

# Using customer data? Start with clean data

 By [Martin Doyle](#)

30 Jan 2015

Office, the high street shoe retailer, has had something of a lucky escape when it comes to its data quality.

Last week, the Information Commissioner's Office decided not to levy a fine on the company for a massive data leak, although luck had its part to play in the outcome.

In May 2014, hackers gained access to Office's servers and stole customer information relating to more than a million past customers. Sensitive data, such as postal addresses, was exposed, and the hackers were also able to steal passwords from the database. Fortunately, none of this data appears to have been used for malicious purposes.

For many businesses, the consequences of a hack are severe. Fines, bad publicity and compensation payments can have serious consequences for profitability. This is why master data management is a key concept in information security, and achieving a state of security and consistency is critical.

## Lucky sscares

The reason the Office hack was so serious was because the database was so large, and so old. It contained out of data, 'dirty' contact information - records that should have been deleted when a new database was brought online. Instead of quickly merging the old database with the new one, Office held back and chose to sit on the old data as it decayed and sat neglected on a server nobody remembered was there.

The old database was taking up space, for one thing, and costing money by the by, but that's not the worst of Office's worries. It was stored in a location nobody monitored, and it was fast becoming a liability for the business without it even realising.

Unfortunately, when the hackers got in to the old database, there was plenty to steal. The data had not been deduplicated, matched and merged with new records. What's more, the old database was unencrypted, and that was the final flaw that exposed this vast dataset to prying eyes.

According to Gartner, businesses have become accomplished at simply coping with bad data, rather than doing something about it. Rather than tackling the data quality challenges of merging data and perfecting it, Office chose to simply hide the data, which meant that customer details were held without proper checks and governance.

## Dealing with Old Data

All businesses are faced with upgrades and data migration at some stage in their evolution, and the need for merging of datasets becomes greater as the business matures. Legacy systems get phased out and replaced, and employees ditch old ways of working when better solutions are brought online. Often, compliance guidelines force a change in the way data is stored and managed, and new team members can bring new systems and fresh ideas.

In Office's case, staff felt that migrating the old data was risky. One of its key concerns was an inability to match the old data to the new data. As such, there was a problem with duplicates from day one. But managing data properly means that the business needs to understand how data is being used. If that means dropping old databases completely, that's what has to be done.

Yes, businesses are right to approach data migration with caution. Invalid entries are a huge source of data quality problems, and merging two datasets can be a source of huge data corruption. This can result in confusion for staff; fields that should contain fixed values may contain all kinds of invalid results, and this can even stop records from saving when they fail automatic validation checks.

## **Customer Control**

Businesses that retain personal data have to work within the law, which compounds the risk that poor data quality presents. The holder of data must make sure that it is accurate, and held for a reasonable period of time. These requirements are not new, yet businesses are still failing to address the risk that poor data quality presents, and failing to spot the obvious danger signs early on.

It's easy to see why the Information Commissioner objected to the legacy system Office was using. It was uncontrolled, unmonitored and completely lacking updates. It could also be argued that customer data was held for far longer than it should have been, given that a newer system had already been brought online.

## **Lessons in Data Management**

The Information Commissioners Office describes data as "vulnerable". Using the same analogy, dirty data is data at its most exposed. This isn't a case of increasing security (although an unencrypted database is clearly going to be vulnerable). It's not simply a case of the IT department stepping up their controls.

In Article 5 of the Information Commissioner's standards principles, there's a clear requirement for data to be deleted as soon as it's no longer needed. Clearly, the Office data breach proves why this is so important, and there needs to be a clear process and policy in place. Old data should be disposed of, new data cleansed at the point of entry, and ageing data regularly checked and managed using automated data quality solutions.

Managing data is also far easier if you don't hold records you don't need. If you delete records that you know are out of date, there are fewer risks to the business when you try to merge them. If you remove known duplicates using data quality software, there are fewer risks of customers being inconvenienced with wasteful duplicate communications.

## **Planning for Success**

Data quality initiatives require strategic planning and concerted effort, and that means treating old data exactly the same as new. Just as new entries are filtered using form fields and validation, old data should be subjected to the same standards and checks.

A customer data warehouse is a key business asset, and it's an asset your customers expect you to value and protect. Data governance requires the right people, the right funding and sustained effort, but the reward is an error-free dataset that does not expose any party to unnecessary risk.

## ABOUT MARTIN DOYLE

Martin Doyle is a Data Quality Improvement evangelist and runs a software development business, DQ Global, which likes to help businesses succeed in their CRM and MDM initiatives through the use of higher quality data. you can contact Martin on [martin.doyle@dqglobal.com](mailto:martin.doyle@dqglobal.com)

- How a Chief Data Officer can make your data great - 27 Mar 2015
- Using customer data? Start with clean data - 30 Jan 2015
- Is the data in Your CRM a ticking time bomb? - 15 Jan 2015
- Is your CRM data the elephant in the room? - 5 Jan 2015

[View my profile and articles...](#)

For more, visit: <https://www.bizcommunity.com>