



Data Health Check

CONTENTS

1	INTRODUCTION AND DEFINITIONS	2
2	DO WE KNOW ENOUGH ABOUT OUR DATA?	2
3	DATA EXISTENCE AND USAGE	2
4	DATA SOURCING AND PROCUREMENT	3
5	QUALITY OF DATA	3
6	DATA DUPLICATION AND INCONSISTENCY	4
7	DATA STRATEGY AND ARCHITECTURE	4
8	DATA LEGISLATION	5
9	COMMERCIAL SOFTWARE	5
10	SOME TRICKIER QUESTIONS	5

This paper has been written by an independent consultant, John Amiry, and is intended as a guide to illustrate how progressive organisations can make real and tangible progress in exploiting their strategic resource known as data. The author is a recognised subject matter expert in the field of data architecture and strategic data management. Assessing a company's maturity and health in how it strategically manages and exploits its data resource is precisely an area where expert advice and guidance should be sought. Clients wishing to explore this area should contact Virtual Resources. The copyright in this work is owned by and remains the property of its author.

1 INTRODUCTION AND DEFINITIONS

By considering the questions in this brief paper an organisation can rapidly gain an idea of how "healthy" or "unhealthy" it is in relation to its management of data and information.

By Data we mean the raw facts that we collect during our normal business activity. This is distinct from "Information" where some processing of data has taken place with an audience in mind.

These questions do not address the area of "Knowledge".

2 DO WE KNOW ENOUGH ABOUT OUR DATA?

- 2.1 Is there a model of all the data the organisation uses, showing its nature and structure?
- 2.2 Is the model available to anyone within the organisation?
- 2.3 Does the data have a definition?
- 2.4 Does we record and explain variations of definitions within the organisation?
- 2.5 Are synonyms and homonyms recorded (including the usage)?
- 2.6 Are these definitions and descriptions available to anyone in the organisation?
- 2.7 Do we know what the required quality of data we need is and any variations to these quality standards?
- 2.8 Do we keep the data about our data up-to-date?

3 DATA EXISTENCE AND USAGE

- 3.1 Can people in our organisation easily determine all our stores of data, whether in databases, spreadsheets, documents, PDA's etc?
- 3.2 Do we know which systems use these stores of data and what actions they take on the data?

- 3.3 Have we mapped the Logical Data Model against datasets or do we rely on the documentation of the dataset to determine what data is contained?
- 3.4 Do we know who uses the systems that access the data?
- 3.5 Is it easy to find out what the plans are for any particular system?
- 3.6 Do we know what our data is used for? (E.g. what reports or information streams it feeds?)

4 DATA SOURCING AND PROCUREMENT

- 4.1 Do we know whom we consider to be the true owner of the data we are considering?
- 4.2 Have we defined the minimum acceptable quality for all users of a supply of data?
- 4.3 Have we defined the maximum quality of data we need?
- 4.4 Have we decided on an acceptable approach to the supplier of the data updating the data in terms of frequency and style (update in place, replace etc)?
- 4.5 Do we coordinate our Data Procurement to maximise re-use of data and minimise duplicate purchase?
- 4.6 Does someone purchasing data do so on behalf of the organisation or on behalf of a smaller organisation unit?
- 4.7 Is it easy to find out any existing contracts we have for the supply of data?
- 4.8 Do we have a formal procurement process for data?
- 4.9 Do we check what legal rights the supplier of the data has to supply it?

5 QUALITY OF DATA

- 5.1 Is someone in our organisation directly accountable for the quality of data in a system?

- 5.2 If someone in the organisation detects an error in data, how would that error be corrected?
- 5.3 If someone outside the organisation detects an error, how would that error be corrected?
- 5.4 If there were a debate about the validity of a data value, whose side would we take?
- 5.5 Do we measure the quality of our stored data?
- 5.6 Do we devise and implement means of correcting the overall quality of our data?
- 5.7 When addressing a quality issue with data do we address the cause of an error as well as correcting the error?

6 DATA DUPLICATION AND INCONSISTENCY

- 6.1 Do we manage and track duplication of data?
- 6.2 Do we avoid unnecessary duplication of data?
- 6.3 Do we aim to minimise necessary duplication of data?
- 6.4 Do we actively seek to reduce existing duplication?
- 6.5 What restrictions are placed upon people creating new stores of data?
- 6.6 Do we ensure where there is duplicate data that changes in the values of one set of data are reflected in the other sets of data within an acceptable timeframe?
- 6.7 Do we audit and measure variations between sets of duplicate data?

7 DATA STRATEGY AND ARCHITECTURE

- 7.1 Is Data the subject of Strategic Planning and Thinking?
- 7.2 If there is a proposed change in business circumstances, such as a new Business Plan, does anyone assess the effect this will have on our data needs and existing data capability?

- 7.3 Do we express a strategic direction for our data or do we see the data as a result of something else, such as Systems Strategies or Business Processes Strategies?
- 7.4 Do we have a strategy for all data or do we have more detailed strategies for each area of data?
- 7.5 Can we have several strategies for the same area of data?
- 7.6 How does a strategy for an area of data affect the strategy for the store of data or for the business strategy for the user of the data?

8 DATA LEGISLATION

- 8.1 Whose responsibility is the Data Protection Act?
- 8.2 What other legislation that we are aware has serious data implications?
- 8.3 What is impact on an organisation of legislation such Sarbanne-Oxley?

9 COMMERCIAL SOFTWARE

- 9.1 Do we assess software for purchase on the basis of its data structures?
- 9.2 Do we ensure that purchasers of packages are aware of the Total Cost of Ownership of the proposed package, specifically the cost of introducing a data store (s) that conflicts with our existing data?
- 9.3 Do we assess the true costs of any proprietary code inherent in a package?

10 SOME TRICKIER QUESTIONS

- 10.1 Do we think that data quality is the responsibility of the IS group? (Corollary, does the IS group think that Data Quality is the responsibility of the IS group?)
- 10.2 Do we think that managing our data should be a project with a start and an end date?
- 10.3 Do we think the major cost of devising systems is in developing the software and we can save all of that by buying an off-the-shelf package?

- 10.4 Do we think that XML is the answer?
- 10.5 Do we think that the route to quality data is to have junior staff working all weekend to correct mistakes that should never have been made in the first place?
- 10.6 Do we think the best way to get good data about our clients is to buy a CRM package?
- 10.7 Do we think interfaces are a good thing?
- 10.8 Do we think things can never be perfect and therefore we may as well mess them up to start with?
- 10.9 Do we understand the difference in behaviour between transmitted and reflected light?
- 10.10 Do we know any of the Laws of Thermodynamics and do we understand them?
- 10.11 Do we think that since no one can check a lot of the information we generate that it doesn't really matter if it's wrong?
- 10.12 Do we think we can add lots of little strategies together and get a big strategy?
- 10.13 Do we think if we take pretty good data and combine it with some other pretty good data we will get lots of pretty good data?