EXECUTIVE INSIGHT PAPER

Managing Customer Data Quality

A View for IT Leaders and Architects

Freeform Dynamics Ltd

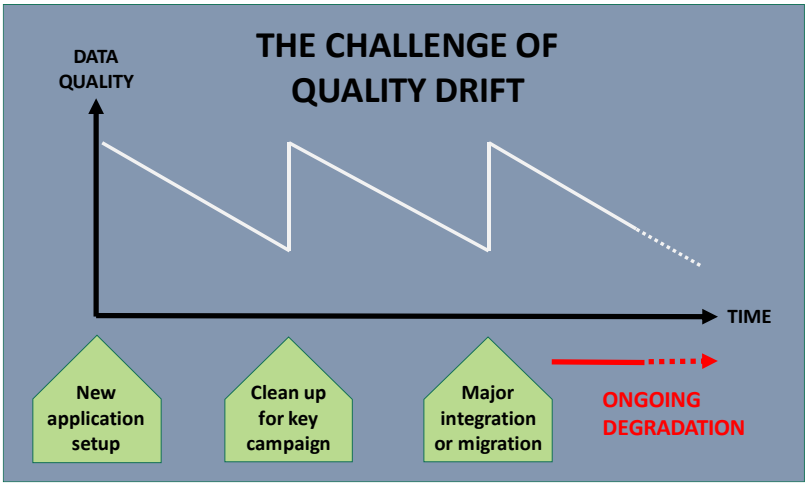
November 2013

This paper was authored on an independent basis by the analyst team at Freeform Dynamics Ltd, with sponsorship from Experian. The content is based on industry intelligence gathered and analysed by the authors and does not necessarily reflect the views of the sponsor.

Introduction

When was the last time you could look at one of your customer databases and have confidence that it was largely accurate and complete? If your organisation is anything like most of its peers, then the answer is likely to be immediately after your last big application implementation, integration or upgrade project. And if that project was a while ago, the chances are that the quality of the information in the database concerned has since degraded significantly. The same is also likely to be true of other customer databases that exist across the business.

So why is this?



Well the problem is that customer data quality is typically managed through a series of periodic discrete exercises. We might clean a dataset before importing it into a new system, but then it's left alone until an event occurs that prompts looking at it again. Triggers here include things like a system upgrade, preparing for a major campaign or market initiative, the results of some kind of compliance audit, or simply poor data quality impairing business performance to a degree that someone with influence declares to be intolerable.

Meanwhile, a number of factors contribute to the gradual but inexorable downward drift in quality:

- Data entry errors on the part of employees or business partners
- Erroneous or deliberately inaccurate information entered directly by customers and prospects via your website
- Inaccurate or duplicate data sourced externally and loaded into your systems
- Failure to detect and record changes in customer details
- Bugs, limitations and out-dated functionality in applications and/or integration software

The mention of applications and integration software here is particularly important because it's often the sheer complexity and dynamic nature of the systems landscape that undermines any attempts to tackle the data quality problem in a sustainable manner.

If your organisation is medium or large in terms of size, for example, you will have a number of databases in place that hold customer information[1], and various pieces of middleware and custom integration components for moving and synchronising data between systems. And where direct integration between applications isn't in place, the situation is aggravated by error-prone manual processes that involve the rekeying of information from one system into another.

With the best will in the world, it is difficult to ensure that data validation rules and information update policies are applied consistently across any complex landscape, especially as the environment, rules and policies are all subject to change. Put all this together, and these are the reasons why customer data quality tends to drift downwards over time. Then of course, fixing it in any particular system becomes quite a significant undertaking, so with limited resources and competing priorities in more urgent or glamorous areas, it continually gets deferred.

If any or all of this sounds familiar, then we encourage you to read on. There is another way of thinking about and tackling the problem that allows data quality to be sustained on an ongoing basis.

An alternative approach

In order to maintain a high level of accuracy, completeness and consistency within your customer databases on an ongoing basis, it is necessary to break some old habits. You need to get away from dealing with data quality through discrete projects, and implement quality management as a continuous process. This in turn needs to be underpinned by a robust enabling architecture and set of services.

To make this switch, however, you must first accept the fallibility of both people and systems. Whether it's employees or customers, you cannot get away from the fact that they will make mistakes and sometimes deliberately introduce inaccuracies. And from an IT perspective, things are never going to be perfect in such a complex and continually changing environment.

The upshot is that it is unrealistic to expect to solve the problem by 'fixing' people and core business systems completely. So let's do what we can through policies, training and best practice, but in the meantime get on with tackling the data quality problem more directly.

Of course an important pre-requisite for spending time and effort developing an alternative approach is a clear understanding of the rationale for doing so. The key question is why business stakeholders should care about data quality issues in the first place.

This is something we discuss in some depth in our companion paper[2] entitled 'Customer Data Quality in Context: A Business Perspective'. Suffice it to say for now that poor quality customer data can undermine activities across the whole of the business cycle, from marketing, through sales, to customer services. Along the way, there can be a negative impact on customer acquisition/retention and operational efficiency, along with IT responsiveness and flexibility. Creating an environment in which a high level of customer data quality can be efficiently maintained therefore results in significant tangible business benefit.

Before getting into the mechanics of implementing a more process-driven and architecture-enabled approach, however, it's worth spending a few minutes on some important practical considerations.

Practical considerations

To begin with, let's be a bit more precise about the nature of the quality issues that arise in the context of customer data, specifically when the customers concerned are consumers:

- Incomplete records (e.g. missing address details, gaps in personal data)
- Inaccurate records (e.g. wrong address details, incorrect date-of-birth)
- Duplicate records (e.g. same person, but possibly with different details attached)
- Inappropriate records (e.g. prospects 'gone away' or with marketing preference flags)

These examples are here to give a flavour of the basic quality and integrity checks that are generally applicable to most types of consumer-facing businesses. Other information that might be relevant in the context of your organisation could include bank details, contact methods, social media identities, relationships to other people (family members, same household, etc.) and markers relating to demographics, occupation and other indicators useful for analysis and segmentation.

Records relating to the same customer or prospect are likely to exist in multiple databases within your organisation, and in an ideal world, all of these would be synchronised. The problem is that each record is likely to hold a different subset of the information you would like on file. Indeed some desired information may not be held in any of your systems and databases at all, which is something we pick up on a little later when we talk about the concept of 'data enrichment'. Also, of course, each database will have its own set of quality issues associated with the systems, processes and people involved in populating and updating it.

Given these factors, one of the biggest challenges when tackling data quality and integrity is deciding what represents the 'truth' when it comes to any given customer record. It's rarely the case that you can point to one database as the reference for everything because there will almost certainly be information that's important elsewhere that is not included in its record structure. Even if you do identify a single reference point for certain attributes (e.g. current address), you need to be careful not to propagate the inaccuracies that creep in via the mechanisms previously discussed.

SOME IMPORTANT QUESTIONS

How to determine the best point of reference for assessing accuracy and completeness?

How to keep databases synchronised without propagating errors and inconsistencies?

How to gain programmatic access to databases to facilitate routine cleansing efficiently?

The last practicality we will discuss before looking at the solution side of the equation is how data is accessed for cleansing purposes. An obvious way of achieving this is through exporting whole data sets, cleaning them up, then reimporting them. This may be the best way to do it in the context of a big migration project, but is far too cumbersome for routine use. This matters because if we are going to adopt a process-driven approach to maintaining data quality on an ongoing basis, we need a more elegant and efficient way of going about things.

With that, let's move on to consider how we can deal with some of these practicalities.

Blueprint for a sustainable solution

It's beyond the scope of this paper to go through the technical specifics of different solution elements, but we can define a high level blueprint that will help you to identify and engage relevant IT vendors and service providers. First, however, we need to be clear on what we are trying to achieve:

OVERALL OBJECTIVE

To create an environment in which customer data is validated and enriched at the point of entry or capture, then kept in shape thereafter through regular and frequent scanning, cross referencing, correction and completion.

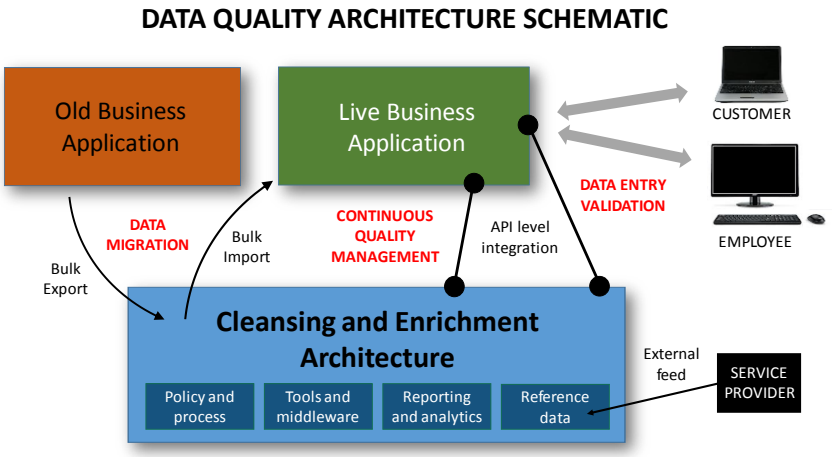
One of the keys to sustainable quality is the injection of external reference data into the mix from a service provider whose core business revolves around maintaining accurate and complete customer information. Beyond reference data, the right partner will also provide tools and knowhow to deal with the way in which data is accessed and moves around. Elements here include:

- Specialist knowledge and skills so you don't have to work everything out from first principles when assessing your current situation and implementing appropriate technology and best practices.
- Tools for working on customer datasets in bulk, e.g. when records need to be cleaned as part of a data migration project, or enriched with additional information to drive a marketing campaign.
- Software components that can be embedded behind web forms or other application interfaces that validate/complete customer data at the point of entry to minimise the accumulation of errors, gaps and inconsistencies in live systems.
- Rules based software that integrates with key applications such as ERP and CRM through standard APIs to facilitate regular and frequent scans of live customer databases, comparing records against reference data and correcting/extending as necessary.

An example of how this last type of software might be used would be a batch job that executes nightly which runs through all new or updated customer records in an SAP system. As it executes, it will

correct errors and add missing information based on lookups to a combination of other internal databases and/or external reference data. Any conflicts, duplicates or inappropriate records may be logged as part of this process for dealing with later by the data management team.

A simplified picture of what you might end up with when you put all this together is shown in the schematic to the right.



Of course when it comes to implementing something like this in practice, the Devil is in the detail. The good news, however, is that you don't have to work it all out from first principles. Advice and guidance is available from vendors and service providers to help you with audit and assessment, requirements definition, architecture design, solution implementation and adoption of relevant best practice.

And you don't have to try to do everything at once. In fact, it is often better to begin working with one or two key systems that are of particular concern to begin with, then once you have experience with the tools and processes, start to embrace other customer databases. The ultimate aim is to create a process driven framework designed to counter that negative drift in quality and make sure your customer information remains in good shape on an ongoing basis.

Final thoughts

One of the biggest impediments to tackling customer data quality is lack of attention and priority. This is often down to a combination of inertia and not realising what can be achieved with the latest technology and techniques. However, it's also easy for initiatives to be starved of investment as attention is drawn to more glamorous areas such as Big Data and Social Media, which are often positioned as the latest enablers of better customer engagement.

An aggravating factor here is the frequently heard notion that it's not worth worrying about data quality anymore. Big Data evangelists point out that parallel processing systems such as Hadoop can crunch though any data set, no matter how dirty, incomplete and inconsistent, and still produce useful business insights. That's as may be, but there is a world of difference between looking for trends and correlations at an aggregate level for review and planning purposes, and providing clean and accurate information to enable the day to day operation of the business. Furthermore, regardless of the rhetoric, operational analytics based on specific customer activity, e.g. to help with decision-making on the front-line, or to plan the detail of a marketing campaign, still requires good data.

But overriding such practical matters, there is the basic question of ROI. No matter how seductive the promises of Big Data and Social Media might be, the truth is that benefits are typically unpredictable and hard to quantify. Meanwhile, a well-designed and executed customer data quality initiative will deliver tangible and sustainable returns in a short space of time across many areas of the business.

So, while we would not discourage investigating the latest 'hot' technologies, this should not be at the expense of addressing the fundamental business information requirements discussed in this paper.

References and further reading

1. Storage Anywhere and Everywhere: Dealing with the Challenges of Data Fragmentation
2. Customer Data Quality in Context: A Business Perspective

The above papers and reports are available for download at www.freeformdynamics.com.

About Freeform Dynamics

Freeform Dynamics is a research and analysis firm. We track and report on the business impact of developments in the IT and communications sectors.

As part of this, we use an innovative research methodology to gather feedback directly from those involved in IT strategy, planning, procurement and implementation. Our output is therefore grounded in real-world practicality for use by mainstream IT and business professionals.

For further information or to subscribe to the Freeform Dynamics free research service, please visit www.freeformdynamics.com or contact us via info@freeformdynamics.com.

About Experian and Data Quality

Experian QAS has built up exceptional market coverage assisting customers with their unique data quality challenges. We provide a comprehensive toolkit for data quality projects combining our market leading software with a vast scope of reference data assets and services. Our mission is to put our customers in a position to make the right decisions from accurate and reliable data. The size and scope of data management projects varies considerably but the common factor in all ventures is unlocking operational efficiency and improving customer engagement. We see the potential of data.

Our solution portfolio includes:

Data quality tools
Take greater control of your data accuracy with our comprehensive range of data quality tools for public and private sector organisations.

Data quality services and consultancy
Struggling to implement a data quality strategy? Our data consultants have a wealth of experience executing data related projects.

Data quality for enterprise applications
Improve the return on investment from your enterprise applications with our seamless data quality integrations.

For more information visit www.qas.co.uk.

Terms of Use