Best Practice for Data Quality Enables Asset Management for Rail

T.M. King & D. Crowley-Sweet†*

** LSC Group, GBR, TMK@LSC.CO.UK*

† Network Rail, GBR, DAVIN.CROWLEY-SWEET@NETWORKRAIL.CO.UK

Keywords: ISO 8000, conformance, computer-processable data specifications, bowtie risk model

Abstract

International standards are now available for organisations that are wishing to implement a systematic approach to asset management. These standards include ISO 55001 ("Asset management -- Management systems – Requirements") and ISO 55002 ("Asset management -- Management systems -- Guidelines for the application of ISO 55001").

One section of ISO 55001 identifies the important role of information requirements in supporting asset management. In addition, ISO 55002 includes guidance on the scope of the necessary activities to determine those requirements. Neither standard, however, provides a detailed explanation of the technicalities of achieving data quality when addressing information requirements.

Owners of physical assets must take a series of decisions to achieve effective and efficient management of those assets. Each decision is fully robust only when information is available to provide objective evidence about all relevant aspects of the asset and the asset management system. This information will be fit for the purpose of the right person taking the right decision at the right time if the underpinning data has characteristics that conform to all applicable requirements. This conformance to requirements is the formal definition of quality as per ISO 9000 ("Quality management systems -- Fundamentals and vocabulary").

Network Rail has embarked on a journey to achieve data quality by identifying best practice in the form of ISO 8000 ("Data quality"), which builds on the general approach to quality management within ISO 9000 by addressing how specific features of data affect this approach. Network Rail has, in particular, adopted ISO 8000-150 ("Data quality -- Part 150: Master data: Quality management framework") and as a result:

- supported the strategic objective of treating data as an asset with as much value as physical assets;
- addressed classes of asset information that enable the full range of asset management activities;
- adopted the lower level processes for managing data quality within the standard as the basis on which to determine how data quality is a risk for consideration at the strategic level of Network Rail;
- implemented roles and responsibilities for the individuals who perform activities within the lower level processes for managing data quality;
- demonstrated the complementary nature of the ISO 55000 series of standards and ISO 8000 in achieving a comprehensive and coherent approach to meeting the information requirements of asset management.

By deploying ISO 8000 in support of asset management, Network Rail has already achieved a series of benefits, covering cost and risk reduction and improved compliance. In particular, the organisation has been able to address concerns from the Office of the Rail Regulator with respect to data quality within Network Rail. These concerns were mirrored by low confidence in data; using ISO 8000 and the ISO 55000 series of standards in tandem has begun to increase this confidence across Network Rail.

1 Introduction: Why data quality is an issue

On 23 September 1999, the Mars Climate Orbiter unexpectedly failed to communicate with mission control at National Aeronautics and Space Administration (NASA) **[1]**. The Orbiter never resumed communications, leaving little to show for a total cost of just under USD 330 million for the whole mission.

Subsequent investigations determined that the Orbiter had approached too close to the surface of Mars. This approach would have exposed the Orbiter to excessive and ultimately destructive stresses from the atmosphere of the planet.

Within days, the investigators were able to identify the fundamental root cause of the incident **[2]**. This cause arose when one team submitted a data file to another team [see **Figure 1**]. This file included a value for a length where the applicable interface specification required the corresponding unit of measure to be metric (Newton-second). The file was received by another team, which interpreted the value for angular momentum as having the unit in the specification.

The value, however, was actually correct for a unit of measure of pound-second.

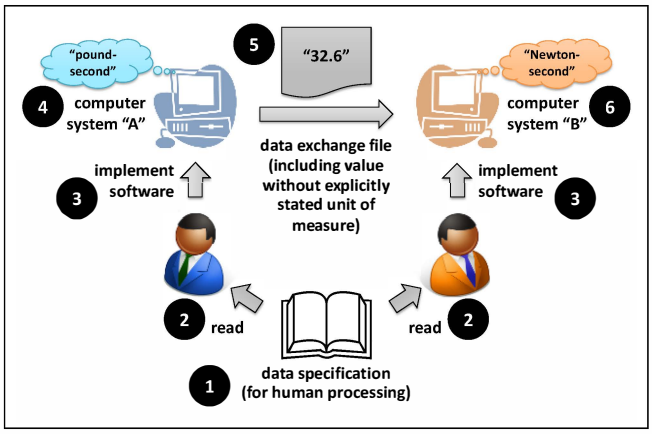

Figure 1: An overview of the root cause for the Mars Climate Orbiter incident.

The development of ISO 8000 (Data quality) has been a response to the recognition that data can cause incidents such as the one befalling the Mars Climate Orbiter, as well as causing organisations many less high profile issues.

When an organisation is not one receiving public money then commercial sensitivities can reduce willingness to share the specific details of issues arising from non-conforming data. Such issues, however, have been a motivation for participants in the development of ISO 8000. One participant was able to report personal knowledge of issues at a major global oil and gas company **[3]**. These issues included:

- inaccurate data caused the collision of two oil wells, halting production for more than one month;
- one paint manufacturer was supplying ten different combinations of the colours for the logo of the oil and gas company;
- different parts of the company were using different software for the same tasks, causing unnecessary inefficiency;
- different parts of the company were producing incompatible financial reports, requiring costly and time-consuming reconciliation;
- inconsistent master data led to costs of several million United States dollars;
- drawings went missing, resulting in a fine from regulators of two million United States dollars.

Furthermore, the United Kingdom Council for Electronic Business (UKCeB) has identified similar issues in Defence **[4]**. Such issues have increased costs or risks within various organisations.

The following section explains how systematic and systemic data quality is the means by which to reduce the risk of data becoming a source of issues within organisations.

2 The vision for systematic & systemic data quality

The Mars Climate Orbiter demonstrates the key features necessary to achieve a sustainable, effective approach to data quality.

The first feature is the role of a rigorous specification that is processable by computer. Such a specification is the basis for data quality and, in the case of the Mars Climate Orbiter, would have been rigorous by directing engineers to state explicitly the unit of measure for the problematic length in the file. The same specification would also be the basis on which the receiving team would have been able to test that the unit of measure was actually the required unit (in this case, Newton-second).

Since 1984, ISO/TC184/SC4 (Industrial Data) has been developing standards that contain rigorous, computer-processable data specifications. These standards have also all recognised that every data file shall always explicitly state the units of measure where applicable to any values in the file.

The second feature is that a computer-processable specification is not sufficient to guarantee any particular organisation is capable to produce quality data.

In response to the loss of the Mars Climate Orbiter, NASA implemented an action plan that resulted in changes to a wide range of different elements involved in NASA prosecuting space exploration. These elements included **[5]**:

- implementing additional management support;
- improving lines of communication;
- co-locating some technical teams;
- executing additional systems engineering activity;
- making changes to propulsion systems on spacecraft;
- increasing the use of peer review of tasks;
- validating existing risk assessments across the portfolio of missions.

Such elements express the breadth of how data ultimately serves as the fundamental basis for effecting the activities and outputs of organisations. This basis is the role of data in enabling decision making.

The following section explains how ISO 8000 is the emerging means by which to achieve systematic and systemic data quality.

3 The development & application of ISO 8000

3.1 Overview

ISO 8000 is a multi-part standard under the control of ISO/TC184/SC4. The following sections explain:

- the development history and current status of ISO 8000 [see **Section 3.2**];
- how ISO 8000 relates to asset management [see **Section 3.3**];
- how Network Rail has begun to exploit ISO 8000 [see **Section 3.4**].

3.2 The history & status of ISO 8000

In executing a programme to produce standards such as ISO 10303 (Industrial automation systems and integration -- Product data representation and exchange) and ISO 15926 (Industrial automation systems and integration -- Integration of life-cycle data for process plants including oil and gas production facilities), ISO/TC184/SC4 has met the need for rigorous, computer-processable specifications to enable data operations in support of engineering. These operations include data exchange, sharing and archiving. The programme has developed the individual data specifications but also led to the discovery of common features in respect of setting requirements on and testing the conformance of data.

In 2005, ISO/TC184/SC4 initiated the development of ISO 8000 as the standard to capture the generic route to achieving data quality. This standard sits as the bridge between the overall approach to quality management within the ISO 9000 series of standards and individual data specifications , which can be within other standards from ISO/TC184/SC4, within standards from other committees and organisations or even just specific to the needs of a particular set of circumstances.

The appropriate scope of ISO 8000 became apparent from analysing ISO 9000 and the existing standards of ISO/TC184/SC4. This scope includes the following key elements (with full details available in Part 1 of ISO 8000):

- the principles of data quality;
- the characteristics that determine the quality of data;
- the necessary elements that support achieving data quality;
- representing data requirements, measurement methods and inspection results;
- frameworks for measuring and improving data quality.

The development programme has resulted in the following parts, which are currently at different stages of progress:

- ISO/TS 8000-1:2011, Data quality -- Part 1: Overview;
- ISO 8000-2:2012, Data quality -- Part 2: Vocabulary;
- ISO/DIS 8000-8, Data quality -- Part 8: Information and data quality: Concepts and measuring;
- ISO/DTS 8000-60, Data quality -- Part 60: Information and data quality management process assessment;
- ISO/CD 8000-61, Data quality -- Part 61: Information & data quality management process reference model;
- ISO/NP 8000-62, Data quality -- Part 62: Process maturity assessment model;
- ISO/NP 8000-63, Data quality -- Part 63: Measurement framework;
- ISO/TS 8000-100:2009, Data quality -- Part 100: Master data: Overview;
- ISO 8000-110:2009, Data quality -- Part 110: Master data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification;
- ISO/TS 8000-120:2009, Data quality -- Part 120: Master data: Exchange of characteristic data: Provenance;
- ISO/TS 8000-130:2009, Data quality -- Part 130: Master data: Exchange of characteristic data: Accuracy;
- ISO/TS 8000-140:2009, Data quality -- Part 140: Master data: Exchange of characteristic data: Completeness;
- ISO/TS 8000-150:2011, Data quality -- Part 150: Master data: Quality management framework;
- ISO/TS 8000-311:2012, Data quality -- Part 311: Guidance for the application of product data quality for shape (PDQ-S).

This programme has led to being able to identify the following features of systematic and systemic data quality:

- data quality is nothing more or nothing less than the characteristics of the data meeting requirements;
- the ISO 9000 series of standards establishes a suitable basis for performing data quality management;
- systems engineering is the means by which to achieve a coherent and effective connection from the business need for information through to the technology and services that enable all of the required activities to process and manage data (ISO 15288 provides further details of a general approach to systems engineering);
- data cleansing is a useful technique to address specific issues in sets of data but is not sufficient because such cleansing fails to identify the root causes of non-conforming data;
- the ultimate business value of data quality is to enable the right person to make the right decision at the right time.

These features lead to a diagrammatic overview of the overall scope of ISO 8000 [see **Figure 2**]. This overview

encompasses all the parts of the standard within the current development programme. The overview can also support identification of where additional new parts could extend the coverage of the standard.

The diagrammatic overview is the basis for identifying a wide range of different constituent means to achieving systematic and systemic data quality. Such means are as varied, for example, as the role of enterprise architecture through to the role of training. Enterprise architecture establishes a coherent landscape within which data quality can address strategic objectives within the organisation. Training ensures people possess appropriate skills to perform effective data quality management.

The overview also applies to all types and aspects of information and communications technology, covering:

- software and hardware;
- all representations of data, including databases, documents and websites;
- all types of data, including reports, transactions, master data, reference data, meta data and protective marking (for security purposes);
- all types of process, including database operations, data exchange and sharing, data migration, data cleansing, archiving and predictive analysis;
- all types of solution, including business intelligence, enterprise data management, knowledge management and domain applications (enterprise resource planning, computer aided design and so on);
- all solution elements, including access (for example, mobile devices), search, information delivery (for example, portals), cloud-based services, security and support.

The following section explains how ISO 8000 complements the content of PAS 55 ("Asset management") **[6]** **[7]** and the ISO 55000 series of standards.

3.3 ISO 8000 & asset management

PAS 55 and the ISO 55000 series of standards establish best practice approaches to asset management. These standards make a clear reference to the underpinning role of data in achieving effective and efficient asset management. This role, however, depends on the nature of data and the standards do not provide detailed observations on that nature.

ISO 8000 provides the detail by which to achieve

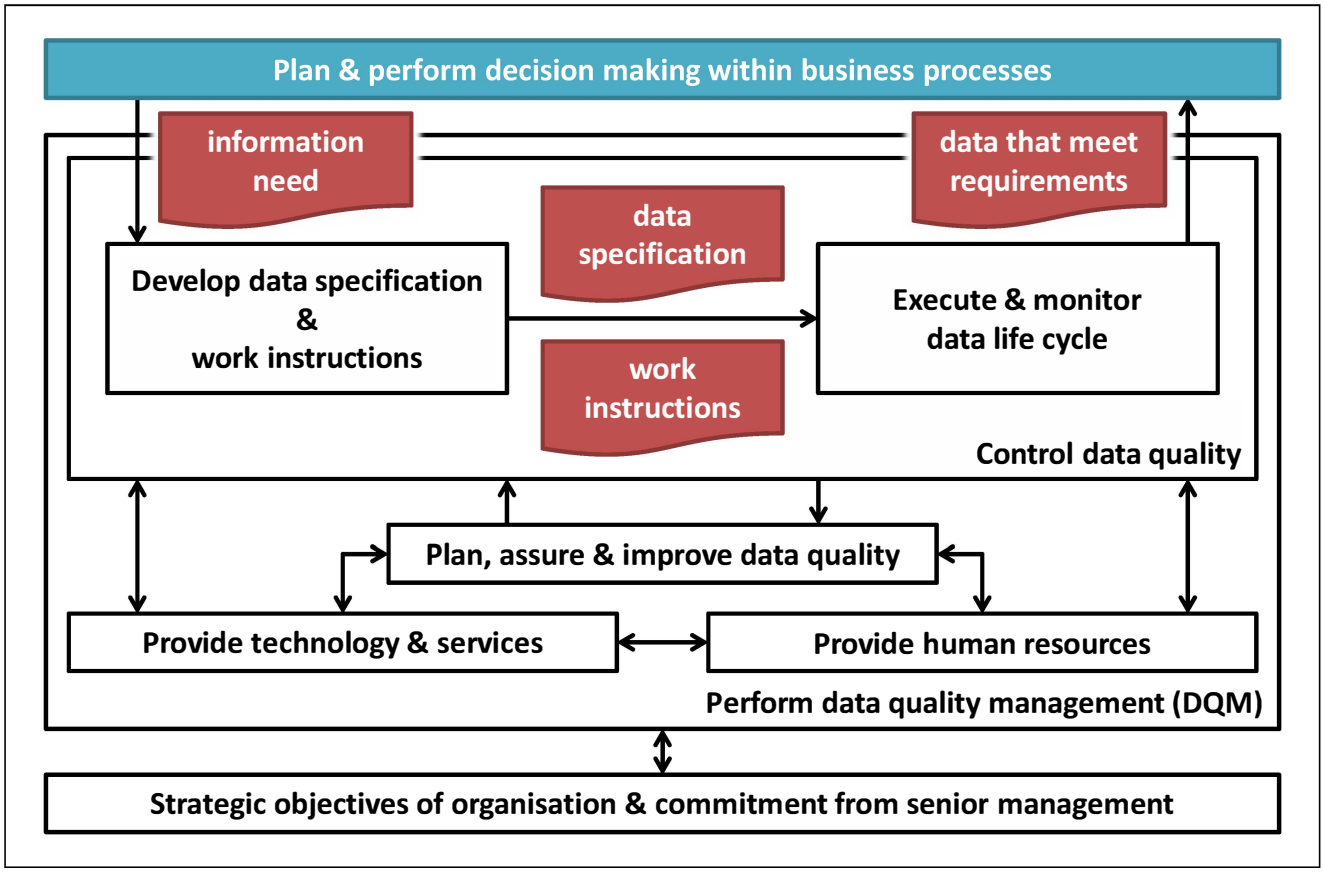

Figure 2: An approach to systematic & systemic data quality.

appropriate data quality to support the asset knowledge enablers as identified by PAS 55 and the ISO 55000 series of standards. These enablers (and the mapping to ISO 8000) are:

- asset information strategy (covered by planning within the "plan, assure & improve data quality" box in the diagrammatic view of ISO 8000 [see **Figure 2**]);
- asset knowledge standards (covered by "Develop data specifications & work instructions", where re-use of standards avoids nugatory work and establishes a foundation for interoperability with other organisations that adopt the standards);
- asset information systems (covered by "Provide technology & services");
- asset data and knowledge (covered by "Execute & monitor data life cycle).

Although the ISO 55000 series of standards is not an explicit part of current best practice within the Ministry of Defence (MOD), the approach to asset management aligns with the content of the standards. This approach is, by definition, also as equally reliant on data quality. The MOD has supported the development by UKCeB of a Data Quality Assessment Tool (DQAT) to provide a means by which to determine a holistic view (processes, data, performance measurement, governance and culture) of data quality rather than just focussing on the issues with individual data values within databases of asset information.

The DQAT has enabled the MOD to take a more strategic approach to identifying how data quality impacts the safe, effective and efficient execution of asset management across a complex enterprise (large supplier base, highly mobile assets, cutting-edge technology and global operations). The tool also aligns with the emerging content of Part 60 series of standards within ISO 8000.

The following section explains how Network Rail has taken the content of ISO 8000 and used the standard to underpin improvement of information to support management of a complex portfolio of assets across the organisation.

3.4 Exploitation of ISO 8000 by Network Rail

In order to address the important role of data in enabling effective and efficient asset management, Network Rail is executing a programme of transformation **[8]**. This programme has the name ORBIS (Offering Rail Better Information Services).

ORBIS has the primary purpose to improve the capabilities of the Asset Management organisation within Network Rail. This organisation fulfils three main objectives in order to transform data into intelligence:

- align information to business processes;
- deliver trusted and sustainable quality information;
- enhance and exploit five related information types.

These objectives have established the destination for data management across Network Rail but rely on a detailed approach to data quality in order to reach that destination. Network Rail required a suitable source of best practice to determine this detail.

Despite not having been involved in the development of the ISO 8000 standard, Network Rail discovered the availability of ISO 8000-150 and was able to identify the relevance and suitability of that document to provide information on best practice for managing data quality. ISO 8000-150 describes a set of three roles and three top-level processes; the resulting matrix contains nine lower level processes that are the focus of the standard [see **Figure 3**].

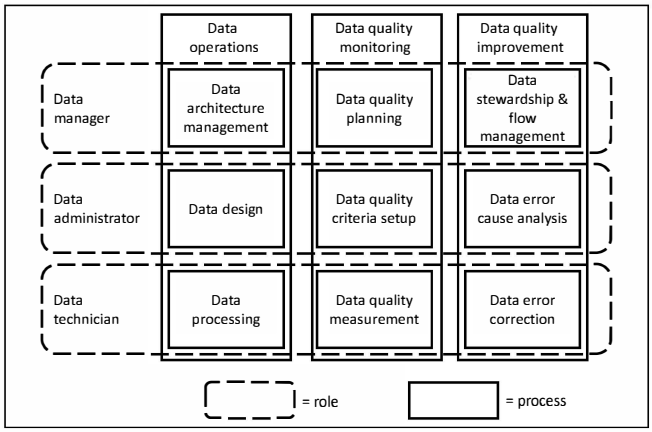

Figure 3: The quality management framework from ISO 8000-150.

Network Rail now had a means by which to establish governance against a definitive framework of best practice, enabling a capability of bench marking on an ongoing basis. The framework has also brought proper focus on the health of the management system for data and how this system drives the accuracy of data across the organisation.

Network Rail identified one important principle with respect to assessing the potential impact of data quality on the safe, effective and efficient operations of the organisation. This principle is not to try to claim the positive effects of data quality because data enables decisions, which are integral to processes across Network Rail. These processes will already have claimed the positive benefits. Data quality, instead, is a potential threat to being able to achieve those benefits.

By identifying the role of data quality as a threat, Network Rail has established the basis on which to address the strategic significance of data quality alongside other strategic considerations across the organisation. Threats are the source of risks with negative consequences and, thus, fit within the overall approach to risk management by Network Rail.

The bowties approach is a tool used by Network Rail to perform risk management. This approach depends on identifying the failure modes leading to an event that is

hazardous to the operations of an organisation. One such event is inadequate data quality, which leads to a clear set of consequences in terms of inability to make safe and effective decisions.

Network Rail needed to identify the failure modes that lead to the event of inadequate data quality. These modes are the lower level processes that appear in ISO 8000-150.

On the basis of the above, Network Rail has been able to establish a systematic approach that consists of the following steps [see **Figure 4**]:

- taking the lower level processes within ISO 8000-150;
- creating a system model that has leaf nodes corresponding to the lower level processes;
- populating the bowtie model with the leaf nodes as the failure modes;
- analysing the bowtie to generate results by which to report data quality as a risk alongside other risks across the organisation.

The system model includes a structured definition of each node. This definition consists of:

- the requirement, identifying the purpose of the node;
- the specification, providing detail of the activity and the applicable inputs, outputs, controls and resources (as per an IDEF0 model **[9]**);
- the threat, describing how the node can fail to deliver appropriate effect;
- the consequence, describing how the node can cause a negative impact on data quality;
- the roles applicable, covering the responsible, accountable, consulted and informed parts of the organisation (as per a RACI matrix **[10]**).

Having identified necessary resources in the specification, Network Rail is now in the process of creating a competency framework for all the roles that successfully deliver data quality across the organisation.

Furthermore, by understanding data to be an asset **[11]** alongside the physical assets across the organisation, Network Rail has identified how data is as dependent on appropriate maintenance activity as a physical asset. Such activity is an important part of the overall approach to managing data quality.

The following section describes how ISO/TC184/SC4 and Network Rail can further build on the development and exploitation of ISO 8000 to date.

4 Further work & next steps

ISO/TC184/SC4 has yet to agree whether and how to include the diagrammatic overview [see **Figure 2**] of ISO 8000 into the actual standard. The overview is potentially suitable for a new part that explains the overall structure of the standard.

ISO 8000-150 currently includes "master data" in the title but Network Rail has demonstrated that this part of the overall standard is suitable for multiple types of data (for example, master, meta, reference and transactional). This demonstration is potentially sufficiently compelling to persuade ISO/TC184/SC4 to create a new version of Part 150 with a suitable modification to the title and contents. These modifications would indicate the wider applicability of the standard.

ISO 8000 does not currently include any detailed content to describe how to treat data quality as a risk. Such a description could benefit from the experience of Network Rail and ISO/TC184/SC4 should review the risk-based approach at Network Rail to decide how to further enhance ISO 8000.

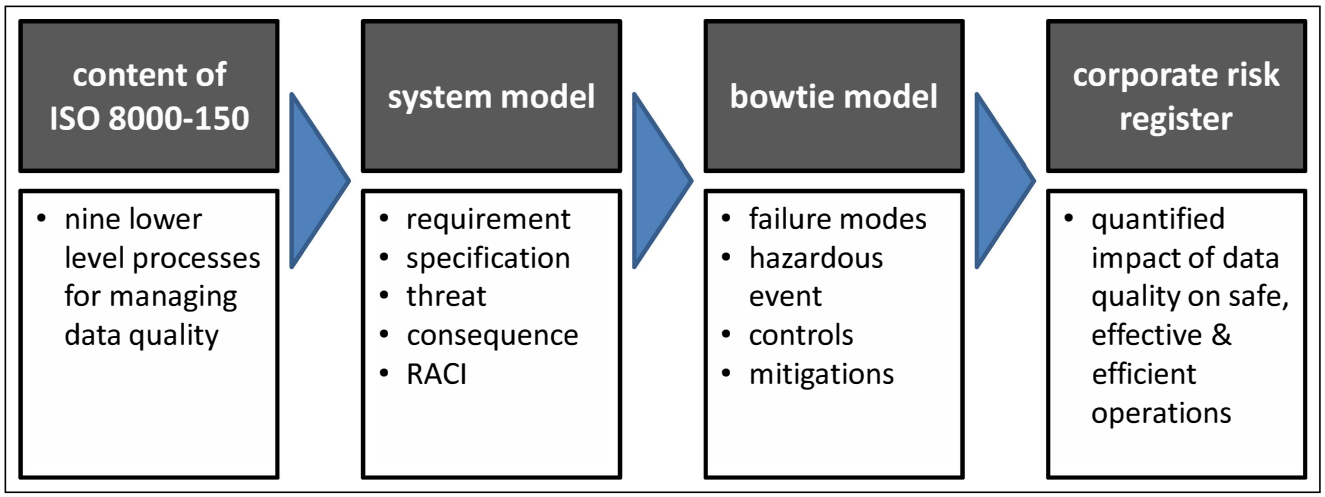

Figure 4: The Network Rail approach to managing data quality.

Network Rail is looking to develop a richer analysis of the network model that is a logical derivation from the system model of the lower level processes in ISO 8000-150 [see **Figure 3**]. This network model captures how artefacts flow between the processes (for example, the process "data design" creates the artefact "data schema" that is the basis for the process "data processing"). Networks possess characteristics such as resilience that are a function of the potential weakness of the individual flows. Network Rail would be able to enhance the bowtie model by having a better quantification of the network for managing data quality. This quantification could, for instance, be on the basis of the output from using the DQAT from UKCeB to provide insight into both hard (for example, the existence of formal schemas for critical data) and soft (for example, the perception among employees as to the value of data quality) aspects of data quality across the enterprise. Quantification also helps to understand the relative balance of resource that is necessary for the different lower level processes (for example, by increasing the resources for "data architecture management", all the other processes would potentially require fewer resources).

The following section concludes this paper by drawing attention to the benefits arising from successful exploitation of ISO 8000.

5 Conclusions: The benefits of systematic & systemic data quality

Data quality serves one fundamental purpose: if the characteristics of data meet requirements then the right individual can make the right decision at the right time. These decisions are the life blood of all processes and are the only basis for safe, effective and efficient operations.

ISO 8000 establishes a framework for systematic and systemic data quality (including computer-processable data specifications and a holistic approach to managing data quality). The standard offers interacting parties the opportunity to recognise the adoption of appropriate best practice and to achieve interoperability.

By achieving systematic and systemic data quality, organisations can also look to gaining from strategic benefits. These benefits include regulatory compliance and also a pair of complementary characteristics of the enterprise: agility (the ability to respond to opportunities); and resilience (the ability to withstand threats).

Network Rail is on a journey of exploiting ISO 8000 and is beginning to benefit from systematic and systemic data quality.

Acknowledgements

The authors are grateful for the support and stimulation of colleagues across Network Rail, LSC Group and ISO/TC184/SC4 who have participated in many of the experiences that appear in this paper.

References

[1] "Mars Climate Orbiter Mishap Investigation Board: Phase I Report" (1999-11-10).

[2] "Mars Climate Orbiter team finds likely cause of loss", Release 99-113, National Aeronautics and Space Administration (NASA) (1999-09-30).

[3] Private correspondence with T. M. King (2012-03).

[4] King, T.M., Dixon, R.A.J., Galeozzie, R.J. "Decision support for asset management: A Defence perspective", IET and IAM Asset Management Conference 2013 (2013-11-27 to 28).

[5] "Mars press conference: Dr. Edward Stone's Presentation", National Aeronautics and Space Administration (NASA) Jet Propulsion Laboratory (JPL) (1999-11-10).

[6] PAS 55-1:2008. "Asset management -- Part 1: Specification for the optimized management of physical assets", Publicly Available Specification, British Standards Institution (2008).

[7] PAS 55-2:2008. "Asset management -- Part 2: Guidelines for the application of PAS 55-1", Publicly Available Specification, British Standards Institution (2008).

[8] Crowley-Sweet, D. "Offering Rail Better Information Services", workshop, Institute for Asset Management Annual Conference 2013 (2013-06-25).

[9] "Information technology -- Modeling Languages -- Part 1: Syntax and Semantics for IDEF0", ISO/IEC/IEEE 31320-1 (2012).

[10] Cadle, J., Paul, D., Turner, P.: "Business analysis techniques: 99 essential tools for success", 2nd edition, British Computer Society (BCS) (2014-09-23).

[11] Laney, D. "Introducing infonomics: Valuing information as a corporate asset", Gartner (2012-03-21).