompassing the design & 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	SCSC Data Safety Initiative Working Group
E		
End of Life Data	Data about how to stop, remove, replace or dispose of the system: this is data covering all activities related to taking the system out of service or mothballing / storage / dormant phases	SCSC Data Safety Initiative Working Group
(data) Error	Discrepancy with the universe of discourse	ISO 19138:2006 [17]
	Discrepancy between a data value and the true, specified or theoretically correct value or condition.	P. Ensor [3]
Essential Data	Data with an integrity level as defined in Chapter 3, Section 3.2 point 3.2.8(b) of Annex 15 to the Chicago Convention, i.e. integrity level one in one hundred thousand: there is a low probability when using corrupted essential 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 [16]
Evolution Data	Data about changes after deployment, i.e. data that cover enhancements, formal changes, workarounds, and maintenance issues. It also covers data produced by configuration management activities, such as baselines or branch data	SCSC Data Safety Initiative Working Group
F		
G		
H		
(data) Hazard	A data error that has the potential to lead to an accident	P. Ensor [3]
I		
Information	Knowledge communicated concerning some particular fact, subject, or event; that of which one is apprised or told - intelligence, news - as contrasted with data.	Oxford English Dictionary (OED)
	Knowledge that has a contextual meaning	ISO/IEC 2382 [22]
Information (aeronautical)	information resulting from the assembly, analysis and formatting of aeronautical data	(EU) No 73/2010 [16]
(data) Integrity	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 [1]
	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 [16]
	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 [26]
	The property of protecting the accuracy and completeness of assets, i.e. that which has value to the organisation	ISO27001:2005 [20]
Instructional Data	Data used to warn, train or instruct users about the system: this is data that explains to users the risks of the systems and gives any mitigations that may be required to be implemented by users, e.g. by process, procedure, workarounds, limitations of use	SCSC Data Safety Initiative Working Group
Intent (data)	Data describing how a system will behave	J. Inge [2]
(data) item	Single attribute of a complete data set, which is allocated a value that defines its current status	(EU) No 73/2010 [16]
Interface Data	Data used to enable interfaces between systems: for operations, initialisation or export from the system: data that exists to enable exchange between systems. Covers start-of-life operations (data import or migration), end-of-life operations and ongoing operational exchange of data between systems.	SCSC Data Safety Initiative Working Group
Investigation Data	Data to support accident or incident investigations (i.e. potential evidence: this is data collected or produced during an accident investigation which may be used in investigation reports, lessons learnt or prosecutions. This can be source data (e.g. photographs of crash site) or may be derived (accident simulations, analyses, etc)	SCSC Data Safety Initiative Working Group
J		

Justification Data	Data used to justify the safety position of the system: data used to justify, explain and make the case for starting or continuing live operations and why they are safe enough. Often passed to external bodies (regulators, HSE, ISAs) for their review.	SCSC Data Safety Initiative Working Group
K		
L		
M		
Meta-data	Data that represents information about data itself. Note: One should distinguish between "Structural Meta-data", which is data about the design and specification of data structures (and is more properly called "data about the containers of data") and "Descriptive Meta-data", which is about individual instances of application data, the data content.	J. Inge [2]
N		
O		
Objective (data)	Data describing a system's environment	J. Inge [2]
(data) Originator	Entity responsible for data origination	(EU) No 73/2010 [16]
Operational Data	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 may be the results of various phases of introduction and may include trend analysis to look for long-term problems.	SCSC Data Safety Initiative Working Group
(data) Origination	Creation of a new data item with its associated value, the modification of the value of an existing data item or the deletion of an existing data item	(EU) No 73/2010 [16]
P		
Passive (data)	Data acquired from records collected for some other purpose.	online medical dictionary
(data) Product	Dataset or dataset series that conforms to a data product specification.	BS EN ISO 19131:2008 [27]
Prediction Data	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	SCSC Data Safety Initiative Working Group
Q		
(data) Quality	A degree or level of confidence that the data provided meet the requirements of the user. These requirements include levels of accuracy, resolution, assurance level, traceability, timeliness, completeness, and format	RTCA/DO-200A [1]
	Process by which the Electronic Chart Systems (ECS) Database is produced, the source materials, the resolution and reproduction accuracy of chart features, and the correctness and completeness of data.	ISO 19379:2003 [28]
	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 [16]
(data) Quality Attributes	accuracy, resolution, assurance level, traceability, timeliness, completeness and format	RTCA/DO-200A [1]
R		
Release Data	Data used to ensure safe operations per release instance: explanation of particular features or limitations of a release or instance. May include specific time-limited workarounds and caveats for a release.	SCSC Data Safety Initiative Working Group
Requirements Data	Data used to specify what the system has to do: data encompassing requirements, specifications, internal interface or control definitions, data formats, etc.	SCSC Data Safety Initiative Working Group
Resolution	The smallest difference between two adjacent values that can be represented in a data storage, display or transfer system	RTCA/DO-200A [1]
	a number of units or digits to which a measured or calculated value is expressed and used	(EU) No 73/2010 [16]
Routine Data	Data with an integrity level as defined in Chapter 3, Section 3.2 point 3.2.8(b) of Annex 15 to the Chicago Convention, i.e. integrity level one in one thousand: there is a very low probability when using corrupted routine 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 [16]
S		
(Data) Set	Identifiable collection of data. NOTE 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[27]
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 [19]
Staffing and Training Data	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 may include training records, competency assessments, permits to work, etc.	SCSC Data Safety Initiative Working Group
Standards and Regulatory Data	Data that governs the approaches, processes and procedures used to develop safety systems: this is 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.	SCSC Data Safety Initiative Working Group
System Data	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 may cover hardware, software and data.	SCSC Data Safety Initiative Working Group
T		
Third Party (data)	Data of no direct relevance to a system	J. Inge [2]

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 [3]
	The difference between the time of output of a data item and the time of applicability of that data item	(EU) No 1207/2011 [26]
Traceability	Ability to determine the origin of the data	RTCA/DO-200A [1]
Trace (data)	Data providing evidence of traceability of development and verification processes software life cycle data without implying the production of any particular artifact.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 life cycle data.	RTCA/DO-178C [6]
U		
V		
(data) validation	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 [1]
	Process of ensuring that data meets the requirements for the specified application or intended use	(EU) No 73/2010 [16]
validity (period of)	Period between the date and time on which aeronautical information is published and the date and time on which the information ceases to be effective	(EU) No 73/2010 [16]
(data) verification	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 [16]
Verification Data	Data used to test and analyse the system: this is data comprising the test values and test data sets used to verify the system. It may include real data, modified real data or synthetic data. It includes data used to drive stubs, and any data files used by simulators or emulators.	SCSC Data Safety Initiative Working Group
W		
X		
Y		
Z		

Annex E: References

[1] RTCA/DO-200A, EUROCAE Document ED-76, Standards for Processing Aeronautical Data, September 1998.
[2] Improving the Analysis of Data in Safety-Related Systems, James Inge, 12 September 2008, http://www.safety.inge.org.uk/20080912-Inge2008b_Improving_the_Analysis_of_Data_in_Safety_Related_Systems-U.pdf
[3] Safety Analysis of Navigational Data, Paul Ensor, September 2009
[4] Integrity – an often-ignored aspect of safety systems, Alastair Faulkner, 2004, (EngD thesis), http://wrap.warwick.ac.uk/1212/
[5] BS EN 61508-3:2010, Functional safety of electrical/electronic/ programmable electronic safety-related systems. Software Requirements, June 2010.
[6] RTCA/DO-178C, EUROCAE Document ED-12C, Software Considerations in Airborne Systems and Equipment Certification, January 2012.
[7] DEF.STAN.00-56/4, Defence Standard 00-56, Issue 4, Safety Management Requirements for Defence Systems, June 2007.
[8] ISO31000:2009, Risk Management - Principles and Guidelines. First Edition, 2009-11-15.
[9] BS EN 50128:2011. Railway applications. Communication, signalling and processing systems. Software for railway control and protection systems, July 2011.
[10] RTCA/DO-330, EUROCAE Document ED-215, Software Tool Qualification Considerations, January 2012.
[11] Mars Climate Orbiter Mishap Investigation Board Phase I Report November 10, 1999, http://sunnyday.mit.edu/accidents/MCO_report.pdf
[12] Disastercast Episode 15 Quantitative Nonsense [including Mars Climate Orbiter], Drew Rae, http://disastercast.co.uk/transcripts/episode-15-transcript/
[13] American Airlines Flight 965 http://en.wikipedia.org/wiki/American_Airlines_Flight_965
[14] Disastercast Episode 18 Friendly Fire [including Data Safety], Drew Rae, http://disastercast.co.uk/transcripts/episode-18-transcript/
[15] BS EN ISO 19113:2005, Geographic Information. Quality Principles.
[16] 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 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:023:0006:0027:EN:PDF
[17] ISO/TS 19138:2006, Geographic Information. Data Quality Measures.
[18] Commission Regulation (EC) No 482/2008 of 30 May 2008 establishing a software safety assurance system, as amended http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:141:0005:0010:EN:PDF
[19] EUROCAE Document ED-153, Guidelines for ANS Software Safety Assurance – use for definitions only
[20] BS ISO/IEC 27001:2005, Information Technology. Security Techniques. Information Security Management Systems. Requirements.
[21] DEF(AUST)5679, Issue 2, Safety Engineering for Defence Systems – Standard, October 2008.
[22] ISO/IEC 2382-1:1993, Information Technology. Vocabulary. Part 1: Fundamental Terms.
[23] McGraw-Hill Dictionary of Scientific and Technical Terms, 6th Edition, November 2002. ISBN-10: 007042313X
[24] BS EN 61508-4:2010, Functional safety of electrical/electronic/ programmable electronic safety related systems. Definitions and Abbreviations, June 2010.
[25] The Characteristics of Data in Data-intensive Safety-related Systems, Neil Storey & Alastair Faulkner. Lecture Notes in Computer Science, Volume 2788, 396-409, 2003
[26] 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 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:305:0035:0052:EN:PDF
[27] BS EN ISO 19131:2008, Geographic Information. Data Product Specifications.
[28] BS ISO 19379:2003, Ships and Marine Technology. ECS databases. Content, Quality, Updating and Testing.

Annex F: Contributors

Contributors to this document include:

- Mike Ainsworth, Altran Praxis
- Dave Banham, Rolls-Royce plc
- Ian Bingham, CGI IT UK Ltd
- Simon Brown, QinetiQ
- Dale Callicott, BAE SYSTEMS
- John Carter, General Dynamics
- Martyn Clarke, RPS
- Duncan Dowling, DARD
- Paul Ensor, Boeing Defence UK Ltd
- Alastair Faulkner, Abbeymeade
- Ken Frazer, KAF
- Rob Green, NATS
- Amira Hamilton, CGI IT UK Ltd
- Paul Hampton, CGI IT UK Ltd
- Ali Hessami, Vega Systems
- Pete Hutchison, RPS
- Gavin Jones, Raytheon Systems Ltd
- Andrew Kent, Thales
- Julian Lockett, Frazer-Nash Consultancy Ltd
- David Lund, CGI IT UK Ltd
- Mark Nicholson, University of York
- Robert Oates, Rolls-Royce plc
- Mike Parsons, CGI IT UK Ltd
- David Perrin, Virtual PV
- Andrew Rankine, NATS
- Felix Redmill, SCSC
- Sam Robinson, EDF Energy
- Tim Rowe, EC Harris
- Alan Simpson, Ebeni
- John Spriggs, NATS
- Mark Templeton, QinetiQ
- Lesley Winsborrow, EDF Energy

Annex G: Acknowledgements

The document contributors would like to thank:

1. The Safety Critical Systems Club for support and encouragement
2. Brian Jepson of the SCSC for web hosting support and technical help with the SCSC web site
3. All the organisations that have hosted (or will be) hosting working meetings: Boeing, CGI, Ebeni, EDF, Frazer-Nash Consultancy, NATS, Rolls-Royce, QinetiQ, UKHO
4. Those that have been unable to attend meetings but have made supporting contributions
5. CGI for recognizing the importance of this activity including supporting the KO meeting