Web Enablement &
Information Quality
White Paper

Web Enablement

and

Information Quality

White Paper

Contents

Is your organisation's data fit for purpose? How do you know? In a recent survey of 600 senior CIO and IT Directors, 75% reported that they had data quality problems. Only 30% were "very confident" in their data and many reported direct & consequential costs arising from Data Quality issues e.g. 1 in 3 have delayed or cancelled new systems because of poor data quality (Source PriceWaterhouseCoopers, 2001).

In April 2001, Computer Weekly reported that dirty data costs UK organisations £1bn.

These are clearly headlines designed to gain attention but what lies behind the headlines? Undoubtedly, there is there a problem but what do we know about it? How could IT allow it to happen? How big is it? Most importantly – what is the impact and cost of rectification? Given that the problem exists, what can be done now to prevent the embarrassment of poor quality and mis-matched data reaching the scrutiny of presentation on the internet?

Web Enablement & Information Quality White Paper

Background

There are many hurdles to overcome in the quest to make information and data available to all those that should rightly have access. Hot issues at the top of the list are security, encryption, presentation, protocols, server performance under stress, browser compatibility. What is often close to the bottom or missing altogether is the data itself. The assumption is that data is there (which is largely valid but not always) and available which is a giant leap of faith.

Web enablement is the right concept for dissemination of information provided the right data gets to the right people and when it gets there it is complete and accurate. No longer does the Call Centre act as a filter between requester and provider to paper over the cracks between disjointed systems and ask those awkward questions about mother's maiden name. Now the **data presented** through the internet **must be complete, appropriate and accurate**. Anything else is, at best embarrassing and at worst, producing dissatisfied customers that could ultimately result in legal action.

Data that is central to achieving the goal of achieving web-enablement is likely to be sourced from the organisation's operational systems and, more often than not, is: -

- **Buried** data beneath millions of lines of legacy code that nobody understands and with no or out-of-date documentation
- **Formatted incorrectly** for use by modern applications
- Of **dubious quality** with gaps, partially and incorrectly populated fields
- **Replicated** across many systems and poorly synchronised
- **Inconsistent** formats
- **Poor consistency** of data values that make the task of matching various systems together to provide a 'customer-centric' view difficult to achieve without very labour intensive procedures

These typical Information Quality issues generally gain visibility to senior management when organisations are trying to use their data for purposes other than that for which it was originally intended. At that stage it is too late. The main project is already behind schedule and **Information Quality is already on the critical path**.

Web enablement is the most recent cause that uncovers these issues. Also valid and likely to cause the same level of concern in an organisation are movement to other operational platforms, packages or to feed analytical data warehouses for customer analysis, segmentation and to permit amongst other things, targeted cross selling. A recent example of an obvious data quality problem is –

Scottish Widows offers tax-free health cover to 100 dead people

Dignified image: Scottish Widows has apologised

By Simon de Bruxelles

A COMPANY called Scottish Widows might have been expected to know better.

The insurance firm, based in Edinburgh, apologised yesterday for offering health cover to 100 people that it knew to be dead.

The administrative error came to light when a widower received a letter addressed to "Mrs Mair Watkins (Deceased)".

David Watkins said yesterday that he had just been starting to get over the loss of his wife Mair, who lost her battle against cancer last summer.

The letter asked Mrs Watkins: "If you became too ill to work, how would you cope financially?" The mail shot included a "tax-free" mock cheque also made out to "Mrs Mair Watkins (Deceased)".

Scottish Widows admitted that similar letters had been sent to 100 other deceased people across Britain and said it would be apologising to their families.

Mr Watkins, 67, said: "I came back from holiday when this letter came through. It was so upsetting and completely insensitive and brought the awful memories flooding back." Mr Watkins, of New Inn, near Pontypool in South Wales, nursed his wife through breast and liver cancer before she died at the age of 65.

"She was a lovely, kind, sensitive woman," he said, "and her death came as a huge blow to me. This company were obviously aware that my wife had died. I can't understand how they can make this sort of mistake."

A spokeswoman for Scottish Widows said yesterday that more than 400,000 letters had been sent out as part of a new marketing campaign, including 100 to people recorded as deceased in company files.

She said: "There has been an administrative error. We normally make checks to ensure this cannot happen but that has failed in these cases. We are sending out follow-up correspondence to the families of the people concerned to apologise for our mistake."

The Times - Friday 12th July 2002

Most organisations are doing something is these areas. Is yours?

All of these issues are correctable - the problem for organisations is to know how large (or small) the problem is and consequently how costly and time-consuming it will be to fix it. **The first step towards correction must be to understand the size of the problem.**

Origins of the problem

Whilst sounding like a long list of feeble excuses, the reasons are perfectly valid and common causes of the issues mentioned above have arisen from history – the way systems were cost justified, approved, analysed, designed and constructed in the early days of IT (DP in those days) and ever since implementation, have been enhanced, corrected and extended beyond their original purpose.

The origins of Information Quality issues are manifold, the most common being: -

a) Data entered when validation was lax

b) Faulty program logic generating corrupt data

c) Company Culture – emphasis on productivity (never mind the quality, feel the width)

d) Stove Pipe IT systems store the same data multiple times and differently

In its cosy home-system, these Information Quality issues may not cause problems but lie in wait until the data is put to other purposes that which it was first designed. In today's world this vital company asset is used to understand the organisation's customer base, competitive marketplace, sales opportunities, follow and predict trends and often determine expensive marketing strategies. This erroneous data will provide a flawed base on which to base company decisions.

Cost Analysis

Lax Validation

As IT systems evolve, data has been loaded through many different validation rules and faults in processing logic have caused erroneous data to abound. The validation and processing rules may have been tightened but the old data remains in the system unchanged. IT departments are often amazed when confronted with data from live systems uttering words like "don't know how that got there – definitely a DBA problem!".

Flawed Program Logic

Programs that fail are an annoyance but their impact is obvious as the system grinds to a halt. Other programming faults go unnoticed or unreported but, nonetheless, data is being left corrupt in your databases. Even if reported and the offending program is corrected, the corrupt data is left unchanged if the system runs. 'If it ain't broke don't fix it'. The legacy of the fault lives on.

Company Profits before all else

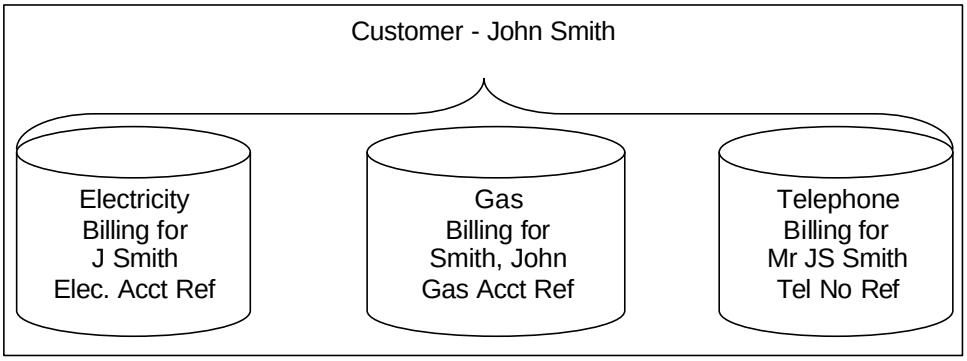
Nobody can argue that companies have to make profits to survive. The cost of poor Information Quality has already been discussed and directly conflicts with the short term gain of immediate profit. The longer term legacy is costing you now.

Question - Why is it that an insurance company client's accident claims are 90% in the Ankle Injury category? Answer - Ankle Injury is represented by the code 'A' and that key is closest to the Tab key on the keyboard. It's the quickest valid option for time-pressed data entry clerks. This highlights a 'Company Culture' issue that is driven from high in the organisation. Data entry staff incentivised on productivity will be focussed on doing their job in the minimum time and few keystrokes as possible. Data quality will be compromised with missing data (typing that lot slows me down), misleading data (financial amounts of pence rather than pounds) and valid but incorrect data (keys grouped around the 'Enter' and 'Tab' keys get hit a lot accidentally). All result in information being lost forever.

Arguably this is soft money, but incorrect and inaccurate data can cost hard money as well. Names and addresses are a common area of concern when organisations try to mail-shot their customers and take full advantage of Post Office discounts sorting letters according to post code only to find that many post codes stored in their system are missing, stored in the wrong field or are wrong. This is often only noticed when expected discounts are not granted.

Stove Pipes

The historical design and implementation of IT System effects data quality as well. The early concept of building stand-alone IT systems to support a part of the business has led to the 'Stove Pipe' scenario where one organisation now has a plethora of customer information stored in many diverse IT environments. The net effect is that each system stores what should be common data but in its own database and probably in different formats and certainly with different values.



Our example shows that John Smith has 3 accounts with his utility provider - an electricity account under the name of J SMITH, a gas account under the name of Smith, John and a telephone account under the name of Mr JS Smith.

To compound the problem, in our example, Mr John Smith will have different customer references and unique keys so linking the three John Smith accounts is near impossible without significant levels of manual effort and individual judgement. To achieve a 'Customer Centric View' of the business, this disparate and duplicate data has to be matched, de-duplicated and merged into a central standard format. Matching is no mean task – what do you match against what? Name is not enough – how many John Smith's are there? Name and address will refine the criteria but addresses are a source of notoriously poor quality data. Post Code can be an accurate fragment of data but how do you distinguish Joanna Smith and John Smith living within the same post code or house. There may be very good reasons why Joanna does not want John to know her business.

Ithaca Information Quality Services

To address the need for accurate and consistent data being fed to internet applications, the first step is to understand the size of the problem. Ithaca IQ Services cover this and two subsequent stages: -

- Analysis of Data Quality and Data Duplication (what is the problem now)
- Strategy and Plan (where do you want to be and how to get there)
- Cleanse and Consolidation Implementation (achieving the objective).

Analysis of Data Quality and Data Duplication

Analysis of the business data is a pre-requisite to all other steps and analyses each data store and system that accesses it and highlights where quality issues exist, where duplication issues exist. Each issue raised will have a recommendation on how it should be resolved taking into account: -

a) Requirements of the system that can be disturbed by re-introducing cleansed data;
b) Where data is used both in-situ and downstream e.g. in warehouses.
c) Volumes of data that require cleanse - whether cleanse can be automated manual
d) What options are available for matching and merging data

Strategy and Plan

Once analysed, and a thorough understanding of the problems is gained, Ithaca will help you determine a Data Cleansing Strategy to get to the organisation's goal whatever that may be.

This service is aimed at defining the most effective strategy for: -

a) Defining the environments required to implement the Cleansing Strategy
b) Defining what activities are required and in what sequence
c) Which software tools can be used effectively to assist the cleanse
d) Where the cleansed data is stored
e) What changes are required to the source system(s) to prevent further data corruption
f) What audit and reconciliation is required on the changing data.

Cleanse and Consolidation Implementation

Once the Plan is established, Ithaca can manage the Cleanse and Consolidation process according to the plan. It includes all the professional project management control activities such as: -

a) Progress Management against plan
b) Progress Reporting
c) Risk Management
d) Change Management

In parallel with the Management activity, execution of the plan can be performed with the assistance of Ithaca's consultants to guide, help and advise your staff at the desktop.

Contact

If you would like to talk further about how Ithaca can help you determine a workable strategy to realise some or all of the above contact us at:

Ithaca Solutions Ltd,

London Office
5 – 11 Lavington Street,
LONDON,
SE1 0NZ
England

Telephone+ 44 (0) 2079 456296

Belfast Office
Unit 27
Ormeau Business Park
8 Cromac Avenue
Belfast
BT7 2JA
Northern Ireland

Telephone+ 44 (0) 2890 232023

Email sales@ithaca-solutions.com
Web www.ithaca-solutions.com