Network Rail

OFFICE OF RAIL REGULATION

Scott Wilson

SCOTT WILSON BUSINESS CONSULTANCY

Improving Rail Data

Overview Report – September 2006 v1.0

September 2006

Scott Wilson Business Consultancy

We work with clients to develop, implement and evaluate projects, programmes and change initiatives to improve performance and reduce risk.

Overview Report – September 2006

Authors: J Macnab

Reviewed by: C Palmer

Approved by: S Brundle

Scott Wilson Business Consultancy
Greencoat House
15 Francis Street
London
SW1P 1DH

Tel: +44 (0)20 7798 5200
Fax: +44 (0)20 7798 5201
Email: consultancy@scottwilson.com

Document History

Version Number	Version Date	Reason for Issue	Approval Initials
1.0	25/09/06	First release	SB

Table of Contents

Glossary of Terms

Term	Description
ACTRAFF	Actual Traffic
CIF	Common Interface File
CSV	Comma Separated Variable, a text file format for importing into spreadsheets
DA	Delay Attribution
FRAME	Fault Recording and Monitoring of Equipment, system for recording infrastructure faults and progress on fixing them
FOC	Freight Operating Company
IMC	Infrastructure Maintenance Company
JPIP	Joint Performance Improvement Plan
NETRAFF	Network Traffic
NR	Network Rail
ORR (The)	The Office of Rail Regulation
PALADIN	Performance & Loading Database, holds historical train and incident data
PAT	Performance Action Tracking, system for creating and progressing Initiatives
PEARS	Paladin Extract and Reporting System, system for commercial (Schedule 8) settlements
PfPI	Process for Performance Improvement, workbooks of aggregated performance data
PMR	Performance Management Reporting, system populating PfPI from PUMPS and PAT
PMRS	Performance Management Reporting Systems (PAT, PUMPS and PMR)
PPM	Industry performance measure
PPS	Possession Planning System
PSR	Passenger Service Requirement, trains which are obliged to run as part of the TOC's Franchise conditions
PSS	Performance Systems Strategy
PUMPS	System contains extracts of train delay & incident data from Paladin
RotR	Rules of the Route

RotP	Rules of the Plan
RSSG	Rail Statistics Steering Group
SLA(s)	Service level agreement(s)
TMS	Train Management Systems (in the context of PSS are TRUST and/or TSI).
TOC	Train Operating Company
TOPS	Total Operations Processing System
TRUST	Train Running System TOPS, operational system recording movements, incidents and delays
TSR	Temporary Speed Restriction

1 Executive Summary

1. Scott Wilson has been appointed as the Independent Reporter for Rail Data (The Reporter) on behalf of the Office of Rail Regulation (ORR) and Network Rail (NR), to review the fitness for purpose of data within Network Rail systems. As the data held in these systems is closely connected with other rail industry data, the Reporter has also considered the wider industry picture.

2. The overall aim is to support the industry in enhancing the reliability of its information for decision making by describing best practice data management and data quality in an industry 'Data Code', and independently assessing actual practice on high priority datasets against the Data Code. This will not only provide assurance on data management and quality, but enable improvements to be made and demonstrated.

3. The Reporter's overall plan of work can be seen in four main stages:

    a. Analysis of Network Rail systems, initial prioritisation of industry information needs, datasets and proposals for assessing the datasets using the Data Code;

    b. Consultation with the industry on priorities, assessment and the Data Code;

    c. Final agreement of the Data Code and the order of assessment of datasets;

    d. Implementation of the assessment programme of agreed priority datasets.

4. The output from the first stage is this Overview Report. Its purpose is to report on the work carried out so far and to outline the future work programme. Its purpose is also to provide an industry consultation document that:

    a. gives an overview of Network Rail systems and industry processes and data flows that provide information for industry wide decision making;

    b. proposes a prioritised list of industry needs related to key industry decisions from which the highest priority datasets can be agreed for assessment;

    c. proposes how the agreed datasets will be assessed against the Data Code.

    The work programme in this first stage has consisted of three activity streams: a 'bottom up' analysis of Network Rail systems; a 'top down' review of industry decision making and information needs; and a study of best practice in data management and data quality, considering available government and commercial standards.

5. The main observations from the initial analysis of Network Rail systems are:

    a. The complexity of the system environments and interfaces, the level of manual intervention and the different standards of data management are barriers to achieving reliable information. Many systems have long histories and limited available documentation.

    b. Network Rail is seeking ways to reduce this complexity and to integrate data through its 'Information Management Vision' [1]. A new integrated operational planning system, amongst other initiatives, is underway in support of this vision. The Reporter supports the intent of these programmes and proposes that new systems that relate to the agreed priority datasets should be brought within the scope of The Reporter's work.

c. Enhanced data management will not only help improve data reliability, but will also inform the design and development of data structures and interfaces in new or upgraded systems.

d. The main functional processes and associated systems and data flows have been derived from systems analysis and these will be used to inform the prioritisation of the datasets.

6. The Reporter now proposes that the prioritisation of datasets for assessment is based on industry strategic (business), mandatory (legal/regulatory) and operational (functional process) needs in the form of a top ten 'critical industry outputs' for decision making, and has developed these into a 'straw man' with each output related to its constituent datasets. This is presented for industry consultation.

7. The Reporter is proposing an approach to assessment of datasets that is simple, comprehensive and repeatable. In particular, it examines data management processes so that sensitivities around confidential information can be accommodated. It promulgates best practice in data management through the Data Code, provides a scorecard for comparisons over time and between datasets, and provides recommendations for improvements that can be later reviewed. Reports and scorecards can be aggregated to demonstrate overall improvements in data management practice and in data quality over time across the industry.

8. The active engagement of the main industry participants is critical to success because of the complexity of information flows, constant industry change and sensitivities around confidential information. The shared goal is to improve information for decision making across the industry. The Reporter sees the Rail Statistics Steering Group (the RSSG), which has been established by ORR to oversee this and associated work, playing an important role and proposes specific terms of reference for this group in relation to the consultation plans, the Data Code and the assessment programme.

9. The Reporter is seeking feedback from the industry on the Overview Report, in particular:

a. The proposed 'straw man' of critical industry outputs and associated datasets;

b. The proposed assessment of datasets and development of the Data Code.

10. The Reporter welcomes feedback on this Overview Report published on ORR's web site. It also recognises that the subject matter requires input from a selected sample of decision makers, those involved in the production of critical industry outputs, and information professionals, covering policy making, regulation, infrastructure and operations across the industry; thus:

a. the Report should be sent to main industry participants for comment;

b. interviews should be held with the selected sample to brief them and elicit views;

11. After consultation, The Reporter plans to publish a feedback report via ORR's web site, and then review and confirm the priority datasets and approaches to the assessments and Data Code reflecting industry views. It is intended to run initial field trial assessments with industry participation to verify the approach so that the later programme of assessments meets the recognised industry needs. After completion of the assessment of the highest priority datasets, The Reporter will be in a position to give an overall review of performance and recommended improvements.

2 Introduction

2.1 The Overview Report as a Consultation Document

This report is presented for consultation within the rail industry by Scott Wilson in its role as the Independent Reporter for Rail Data in support of industry objectives to enhance the quality of information and data management for decision making.

Feedback is being sought through industry consultation from decision makers, information providers and information professionals across the industry in order to hone the proposals contained in The Report and to agree the best way forward.

This report additionally seeks to inform the reader on the processes and mechanisms to be employed in the assessment of datasets within the industry specifically:

- The prioritisation of datasets for assessment;
- The process for assessing datasets;
- The development of the Data Code of best practice in data management and data quality.

This introduction sets the context for the Report. The detailed proposals on which feedback is sought are covered in Section 4.

The Reporter would like to thank all those who contributed to this work especially those in Network Rail who were very helpful and open in their responses.

2.2 The Reporter's Role in Enhancing Industry Information and Data Management

The Government White Paper, The Future of Rail [2], recognised the need for robust information with which to plan, take decisions and monitor progress within the industry. ORR, as part of its remit, is therefore seeking to facilitate better industry decision making through improved reliability and access to that data.

In this context, Scott Wilson has been appointed as the Independent Reporter for Rail Data (The Reporter) on behalf of ORR and NR to examine and provide assurance upon a number of Network Rail systems, concentrating on those systems that are key to the industry and produce information used by the wider industry.

Information flows back and forth across all the main industry participants from Network Rail as the infrastructure provider, to the operating companies and the industry suppliers and to the regulator and Government in England, Wales and Scotland. Thus, whilst Network Rail systems are at the heart of this information flow, data reliability and accessibility is seen as a cross industry concern. This means that the main participants need to be actively engaged as the assessments of datasets will include data owned and held not only by Network Rail or ORR, but also by other parties; the whole industry should

therefore be able to recognise the benefits of the data code, and The Reporter needs to take account of this industry perspective in fulfilling its role.

The Reporter's overall aim is to support the industry in enhancing the reliability of its information, and the data from which that information is derived, for decision making and, in particular:

- Work with the industry to describe and update best practice in managing data and achieving data reliability;
- Agree priorities for its work with the industry;
- Independently assessing how actual practice compares with best practice;
- Make recommendations for how data management and data quality can be improved across the industry;
- Provide assurance by enabling the performance and improvements to be demonstrated.

2.3 The Scope of The Reporter's Work

As Reporter, Scott Wilson initially focused on the systems and underlying datasets within Network Rail as the major information provider and user within the industry. To provide a starting point for the work, an initial list of systems was drawn up and agreed between ORR and Network Rail. Subsequently, additional systems were identified by The Reporter as important, and one of the systems on the original list was identified as obsolete. These systems are shown in Annex A and referred to as the pre-identified and newly identified systems within this report.

It was noted that a contract had been let in March 2006 to Selex SI as prime contractor for a new Integrated Train Planning System (ITPS). This new system is replacing identified systems, most notably TSDB and Aplan and is aiming to create a new integrated planning system based on a single, unique, rationalised database. It was agreed that this development should, in due course, come within the scope of work.

Scott Wilson is required to set out how it proposes to assess the datasets within these systems for completeness, reliability and accuracy of data in relation to their use for decision making, and the quality of the processes by which the data is compiled and whether these comply with agreed procedures. In order to do this, it was agreed that analysis should be undertaken to establish industry needs in relation to key industry decisions and to establish an identification and prioritisation process for industry datasets in this context and in relation to the main functional processes across the industry.

Any assessment process necessarily needs to be made against an agreed set of criteria. It was therefore intended that a Data Code should be established as part of the scope of this work, defining standards of best practice and annexing two lists: the first is the agreed list of datasets to be assessed against the standards as a priority (Annex 1 to the Data Code), and the second listing those that had been assessed with a summary of results (Annex 2 to the Data Code).

2.4 Structure of the Remainder of the Overview Report

The remainder of this report explains how The Reporter has gone about its work, what has been delivered to date and proposes the way forward and how it intends to engage the industry through the consultation and beyond. The remainder of the Report is structured as follows:

- Section 3 summarises the **Overall Plan** devised by The Reporter in order to fulfil its role, and the approach taken to the work to produce this Overview Report for consultation;
- Section 4 outlines the **Outputs** as the main proposals from this stage of the work on which industry views are sought:
    - Overview of Network Rail systems;
    - Identification and prioritisation of datasets;
    - The assessment process;
    - The Data Code;
    - The consultation plan;
- Section 5 details **Next Steps** taking account of industry feedback on the proposals in the context of the overall plan;
- Section 6 gives the **Conclusions and Recommendations.**

Supporting detailed work and references are attached in the **Annexes** to the Overview Report.

3 Overall Plan

3.1 Introduction

This section explains how The Reporter has conducted its work to date and is planning to approach its future work and the reasons for that approach. It explains how this first stage, up to the consultation, fits into achieving the overall objective.

3.2 The Decision Making Context

The overall plan for The Reporter's work needs to be seen in the context of the wider influences on decision making in the industry. Generically the 'decision-making scene' can be pictured as in Figure 1. This provides a picture of the overall scene that is being surveyed and gives a continuing emphasis on the ultimate goal of improved decision making using robust information derived from high quality and accessible data in order to plan, make decisions and monitor progress within the industry.

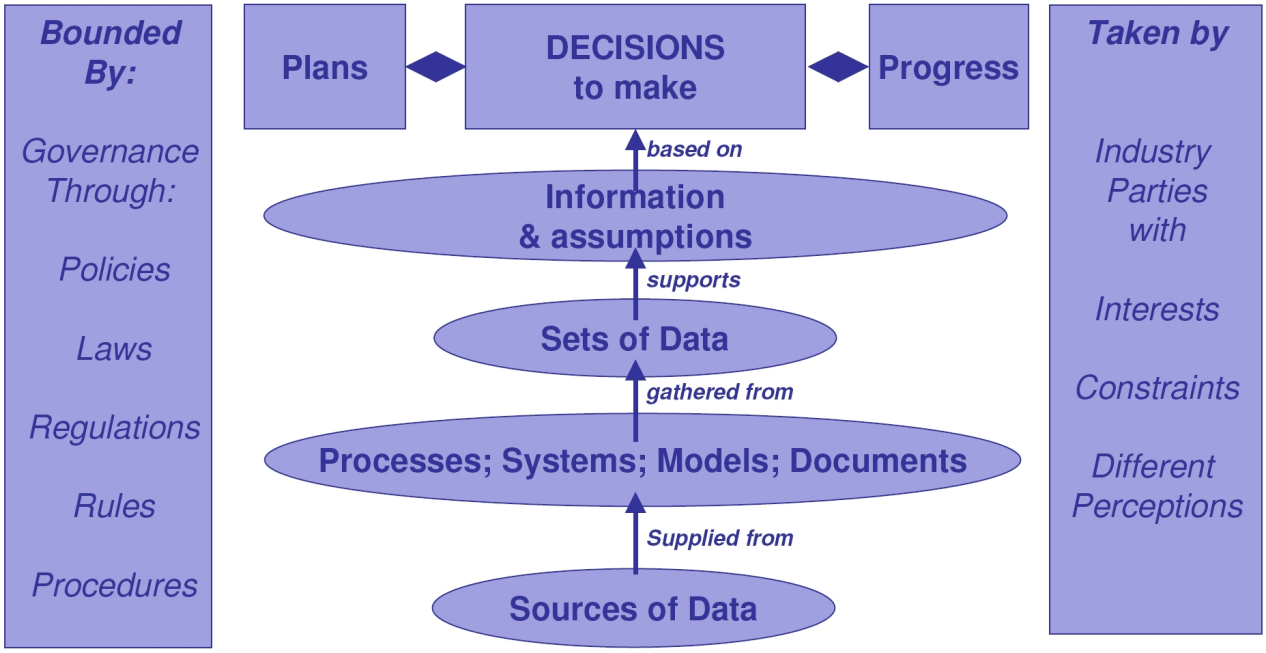

Figure 1 : Industry Decision Making Scene

Decisions within the industry are made in relation to plans and progress, based on information and associated assumptions derived from sets of data (datasets) gathered

from various sources held within processes, systems, models, and documents. Challenges arise because:

1. The decisions are bounded by a set of policies, laws, regulations, rules and procedures that are largely apparent, but which constantly change;
2. Decisions are taken by different industry parties (e.g. Government, ORR, Network Rail, Train Operating Companies (TOCs) and Freight Operating Companies (FOCs)), together and independently, with different interests (e.g. shareholder returns), constraints (e.g. contractual obligations) and different perceptions (e.g. belief in free market).

The Reporter role initially covered a sub-set of this scene: 'Network Rail systems' (including electronic and paper systems) and the associated datasets that underpin the main industry processes and models, and particular industry rules (the Rules of the Route and the Rules of the Plan).

Given the focus of attention on decision making and information derived from datasets, it is important to consider the wider aspects of the decision making scene to take account of the political, legal and regulatory environment as well as the vital part that the main industry participants play. This wider engagement is critical to success and this view of the decision making scene underpins the overall plan and approach of The Reporter.

3.3 The End Goal and Demonstrating Progress

The Reporter is providing assurance that:

- The information underpinning key decisions can be relied upon as conforming to defined standards and levels of quality;
- The process by which this information is compiled is robust.

Where this is not the case, the Reporter intends to work with the industry on agreeing and implementing practical improvements that bring real benefits.

In order to achieve these ends, The Reporter is planning to assess individual datasets against the standards agreed in the Data Code of best practice for data management and data quality. This will enable a scorecard to be produced alongside any recommendations for improved performance. This is illustrated in generic terms in Figure 2.

Thus, the Data Code becomes the repository of best practice standards and the reports of individual assessments. These will show the current and previous status of the dataset under review alongside any recommendations, a benchmark can also be incorporated. This will allow the industry to demonstrate overall improvements in performance over time by aggregating the scores and summarising the improvements and recommendations in comparison with past results. As shown in Figure 2.

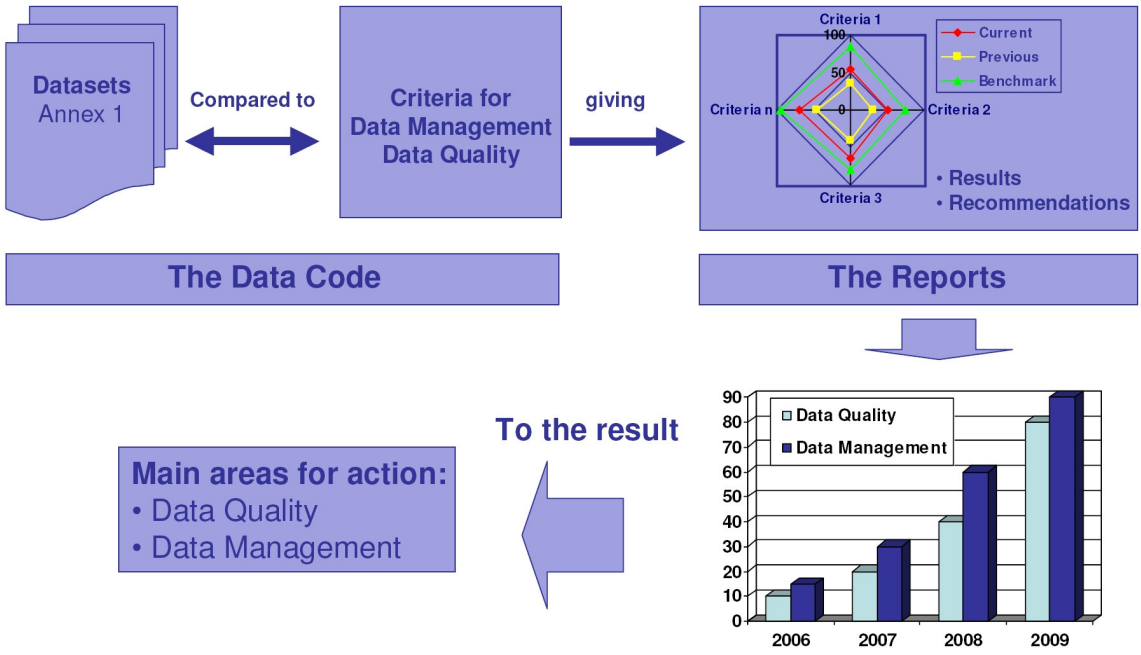

Figure 2 : Reporting Overall Dataset Performance

3.4 The Reporter's Overall Plan

The overall plan has been designed to engage the main participants in the industry in the common purpose of enhancing the quality of information and data for decision making and to achieve the end goal previously described. Therefore, the four stages envisaged not only include a major industry consultation but also the active oversight of the RSSG. This is described in detail in the next section of the Report. The four stages envisaged are:

1. Initial analysis of Network Rail systems, initial prioritisation of industry information needs and datasets, and proposals for assessing the datasets using the Data Code;
2. Consultation with the industry on priorities, assessment and the Data Code and the publication of the feedback from the consultation;
3. Final agreement, taking account of industry views, of the Data Code and the order of assessment of datasets through the RSSG;
4. Field trialling of the assessment process and Data Code with the industry followed by the implementation of the assessment programme of agreed priority datasets and reporting of results and recommendations.

This is illustrated in Figure 3.

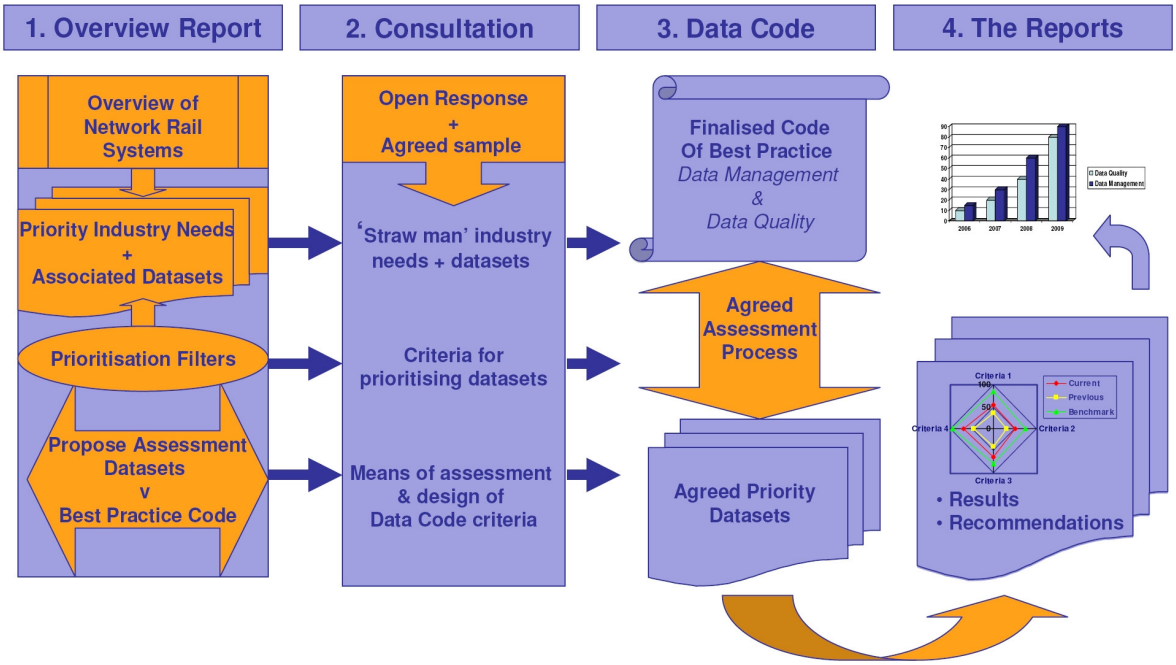

Figure 3 : The Reporter's Four Stage Plan

3.5 The Overview Report

The Overview Report completes the main work from the first stage of The Reporter's overall plan. The purpose of the Report is to provide an industry consultation document which:

1. Gives an overview of the Network Rail systems (and associated processes and data flows) that provide information for industry wide decision making;
2. Proposes a prioritised list of industry needs related to key industry decisions and the criteria from which the highest priority datasets can be identified for assessment;
3. Proposes how the final agreed list of datasets will be assessed by The Reporter against the Data Code.

The Report has been developed with the active involvement of ORR, Network Rail and with representatives from the TOCs and FOCs and other industry parties. This has been important in ensuring that industry needs are met, that the challenges facing the industry are recognised and that the current work going on in the industry in pursuit of improved decision making and information is acknowledged. It has also helped in taking account of

sensitivities around confidential information and the complexity of the various relationships.

In this context, it was important to look at the information and data management issues from a number of angles to get as clear a view as possible and a set of recommendations on the way forward that were credible and achievable. The main activity streams to produce the Overview Report were undertaken from three angles:

- A 'bottom up' analysis of Network Rail systems taking account of associated rules, so as to get a detailed understanding of industry functional processes and information flows, their impact on critical decisions and the barriers to effective information and data management in the industry;
- A 'top down' review of industry decision making and information needs including legal and regulatory requirements as a way of prioritising the large number and diversity of datasets;
- An 'in out' study of external approaches to assessment and best practice in data management, considering available government, commercial and industry standards and their appropriateness to the industry; this was the basis for the design of the assessment process and the Data Code.

This approach is illustrated in Figure 4.

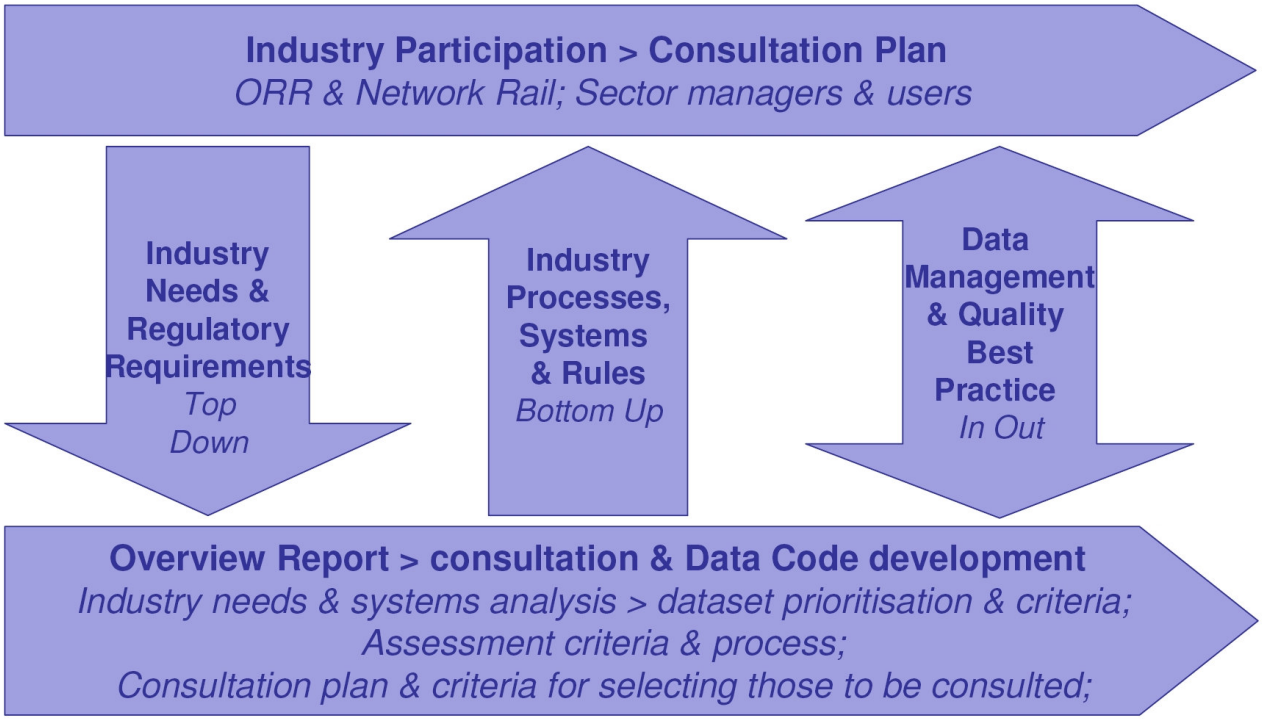

Figure 4 : Approach to Developing the Overview Report

The outputs from these activity streams are presented in the next section of the Report along with the approach to the consultation stage and the specific feedback that is being sought from the industry.

4 Overview Report Outputs

4.1 Introduction

The work of The Reporter has produced the following outputs that are described in full in the next section:

- Overview of Network Rail systems;
- The identification and prioritisation of datasets;
- The dataset assessment process;
- The development of the Data Code;
- The industry consultation plan.

The work on the assessment process and the Data Code has taken account of external approaches, methodologies, best practice and how these can best be applied to the specific industry needs.

The work alongside, and following, the initial Network Rail systems analysis was broken down into the following tasks:

- Engagement of Network Rail to identify any existing information or analysis, and to review this in order to avoid duplication of effort;
- Engagement with other industry participants to establish additional needs for information outside those of Network Rail:
  - Mapping of functional processes, systems interfaces and data flows across the industry starting with the pre-identified Network Rail systems and datasets;
  - Identification, description, recording and analysis of the datasets from which key industry information is derived and compiled;
  - Development of criteria for prioritisation of industry needs and supporting datasets;
- Review of data management and data quality standards in the light of the Network Rail systems analysis and through gathering known best practice standards;
- Review of approaches to assessment and designing an overall structure for the Data Code standards;
- Development of the consultation process to get the clearest possible view on the priority of industry needs, the criteria for finalising dataset priorities, the assessment process and design of and approach to the development of the Data Code.

4.2 Overview of Network Rail Systems

4.2.1 The Scope and Purpose of the Systems Analysis

This section covers the results and conclusions of The Reporter's initial analysis of Network Rail systems and underlying datasets. It is presented as important background to the consultation process, describing the process and systems overview and the nature of the challenges facing the industry in achieving robust and reliable information.

The rail industry relies upon a very large number of systems and data flows to support its operations. To facilitate understanding at a high level, The Reporter was asked to look at a set of pre-identified systems. These systems were viewed as being those that produce information that is used industry wide and therefore underpin the main operational processes of the industry. Additional systems were identified as significant repositories or generators of industry data and they are listed as newly identified systems. Both lists are shown in Annex A.

JPIP, BIFS, TRUST, TSDB and Trainplan were identified, as systems likely to be the most critical within the Network Rail information infrastructure. However, it transpires that some of these systems are due for imminent replacement (c.f. Annex E, Network Timetable (TSDB) and APlan).

The purpose of this analysis was for The Reporter to:

- Develop an overview of the industry's operational processes and associated systems and information and data flows;
- Make an initial assessment of the issues facing the industry in achieving enhanced reliability of information and data.

This work will subsequently inform the prioritisation of datasets, especially as this relates to operational process needs, and the approach to assessment and design of the Data Code.

4.2.2 Network Rail and Industry Operational Processes, Systems and Data Flows

The main operational processes, systems and data flows are shown in Figure 5 below. The main processes supported by these systems are operational planning, traffic monitoring, timetabling, billing and performance monitoring. The industry is therefore wholly reliant on the effective operation of these systems and there is a very high level of connectivity between the pre-identified systems. This level of dependency and connectivity is further illustrated in Figure 6 later in this document with the inclusion of the main output links to Government (DfT, Transport Scotland and Assembly for Wales), ORR and the TOCs and FOCs.

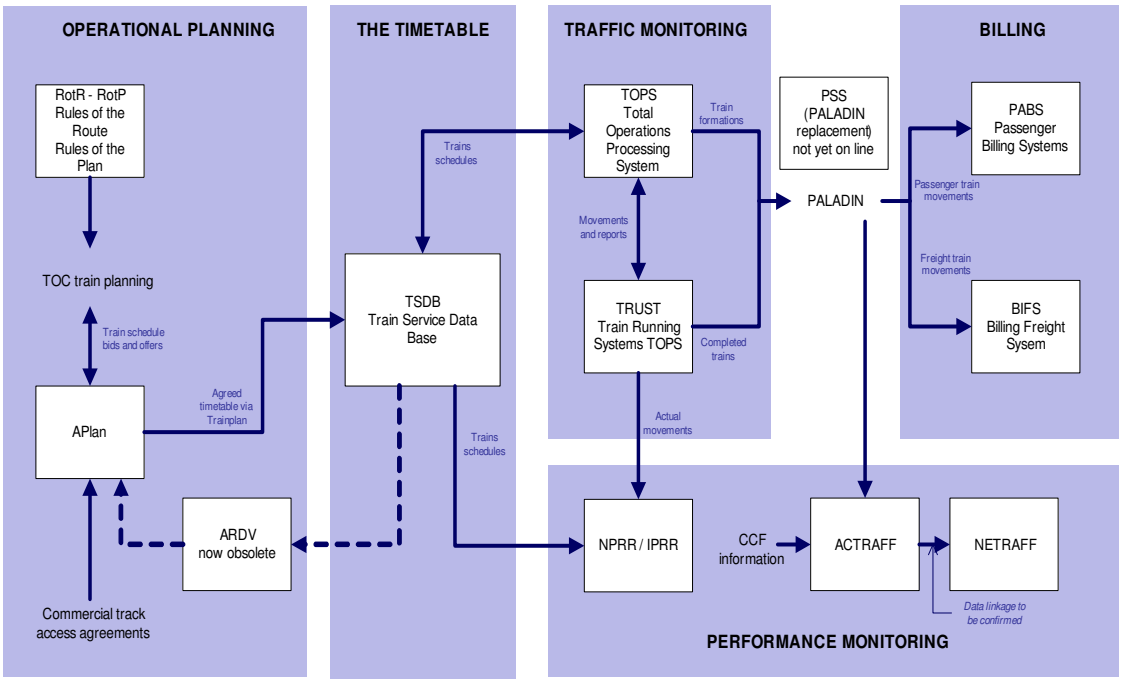

Figure 5: Network Rail Process and Systems Connectivity

In order to better understand, the levels of dependency as a precursor to dataset prioritisation, The Reporter also completed an analysis of the impacts of failure of these systems in relation to their connectivity and the results are shown in Annex B. What this particularly illustrates is the criticality of certain legacy systems within Networks Rail's infrastructure, notably those systems concerned with timetabling and performance against that timetable.

4.2.3 Initial Assessment of Network Rail Systems

This work consisted of a series of meetings with Network Rail systems owners and users on the function of the pre-identified system, its connectivity, its fitness for purpose, what information it processes, the perceived quality of data and any associated issues. Minutes of all meetings were taken and subsequently agreed with the owners. The detailed descriptions in Annex E are summaries of the findings agreed between The Reporter and the systems owners. The descriptions are not detailed analyses of the workings of each part of each system.

The Reporter's overall observations from this work are that:

- The Network Rail systems form a group of current and legacy distributed multi-platform, multi-environment entities with a number of different logical and physical information datasets;
- These systems largely do not have a single, defined and agreed interface or defined operational agreements with industry partners;
- A high level of manual intervention is required throughout Network Rail information systems;
- Multiple versions of the same logical information exist in more than one system and, in some instances, these do not correlate with each other;
- Multiple interfaces exist for the same information entering or leaving a given system; for example, the interface between the TOCs, who send back information to the Train Scheduling system (TSDB), do so in a variety of formats and with inconsistent data content;
- There is a lack of documentation available for some key systems within Network Rail. Some of these systems are both legacy and outsourced to a third party but detailed documentation should still be available.

These issues are widely recognised within Network Rail. They have developed a ten year Information Management Vision and have initiated a programme to replace TSDB and Aplan with a new Integrated Train Planning System (ITPS) based on a single, unique, rationalised database as previously mentioned. The Reporter endorses the intent of this Vision and the associated systems replacement strategy with its inherent standardisation and reduction in the number of environments.

The implications of these findings are that the most appropriate and beneficial way to tackle the assessment of datasets is initially by looking at data management issues. This approach will tackle the multiple interface issues and the standards for managing these as a priority and is the most likely to deliver the greatest and earliest benefits, given that the systems replacement programme will necessarily take time. It is also important that the new systems need to be brought within the scope of work, as they materially affect the quality of decision making within the industry into the future.

4.3 Identification and Prioritisation of Datasets

4.3.1 Objectives and Overall Approach

The objective of this part of the Reporter's work is to propose which datasets are selected for assessment from the enormous number within the rail industry so that the first assessments can swiftly target those areas likely to have the biggest impact on information for key industry decisions. The challenge was to find a way of doing this in the shortest possible time as the task of identifying and prioritising each one would be uneconomic and very time consuming.

It is also important that The Reporter establishes a transparent process for identifying and then prioritising these datasets so that the industry can be clear not only that the assessments are aimed at the highest priority datasets, but also that there is a logical and credible approach to the selection.

The Reporter has broken down the task of dealing with this large number of datasets into a series of manageable steps. The first job of identifying the most important datasets relied on the 'top-down' analysis of industry needs and the 'bottom up' systems analysis. For the prioritisation process, the Reporter has developed an overall 'filter mechanism'. This mechanism relates datasets directly to their importance in the main decisions taken within the industry by looking at the strategic (business), mandatory (legal and regulatory) and operational (functional processes) needs.

The main output for consultation from the Reporter's work on dataset identification and prioritisation to date is a 'straw man' (Annex D) of 'critical industry outputs' and their defined inputs and associated datasets, presented in priority order; these support the key industry decisions.

The consultation is specifically seeking feedback on the prioritisation of these 'critical industry outputs' and any modifications to or omissions from the associated datasets listed.

Once the industry has fed back its views of the top 'critical industry outputs' the Reporter will finalise the list of highest priority datasets to be included in Annex 1 of the Data Code. This will be done through the 'filter mechanism' to ensure that the process is logical, systematic, transparent and uses clear criteria for ordering datasets.

This section of The Report explains the proposals on the identification and prioritisation of datasets for consultation and how those proposals have been arrived at.

4.3.2 The Straw Man of Critical Industry Outputs

The Reporter is proposing a 'straw man' of 'critical industry outputs' for industry consultation as the means by which the highest priority datasets can be identified. It is shown in full in Annex D where the 'critical industry outputs' are listed from 1-24 in proposed priority order. The header line of this Annex is shown in Table 1 below in order to explain the content of the Annex and its importance in the prioritisation process; the item ranked 5 has simply been selected for purposes of illustration.

Rank	Owner	Critical Industry Outputs	Description	Defined Inputs	Regulatory / Legal requirement
5	Network Rail	PPM	PPM combines figures for punctuality and reliability into a single performance measure. It is measured in terms of % trains arriving within a defined time band.	Network Rail punctuality and cancellation data for each TOC. Data from each TOC on trains cancelled and planned.	Network Code

Table 1 : Header line from 'straw man' of 'critical industry outputs' (full version in Annex D)

The 'critical industry output' is a set of information, normally a major industry document, that is produced to communicate the status of some major part of the industry's strategy and operations and is the basis on which main decisions are made by the major industry participants. The 'critical industry outputs' are listed in the third column. Below are the explanations of each column reading across from the left.

1. 'Rank' for the priority proposed by The Reporter for the 'critical industry output.
2. 'Owner' every critical industry output has a single owner responsible for its production and maintenance; this is currently represented by the organisation rather than the specific person in order to keep the information within the table at an appropriately high level. It will not be until The Reporter assesses the datasets that individuals will need to be identified.
3. 'Critical Industry Output' is the name given to the identified output.
4. 'Description.' A brief explanation of the output and what it is for.
5. 'Defined Inputs.' These are the main sets of information and documents that go into producing the critical industry output and on which the Reporter will be seeking feedback to ensure there are no omissions or errors.
6. 'Regulatory/Legal Requirement.' These are the mandatory requirements that largely drive the need for the production of the output.

The datasets can be derived directly from the defined inputs, the defined inputs can be either datasets or, as in the above example, classifications of datasets. It has been represented in this way for presentation purposes, listing all datasets would result in a very large document and would add little value at this stage because the work has focussed on critical industry outputs rather than datasets. Equally, when presenting the information in Annex 2 of the Data Code it might be appropriate to group the datasets for presentation purposes and to aid understanding but for assessment purposes the individual datasets will need to be considered. As a way of illustration the datasets contained within the above defined inputs are:

- Network Rail punctuality and cancellation data for each TOC – the 6 datasets within this are listed below:

1. P0606iac All trains cancellation 2. P0606iap All trains punctuality	3. P0606iic ONE Intercity cancellations 4. P06006iip ONE Intercity punctuality	5. P0606ipc Peak cancellations 6. P0606ipp Peak Punctuality

- Data from each TOC on trains cancelled and planned – The 56 datasets within this are listed below:

1. Arriva Trains Wales – Trains Planned 2. c2c – Trains Planned 3. Central Trains – Trains Planned 4. Chiltern Railways – Trains Planned 5. Enterprise – Trains Planned 6. Eurostar – Trains Planned 7. First Capital Connect – Trains Planned 8. First Great Western – Trains Planned 9. First ScotRail – Trains Planned 10. First TransPennine Express – Trains Planned 11. Gatwick Express – Trains Planned 12. Grand Central Trains – Trains Planned 13. Great North Eastern Railway (GNER) – Trains Planned 14. Heathrow Connect – Trains Planned 15. Heathrow Express – Trains Planned 16. Hull Trains – Trains Planned 17. Island Line – Trains Planned 18. Merseyrail Electrics – Trains Planned 19. Midland Mainline – Trains Planned	20. Northern Rail – Trains Planned 21. Northern Ireland Railways – Trains Planned 22. 'one' – Trains Planned 23. Silverlink – Trains Planned 24. Southern – Trains Planned 25. Southeastern – Trains Planned 26. South West Trains – Trains Planned 27. Virgin Trains CrossCountry – Trains Planned 28. Virgin Trains West Coast – Trains Planned 29. Arriva Trains Wales – Trains Cancelled 30. c2c1 – Trains Cancelled 31. Central Trains – Trains Cancelled 32. Chiltern Railways – Trains Cancelled 33. Enterprise – Trains Cancelled 34. Eurostar – Trains Cancelled 35. First Capital Connect – Trains Cancelled 36. First Great Western – Trains Cancelled 37. First ScotRail – Trains Cancelled 38. First TransPennine Express – Trains Cancelled	39. Gatwick Express – Trains Cancelled 40. Grand Central Trains – Trains Cancelled 41. Great North Eastern Railway (GNER) – Trains Cancelled 42. Heathrow Connect – Trains Cancelled 43. Heathrow Express – Trains Cancelled 44. Hull Trains – Trains Cancelled 45. Island Line – Trains Cancelled 46. Merseyrail Electrics – Trains Cancelled 47. Midland Mainline – Trains Cancelled 48. Northern Rail – Trains Cancelled 49. Northern Ireland Railways – Trains Cancelled 50. 'one' – Trains Cancelled 51. Silverlink – Trains Cancelled 52. Southern – Trains Cancelled 53. Southeastern – Trains Cancelled 54. South West Trains – Trains Cancelled 55. Virgin Trains CrossCountry – Trains Cancelled 56. Virgin Trains West Coast – Trains Cancelled

4.3.3 Development of the 'straw man' of 'critical industry outputs'

The method for producing the list of critical industry outputs was through consultation with industry experts and scrutiny of regulatory and legal documents that govern the relationships between the main industry parties.

It became clear from these discussions and the review of key documents that these outputs are pivotal in the decisions that most affect the running of the industry. The production of these critical industry outputs is subject to well-documented and prescribed processes driven by the various regulatory requirements and legal agreements. Inputs described in these documents will be used to derive datasets that will be included in Annex 1.

4.3.4 Identifying Datasets from the Critical Industry Outputs

In order to get a more detailed analysis of the way in which these critical industry outputs were assembled and to identify the constituent datasets, it was necessary to understand the flow of information that supports the production of the outputs.

The 'top down' analysis of the industry produced the industry information flow needs (Figure 6). Analysis of this throws valuable light on the critical information flows that the industry relies upon to facilitate effective and efficient decision making through the 'critical industry outputs'.

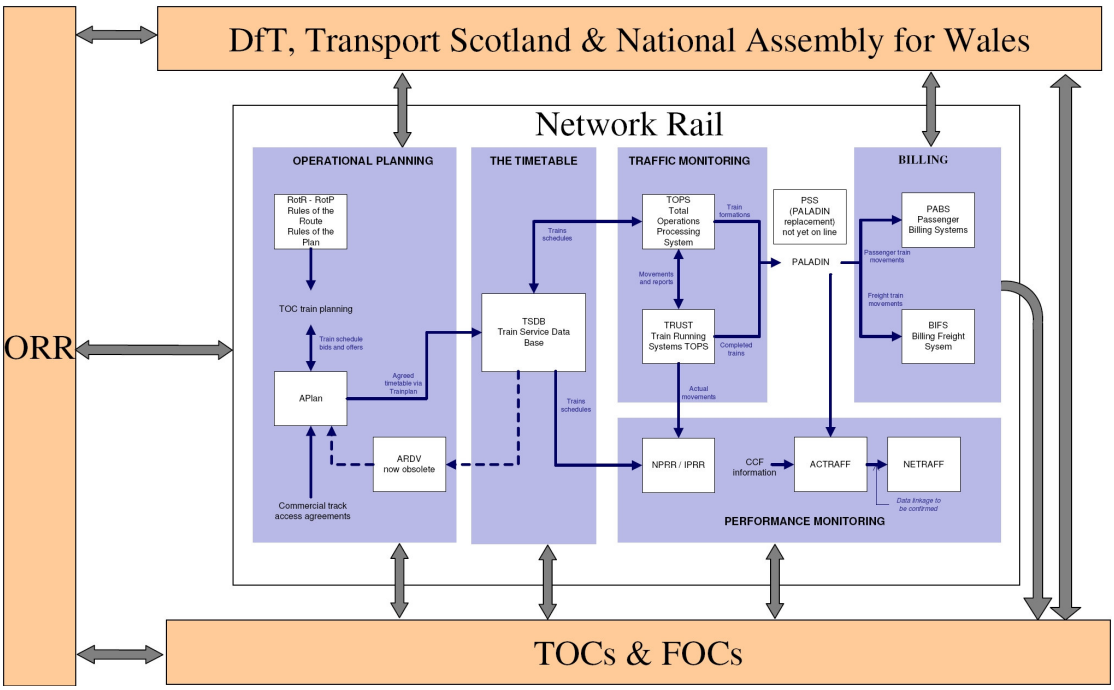

Figure 6 : High Level Cross Industry Information Flows

The datasets were identified and verified using the 'bottom up' systems analysis showing data flow through the industry and also by referencing the industry produced best practice guides on producing the relevant 'critical industry outputs.'

4.3.5 Prioritising the Datasets: The Filter Mechanism

The ultimate goal for this section of work is to obtain a prioritised list of datasets. Given that The Reporter has identified a proposed list of datasets developed from the defined inputs to the 'critical industry outputs', there is a need for a mechanism to decide on the highest priority datasets from the still very large number i.e. those that most affect the 'critical industry outputs' and are therefore the best place for The Reporter to begin assessment.

Thus, The Reporter developed what can be described as a 'filter mechanism' to analyse and ultimately list the datasets in a priority order for inclusion in Annex 1 of the Data Code.

There are a number of separate steps in the 'filter mechanism' as shown in Figure 7; each filter has been given a number shown in the right hand edge of the diagram to correspond with the explanation below.

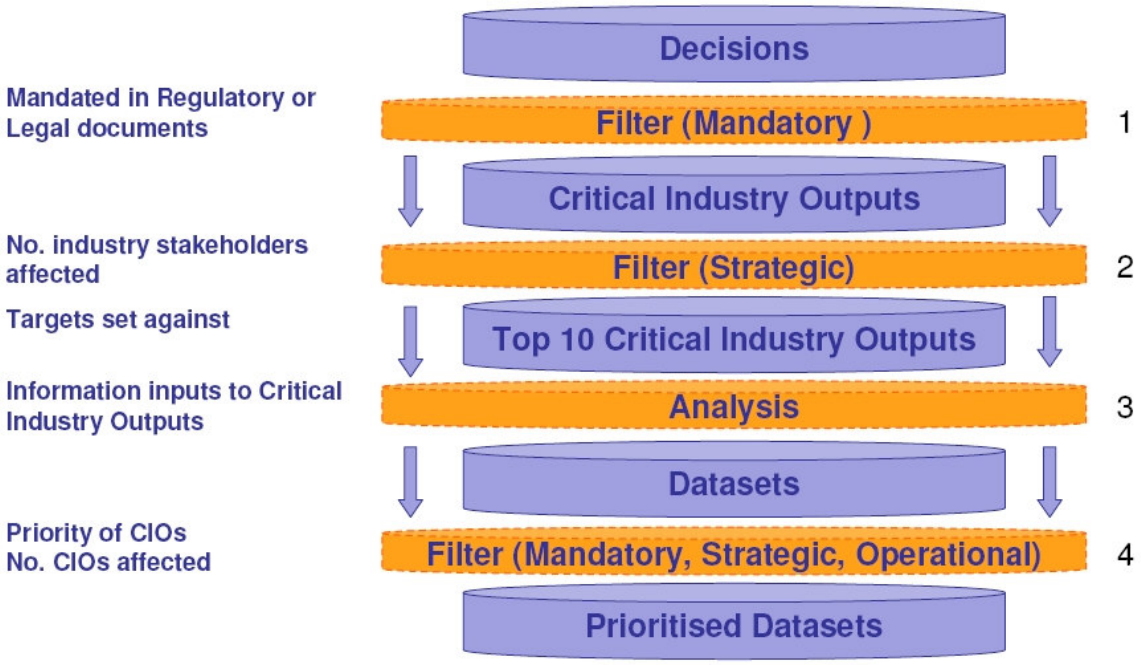

Figure 7 : Filter Mechanisms

These filters can be described as follows:

1. Mandatory legal and regulatory requirements drive out the main strategic documents for high level policy and funding decisions; this results in a list of 'critical industry outputs' such as the HLOS in both England & Wales and Scotland as described earlier in this section;
2. The top ten 'critical industry outputs' are proposed from the total list of 22 by analysing strategic issues, including the number of different industry players affected by the decisions and the extent of target setting against these outputs;
3. The top ten 'critical industry outputs' are then analysed to show the main information inputs to each and the datasets that underpin those;
4. The priority of each dataset can then be assessed by considering the importance and number of critical industry outputs it impacts.

In the final stage of the 'filter mechanism', further analysis is needed to establish the priority order in which the datasets are listed. This is because many datasets support more than one key industry need or output, and some outputs are themselves inputs, raising the importance of the datasets underpinning them. Thus, the criteria at this stage become more granular and include the strategic links with 'critical industry outputs', the mandatory requirements under the regulatory and legal documents and the operational impacts upon the main industry functional processes.

The detailed criteria used for prioritisation in steps 2 and 4 of the 'filter mechanism' above are described in Annex C.

4.3.6 Consultation

The Reporter plans to invite a selected group of key industry participants to comment on: the content of the 'straw man'; the 'critical industry outputs' they feel should be in the 'top 10'; their views on prioritisation of the 'top 10'; and the accuracy of the list of identified inputs. Whilst The Reporter has developed a systematic approach to this, the participants in the consultation are being asked to give it a 'reality check' and to suggest their own criteria if they think it appropriate. The consultation plan can be found in Section 4.6.

4.4 Dataset Assessment Process

4.4.1 The Objectives of the Dataset Assessment Process

As explained in the Overall Plan (section 3.3), The Reporter is providing assurance that:

- The information underpinning key decisions can be relied upon as conforming to defined standards and levels of quality;
- The process by which this information is compiled is robust.

Where this is not the case, The Reporter intends to work with the industry on agreeing and implementing practical improvements that bring real benefits.

As previously noted (section 2.3) the intent is that datasets listed as priorities for assessment in Annex 1 of the Data Code will be moved to Annex 2 after assessment. The report of each assessment and a report of aggregated results will be published on ORR's web site so that the industry has visibility of the status of its data and of actions to improve data management and quality.

This section describes how The Reporter proposes to carry out these assessments using the Data Code. This includes the overall approach, methods and outputs. The design of the Data Code of best practice is described in more detail in section 4.5.

4.4.2 Approach to Dataset Assessment

The Reporter is well aware that, for the successful implementation of the assessment programme and the promulgation of the Data Code best practice across the industry as a whole, the dataset owners need to contribute in the form of the availability of resources and information sought by The Reporter. To that end, the assessment process proposed is intended to actively engage the dataset owners in applying the Data Code to the management of particular datasets and the achievement of data quality. The assessment process and reports will also be overseen by the RSSG in support of the process.

The overall dataset assessment process is described below in Figure 8.

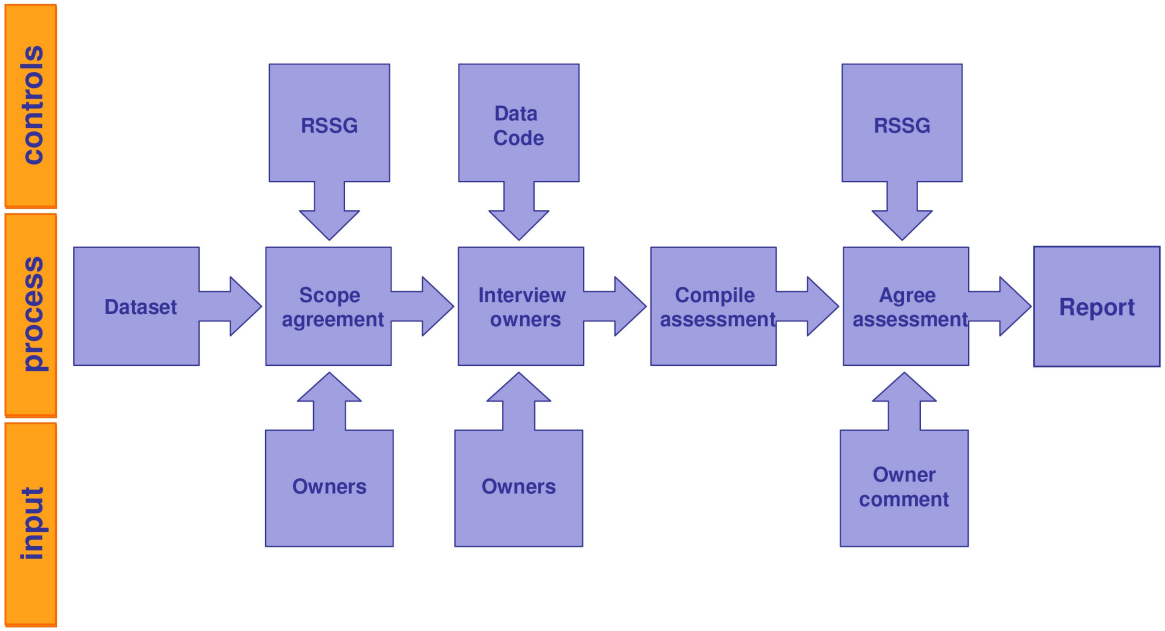

Figure 8 : Dataset Assessment Process

Thus, The Reporter is proposing that the owners will be party to the agreement of the scope, interviewed during the assessment and given the opportunity to comment on the assessment. The RSSG will have oversight of the scope, assessment process and assessment report.

It is proposed that dataset assurance is approached by the use of managed interview templates and questionnaires defined within the Data Code, covering data management and data quality as separate disciplines. Thus the complete assessment for any dataset will be made up of the sum of two parts undertaken as parallel assessments brought together at the end for the completion of the report.

Responses to the interviews and questionnaires will be reviewed by The Reporter in consultation with the dataset owner and users and recommendations made based on the findings. Overall rankings will be given for each data management and data quality element in the Data Code. In each case the score is awarded by The Reporter, not the dataset owner though their comments will be taken into account. A comprehensive ranking for the dataset will be constructed once the assessments for both data management and data quality have been completed and recommendations made.

The exact scoring system is yet to be agreed and must relate closely to the detailed design of the Data Code of best practice. The Reporter is considering several options for scoring systems:

1. Binary based systems, though providing very strong indications of exact status, do not reflect well in circumstances where perfectly effective workarounds have been implemented to solve problems;
2. Scoring systems incorporating large sliding scales (score 1 to 10) that give much greater granularity but often add complexity and can even mask true situations from the reader when aggregated;

3. Scoring systems incorporating three or five possible responses give some granularity and the ability to encompass a range of situations; this is currently The Reporter's preferred option.

The Reporter also feels it is important to identify improving or deteriorating positions where they exist, thus intends to provide a score 'as is' with a score for anticipated future position provided certain actions are taken.

4.4.3 Design of the Dataset Assessment

The Reporter has been seeking a clear and cost effective way of designing the assessment process using the Data Code. The proposal is based on the concept of a dataset as data that has been stored or modified and passed between one or more 'systems' (whether they are electronic, paper, models or processes) to give a particular output for decision making. This is illustrated schematically in Figure 9, and highlights these facets of a dataset:

1. The data from which it is derived had its source as 'raw data' at some point;
2. The data has a describable journey through one or more 'systems' from raw data to industry output;
3. A dataset is the data existing on a part of that journey and can be defined as a collection of data passed between two systems.

Each 'system' from which a dataset is passed requires a level of 'data management', and inherent within that system is a level of 'data quality' (shown in the figure as 'data management layer' and 'data quality layer').

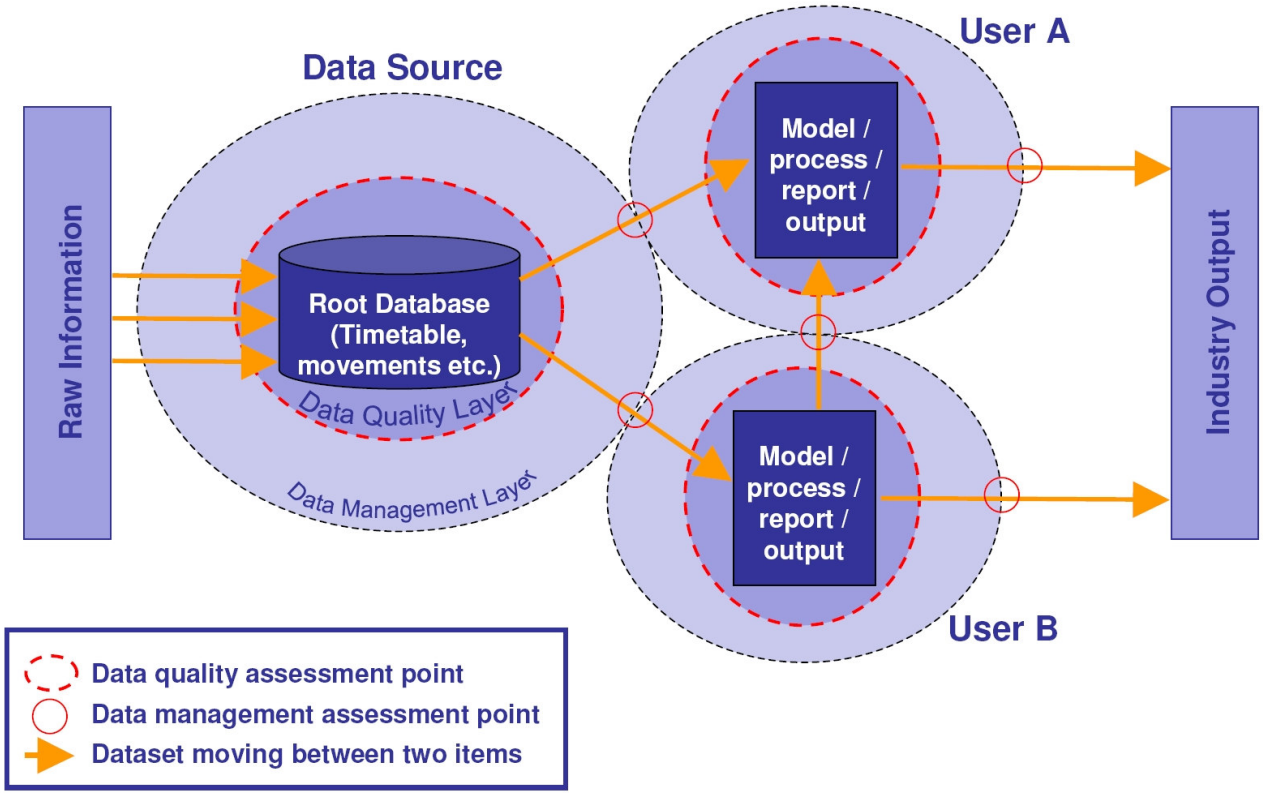

Figure 9 : Dataset Assessment Points

This means that the assessment of any dataset needs to consider the following:

1. The definition of the dataset and its journey to industry output;
2. The data management in place for the dataset as it passes through;
3. The quality of each 'system' through which the data passes.

The point at which the dataset passes from one 'system' to the next is the point at which to asses the data management of the dataset.

Two options were considered in designing the assessment process:

1. Assess the management along the data journey and the quality identified in each system that the data passes through; or
2. Assess each system considering the management and quality characteristics of the complete system.

The initial work undertaken by The Reporter of the systems within Network Rail has identified that there are often multiple interfaces to any particular system. If an assessment of a single system were undertaken, it would result in having to consider many interfaces. This is not an effective method of study as many interfaces will feed information of little consequence within the larger decision making framework.. In addition, several systems were identified as having tens of interfaces. This is shown diagrammatically in Figure 10 below.

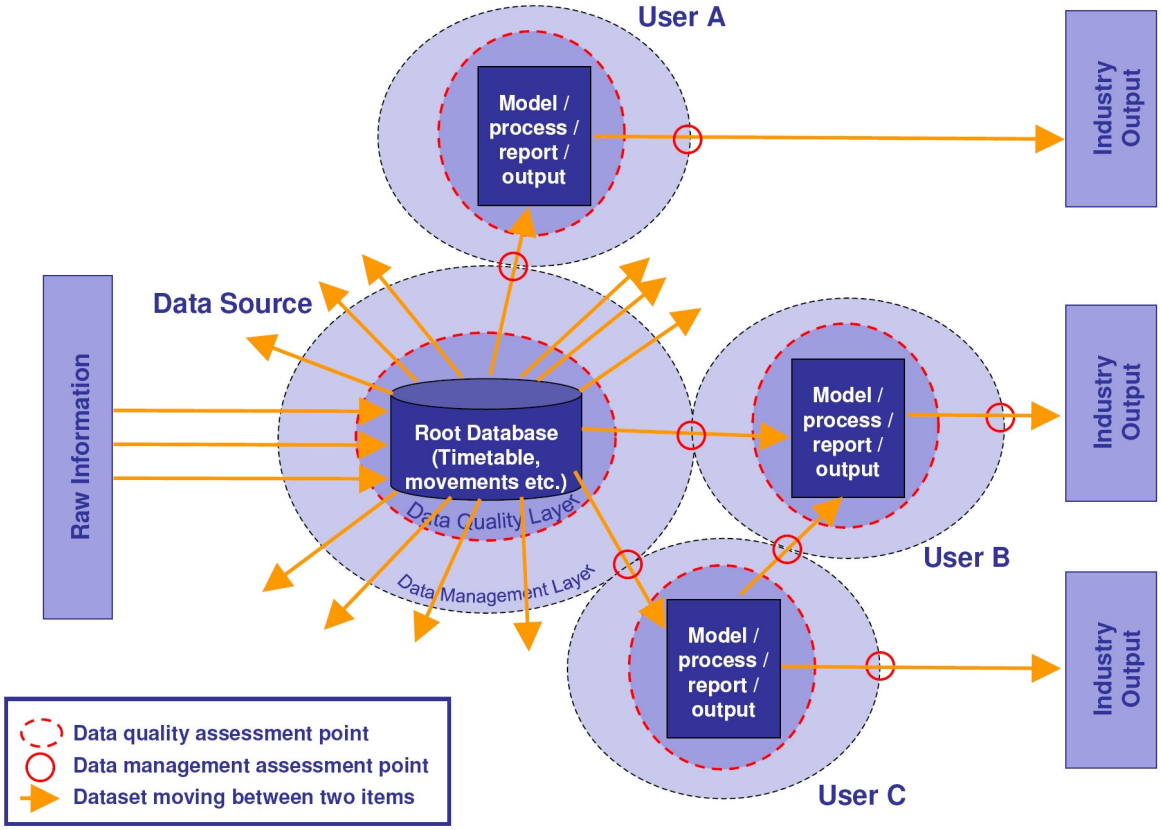

Figure 10 : Assessment prioritisation of systems with multiple data connections

Therefore The Reporter will consider the dataset journey for each dataset and not perform assessments on a system by system basis.

Data management assessment will focus only upon the interfaces encountered by a single dataset. Data quality will be assessed as the quality of the system through which that dataset originated. This will target the work to specific dataset extract points within a system rather than at the system as a whole.

The Reporter will use a pragmatic approach in the planning of assessments to ensure that the maximum value can be delivered cost effectively. Thus, where a system has a standardised data management process for extracting data, which supplies more than one dataset, a single management assessment will be undertaken for all datasets derived from that system. Similarly, data quality assessment for a system will be utilised where appropriate for all datasets derived from that system. This means that the volume of the work will continue to be on data management.

4.4.4 Outputs of the Process

The Reporter proposes that a report on each dataset assessed is compiled from the rankings within the assessments. The report, as entered into Annex 2 of the Data Code, should consist of:

- Dashboard summary page for the dataset:
  - Summary description of dataset:
  - Data management;
  - Data quality;
- A description and overview of the Dataset itself:
  - Dataset Lifecycle (Source of data through to retirement of data);
  - Main uses;
  - Inputs / Outputs;
  - Dataset Ownership;
  - Access methods and availability;
- Data Management:
  - Data management findings summary: conclusions and recommendations;
  - Data management detailed responses, Reporter comments and additional information (future developments, known problems, unique attributes etc.);
- Data Quality:
  - Data quality findings summary: conclusions and recommendations;
  - Data quality detailed responses, Reporter comments and additional information (future developments, known problems, unique attributes etc.).

The overall structure of the dataset report is shown in Figure 11 below. It is anticipated that all reporting will be accessible electronically through ORR's web site. Work is ongoing in this area as the detailed Data Code design and reporting requirements are finalised.

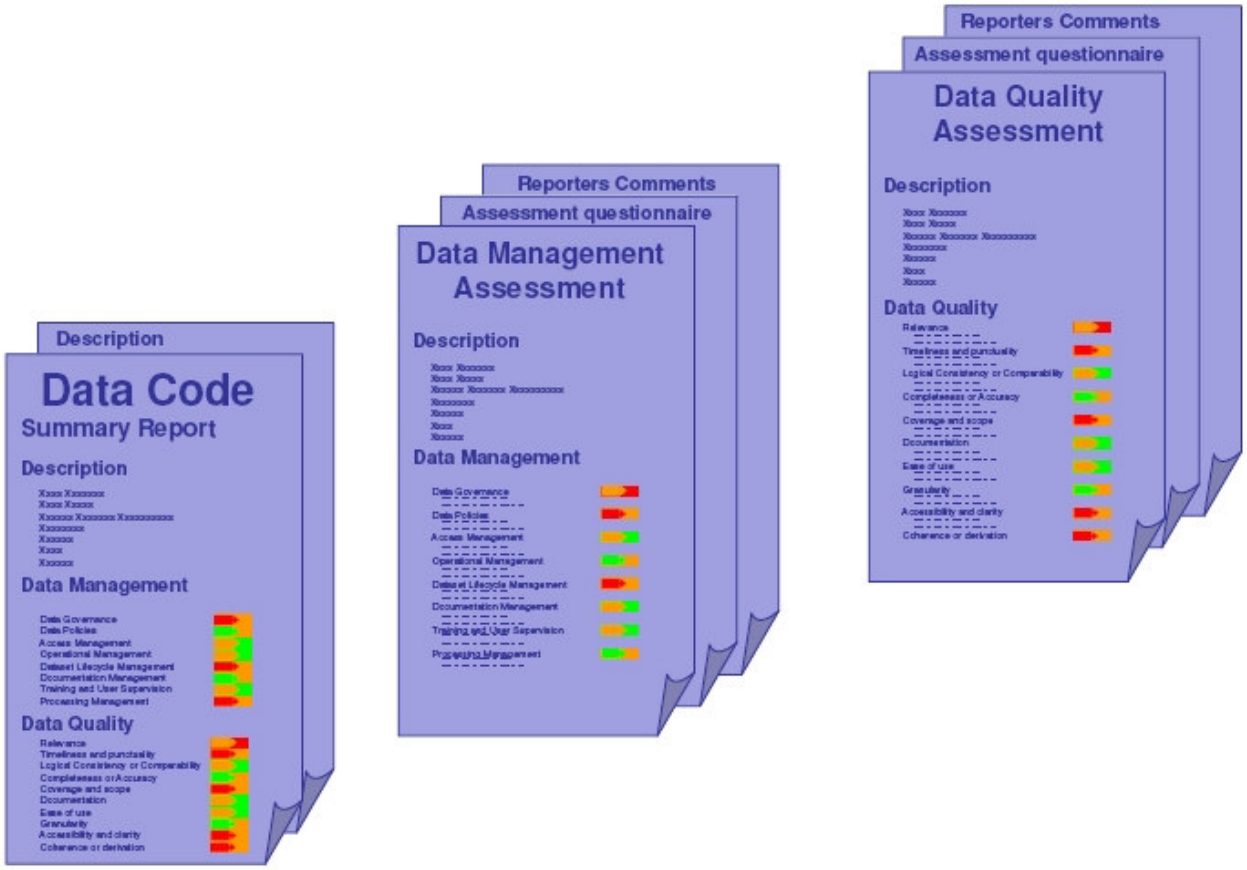

Figure 11 : Data Code Assessment Report Construction

4.5 The Data Code

4.5.1 Data Code Purpose and Benefits

The purpose of the Data Code is to provide a framework of best practice against which to assess datasets within the rail industry to establish their reliability.

The Data Code will encapsulate standards of best practice for the rail industry in data management and data quality and provide the structure and controls for the continuous assessment of information. The Data Code will also become the repository of the assessment questionnaires used as a basis for the assessment of datasets. There will be two annexes. Datasets identified for assessment, once defined and prioritised, will be listed in Annex 1 of the Data Code. Datasets that have been assessed against the Data Code will be listed in Annex 2 with their assessments.

Implementation of the Data Code will, over time, provide the industry with a number of benefits:

- Assurance of the reliability of a given dataset;
- A visible and usable baseline of industry dataset standards and definitions;
- A goal for improving industry dataset standards;
- Visible improvements in dataset reliability;
- Visible evidence that management of the datasets is in place.

4.5.2 Capturing Best Practice

International and UK best practice has been closely scrutinised for the structuring of the Data Code. The Reporter recognises that Network Rail's internal audit function targets British Standards (BS 7799) as best practice within Network Rail (BS 7799 has been superseded by ISO 27001 and Network Rail have adopted this for information systems management and control).

The Data Code is being developed from best practice identified within several sources of information including ISO 27001. The key elements of the best practice used in the development of this Data Code are listed in Table 2 below.

Title	Source
Information technology — Reference Model of Data Management	ISO/IEC TR 10032:2003(E)
Information technology — Security techniques — Information security management systems — Requirements	ISO/IEC 27001:2005(E)
The Principles of good practice for information management	BSI
The Principles of Good Data Management	IGGI
Protocol on Data Management, Documentation and Preservation	National Statistics
Standards for Information Systems Auditing	Standards Board of ISACA®
IS Auditing Standard Audit Charter	Standards Board of ISACA®
Quality management systems —Fundamentals and vocabulary	ISO/IEC 9000:2005(E)
Guidelines for quality and/or environmental management systems auditing	ISO/IEC 19011:2002(E)

Table 2 : Sources of Data Quality and Management Best Practice

The Data Code standards and any subsequent changes and development will be overseen by the RSSG. This oversight function is important as The Reporter needs the Data Code to remain constant throughout a given 'programme' to ensure consistency of assessment and comparability of results. Any updates to the Data Code should therefore be made only between 'programmes' and in a way that protects comparability and consistency whilst recognising industry needs and upgrades in best practice. The overall and ongoing development of the Data Code is described in Figure 12 below.

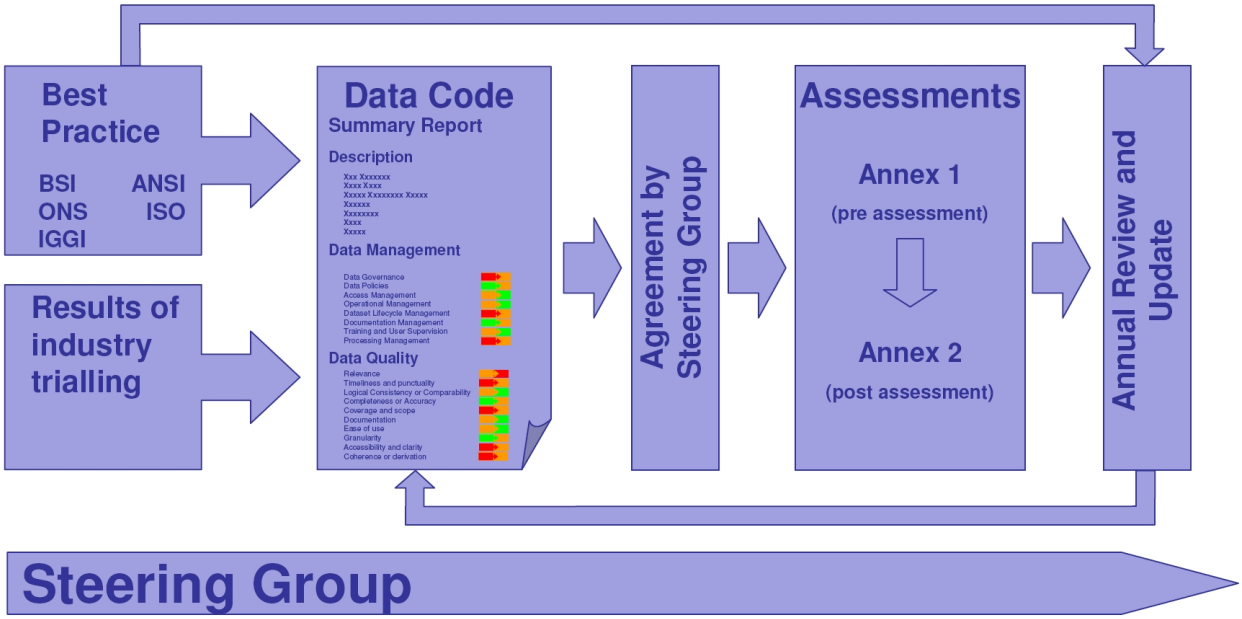

Figure 12 : Ongoing Development of the Data Code

4.5.3 The Rail Statistics Steering Group (RSSG) Within the Data Code Development

The Reporter proposes that the RSSG own and manage the Data Code. Membership of the steering group is set out on ORR's website. In relation to the Data Code, the role of the RSSG is as follows:

- Oversee development of the Data Code;
- Overseeing any changes to the Data Code;
- Agreement with The Reporter of the contents of Annex 1;
- Ownership of the schedule of assessment.

The terms and scope of the Data Code will be subject to ongoing review. An annual review will be carried out by the RSSG and The Reporter who, together, will agree changes to the Data Code.

4.5.4 Data Code Principles and Standards

The rational for the development of the Data Code are:

- Establishment of a baseline of data reliability in the support of decision making within the industry;
- Identification and dissemination of best practice in data reliability will be in the form of data management (levels of control) and data quality (levels of accuracy) across the industry;
- Continuous monitoring of reliability over time to ascertain improvements in data standards;
- Construction of a central record base for all information available within the industry under the data code outlining its uses and reliability.

The Reporter is proposing that data in the form of datasets will be assessed against two sets of standards derived from current best practice: data management standards and data quality standards.

The Reporter suggests that data management standards and associated criteria for the assessments should be, subject to discussion, applied across the industry and will cover the following:

- Roles, responsibilities and governance;
- Dataset policies;
- Organisation of information;
- Physical and environmental;
- Communications and operations management;
- Access control;
- Information systems acquisition, development and maintenance;
- Reporting information events and weaknesses;
- Business continuity management;
- Compliance.

The Reporter also proposes that data quality standards and associated assessment criteria should be consistent across data types and will cover:

- Relevance;
- Timeliness and punctuality;
- Logical consistency or comparability;
- Completeness or accuracy;
- Coverage and scope;
- Documentation;
- Ease of use;
- Granularity;

- Accessibility and clarity;
- Coherence or derivation.

The code will not alter arrangements for access to confidential data. It will however highlight where data is available.

As a result of the application of the Data Code, it will be possible to publish dataset reliability rankings with the definitions, assessments and corresponding findings for any particular dataset.

4.5.5 Implementation

Implementation of The Data Code will be undertaken in three stages:

- Initial trialling of the Data Code and assessment methodology;
- Application of dataset management assessments;
- Application of the dataset quality assessments.

The trialling of the Data Code standards of best practice through the assessment process will be overseen by the RSSG and The Reporter will produce a report on the effectiveness of the process and the application of the standards with any recommendations as to how they should be modified for the main programme.

Once trialling has been completed successfully, data management assessments will be undertaken on the highest priority datasets so that the greatest and earliest benefits can be gained. The results from these assessments will be incorporated into the priority list for the future and the quality assessments will begin in the order defined by the reviewed priority list.

Whilst the programme of assessments will be carried out to a pre-agreed plan overseen by the RSSG, The Reporter recognises that the assessment programme needs to be adaptable to allow re-assessment of a given dataset or the inclusion of a new priority to meet industry needs. This will help ensure that maximum benefits continue to be gained from the ongoing programme.

4.5.6 Data Code Development

The Data code is currently under development by The Reporter

The main consideration is to provide a framework within which the industry can operate and which recognises the different ways in which different organisations achieve appropriate data management and data quality. The ongoing design and management of the Data Code is expected to develop and be refined as the industry itself successfully improves the management and control of information and datasets over time.

It makes sense for this best practice to be based on a standard such as a generalised Quality Management System (QMS) (ISO/IEC 9000). The Reporter would therefore expect to see the elements of a QMS related to data management and data quality. Nevertheless, The Reporter will also recognise the different ways in which these elements can be in place.

Care is being taken throughout the development to make the terminology as appropriate and helpful as possible to the industry.

The Data Code will independently assess the dataset against best practice derived from within the documentation available from ISO, BSI, ONS, IGGI and ANSI or whatever other standard. Dataset quality components are highly dependent upon the specifications and definitions developed by the dataset owners themselves and thus rely upon a suitable level of control over the dataset being available in the first instance.

Considerable progress has been made towards the production of a working data code suitable for initial trialling within the industry.

4.6 The Industry Consultation

4.6.1 Objectives

The main objective of the consultation process is to obtain the agreement of the main industry participants on the prioritised list of 'critical industry outputs' identified in Annex D of the Report and to open discussion on the process and characteristics of the development of the Data Code. The specific questions that The Reporter is seeking answers to are:

1. Are there any identifiable exceptions or omissions within the initial prioritised list of 'critical industry outputs' presented within the Overview Report?
2. Do you agree that the top ten 'critical industry outputs' within the initial list are the top ten and if not then why not?
3. Within the top ten identified critical industry outputs in the initial list are there any that you consider to be in the incorrect position?

The agreed list of 'critical industry outputs' will be the basis of the identification and prioritisation of key datasets within the industry in support of the development of an industry wide Data Code.

This section of the report presents The Reporter's proposals regarding the format, contents and the outputs of the consultation process.

It is expected that the benefits amongst the key industry players will be:

- A general consensus on the need to establish standards of data management and quality assessment and the benefits of doing so;
- Increased awareness and agreement within the industry of issues related to data management and quality;
- Agreement of the top industry needs and the way of proceeding to prioritise and assess datasets.

4.6.2 Outcomes

An important part of the consultation process is to ensure the involvement of the key industry players in the development of the prioritised list of 'critical industry outputs' and discussion of the Data Code. This will facilitate the increased awareness of the data control and management issues within the industry.

The Reporter has identified and proposed the top twenty two industry needs (Annex D). As a result of the consultation, ten will be selected and prioritised. The Reporter can then apply the 'filter mechanism' and criteria described in section 4.3.5 of this report to derive a prioritised list of the key datasets underpinning the needs.

The outputs of the completed consultation process will therefore be:

- Prioritised list of critical industry outputs;
- Prioritised list of datasets underpinning those critical industry outputs.

4.6.3 Overall approach

The Reporter proposes that the consultation process will involve a group of carefully selected participants from within the industry. The Report will be published on ORR's web site and will elicit the views of the wider industry and invite comment for inclusion into the process. However, the main consultation will be with the selected group. This approach reflects the fact that the subject matter of the consultation is highly specific in nature.

The Reporter proposes that the participants be selected in accordance with the criteria defined below and confirmed by discussions between The Reporter, Network Rail and ORR.

4.6.4 Stages of the Consultation Process

It is proposed that the consultation process will be carried out in four main stages:

Stage 1. Identification of Participants:

a) Definition and agreement of participant selection criteria;
b) Selection and invitations to participants;

Stage 2. Consultation:

a) Distribution of the Overview Report that forms the basis of the consultation to the participants;
b) Individual facilitated consultations:
   - Structured discussions (interviews) on the identified critical industry outputs;
   - Identification of specific issues of importance for each of the participants;

Stage 3. Compilation, discussion and agreement of results;

a) Identification and verification of the datasets supporting the prioritised critical industry outputs;
b) Prioritisation of the order of study of the datasets by The Reporter in agreement with ORR and Network Rail;

Stage 4. Publication of the Results.

4.6.5 Identification of Participants

The Reporter proposes that the participants' group must not only represent a cross section of the industry, but must also represent different interests to reduce the impact of localised issues. The proposed criteria listed below reflect the specific nature of the subject matter of the consultation and will allow the selection of a relevant and balanced participant group. The first two criteria are considered to be pre-eminent, reflecting the impact of the Data Code upon the participant.

1. Parties representing major strategic recipients of data,
2. Parties representing major producers of data,
3. Parties representing national rail infrastructure,
4. Parties representing national policy setting and governance groups,
5. Parties representing major users of the network:
   a. Private users (passengers);
   b. Business users.

The list of proposed Participants that fits the criteria above are:

1. Network Rail;
2. TOCs;
3. FOCs;
4. ROSCOs;
5. DfT;
6. Transport Scotland;
7. Assembly for Wales;
8. PTEs;
9. ORR;
10. TfL;
11. Passenger Focus;
12. London Travelwatch.

4.6.6 Compilation and Agreement of Results

Following the individual discussions and any comments received following the publication of the Overview Report, the responses to the consultation questions can be assessed and compiled to form the consultation first stage results. These will be used in the dataset prioritisation process.

The datasets that support the critical industry output, in the form of the defined inputs (Annex D), will be identified by The Reporter. For example under the critical industry output 'PPM' one of the defined inputs is 'DAT Files' this in itself is not a dataset. The datasets are the six DAT files identified in section 4.3.2 which form that defined input.

A prioritisation process, outlined in section 4.3.5, will be applied to the list of datasets to produce an order of study for the datasets that underpin the prioritised critical industry outputs. These will be agreed between The Reporter, ORR and Network Rail.

5 Next Steps

These next steps follow the publication of The Report. the RSSG will oversee these activities and the outputs will be published on ORR's web site.

5.1 Consultation: Preparation, Implementation & Feedback

The Reporter plans to set up and run the consultation in three stages:

1. Preparation: distribution of The Report; detailed planning and setting up of the interviews and workshops; organising the support processes to gather, record and report feedback;
2. Implementation: running interviews and workshops and receiving any comments in writing;
3. Feedback: interpreting the workshop outputs, interviews and written comments; preparing modified proposals; and drafting and distributing feedback to the industry.

the RSSG will be presented with the feedback from the consultation, the final proposals on industry information and dataset priorities, and proposals on the assessment process for discussion and agreement. This will be encapsulated in the final version of the Data Code and Annex 1 listing the high priority datasets for assessment.

The Reporter plans to prepare a report summarising the feedback received and how this has been taken into account in the prioritisation and assessment process. This will be sent to all interested parties and published on ORR's web site.

5.2 Worked Examples of the Assessment Report

In order to clarify the assessment process for the interviews and workshops, The Reporter plans to produce two worked examples of assessment reports with full dataset descriptions. The report examples will include:

- An introduction giving the terms of reference, methodology, context and details about the specific assessment;
- An executive summary highlighting the main areas of achievement and/or concern;
- A main report, setting out details of what was found and giving recommendations for improving service overall;
- A worked scorecard showing performance against the criteria;

- A summary list of findings and recommendations including an overall scorecard;
- An outline example of the dataset dashboard.

5.3 Field Trials Assessment

The Reporter plans to complete two field trial assessments with main industry participants immediately after the consultation process is completed. These will be carried out as formal assessments following the full procedure. The purposes of the field trials are:

- To test the assessment process in action;
- To get feedback from the participants on its practicability and usefulness and how, if at all, this can be improved;
- To review the efficiency and effectiveness of the assessment process from an operational viewpoint and to make modifications as necessary;
- To report back on the field trials to the RSSG.

The datasets for assessment will be chosen from Annex 1 of the Data Code and will cover different functional processes and stakeholders to ensure as representative and realistic a trial as possible. Following the field trials, The Reporter will present a field trial review to the RSSG with proposals as to how the process will be modified in the light of experience. The review will be published on ORR's web site.

5.4 Full Roll Out of the Assessment Process

The Reporter will plan a full programme of assessments for the next agreed period and submit this for agreement to the RSSG, with reporting timescales and the date and format for completion of a summary report at the end of the first year's programme. The details of this programme will be published on ORR's web site.

6 Conclusions and Recommendations

The industry need for robust and accurate information for decision making is widely recognised. A number of initiatives are in place within ORR and Network Rail in particular to enhance the quality of data provided for legal, regulatory, strategic and operational purposes. Scott Wilson has been appointed to support the industry in enhancing the quality of its data: assessing the current state of play, helping to promulgate best practice by developing a Data Code and reviewing performance on datasets against that Code.

This section clarifies the observations from the work completed by The Reporter in it's initial review of Network Rail systems, in prioritising datasets, in designing the Data Code and in proposing the assessment mechanisms. From the insights and understanding gained from this work, the Reporter is able to provide a clear structure and recommendations on the way forward in fulfilling its role to the benefit of the industry as a whole.

The initial 'bottom up' analysis undertaken by The Reporter has provided an overview of Network Rail systems as they affect decision making across the industry at the various levels. The main observations from the analysis of the identified Network Rail systems have been:

- The multiplicity of systems environments;
- The fragmentation of data across different database environments;
- The inconsistencies in data management: varying interface standards, and differing data definitions;
- A high degree of manual intervention.

These problems largely arise because of the legacy systems which were built to meet earlier needs, and which are difficult to change with limited available documentation. As a result, Network Rail has developed a ten year Information Management Vision [1] and within that vision is re-developing its core operational planning system with an integrated and rationalised database. The Reporter endorses the intent of these initiatives.

The Reporter, as part of this initial systems analysis, has mapped the main industry functional processes and associated these with the identified systems and data flows. This demonstrates the high degree of dependency of the industry on the data held within, and passed between these systems, and the potential impact of their loss or failure.

The Reporter's completed initial review of Network Rail systems also highlights the lack of consistency in the way in which the data flows are managed. This leads The Reporter to conclude that priority should be given to data management standards and performance as a precursor to any sampling of the quality of particular data. This also helps accommodate known sensitivities around information confidentiality.

The 'top down' analysis undertaken by The Reporter recognises that the ultimate use of data is for decision making. The Reporter concluded that the complexity of the systems and database environments and the myriad of datasets held in these systems and passed across the industry, required the development of a number of transparent filters to

prioritise the datasets in relation to the strategic, mandatory and operational needs of the industry. This completed work ensures that the datasets to be assessed are those that will have the biggest and earliest impact on the quality of decision making.

The Reporter's initial study of data management standards from UK and international sources has informed the design of the Data Code. Thus the Reporter has concluded that a structure based on these standards will ensure that they are known to be widely accepted and underwritten, yet capable of adaptation to meet particular industry needs. The Reporter also concluded that the design of the assessment process, based on the Data Code standards, requires an approach that is cost effective, simple to run, easy to repeat and provides meaningful comparisons between datasets which, over time, will allow visible progress to be demonstrated. This design will give ORR, Network Rail and the rest of the industry the necessary assurances on data management and data quality and clarity on what is needed to improve performance.

From the outset, the Reporter recognised that there is a shared common purpose in improving the quality of information for decision making amongst the main industry participants. In supporting this common purpose, The Reporter concluded that it was critical to success that the industry was actively engaged in the development of the assessment programme and, in particular, the prioritisation of the datasets for assessment and input to the overall design and development of The Data Code. This led to the design of a consultation programme around the main industry players and their information needs and the active involvement of the RSSG established by ORR to oversee initiatives in this area with specific terms of reference in relation to the Reporter's work.

As a result of this work the Reporter recommends that, subject to industry consultation:

- The prioritisation process for agreeing Annex 1 datasets is based on the 'straw man' of critical industry outputs derived from strategic, mandatory and operational criteria and is then finalised as proposed;
- Newly identified core systems should be brought within the scope of The Reporter's work;
- The Data Code is derived from UK and international standards for data management and data quality in order to provide ORR, Network Rail and the rest of the rail industry with a powerful mechanism for promulgating best practice;
- The assessment process initially focuses on data management processes as a precursor of any sampling for data quality and delivers a scorecard and recommendations for improvement as the way to gain the earliest benefits and to encourage and demonstrate progress;
- The main industry participants continue to be actively engaged in the Reporter's work through the proposed consultation and the RSSG overseeing the work planned; this will ensure industry needs continue to be met and sensitivities recognised and accommodated;
- Changes to the Data Code standards must be properly managed and agreed from time to time so that consistency is maintained and comparative performance can be measured; this again should be managed through the RSSG;
- Field trials are carried out to verify the Data Code and assessment process prior to the roll out of the assessment programme with the oversight of the RSSG.

By implementing these recommendations, The Reporter is confident that it can make a significant contribution to the industry objective of enhancing the quality of information for decision making.

7 References

[1] 10 Year Vision
Version F01 submitted 31 March 2005

[2] The Future of Rail
Government White Paper – Department for Transport
July 2004

8 Annexes

Annex A Network Rail Information Systems

Pre Identified Systems

Item	Systems
(a)	Freight Billing System (BIFS) including tonnage and axle-load data.
(b)	Performance Improvement Plans (PAT, JPIP)
(c)	Train Running System on TOPS (Total Operations Processing System) (TRUST)
(d)	Performance data warehouse (PSS)
(e)	NPPR/IPRR – Summary performance data: outturn and causes (Delay minutes, PPM, Cancellations)
(f)	Rules of the Route/Rules of the Plan
(g)	Network timetable (TSDB) including APLAN
(h)	Database of access rights (ARDV)
(i)	NETRAFF/ACTRAFF systems used for measuring traffic levels (including tonnage)

Newly Identified Systems

Item	System	Description
[A]	PABS – Invoicing passenger journeys	PABS is the passenger billing system.
[B]	PPS – Possession Planning System	PSS feed directly into RotR and the WON (Weekly Operating Notice)
[C]	WON – Weekly Operating Notice	To be considered in conjunction with RotR and RotP.
[D]	TrainPlan – Train Planning	High volume outputs from this system feed directly into TSDB which has already been identified as a system within scope

Annex B Business Risk Assessment Information Network

System or Process		Data Connectivity*	System Status	Business Risk Impact**	Comment
(a)	**BIFS**	Medium	System Replacement Underway (TABS)	**Medium**	Outputs feed into traffic mapping and thus is a key input into higher level business data
(b)	**JPIP**	Low	N/a	**Medium**	Key Service Performance Mechanism – However core measurement data is fed via PPM
(c1)	**TRUST**	High	Train Management and Control Strategy being developed	**High**	Delay Minutes provide core data for Performance Regime/Performance Management
(c2)	**TOPS**	Medium	Subject to Train Management and Control Review	**Medium**	Lead Operational Planning System, risk impact is limited to this industry mechanism
(d)	**PALADIN /PSS**	Medium	PSS to replace Paladin longer term	**Medium**	Data Warehouse systems – subset of
(e)	**NPPR / IPPR**	Low	System Replacement Underway (tbc)	**Low**	Performance Outputs but these are only used as an Industry Performance "Snapshot"
(f)	**Rules of the Route/Rules of the Plan**	Medium	N/a	**Medium**	Operational Planning process, risk impact is limited to this industry mechanism
(g)	**TSDB**	High	System Replacement Underway (Integrated Train Planning System)	**High**	Lead Operational Planning System, risk impact is high as system underpins many other processes and functions
(h)	**ARDV**	--	--	**--**	System retired
(i1)	**NETRAFF**	Medium	N/a	**Medium**	NETRAFF system is used for calculation of

System or Process		Data Connectivity*	System Status	Business Risk Impact**	Comment
					track categories from train running information.
(i2)	**ACTRAFF**	Medium	tbc	**High**	Traffic Mapping outputs feeds into RUS's and NR Business Plans
[k]	**PPM**	Medium	N/a	**High**	A Key Industry Performance Measure that spans industry mechanisms

* Data Connectivity – Relative number of systems that the system/process interacts with

** Business Risk Impact – Impact on Network Rail of complete system failure based on basic connectivity to other core systems.

Annex C Filter Mechanisms Criteria

Criteria for Step 2 of the Filter Mechanism - Critical Industry Output Prioritisation

Critical industry outputs have been prioritised in order of the importance to the industry of the decisions that they affect. The following criteria were considered when assessing the prioritisation:

- Number of industry stakeholders affected by the decision;
- Number of decisions based on critical industry output;
- Priority of decisions based on critical industry output;
- Number of other critical industry outputs this one inputs to;
- Priority of the critical industry outputs inputted to;
- Targets set against delivery.

Criteria for Step 4 of the Filter Mechanism - Prioritisation of the Datasets

The tests comprise the following criteria:

Strategic – Critical Industry Outputs

- Priority and quantity of critical industry outputs affected.

Mandatory - System defined within Catalogue of Railway Code

- As necessary
- As expedient

Operational

- Quantity of operational processes affected
- Dataset contains source information
- Dataset is modified away from the source information (i.e. cleaned)
- The dataset is passed from one system or process to another

Annex D Critical Industry Outputs 'Straw Man'

Major sources of information are quoted under 'Defined Inputs' in brackets and italics.

Rank	Owner	Critical Industry Outputs	Description	Defined Inputs	Regulatory / Legal requirement
1	DfT	HLOS (England & Wales)	High-level strategy the rail industry in England and Wales will adhere to for the next 5 years.	ORR Periodic Review Network Statements Network Rail Business Plan RUS Regional Planning Assessments DfT Strategic Plan *(Briefing note on the development of the HLOS)*	The Railways Act 2005
2	Transport Scotland	HLOS (Scotland)	High-level strategy the rail industry in Scotland will adhere to for the next 5 years.	ORR Periodic Review Network Statements Network Rail Business Plan RUS Regional Planning Assessments DfT Strategic Plan *(Briefing note on the development of the HLOS)*	The Railways Act 2005
3	Network Rail	Timetable	The schedule to which National Rail, other passenger services and FOCs must run their trains.	Rules of the Route Rules of the Plan RUS National Rail Timetable TOC Bids FOC Bids Annual Possession Plan Track Access Agreements TRATIM Data (point-to-point train running data) Existing Timetable	Network Code
4	Network Rail	RUS	Route Utilisation Strategies (RUSs) seek to balance capacity, passenger & freight demand, operational performance and cost, to address the requirements of funders and stakeholders. Network Rail is developing Route Utilisation Strategies to cover the different routes across the rail network, in conjunction with rail industry partners and wider stakeholders	HLOS Capability Data Capacity Data Capacity Utilisation Index Railsys Model Passenger Service Outputs Route Corridor Plan Working Timetable Rules of the Route Rules of the Plan Passenger Counts Train Service Outputs Passenger Demand MOIRA Model PLANET Model Survey and Count Data Ticket Sales Data *(RUS Technical Guide)*	Network Licence

5	Network Rail	PPM	PPM combines figures for punctuality and reliability into a single performance measure. It is measured in terms or % trains arriving within a defined time band.	DAT Files TOC produced train data *(Periodic PPM Statistics)*	Network Code
6	ORR	Evaluated HLOS	Expectation of Rail Outputs Industry to deliver for 5 years covered by Periodic Review	HLOS Network Statements Network Rail Business Plan ORR Periodic Review RUS Regional Planning Assessments *(Briefing note on the development of the HLOS)*	The Railways Act 2005
7	Operating Companies	Timetable Bids	Infrastructure users requests for train usage. Once agreed these become the Timetable	Rules of the Route Rules of the Plan National Rail Timetable Annual Possession Plan Track Access Agreements National Rail Timetable RUS	Franchise Specification
8	Network Rail	Record of persons and causes of any delay/Cancellation	Record of all delays including the causes and party responsible	Delay Attribution Guide Delay/cancellation identified in train management systems *(Network Code)*	Network Code
9	Network Rail	Rules of the Route	Detailed accounts of access available to operators. Provides details of engineering works and No Trains Periods	Possession requirements Possession Plan (compiled in annual Engineering conference) Route based delivery planning unit (Report) National Access Unit (Report) TOCs Counter Proposals Network Statements *(Network Code)*	Network Code
10	Network Rail	Rules of the plan	Detailed accounts of access available to operators. Focuses on network capability (sectional running times, headways, signal box opening times).	Point-to-point running data Station Dwell times Route-by-route analysis of the impact of revised base data Alternative timetable structures Network Statements *(Network Code)*	Network Code
11	Network Rail	Delay Attribution Guide	For all parties to work together to achieve the core objective of delay attribution – to accurately identify the prime cause of delay to train services for improvement purposes"	Network Code	Network Code
12	ORR	Periodic Review	The periodic review (PR2008) will set Network Rail access charges for the five years. It will include extensive work determining the appropriate financial and incentive framework for Network Rail and reviewing the potential for efficiency savings in its expenditure	Access Charges Review NRs Strategic Business Plan (SBP) NR Scenarios (supports SBP) Determinations of possessions policy Strategic Plan	The Railways Act 2005

13	DfT	Strategic Plan	Sets out the strategic priorities for Britain's railway over the next ten years. The DfT is responsible for delivering the Plan, within the resources available.	The Railways Act 2005 HLOS Periodic Review	The Railways Act 2005
14	Network Rail	Network Statements	The Network Statement provides a single source for the information that will be required by a Train Operator wishing to operate train services on Network Rail's network. It is intended to facilitate access to that information and to ensure that comprehensive information is available to all current and potential train operators on a non-discriminatory basis.	Rail Regulators Criteria for approval of TAAs Network Code Rules of the Plan Rules of the Route	Network Code
15	Network Rail	NR Annual Information - Annual Report	Annual Report and accounts	HLOS Strategic Plan PPM	Network Licence
16	Network Rail	Response to 'Spot Bids' (changes to TT)	Network Rails response to any change to the Timetable requested by an infrastructure user.	Rules of the Route Rules of the Plan National Rail Timetable Annual Possession Plan Track Access Agreements National Rail Timetable RUS	Network Code
17	Network Rail	Railway Operational Code	The objective of the ROC is to sustain and, where necessary, restore expeditiously the operation of services in accordance with the Working Timetable in a manner consistent with the ORR ROC Criteria,	HLOS Working Timetable	Network Code
18	Network Rail	Notification of disruptive event	Notification of any planned event which prevents or disrupts the operation of trains on the network	Details of disruptive event	Network Code
19	DfT (Rail)	Franchise Specification	Specification for bidding/award of Franchises	RUS Rules of the Route Rules of the Plan	The Railways Act 2005
20	Network Rail	Response to enquiry regarding access rights	Response to enquiry regarding access rights	TAA	Network Code
21	Operating Companies	Contingency plans in relation to disruptive events	Contingency plans in relation to particular types of disruptive event	Rules of the Route Rules of the Plan National Rail Timetable Annual Possession Plan Track Access Agreements National Rail Timetable RUS	Franchise Specification
22	Network Rail	Network Code	Contract between DfT and Network Rail, specifying procedures to be followed for the running of the network	The Railways Act 2005	The Railways Act 2005

Annex E Overview of Network Rail Systems

(Numbering is as per Annex A)

(a) Freight Billing Systems (BIFS)
Definition **:** BIFS – Billing Infrastructure for Freight System
Network Rail Responses
What is the system used for?
The system is used for billing FOC for use of Network Rail track assets.
Does the scope of the system meet the user's emerging needs?
The system meets the immediate needs of Network Rail to perform freight billing, but falls short of being an integrated system that allows simple reconciliation of discrepancies and requires (not insignificant) manual intervention.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
After a degree of manual intervention the system is considered 99.5% accurate. The billing output to the FOCs is considered fit for purpose.
How useful is the dataset information?
The information is critical for freight billing of the FOCs.
Reporter Comments
The Reporter recommends, that further more detailed analysis is performed on the BIFS system, at this time, to access the actual data accuracy and to enable identification of the critical dataset information. A new system TABS (Route Based Charging) has been identified. This system is under development to replace BIFS longer term. TABS is expected to be implemented in the summer of 2008. As the marginal delays captured within the TRUST system can lead to cumulative delays which are considered to be of a significant level, consideration should be given to building a business case for extracting this information from TRUST more accurately.

[A] PABS – PAssenger Billing System (Passenger only journeys)
Definition **:** PABS – PAssenger Billing System (Passenger only journeys) PABS has been identified by the Reporter as a system to be included within the scope of work, as it is integral to the Network Rail invoicing of passenger journeys.
Network Rail Responses
What is the system used for?
PABS has been identified by the Reporter as a system to be included within the scope of work, as it is integral to the Network Rail invoicing of passenger journeys.
Does the scope of the system meet the user's emerging needs?
The system is used for billing TOCs for the use of Network Rail track assets.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
The system meets the immediate needs of Network Rail to perform passenger billing. However, the operators require pc emulation to allow access to the mainframe menus and the system was described to the Reporter as 'User Un-friendly'. PABS will not handle Performance billing, Depot Leases, Station leases and ISO Station Operator Access.
How useful is the dataset information?
The information is critical for passenger billing of the TOCs.
Reporter Comments
The new TABS system is currently under development and is due to replace PABS in the longer term.

(b) Performance Improvement Plans (PAT, JPIP)
PAT - Performance Action Tracking database
Network Rail Responses
What is the system used for?
The PAT system is used to input subjective information gathered by analysts' assessments of Network Rail improvement activities as it affects the TOCs
Does the scope of the system meet the user's emerging needs?
The system meets the immediate needs of Network Rail to create performance improvement plans
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
The PAT system is essentially a data entry screen, therefore, provided that the analysts type the information correctly, the system is 100% accurate. However, data entry is rarely 100% accurate. The system is not considered user friendly, therefore a new more user friendly precursor system (PrePAT) with minor changes to the current PAT front end, is currently undergoing development using MS Access. It was intended for deployment in March 2006. An additional system is planned to replace the PAT system entirely in the near future with a more comprehensive system called iPAT, however this replacement system is not yet defined.
How useful is the dataset information?
The PAT information is useful but subjective
Reporter Comments
The Reporter does not recommend that further more detailed analysis is performed on either the PAT/PMRS system, at this time, as the PAT system is in the process of being redefined and replaced. The Reporter does not recommend that further more detailed analysis is performed on the JPP/JPIP processes, at this time, as they form a contractual set of requirements and are, as such, not a dataset. It is recognised that a key challenge for the JPP/JPIP will be the requirement that the process needs to encourage behavioural change and to establish and strengthen existing working relationships between the various parties. It has been noted that Network Rail needs to become more focused on PPM and actions to deliver performance improvements over TOC geography. It has been noted that the TOCs need to be more accepting of investigations into TOC attributed train delays and be more engaged in cross industry work

(b) Performance Improvement Plans (PAT, JPIP)
JPIP – Joint Performance Improvement Plan
Network Rail Responses
What is the system used for?
The new JPP will cover actions across the whole industry to improve train performance for passengers.
Does the scope of the system meet the user's emerging needs?
The target is to deliver improved performance, above and beyond that which would have been delivered under previous arrangements. The new JPP will cover the actions of the whole industry to improve train performance. The new JPP will also replace the current LOC process once the changes have been made to part L of the Network Code. (originally planned for 31st March 06)
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
Inconsistency of information has historically been an issue across the industry, therefore, the new use of templates and better analysis (would) enable better comparison of information across the industry. There is no contractual requirement for JPP/JPIP as this requires a change to the Network Code, but are in place at this time by industry agreement as the preferred mechanism to maintain delivery of the Rail Review and to ensure the industry has continuity of plans for improved performance.
How useful is the dataset information?
The information is fundamental to the creation of a Zonal performance budget, the Network Rail Company Performance plan and LOS.
Reporter Comments
The Reporter do not recommend that further more analysis is performed on either the PAT/PMRS systems, at this time, as the PAT system is in the process of being redefined and replaced. The Reporter does not recommend that further more detailed analysis is performed on the JPP/JPIP processes, at this time, as they form a contractual set of requirements and are, as such, not a dataset. It is recognised that a key challenge for the JPP/JPIP will be the requirement that the process needs to encourage behavioural change and to establish and strengthen existing working relationships between the various parties. It has been noted that Network Rail needs to become more focused on PPM and actions to deliver performance improvements over TOC geography. It has been noted that the TOCs need to be more accepting of investigations into TOC attributed train delays and be more engaged in cross industry work.

(c) TRUST
Definition **:** TRUST - Train RUnning on Systems TOPS
Network Rail Responses
What is the system used for?
TRUST identifies the occurrence of train delays and then allows explanation and attribution of these together with Reliability Events
Does the scope of the system meet the user's emerging needs?
The system meets the immediate needs of Network Rail to perform delay attribution, however, the manual corrections required are prone to error due to the dated interface and mainframe command set.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
The volume of errors overall are deemed low. The majority of this information is automated and correct. TRUST and TOPS are legacy mainframe systems performing functions that are currently considered fit for purpose, but they are inflexible, un-user friendly and are long overdue for replacement. There is an issue of manual reporting for rural trains where they are manually signalled. (e.g. western reaches) These require manual reporting via human input and are therefore prone to human error once entered into TRUST. It is estimated that 5% of trains have manual reports.
How useful is the dataset information?
The TRUST information is critical for delay attribution. PSS (Performance Systems Strategy) system is a new system under development that is being designed to (ultimately) replace TOPS and TRUST. PSS will be receiving live TOPS/TRUST data within the next 6 months for performance analysis. Once implemented, this is where the important industry data will reside.
Reporter Comments
The Reporter concurs with the responses above

(d) PSS – Performance data warehouse
Definition **:** Performance Systems Strategy (PSS) Project
Network Rail Responses
What is the system used for?
The system will enable the consolidation and consistency of source data for reporting to enable all parties to work from the same baseline and remove the current need to draw down data snapshots from mainframe databases using different extraction algorithms. PSS is not yet live, but is due to receive live TOPS data within the next 6 months. (The Reporter was not supplied with exact information for this intended implementation.)
Does the scope of the system meet the user's emerging needs?
The existing systems do not meet the users' emerging needs. Once PSS has been implemented, this will move the rail industry towards a single source of train performance data in PSS, reducing the level of unnecessary dispute inherent in the existing performance systems process. One of the benefits realised with the implementation of the PSS system is the reduction of dependency of information stored within the TRUST system. An additional benefit will be the improved performance of other down-line systems that are currently reliant on the TOPS system for train schedule data that will no longer be required. The PSS environment must be able to integrate and feed back the data provided to it, to supply analysis and forecasting data to a range of stakeholders within rail industry. This will require enhancements to the data warehouse environment to optimise and relate data from multiple sources; these will be at varying levels of granularity. Consideration of Sub-threshold delay will be required to identify how much effort is justified in explaining or analysing what is in effect the majority of train delays recorded. This may require methods for aggregating or sampling data based on pre-defined criteria or exception reporting when levels exceed pre-set parameters.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
Once implemented, this new PSS system encourages system rationalisation, reducing risk and potentially reducing IT operating costs
How useful is the dataset information?
This rationalisation of information within PSS will make the PSS dataset critically important and it will also support the much closer working relationships within the rail industry enshrined in the JPIP (joint industry performance initiative process).
Reporter Comments
The Reporter does not recommend a detailed investigation on this system, at this time, as the system is still under development though analysis of the upstream data flow may still be beneficial.

(e) NPPR / IPPR
Definition **:** NPPR Network Period Performance Report IPPR Industry Period Performance Report
Network Rail Responses
What is the system used for?
The system is used for KPI (Delay minutes, PPM, Cancellations) reporting on an on-going basis and are published quarterly in National Rail Trends.
Does the scope of the system meet the user's emerging needs?
The system meets the immediate needs of Network Rail to produce these reports, however, Network Rail are assessing the current user requirements to replace this system with a real-time web-based electronic platform with ability to drill down to relevant information and enable the production of management information. In future these reports may be changed and Network Rail are currently in consultation with this revised process.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
Due to the sheer volume information received, it can only be assumed that the data received is correct and this is therefore only a subjective opinion. The final reports are considered fit for purpose, but as stated above, the accuracy is difficult to establish.
How useful is the dataset information?
The information is required for KPI Reporting.
Reporter Comments
The Reporter does not recommend that further more detailed analysis is performed on either of the NPPR / IPPR systems, at this time, due to the manual, paper based, environment in which they are processed. Analysis would be better placed upstream of these outputs. User requirements are already being assessed by Network Rail to replace these systems and The Reporter would only be duplicating effort were The Reporter to undertake this task too.

(f) RotR / RotP (and [C] WON)
Definition : RotR - Rules of the Route RotP – Rules of the Plan WON (Weekly Operating (possession) Notice)
Network Rail Responses
What is the system used for?
Rules of the Route and Rules of the Plan are not in themselves datasets or systems; rather they are Network Rails Firm Contractual Rights and although they are established annually through the application of Part D of the Network Code, they match a TOCs Firm Contractual Rights set out in Schedule 5 to a franchised Track Access Contract..
Does the scope of the system meet the user's emerging needs?
The system meets the immediate needs of Network Rail to produce these reports
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
The final reports are considered accurate and fit for purpose.
How useful is the dataset information?
All three of the above documents feed planning information into the various planning systems which ultimately update APlan, Trainplan and TSDB.
Reporter Comments
The Reporter does not recommend that further, more detailed analysis is performed on either of the RotR/RotP systems, due to the manual, paper based, environment in which they are processed. The WON is similarly a paper based output and does not warrant further scrutiny though analysis of the upstream data flow may be beneficial.

(g) Network Timetable (TSDB) and Aplan
Definition : TSDB – Train Service Data Base APlan – Access Planning
Network Rail Responses
What is the system used for?
The TSDB system is used for timetabling of trains and associated operating and commercial information. APlan is the main Network Rail system for storing TMS Geography information.
Does the scope of the system meet the user's emerging needs?
Selex SI has recently (March 06) been appointed prime contractor for the provision of a new Integrated Train Planning System (ITPS). The ITPS will replace (amongst others) the TSDB and Aplan systems, creating an integrated planning system based on a single unique, rationalised dataset. The new system will interface with all main existing Network Rail and third party systems (ie It will still receive PIF files and export CIF files) to allow a seamless integration upon implementation. The ITPS will also add new functionality that is not currently available. (For example, the new system will allow timetabling conflict analysis and reporting.) The information that will reside within the new (Oracle) datasets will also be richer and provide more information than currently available.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
The TSDB system is considered accurate but due to its legacy nature it is in the process of being re-developed. The CIF output to the TOCs and CIF User Groups is considered accurate and fit for purpose.
How useful is the dataset information?
The information is critical for the day to day running of the TMS.
Reporter Comments
The Reporter does not recommend that further, more detailed analysis is performed on the TSDB and Aplan systems, at this time, as they are currently being redeveloped and scheduled to be completely replaced by the new ITPS system. This may change as a result of prioritisation work around Network Rail functions. Phase one of the new ITPS system is due to be operational by January 2007. This will initially encompass a system for long term planning. Operational services will be live from ITPS at the commencement of the 2008 timetable, in December 2007.

[D] Trainplan
Definition *:* Trainplan – Train Planning Trainplan is included here as it feeds high volume planning information directly into TSDB, a system identified as key within the industry.
Network Rail Responses
What is the system used for?
The system was originally designed to replace PROTIM (Train planning system that handles the editing of timetables and uploading to TSDB) and is used to deal with LTP processing. The system now also deals with STP and the TSDB upload.
Does the scope of the system meet the user's emerging needs?
The system falls short of being an integrated system and will be replaced when Selex SI's new ITPS system goes live.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
After a degree of Network Rail planners' manual input, the system is considered accurate. However, manual input is rarely 100% accurate. The functionality is considered fit for purpose, but labour intensive.
How useful is the dataset information?
The information is critical for uploading train plans to TSDB.
Reporter Comments
The Reporter does not recommend that further, more detailed analysis is performed on the Trainplan system, at this time, due to its impending replacement by the new ITPS system. This may change as a result of prioritisation work around Network Rail functions. Phase one of the new ITPS system is due to be operational by January 2007. This will initially encompass a system for long term planning. Operational services will be live from ITPS at the commencement of the 2008 timetable, in December 2007.

(h) ARDV
Definition **:** ARDV - Access Rights Database Validator
Network Rail Responses
What is the system used for?
The system is no longer used.
Does the scope of the system meet the user's emerging needs?
No and it has therefore been removed.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
The system was considered accurate but due to the time it took to produce the required information, was not fit for purpose.
How useful is the dataset information?
There is no longer any information on the system. However, archives do exist of the information, should it ever be required.
Reporter Comments
This system has now been archived and is no longer in operational use. The Reporter recommends that no further analysis is performed on the ARDV system

(i) NETRAFF/ACTRAFF
Definition : NETRAFF (Network Traffic) ACTRAFF (Actual Traffic)
Network Rail Responses
What is the system used for?
The NETRAFF system is used for calculation of track categories from train running information. The ACTRAFF system is used to calculate Actual Traffic flow and tonnage data.
Does the scope of the system meet the user's emerging needs?
These systems meet the immediate needs of Network Rail to perform calculation of track categories from the actual values taken from ACTRAFF, but falls short of being an integrated system and requires significant manual intervention to both input and correct information.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
After a degree of manual intervention the system is considered accurate. The final Reports are considered fit for purpose.
How useful is the dataset information?
The information is important for maintenance and renewal regimes, particularly for track assets, in order that regimes can be set that are appropriate for the speed and tonnage of the traffic.
Reporter Comments
The Reporter does not recommend that further, more detailed analysis is performed on the NETRAFF system, at this time, as Network Rail intend to replace NETRAFF with a new system called the Complete Traffic Data System (CTDS), which will have better links to the train operating systems. The intention is to reduce Network Rail's reliance on ACTRAFF, which is also deemed to be lacking. However, further study may be required as a result of prioritisation of Network Rail functions and timescales for replacement.

[B] PPS
Definition **:** PPS - Possession Planning System PPS is included here as it has been identified as integral to the operational planning process and holds an important dataset within that process.
Network Rail Responses
What is the system used for?
The system is used for the recording of national engineering and access requirements.
Does the scope of the system meet the user's emerging needs?
The system meets the current needs and appears to be quite popular.
Do users have any concerns over the accuracy and fitness for purpose of the datasets in question?
The data entry is manual, but the system is considered accurate. The possessions output are considered fit for purpose.
How useful is the dataset information?
The information is critical information that is then fed into the planning process.
Reporter Comments
The Reporter does not recommend that further, more detailed analysis is performed on the PPS system, at this time, regarding data accuracy, as it is primarily a manual data entry of engineering possessions which are then just fed into the planning process.