

Maximise the value of your data

Data is an essential business asset. It may have no physical form, but its value is far greater than any hardware on which it might be hosted.


The problem for many organisations is that while they may hold a lot of data – often, far more data than they realise – it is in a far from optimal condition. It may be incomplete, lack consistency, and be bloated through the inclusion of duplicate records.

Red Olive has helped many organisations, including utilities, charities and multinational publishing corporations, to optimise their data. In doing so, we've increased the data's intrinsic value, and its suitability as a tool for making actional business decisions. Our clients have reduced their ongoing costs, maximised profits, and in many cases contributed to the betterment of society as a whole.

The advice in this guide will help CTOs, CDOs and project managers to better understand how to structure their data, so it can be used to make informed business decisions. Our six step process for implementing and maintaining good data hygiene is designed to help businesses reduce costs, maximise revenues and gain a more effective overview of their live operations.




Red Olive was founded by a team of experienced Analytics, Business Intelligence and Data Management Professionals with a love for creating tangible business improvement. Our aim is to improve our clients' business performance fast, by helping them apply one of their greatest assets, their data, more effectively.



It starts with a question: what tools and resources do you need to make better, more effective business decisions? When a company can answer that question, it can compare the answer with what it has to hand.

A lot of organisations hold more data than they realise but, when they start to examine it, they then discover that its value is low, as it falls below an acceptable threshold. Red Olive is working with a housing association that manages more than 10,000 properties. At the start of our engagement, it held mobile phone numbers for just 20% of its tenants, so was unable to send text alerts to four in every five of the families that relied on it to provide and

A photograph of a modern library interior. Tall, white, modular bookshelves are arranged in long rows, receding into the distance. The shelves are filled with books of various colors. The floor is light-colored wood. The ceiling has exposed concrete beams and some lighting fixtures. The perspective is from the end of a row, looking down the aisle.

maintain their homes. More seriously, this lack of data also stopped its digital journey in its tracks, since the quality of its data was so low that the association couldn't consider any new initiatives until it was thoroughly cleaned up.

It's not the only organisation to find itself in this position.

The housing industry is going through a process of consolidation, which can itself lead to several new complications. Not least of these is the need to integrate multiple data sources. As soon as one association, or any other business, acquires another, it also acquires its existing data. This could be in an entirely different format to that with which it's already working – and, in many cases, could be incomplete or poorly organised, with mal-formed entries, empty fields and duplicate records. Data protection regulations may have prevented it performing a thorough audit on the data before it came into its possession.



Understand what good quality data looks like

1 At Red Olive, when evaluating data quality, we consider six dimensions: accuracy, completeness, consistency, timeliness, validity, and uniqueness. Businesses hoping to get the most from their data need to do all they can to excel in each of these areas. If they don't, the data will be incomplete, and unsuited to helping them make effective decisions.

Each organisation has its own data 'issues' which depend on the data source, size of the data it's maintained, and even the type of data concerned. However, some issues, like problems with dates of birth, arise more often than others.

A subject's date of birth should be checked at the point of entry, through a passport scan, request to DVLA, or

cross-referencing another data source. Failure to do so can result in basic but serious problems, like a housing association sending letters to residents who have died, or to tenancy holders whose date of birth suggests they must be children.

Why does this happen? Because whoever was responsible for entering the data was focused on just a subset of the six pillars of data quality. Entering the data would satisfy the 'completeness' pillar, and it likely satisfied the 'consistency' pillar if it included a day, month and year. Yet, it will have failed on 'validity'. Such basic mistakes can – and should – be picked up by embedded routines, which can be set up to ensure only quality data makes it into the permanent record.

Know when to allow some leeway

2 Data isn't always entirely right or wholly wrong – and incomplete or inaccurate data isn't always a deal-breaker.

Where a date of birth should be easy to get right and an accurate address is often a requirement even though it won't always conform to a strict format, other metrics – like National Insurance numbers – aren't an essential data point for every business. In many cases, therefore, leaving this field empty or incomplete should not pose a problem.

It can therefore make sense to implement a system of grading, where the quality of the data can be assessed through comparison to defined thresholds. A system of red, amber and green traffic lights, for example, may determine whether a record should be rejected upon input, flagged for future amendments or accepted as true. This can prevent dirty data from contaminating an organisation's resources. Similarly, values like bronze, silver and gold can be used to grade the importance of a field. Those ranked gold, for example, might accept only complete and verified data, with increasing leeway given for those ranked silver or bronze.

This allows an organisation to balance the requirement to compile accurate and complete records with its need to start using data in a timely manner to drive business decisions.

Set short deadlines and realistic goals

3 The first step in building a system to improve an organisation's data quality is to perform a thorough audit of the data it already holds, and to gauge how the set conforms against at least three of the six pillars of data quality: accuracy, completeness and uniqueness. Step two is to clean up any records that need further work, and step three is to maintain data hygiene on an ongoing basis.

Organisations frequently understand the need to perform steps one and two because they have identified for themselves that the data they hold isn't entirely fit for purpose. However, it is only after working with Red Olive that they fully appreciate the need to continue the work on an ongoing basis.

To facilitate the third step, Red Olive uses its own dashboard framework to build systems to monitor the quality of an organisation's data. These allow our partners to see on a monthly, weekly or daily basis how the quality of their data is improving or degrading.

Should any processing be required, it will necessarily take place while the organisation remains operational. It will therefore need to continue accessing the data Red Olive is working with, and the data set itself will continue growing. We therefore adopt an agile approach to project management, working in sprints with short deadlines and clearly defined goals to deliver actionable results at regular stages. Improvements are thus incremental, allowing the client to gain value more quickly, and to evaluate whether and where further changes could be made in real time.

Appoint an appropriate internal liaison

4 A decade ago, data quality projects would have been managed by an organisation's chief technical officer or head of IT. Today, however, an initial approach will more often come from sales, marketing, the Chief Information Officer or the Chief Data Officer. Such roles usually have a better understanding of what the organisation could achieve if its data was in better shape.

That's because businesses don't want to improve the quality of their data purely for the sake of it. The process is always driven by a business imperative, which could be greater compliance, cutting ongoing costs or increasing revenue.

What's unique about Red Olive's offering is its understanding of data quality not as an entity, but a process, which is conducted over time. The discussions we hold at the

start of any project don't tend to focus on data – and neither do they concern themselves with technology. Instead, we want to know more about business processes and what the organisation wants to achieve with both the data it holds and the data that's likely to become available going forward.

It's also helpful to know how that data is collected. If it's generated and entered manually, there's greater potential for it to contain errors and inconsistencies, but if it's captured automatically, perhaps using Internet of Things (IoT) devices, any amendments it might require will more often focus on conversion from one format to another, rather than finer-grained checking, completing and correcting.



Automate, where possible

5 Red Olive is working with several organisations operating in the insurance industry, where data quality is a serious and ongoing issue. Many are running antiquated systems and constantly working through a migratory process to find the 'next best thing'. In doing so, they implement multiple upgrades, which frequently requires their data to be restructured, reformatted or rewritten. Much of the work that Red Olive is conducting in this sector is focused on reconciliation, to ensure that the data emerging at the end of the process is representative of that which was fed in at the start.

When moving data between platforms like this, the source system is unimportant, so long as it is accessible using an API (Application Programming

Interface), to which we connect using our in-house Python-based framework, ROSIE DQ. This manages the initial connection, data extraction, and production of a data quality report, which is generated by running algorithms. Once embedded within an organisation, ROSIE DQ can monitor its data on an ongoing basis, and perform remedial tasks, where required, to improve the data's conformity with the three most important pillars for gauging its quality.

Informed by the report, we can then perform a further data extraction and apply any required transformations, before storing it a new repository where we know the contents are clean and high quality. In this way, we can combine the resources of several, often competing CRM systems. This is essential, as it's not unusual for many organisations to be using in excess of 12 data sources, and often dozens more, on a daily basis.



Don't forget your unstructured data

6 Not all data is immediately quantifiable. A lot of it is unstructured, but still needs to be made accessible as it contains valuable business insights that can be used to drive efficiencies, make cost savings, and increase revenue. Frequently, it is stored in formatted documents, like PDFs, Word files and Excel spreadsheets, rather than fields in a database. It wouldn't be appropriate to try and extract this for conversion to a different format, so instead we would look to make it accessible in its original format, with its context preserved, by indexing and tagging with metadata.

If we return to the example of housing associations, they will frequently be in possession of a wide range of materials, like instruction manuals and warranties for equipment installed in their properties, plans, and building materials specifications. Having these not only 'to hand' but easily accessible through indexing,

tagging and organisation is essential as it enabled them to make repairs in a timely manner. It also reduces costs by allowing them to check whether they comply with regulations and inquiries, often without a site visit.

Such a situation arose in the aftermath of the fire at Grenfell Tower. The blaze started on a lower floor, but quickly spread when it ignited the cladding used on the outside of the building, resulting in the deaths of 70 residents. In the aftermath, freeholders were required to check which of their buildings used similar cladding and, where appropriate, remove or replace it.

Having details of the materials used in the construction and subsequent maintenance of each building in its portfolio would allow a housing association to more quickly identify which buildings were affected without the expense of a manual audit. It could then affect repairs in a timely manner to keep its tenants safe.



It's time to get more from your data

Whether you're ready to audit and improve your data, need to develop a data strategy for the future, or you just want to talk through your options, call Red Olive on **020 3745 9790** or email **hello@red-olive.co.uk**

