IC Position Paper

Managing Data Quality

Below is a summary of the IC document types.

Policy Statement
"You are expected to do ...". Contain statements of Shell corporate policy, to which Shell companies are expected to adhere.

Key Issue Briefing
These give guidance on a particular area, and the best practice within it.
Recommendations are usually made.

Guide
"How to do ...". Are aimed at IC practitioners in a particular area.

Product Evaluation
IC evaluation of one or more products.

Meetings/Workshops
Includes documents like "ICS Managers Conference Minutes" and "Data Management Workshop".

News/Information Exchange
Includes the IT Journal, newsflashes, and client specific project reports that may be useful to a wide audience.

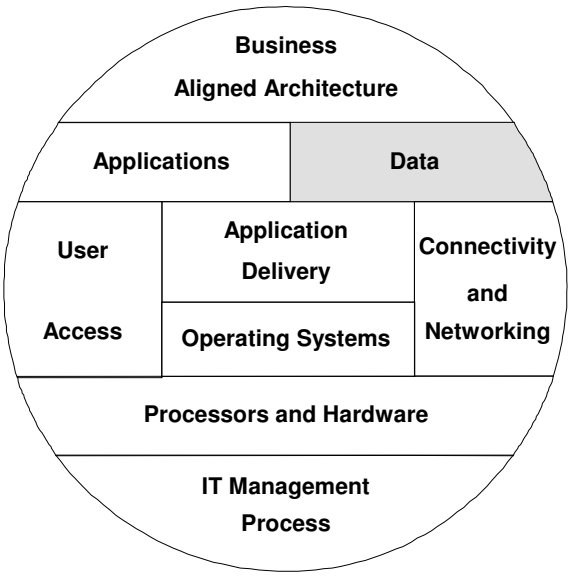
IC documents are also distinguished on the basis of intended readership. Those written for IT practitioners have orange/yellow covers, and those aimed at a non-technical audience have green/yellow covers.

IC Position Paper

Managing Data Quality

Author: M.R. West SIPC-ICT/47

November 1993



Report No. IC92-124

Sponsorship: Global

Recommendations valid until: December 1999

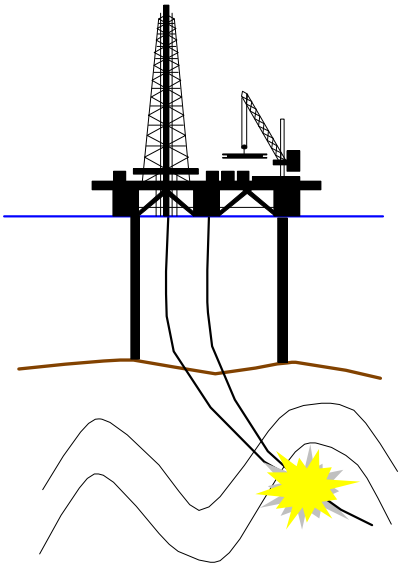
Target Readership

This Position Paper is written for managers at all levels

Acknowledgements

Many Shell employees contributed to this paper.

Contents



On February 24th 1989, a new well was being drilled in the North Sea. It had reached a depth of 1770 metres when it collided with another well. The well had to be redrilled, and there was a one month deferment of oil production. It could have been worse. In another field with higher reservoir pressures, the platform could have been lost.

Although there were many contributory causes, a major factor in this case was that the data for the trajectory of the existing well was incorrect. Thus in deciding the trajectory of the new well, the actual risk of hitting the existing well was unknown.

Following the incident, a number of measures were undertaken to improve the quality of their data, some of which are mentioned later in this document. Although this is an Exploration and Production example, the principles involved are as important to the Downstream.

1 Executive Summary

The value of data comes from its use in contributing to sound decisions. For this reason the following rule is included in "IT Guidelines for Management" IC 93-013.

> Responsibility for data and its use lies with line management.

Data so pervades our business that we tend to take it, and its quality, for granted. However, its quality cannot be assured systematically, unless the definition and quality requirements for the data have been agreed by those who create and use it. Further, procedures must be in place to ensure the requirements are met. Some symptoms that would indicate you have a problem with data include:

- You receive a lot of data you don't need.
- You spend too much of your time finding the data you do need.
- You find it necessary to have your own database, perhaps in a spreadsheet, because the official sources are not trusted, or are inaccessible.
- Mistakes have been made because you misunderstood the data given to you.
- You have to spend time reconciling data from different sources.
- You have to make decisions based on incomplete data because of delays in providing the data you need.

The incident described on the facing page was dramatic, but it is given here to relate what was done to ensure that a similar incident could not happen again. It is used as a thread through the document to show how you can make sure that the data you use, or provide for others, can be relied on.

Recommendations

1. *Ensure the members of your organisation are AWARE of the importance of data quality.*
2. *Identify the data that is CRITICAL to your business.*

For your critical data:

3. *EVALUATE your data to determine if it meets the requirements of its users.*
4. *Ensure that the RESPONSIBILITIES for the definition, creation, validation, and use of data are appropriately allocated and understood.*
5. *Ensure that there are agreed DEFINITIONS and REQUIREMENTS for the data used in your organisation.*
6. *MEASURE the quality of your data and act to IMPROVE it where required.*

2 Purpose

The purpose of this document is to help you understand the impact of data on your business and to show you how to start improving the quality of your data.

3 Why is Data Important?

To perform well, you need the resources to do the job. These resources include people, money, materials ... and data.
Data is involved in every business activity. Indeed, data is sometimes the only or main output of an activity. For instance a market survey is conducted solely to collect data that will be used to develop a marketing strategy. The value of data comes from its use in contributing to sound decisions. If you can't rely on your data then the result can be missed opportunities, or higher costs, as in the well collision example. For this reason the following rule is included in "IT Guidelines for Management" IC 93-013.

> Responsibility for data and its use lies with line management.

It goes on to make the following recommendation.

> *Procedures must be implemented to ensure data is properly controlled.*

There are several aspects to the management and control of data. This document deals with the quality of data and is intended to make you aware of the steps necessary to ensure that you are fulfilling your responsibilities in this area. It does this by:

- showing what quality management means when it is applied to data,
- identifying important properties of data,
- telling you about some consequences of poor quality data,
- showing how to assess the quality of your data, and
- showing how to improve and maintain the quality of your data.

Managing the quality of data is one aspect of data management. A full picture of data management is given opposite.

The Data Management Context[1)]

The *activity* of data management is the planning, implementation, administration, and control of data for the benefit of the business. This should be done within the business (not just the data management department) and is achieved through developing and maintaining:

- business models that are *owned by the business* and used so that new and existing systems (including packages) support the business better,
- databases and the data model that defines them, to facilitate the compatibility of data,
- *agreed definitions, requirements, and responsibilities for data so that data is created and used properly.*

Business Model

A business model is a structured description of the business. It includes descriptions of activities and the information created or used by them. It can be used to analyse and design business processes and to assess how well existing and planned computer systems support them.

Data Model

A data model holds the definition, structure and format of data. A data model provides the basis for ensuring the compatibility and consistent use of data.

Quality Data

Quality data is data that conforms to agreed quality requirements, e.g. for appropriate accuracy and timeliness. It has an agreed definition, and responsibilities to ensure its quality are assigned and acted on.

1) Based on 'Managing Shared Data' IC 91-078.

4 The Quality Management Link

"Quality is meeting agreed customer requirements" (the Quality Blue Book). Figure 1 below illustrates how the quality management message applies to data.

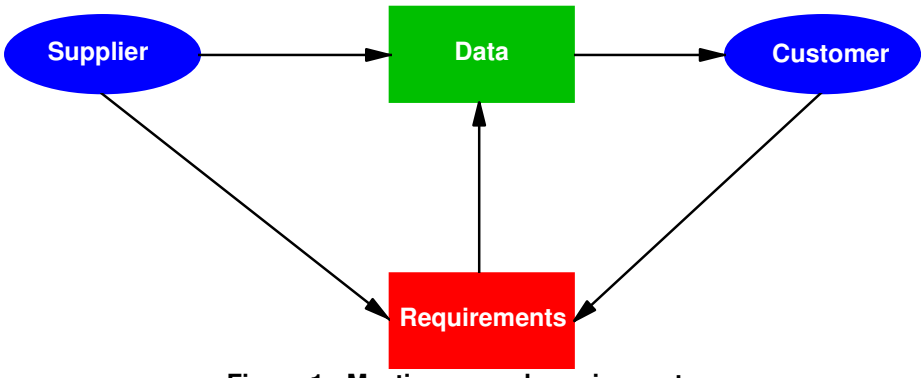
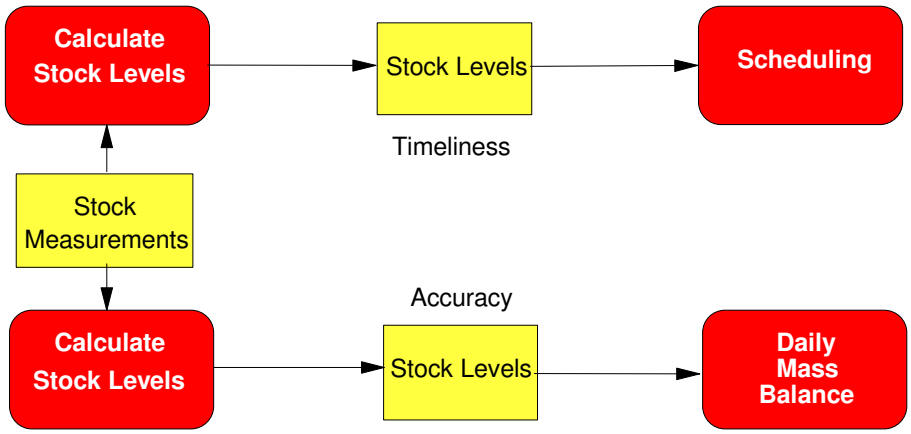


Figure 1 - Meeting agreed requirements

Almost every business activity results in new data being created. The use of this data is not restricted to the activity that creates it; often it is used in other activities. The requirements for the quality of the data may be different for the activity creating the data, and for other activities that use it. Therefore, the requirements for data need to be agreed between its suppliers and customers. The suppliers of the data are responsible for meeting the agreed requirements.

Refinery Stocks - Differing Quality Requirements.



In refineries, stock levels are needed hour by hour for operational reasons. Accuracy is important, but timeliness is critical. However, stock levels must also be calculated with great accuracy for the daily mass balance where timeliness is less critical.

5 Important Properties of Data

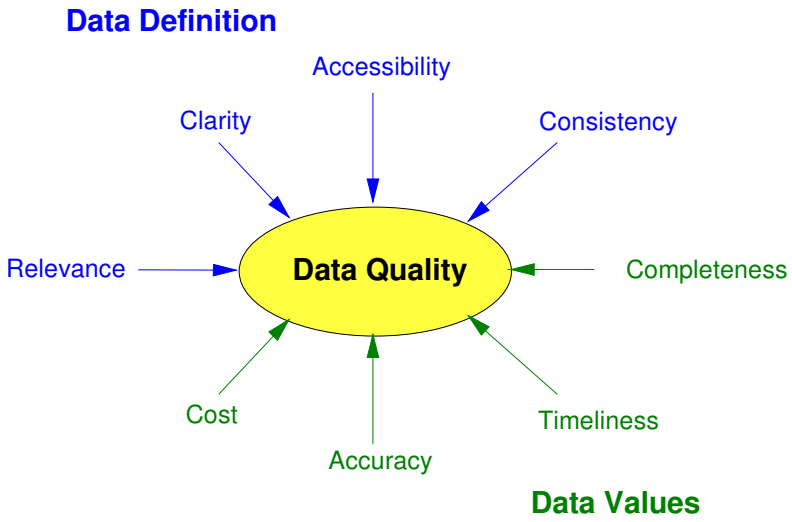


Figure 2 - Some Properties of Data

Some important properties of data for which requirements may need to be agreed are illustrated in Figure 2 and include:

relevance:	the usefulness of the data in the context of your business.
clarity:	the availability of a clear and shared definition for the data.
accessibility:	where, how, and to whom the data is available or not available (e.g. security).
consistency:	the compatibility of the same type of data from different sources.
completeness:	how much of the required data is available.
timeliness:	the availability of data at the time required and how up to date that data is.
accuracy:	how close to the truth the data is.
cost:	the cost incurred in obtaining the data, and making it available for use.

Notice that some properties are more strongly related to the *definition* of the data, whilst others are more strongly related to *data values*.

What Colour is Your Data?

The Shell pecten is valuable to the Shell Group because of the real influence it exerts over customers, suppliers, and governments in their decisions when dealing with the Group. Maintaining the consistency of the pecten, both in terms of shape and colour, is important in achieving this. The selection of colours has historically been on the basis of colour matching. However, a vital piece of data was missing. What was an acceptable match? As a result, one Dutch sign maker was selling three different sets of colours to three different operating companies. In addition there were many costly remixes, because the range of acceptability was not defined.

To overcome the problem, the missing data has been defined. The colours are now specified in terms of a range of spectrometer readings of wavelength and reflectance under standard lighting and viewing conditions. For printing, colour specifications have been created based on standard Pantone colours.

6 Consequences of Poor Quality Data

If you can't rely on your data, the consequences may be severe. Poor data can cost as much as good data to capture, process, and store. In addition, either opportunities will be lost, or additional costs will be incurred in staff time; gathering, reconciling, and correcting data.

- What are the consequences if you are misled by poor data?
- What are the consequences if you provide data that misleads others when it is used?

Some possible consequences of undetected poor data are:

- an incorrect invoice is sent,
- the wrong products are made,
- the wrong materials are bought or used,
- a bill is not paid when due,
- a well is drilled in the wrong location,
- an unsuitable candidate is assigned to a position,
- a distillation column is built the wrong size,
- an application system is acquired which fails to meet the needs of the business,
- a major safety or environmental incident occurs.

Most of these have happened somewhere in the Group, many on more than one occasion.

An Example of Old and Inaccurate Data

A Group refinery entered into a contract to supply a large volume of low sulphur fuel oil. The economics were based on supplying the fuel oil from Nigerian long residue, for which there were measurements of the sulphur content made several years before. Unfortunately, the composition of the crude had changed over time, as a result of which the fuel oil sulphur specification could no longer be met without using expensive blending components. A substantial loss was made.

7 Assessing Data Quality

Identify where You are with Your Data

Figure 3 shows what you must have done to ensure reliable data. For different data, different pieces may be missing. From the missing pieces, you will be able to see what to do to improve your data.

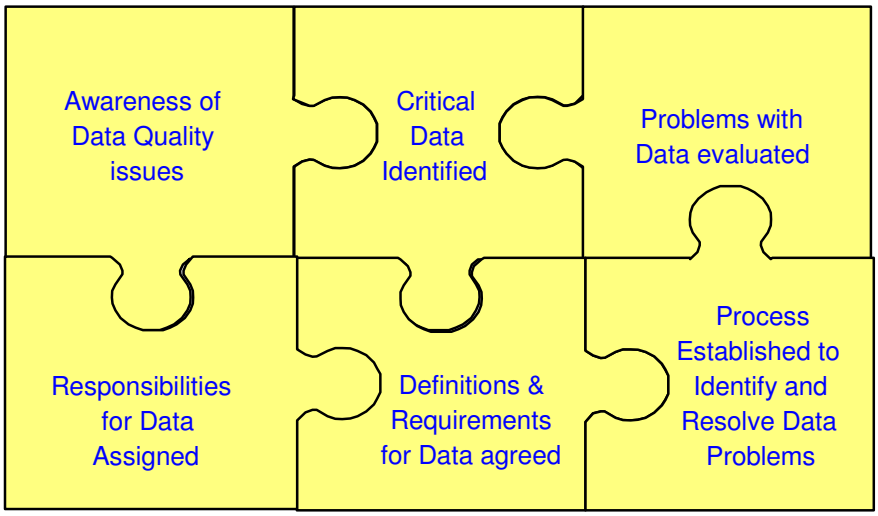


Figure 3 - The Data Quality Jigsaw

Identify the Data that is Critical to Your Business

Your most important data is that which supports your most important activities. Data is critical to an activity if an *uncorrected* problem with the data would have a serious impact. Your company or sector business model can help identify where data is used, and hence its impact on the business. Bear in mind that the same data may be used in many activities with differing requirements. This data must then be traced back to its *original* source to determine its quality. Ensure that all of your data is relevant to your activities, or required, e.g. for legal reasons.

Evaluate Your Critical Data

Having identified your critical data you should determine the properties that are important in its use. The following questions will help to evaluate your critical data.

- Do you have the same definition of the data as the person who created it?
- Is the data easily accessible to you?
- Are there multiple copies or versions of this data available to you or to others? If so, is there a master copy from which the others are derived?
- If the same type of data comes from different sources, is it created in the same way?
- Is all the data you require available, and available when you need it?
- Is the accuracy of the data known and does it meet your requirements?
- Is the cost of acquiring the data less than the benefit achieved from its use?

Improvement of Deviation Data

Following the well collision incident, the quality of existing data for all wells was reviewed.

8 Improving Data Quality

Assign the Responsibilities for Data

There are two levels for which responsibilities need to be assigned and fulfilled:

- *Data Definition,* which includes defining the meaning of data, the basis of any calculation to create the data, and the quality requirements. This is to ensure clarity, relevance, and consistency. This needs to be agreed and understood by all those who create and use the data.
- *Data Values,* which is the data itself. Here, the responsibilities are for creation, validation and use of the data to ensure that the data values are accurate, timely, complete, and used appropriately. In addition, identify those to whom the data should be made available.

Agree the Definition and Requirements for your Data

- Ensure that your critical data has agreed definitions that are documented. This forms part of your corporate data model.
- Ensure that your critical data has agreed (documented) accuracy, availability, timeliness, and completeness requirements.
- Where more than one copy of the same data exists, (often in different systems or departments) ensure procedures are in place to make the copies or versions consistent.

Tackle Problem Data

In evaluating your data you will probably have come across some data that needs to be improved. Set up a project to clean up your critical data, so that it meets the requirements that have been agreed.

Establish Procedures

Ensuring you can rely on your critical data is not a once off exercise. You need to have procedures in place to check that as new data is created it meets the requirements agreed for it. You also need to ensure that when poor data is discovered, the necessary steps are taken to improve the data and prevent a recurrence.
Finally, ensure that an audit of your data becomes a routine part of any review of the business, or quality improvement project.

Data Responsibilities

Following the well collision incident, the responsibilities for survey data were defined, including the procedures that defined the actions that had to be carried out, and the checks that were necessary to ensure the accuracy of the deviation data. These responsibilities were then allocated, and training in borehole surveying quality control was undertaken and maintained. Finally, controls were put in place to ensure that only authorised deviation data could be used for planning purposes.

Data Sharing Agreements

A number of data sharing agreements have been established. In these, customers and suppliers of data make a formal agreement on the definition and quality requirements for data that they share.

Improvement of Deviation Data

When the quality of well data was reviewed, it was discovered that 15 wells needed immediate resurvey, and this work was carried out as a matter of urgency.

Borehole Data Inspection

Following the well collision incident, a PC tool was acquired to check survey data for inconsistencies.

9 Recommendations

1. People will only take steps to improve their data when they realise it is important.

   - *Ensure the members of your organisation are AWARE of the importance of data quality.*

   You could start by circulating this document to other members of your department.

2. Some data is more important than other data. Be pragmatic, tackle your most important data first.

   - *Identify the data that is CRITICAL to your business.*

   Your sector business model can be used to help identify your critical data.

3. If you do not know how good your data is, there may be a shock in store.

   - *EVALUATE your data to determine if it meets the requirements of its users.*

   Start by sampling and testing data in areas that exhibit symptoms of poor quality.

4. To establish control over your data, responsibilities must be defined.

   - *Ensure that the RESPONSIBILITIES for the definition, creation, validation, and use of data are appropriately allocated and understood.*

   Start by ensuring there are documented responsibilities for all data that crosses organisational boundaries in your company.

5. To be able to assess whether your data is good enough, it is necessary to define your requirements.

   - *Ensure that there are agreed DEFINITIONS and REQUIREMENTS for the data used in your organisation.*

   You can use your operating company or sector data model as a start point for gaining agreement.

6. To control the quality of your data, you must first measure its quality.

   - *Measure the quality of your data and act to IMPROVE it where this is required.*

   Ensure that an audit of your data becomes a routine part of any review of the business, or quality improvement project.

Remember, only you can ensure the data you use meets your requirements. However, you can get help from your company or sector data management group. You can benefit from the experience of others in the advice and information they can offer.

The Goal

The goal is data you can rely on.

Appendix A Questionnaire

It is important for IC to measure the quality of its deliverables. For this reason please copy, complete, and return this questionnaire.

Document: **IC Position Paper Managing Data Quality**

Did you read:	summary only	❑
	part of document	❑
	whole document	❑

On a scale of 1-10 (1 = very poor, 10 = excellent), how did you find this document?

Do you have any specific comments? (Please attach extra sheets if required)

Please complete the following so that your comments can be followed up:

Your name: _____

Your reference indicator: Full phone number:

Company: _____ Date: _____