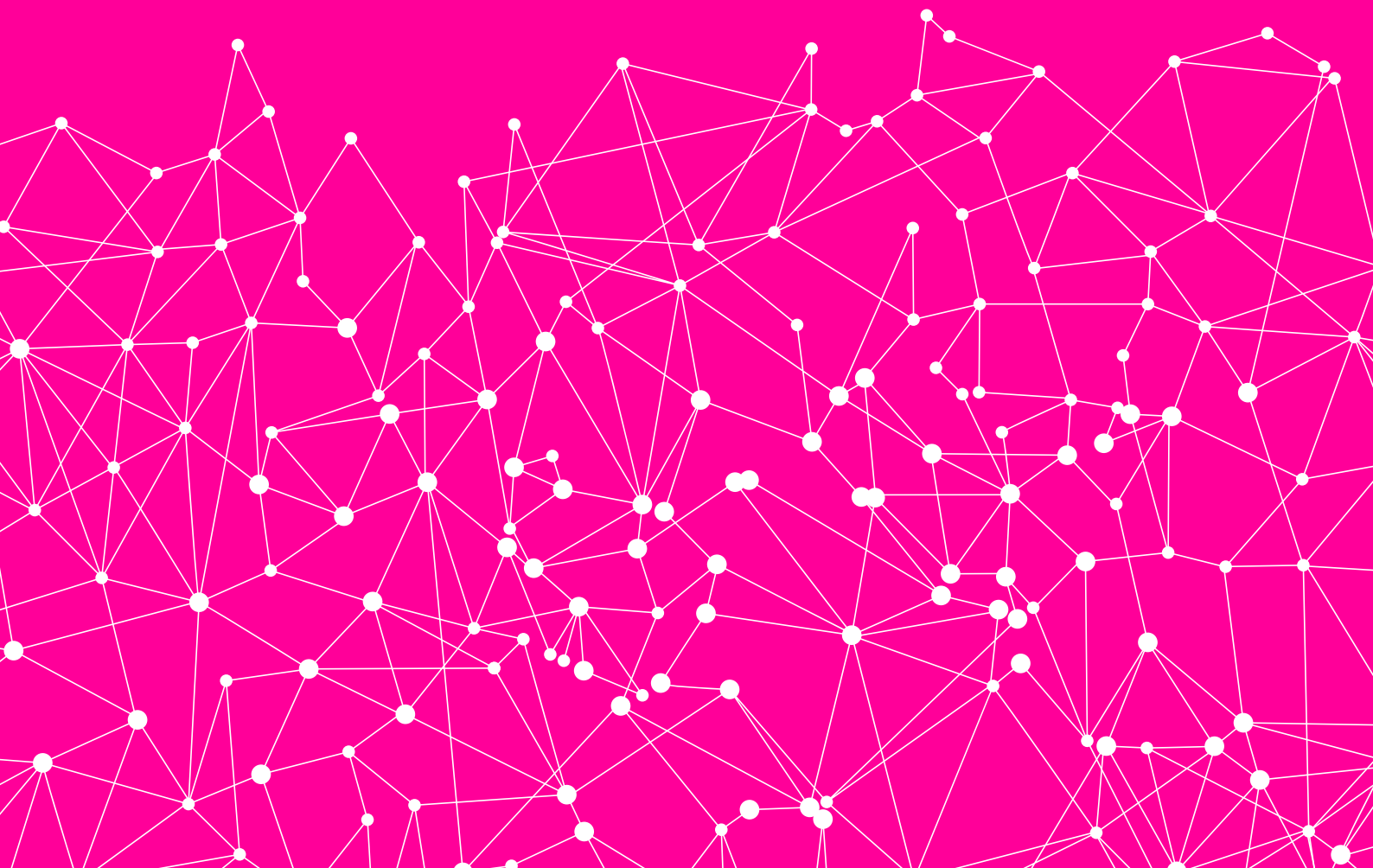
Sagacity

Are you addicted to bad data? Get clean and stay clean

gma Global Marketing Alliance
INSIGHT FOR DATA DRIVEN MARKETERS

Contents



Foreword

We are now well and truly embedded in the 4th industrial revolution. With the transformative development and adoption of technology has come the dramatic increase in the volume and importance of the fuel that drives that technology - data.

Having been working in the data industry for more than 30 years, I have seen first-hand the dramatic growth and impact that data is having on consumers and businesses. The stats are staggering - IBM research from 2019 calculated that 90% of the data that exists today was created in the last two years and the IDC estimate that over 2.5 quintillion bytes of data is created every day at the current pace.

Stop and think about that for a moment. That's a phenomenal amount of data to collect, process and hold, however data savvy you are!

And with this vast – and ever growing - volume of data comes greater responsibility. In the GDPR era, that responsibility has legal implications and the potential for business damaging fines and reputational damage. The old adage 'garbage in, garbage out' has never been more true or more important for businesses. Good quality data means better business decisions, better marketing and more profitable relationships. As a strong proponent of responsible data management and, transparency, I am talking to clients and partners every day about data quality and I am encouraged that this is now very much on the agenda.

I hope you enjoy reading this whitepaper and gain some valuable insights to inform your data strategy now and in the future.

Ultimately, the message is simple - treat your data assets with the respect and care they deserve and reap the many rewards.

Jon Cano-Lopez, Sagacity

Executive Summary

The exponential growth of data and its strategic importance to organisations of all sizes, has thrust the issue of data quality into the spotlight. By drawing on authoritative industry research, this report examines the scale of 'bad data', its implications for businesses and what can be done about it.s are affecting everyone in society, not just lower income groups, and reveals the risk of vulnerable people slipping through the net.

Part 1

The data quality challenge for businesses today:

We uncover the scale of the data quality struggle and examine why clean data has become a business essential:

- The IDC predicts that the amount of data in the world will grow from 33 zettabytes(ZB) this year to 175ZB by 2025.
- Organisations believe inaccurate customer data costs them, on average, six per cent of their annual revenues. Over a third are not sure how much it costs them.
- 19% of businesses have lost a customer by using inaccurate or incomplete information – a loss exacerbated in industries where customers have a high lifetime value.
- 42% of companies have struggled with inaccurate data
- The introduction of the GDPR means that clean data is no longer just a business preference but a legal necessity

Part 2

Building a data-aware organisation:

Solving today's data challenge requires an organisation-wide commitment to data quality. Learn about the key components of a mature data-aware organisation and the key areas where attention needs to be focused.

- Creating a data strategy is key, but it must be linked to the wider business strategy.
- Clean data starts with a robust Data Architecture which governs the process by which data is systematically collected, stored, transformed, distributed and consumed.
- Businesses benefit greatly from quality data, however it deteriorates unless cleaned and updated on an ongoing basis.
- 15,407 major life events occur everyday which degrade data quality.
- Understanding your data's 'half-life' (how long it takes for 50% of your data to be inaccurate on a critical data field) helps frame your approach to data cleansing.
- Creating a “Data Quality Culture” is essential and must be initiated at the top of the organisation.

Part 3:

Getting clean: The key constituents of a data clean-up operation

1. Begin with an audit: Establish the current state of your data across the whole organisation.
2. Set clear goals: Effective data cleaning requires a set of agreed KPI's.
3. Collect the right data, the smart way: Standardising contacts at point of entry saves time and effort in the long run.
4. Integrate data cross-team and cross-platform: Technology and people must work together in harmony to slow down the pace of data decay and ensure that data cleaning isn't performed in a siloed way leaving opportunity for error.
5. Validate your data: The accuracy of data can be verified using different methods including or via automative tools for enriching, appending, cleaning and de-duplicating data. In the absence of technology, ensure there is a clear timeframe/ schedule for data cleaning activities.
6. Source responsibly: Clean data is responsibly sourced data. Ensuring that customer data is collected in a transparent way that respects individual's privacy is a legal requirement in light of the GDPR and errors could be very costly.
7. Make the data talk: Data visualisation tools can play an important role in establishing and maintaining a healthy data culture across the business.

Part 4

Staying clean: The future of data cleansing

The data landscape will grow in complexity as new digital technologies create new data streams to add to the myriad which already exist. AI, VR and voicerecognition - among other tech innovations - are evolving at a rapid pace. Meanwhile regulations will have to adapt – not only to new datatech and the ethical questions they pose, but also to the demands of a more savvy and data literate consumer.

The future of effective data cleaning lies in 'always on' API-driven data cleansing methodologies.

Part 5

Questions to ask your data provider

The GDPR has created some concern over the use of third party data. How can you be sure they're legally compliant? And how can you be sure their data is reliable? Here we provide a checklist of questions to ask your data provider.

Section 1:
The data quality challenge for businesses today

Introduction: Talking the Big Data Challenge

The The world is awash with data. The International Data Corporation (IDC) predicts that it will grow from 33 ZB this year to 175ZB by 2025 [1]. It presents both a daunting challenge and a limitless opportunity for marketers and data-driven organisations. Businesses that most effectively manage and optimise data can provide superior services to customers, improve decision-making, drive greater efficiency and achieve assured compliance with regulators.

There exists a variety of ongoing and unfolding challenges posed by Big Data, including:

- The evolving regulatory climate which requires organisations to
- ensure they source and utilise data in a legally compliant way.
- The practicalities of handling the sheer volume and complexity of data across the business in an integrated way.
- The difficulties of pulling meaningful insights from the vast amount of data collected and acting upon them.
- Adapting to the speed of technological change – such as machine learning, AI and unforeseen technological innovations.
- The proliferation of 'bad data' which is inaccurate, outdated or irrelevant.

"Poor-quality data is a huge problem. It leaves many companies trying to navigate the information age in the equivalent of a horse and buggy."

Bruce Rogers
Chief Insights Officer at Forbes Media

Business benefits of clean data

Marketing: A direct marketing campaign using high quality data reaches the right target contact with relevant offers. This drives up sales leads and campaign ROI.

Sales: A sales representative can reliably contact current customers. With access to complete and accurate data, they can ensure no-one is forgotten or missed.

Compliance: Avoidance of penalties and the related brand damage. Complying with the GDPR and other global data protection regulations require companies to maintain clean data.

Better insights and decisions: Accurate data will deliver more accurate insights and enable better business decisions.

Protect your brand: Applying data quality best practice will help to protect your organisation from brand and reputational damage from contacting deceased contacts for example.

The growth of bad data represents a major impediment for organisations trying to maximise the strategic benefits that data can provide, while also posing a compliance risk.

Indeed, research by Royal Mail Data Services [2] revealed that organisations believe inaccurate customer data costs them, on average, six per cent of their annual revenues. Perhaps more worryingly, over a third were not sure how much it costs them.

Losses caused by bad data can occur for a number of reasons, including lost time spent chasing phantom customers (such as duplicate contacts or redundant email accounts), or misinformed decision-making. A recent data management study by Dun & Bradstreet found that 19% of businesses had lost a customer by using inaccurate or incomplete information – a loss exacerbated in industries where customers have a high lifetime value [3].

If bad data undermines our ability to address the big challenges posed by Big Data – and the potential benefit it provides – then it's 'clean data' which has a fundamental role to play in helping us meet the challenge, grasp the opportunity and safeguard ourselves against tightening data protection regulations.

Clean data can be defined as: accurate, up-to-date, uncorrupted and relevant data which has been appropriately sourced. In today's data-driven age, it's the optimal fuel that companies depend on.

Scaling up: making the business case

The more complex the data needs of an organisation, the more difficult it becomes to keep all data sets clean, and the greater the threat bad data poses.

So, while clean data is vital for businesses of all sizes, larger organisations with complex data requirements are those that stand to gain the most through the combination of more effective marketing campaigns, more efficient spend, and lowered compliance risks and associated brand damage.

For example:

- Handling data for deceased contacts or gone aways
- Hold on to personal data longer than is necessary for the purpose it was processed and you fall foul of GDPR Article 5 (e) [4]
- Collect multiple duplicate data sets and it will skew decision-making or corrupt the insights provided by a machine learning algorithm

According to Dun & Bradstreet, 42% of companies had struggled with inaccurate data while 43% had seen 'some' data-led projects fail. These findings appear to link data inaccuracy with project failure.

The purpose of this report is to outline what organisations must do to establish a mature data culture so that dependable data becomes the norm.

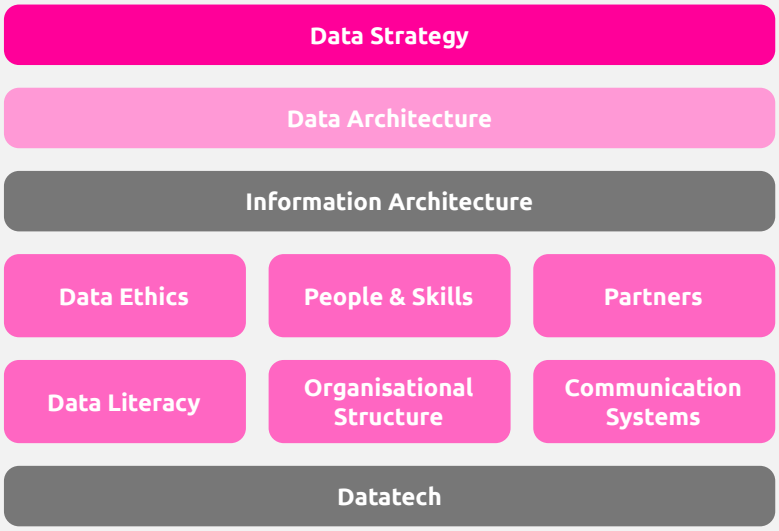
> Clean data is no longer a preference for companies wanting to improve their ROI: it's now a business fundamental.

Section 2: Building a data-aware organisation

To understand the strategic importance of data quality – and by extension the need for effective data cleaning – it needs to be placed in the context of an organisation's wider data management structure.

The process of gaining the full value of data requires businesses to ensure that the constituent parts that make up a mature datainformed organisation are in place, as shown in the Global Marketing Alliance model opposite:

The GMS model for creating a data aware organsation



Data quality is a top priority for 41% of UK data leaders [23]

"If you torture the data long enough, it will confess to anything..."

Darrel Huff
'How to Lie with Statistics'

Creating an effective data strategy is the key starting point. Its importance is reflected by the Insights2020 study [5] which discovered that:

- 67% of execs at over-performing firms were skilled at linking disparate data sources
- 61% of over-performers have 'Insights' involved in all key areas of planning
- 71% of over-performing firms combine analytical and creative thinking

There are many facets to the data strategy, but two fundamentally important areas are the: 'Data Architecture' and 'Information Architecture'. Both play a key role in ensuring that your data is primed for optimal use.

Defining your 'Data Architecture' and 'Information Architecture'

According to the Harvard Business Review a company's Data Architecture describes: "how data is collected, stored, transformed, distributed and consumed" [6]. Its underlying purpose is to ensure your data is ready to meet the needs of the business and is fit for processing. In other words, is your data clean, accurate and up-todate?

It is as much a business decision as it is a technical one, as new business models and entirely new ways of working are driven by data and information. The Data Architecture includes an organisation's governance and compliance frameworks and it provides the foundation for the 'Information Architecture' which is the process of transforming raw data (names, addresses, metrics) into contextually useful and actionable information. Or as David Loshin and Charles Roe from DATAVERSITY [7] summarise it:
"Information is data put into action."

The Data Architecture should always be a Single Source of Truth – there is only right data and wrong data at a field level. However, your Information Architecture could display multiple truths depending on how it's been processed or consumed. They may or may not be accurate and part of your data literacy training should make it clear that all processing innately has bias and can often show what the creator was "looking for".

Clean data starts with a robust Data Architecture

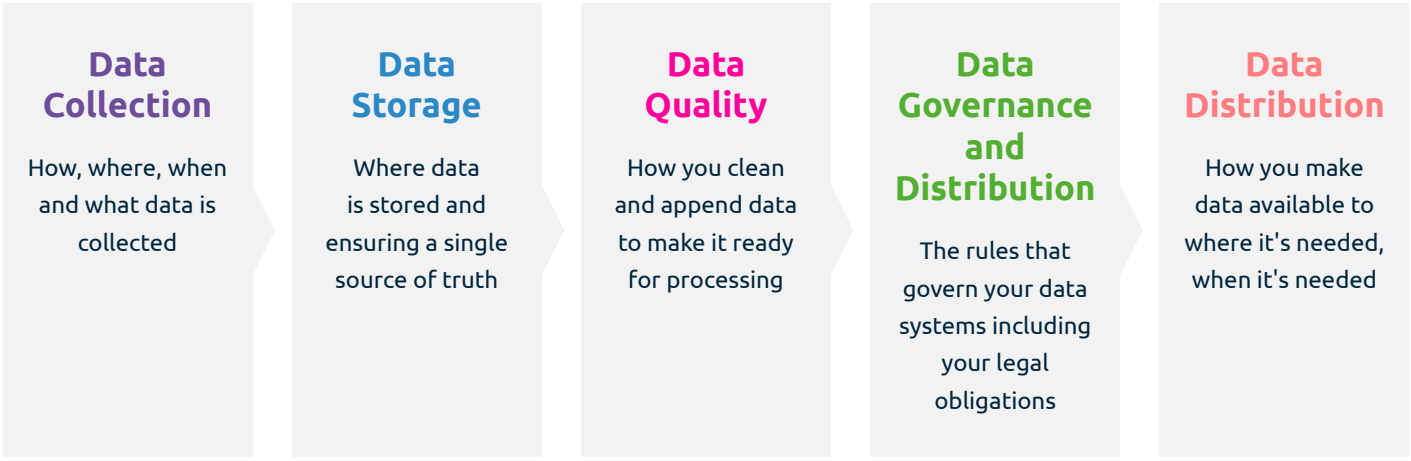
While it's clear that all the constituent parts in the Global Marketing Alliance model are crucial, some, due to the dependency others have on its outputs, are more critical than others. Having a robust Data Architecture is one of these as it is both a prerequisite of delivering your data strategy and for enabling almost every other part of the model to work. As the well known expression goes: "garbage in = garbage out" and it's the role of your Data Architecture to deliver this.

However, in our experience of working with hundreds of companies, data architecture is often poorly resourced compared to other areas which have more roadside appeal to the business. After all, visual dashboards are more appealing than business models!

Our framework for Data Architecture

We look at data architecture as having five core pillars which when combined enable your data for optimal use within the organisation. This high quality, reliable data feed can then be passed into your Information Architecture layer for processing with a high degree of certainty that it is accurate and compliant.

The 5 pillars of a best practice data architecture capability:



This report is focused on the third pillar – data quality and specifically the best practices around data cleansing. If you want to find out more about the other areas then check out the resources and blog sections of our website where you'll find more information.

Defining Data Quality at your organisation

One of the issues that organisations have when it comes to defining data quality is that there is some degree of interpretation and context. Every business has a different data-set, a different product set and a different buyer persona. What is right for one company, product or customer is not necessarily right for another. So it's important to understand whether the data you possess is fit for its intended use in decision-making, operations and planning.

Quality data is useful data. It must be consistent and unambiguous. Data quality issues are often the result of database merges or systems/cloud integration processes in which data fields that should be compatible are not due to schema or format inconsistencies.

We always say to our clients *"If you did XYZ with their data do you think your customers would be delighted or would they be upset?"* Asking this question every time you consider capturing, processing or using personal data gives you the right mindset. Collecting as much data as possible should not be the default option as has traditionally been the case – and too often still is.

Businesses benefit greatly from quality data, however the downside is that it deteriorates unless it's cleaned and updated on an ongoing basis...

Understanding the half-life of your data

In a similar way to radioactive material, data has a halflife. Understanding what the half-life is for your data helps frame your approach to data cleansing. It also impacts a number of other areas of your data strategy.

Your data half-life is: **"the time it would take, if there was no remedial action, for 50% of your data to be inaccurate on a critical data field".** This clearly depends on a number of factors such as:

- Sector: for example, food retail vs banking
- Job roles: some roles turnover very quickly, some less so. The average tenure of a marketing manager is 24 months whereas a CFO is more than 4 years
- Life stage: under 30's are much more likely to move house, change job, get married than someone in their 50's
- Depth of data: the more critical data you hold the more opportunity there is for data to decay

Daily life events affecting customer data accuracy:

9,590 households move
1,496 people marry
2,011 people retire
1,500 people die
810 people divorce

Research by: Royal Mail Data Services (RMDS)
Source: Office of National Statistics [2]

Research by Royal Mail (adjacent) demonstrates that 15,407 life events take place each day which degrade data accuracy. That scales up to 107,849 per week, 465,291 per month and 5,623,555 per year – amounting to 8.5% of the UK population. However, this doesn't include other potential sources of data inaccuracy, such as the proliferation of incorrect form entries or duplicate contacts.

Performing the data half-life exercise will lead you to defining key aspects of your data cleaning processes. For example do you need to have an always-on service such as Connect which uses APIs to make sure that your data is updated at close to real time? Or is it sufficient to manually clean quarterly?

Questions to consider when defining your companies **clean data state:**

• What is your definition of 'is it clean now?'?
• What is the margin of error that is allowable for your business, if any?
• Do you know where all your data is and who's responsible for it?
• Do you have an actionable data maturity model in place?

How you answer these questions should provide a clear sense of where you need to start.

> *"To genuinely understand the value of data, business leaders must ask these questions, within the context of their environment and business objectives, and use the answers to develop a data strategy to drive responsible and profitable decisions. Once tangible value can be attributed to data, then the business case to justify the time and investment for maintaining the accuracy and quality of that data, on an ongoing basis, should write itself."*
>
> **Firas Khnaisser**
> Head of Decisioning at Standard Life and Chairman at DMA Scotland

Creating a Data Quality culture

"Data and analytics leaders need to understand the business priorities and challenges of their organisation. Only then will they be in the right position to create compelling business cases that connect data quality improvement with key business priorities," Ted Friedman, vice president and analyst at Gartner [8].

This point stresses the importance of a data strategy which feeds off of the business strategy – and vice versa. Linking the two highlights data's strategic importance to the wider organisation.

Maintaining the accuracy and integrity of the data collected, held and processed by businesses should be fundamental. As of May 2018 it is also a legal requirement under the GDPR. However, whenever we're asked about data quality, our response is always to ask where this important aspect of responsible data management sits in a wider data strategy. Without a strategy, how can any business attribute and derive real value from their data?

So it is surprising how many businesses do not have a formal data strategy in place. They collect and hold unprecedented volumes of data - and might sometimes acknowledge that data is an important business asset (even though they can't value it) - but do not ask themselves what is the value of this data? Why do we collect it? Why do we maintain it? And are we using it to deliver value to the business?

Perhaps counter-intuitively, getting clean and staying clean is more to do with culture and less to do with process and technology. The hardest part is **staying clean.**

It's easy to perform a data "spring clean" as a one-off project. But this approach simply resets your data halflife and fails to provide the long term solution that datadriven organisations require. Fast forward just a few months and you'll likely be back to the start again.

Creating a "Data Quality Culture" is therefore essential and it must be initiated at the top of the organisation. It is not just a matter of implementing strong validation checks on input screens because no matter how strong these checks are, they can often still be circumvented by the users.

Industry pioneer, Ralph Kimball outlined the general actions you should consider implementing to help build a data quality culture [9]:

1. Declare a high level commitment to a data quality culture
2. Drive process re-engineering at the execution level
3. Invest to improve the data entry environment
4. Invest to improve application integration
5. Invest how to change how processes work
6. Promote end-to-end team awareness
7. Promote interdepartmental cooperation
8. Publicly celebrate data quality excellence
9. Continuously measure and improve data quality

Section 3: Getting clean: the key constituents of a data clean-up operation

A data cleaning project helps unlock the full value of data. Cleaner data leads to increased business efficiency and productivity, improved marketing efforts and greater ROI across the business. But as we've already mentioned: postGDPR, it's also a vital aspect of regulatory compliance.

Yet a worrying number of studies show that organisations have been failing to wrest control of their data. Research by IBM [10] revealed that 27% of business leaders were unsure how much of the data they used was accurate. Meanwhile, bad data has been said to cost the US economy roughly £2.3 trillion dollars each year according to the same report.

While the stakes may be higher for multi-national corporations, having the right processes in place to ensure data remains accurate and updated is of increasing importance for SMEs too. Data protection regulations do not discriminate and the value of quality data can help businesses of all sizes make better decisions and ultimately achieve better results.

In this section, we lay out the essentials for launching a data clean-up operation.

Data cleaning saves £800k

A large utilities business saved more than £800k in direct mail costs in one year by removing deceased, gone-aways and inaccurate data from their mailing list. By using our cleansing technology, the business was able to remove an average of 10% of their contact data per clean-up.

What happens is people relocate? GAS them

Not as sinister as it sounds (honestly!) When people move house, it can cause data to become incorrect and misleading. Our Gone Away Suppression file (GAS) solves this problem by gathering consented data sets from organisations that have been notified about someone's change of address. These are regularly updated so that organisations are equipped with data they can continue to trust.

1. Begin with an audit

The first step is to understand the current state of your data across the whole organisation. In auditing your data, you should:

- Assess data inputs: Can you cope with the amount of data you're gathering? Are your data inputs actually helping the business or simply creating more noise? Prioritise the data sets that are useful and manageable - and jettison the rest. This enables you to focus your data cleaning on the areas that actually matter, rather than wasting resources on data inputs which offer little value.
- Assess data hygiene: Now that you've identified the most important data inputs, you need to check if they're clean. Is the data correct? Is it representative? Data hygiene must be assessed across the whole organisation.
- Evaluate business-wide: Check all systems that your company relies on for in-house customer information. This includes all platforms which gather information about customers, including: CRM, web analytics, registration forms and survey forms.
- Update the audit: Perform audits regularly. Frequency will depend upon the size and scale of your organisation as well as your specific data needs.

2. Set clear goals

Effective data cleaning requires a set of agreed KPI's. For example, you could aim to clean a set number of contacts in a given month. Data cleansing tools can automate this process or individual team members can be given specific targets, such as removing or unifying a set number of duplicate contacts each month.

KPIs can also be linked to your wider business goals. For example, by having more accurate data your sales team will hold more accurate information about potential leads. They will also be more efficient and productive (i.e. they won't be chasing duplicate contacts); all of which makes the sales team more capable of driving a higher number of sales conversions.

3. Collect the right data, the smart way

Standardising contacts at point of entry will save a lot of time and effort in the long run. By creating a single, consistent method for importing data into your system, you can avoid messy contact forms and/or the creation of duplicate contacts for the same individual.

This requires intelligent filtering. For example, when someone enters information on a form (whether a customer or employee) a process can be put in place to ensure that erroneous values are not entered in the first instance and duplicates are harder to create. Ultimately, standardisation minimises the risk of gathering bad data.

"Without a systematic way to start and keep data clean, bad data will happen."

Donato Dorio
CEO at DataZ.

4. Integrate data cross-team and cross-platform

Platforms:
Every organisation utilises a range of platforms, such as: CRM, mobile app & web analytics, marketing automation and social media among others. Wherever possible, these platforms need to be integrated: systems need to speak to each other so that when a customer's details are updated in one system, it's automatically updated in another. Or, in the case of tracking user journeys, a prospective customer can be accurately tracked and mapped across multiple devices giving a holistic view of their behaviours.

Our Single Customer View (SCV) solution is an example of technology which maps and unifies what would otherwise be disparate user journeys while ensuring data is cleansed in the process.

People:
Ensure clear ownership of data across the organisation. In each area of the business where data is collected an individual should be entrusted with maintaining the data and ensuring it stays clean. These people are responsible and accountable for making sure any data collected is purposeful and accurate. Good data hygiene requires leadership across the business.

5. Be a responsible data handler

Clean data also needs to be responsibly sourced data. Arguably, it's good business practice to ensure that customer data is collected in a transparent way that respects individual's privacy, but it is now also a legal requirement in light of GDPR.

Consequently, a risk-assessment process needs to be in place. Simon Blanchard, Senior Associate at Opt-4 and Deputy Chair at the Data Protection Network lays out four key areas to consider as part of a risk assessment:

Data Discovery

- What data is held in which systems?
- Data mapping inflows & outflows
- Create a data inventory
- Who is accountable for each processing task?

Risk Assessment

- Identify & engage the functions which process personal data
- Feed this into your Record of Processing Activities (ROPA)
- Identify data security & privacy risks
- Document risks & issues into action plan, appoint action owners

Measure & Monitor

- Operational dashboard
- Governance programme reports
- Regular compliance assurance

Action Plan

What shall we do about it? There are 4 main options:

- Treat it - take action to reduce the severity and/or likelihood of the risk
- Tolerate the risk
- Terminate the activity
- Transfer that risk to another party

- Evaluate your options, estimate costs and resources needed
- Approval process, timing and delivery

A data governance programme should provide tools to help your teams to identify both existing and emerging risks, so they can be properly assessed, and risks can be mitigated.

The 'measuring and monitoring' process can be assisted by automated tools which can automatically highlight data which is held too long or is inaccurate. It helps ensure that data cleansing is not a one-off event but part of an ongoing data risk assessment to protect themselves and safeguard individual's data.

6. Validate and enhance your data

Data's value is dependent on its quality. If it's accurate, detailed and representative; it can be relied upon to help inform decision-making across all areas of the business. If it's inaccurate, then it's likely to mislead and contribute to poor decisions.

Data hygiene tools can help. The accuracy of data can be verified manually or via automative tools. For the sakes of speed, accuracy and cost; it's worth investing in systems that automatically enrich, append, clean and/or de-duplicate data.

In the absence of this technology, ensure there is a clear timeframe/schedule for data cleaning and enhancement activities.

Auto-validation: Macmillan Group

As a cancer charity, Macmillan requires accurate databases in order to contact those in need of its services. But also to ensure communications are not directed to the wrong people.

Using our automated bespoke data cleaning service, Inline, every campaign data selection taken from Macmillan's database was evaluated for accuracy of contact details – in terms of deceased, gone-aways and, where available, new addresses for relocations.

This not only ensured that their data was accurate and up-todate but also allowed them to set some challenging, high value objectives for the future.

Data aggregation:
Ensure that all the aggregated data you have about customers, leads and contacts is up-to-date. Prioritise existing customers before moving on to leads – as these provide the most value to the business.

Data enrichment:
This is the process of merging third-party data with your existing database of customer data to create a more detailed – or holistic - picture. This will also give you a true understanding of the customer, their spend, transactions, etc. Remember: the data gathered needs to meet regulatory requirements and respect the individual's right to privacy.

De-duplication:
Describes the process of ensuring that duplicate copies of the same information are deleted within your database. As well as potentially saving calls/emails/letters to the same contact (and the time that costs both parties), duplicate datasets can run-up a significant cost in bandwidth and storage - up to 80% according to a data de-duplication whitepaper by Druva [11].

7. Make the data talk

> 45% of data leaders cite a lack of data visualisation as a barrier to achieving business goals [23]

Finally, you need to make sure that data is distributed in such a way that every individual can make sense of it. The more people that can access and understand data, the more it will empower each individual and team, and the greater the stake they will have in the data cleansing process.

Data visualisation tools can play an important role in establishing and maintaining a healthy data culture across the business.

8 key questions all businesses should ask about data quality

1. What processes do I have in place to ensure ongoing data quality across the business?
2. How can I measure and improve the state of data literacy across the business?
3. How can I break down siloed data across the business?
4. How often do I need to audit data?
5. What's the process for auditing data, and who's responsible?
6. Is data quality covered in our data strategy?
7. Have we made a high-level commitment to data quality?
8. How do we celebrate and encourage data quality excellence?

Clean data and GDPR compliance

The introduction of GDPR in May 2018 brought the issue of data quality to the fore. Prior to its launch, the practice of data cleaning was considered to be best practice, now it is a necessary part of legal compliance.

According to the ICO, fines can apply to any "failure to comply with any of the data protection principles, any rights an individual may have, or in relation to any transfers of data to third countries." [12]

The maximum fine under the GDPR is 20 million Euros (or the equivalent in sterling) or 4% of the total annual worldwide turnover in the preceding financial year, whichever is higher - which should make any business pay attention!

Three key areas of concern in regards to bad data

1. Time-limited storage: GDPR Article 5 (e) [4] states that companies must only hold on to personal data for no longer than is necessary for the purpose it was processed. Consequently, even if an organisation has collected data in compliance with the law, it may become unlawfully held if it is deemed to have been stored for an unnecessarily long period.

   This requires companies to draw up a data retention schedule to ensure there's a systematic way of deleting contacts and personal details when they become outdated, as stipulated by the company policy.

2. Subject Access Requests: If any individual asks to know what information an organisation has about them, you must be able to deliver within a calendar month [13]. An organisation's ability to successfully locate this information requires a clean and maintained database.

   This might become a problem if you have duplicate contacts for the same person, for example Jo Smith and Joanna Smith. If that person asks for the data you hold on them, you're likely to only provide one part of the data you hold on that person. This may later come to the contact's attention when they are contacted via an address or phone number which a company claimed to not have stored.

3. Third-party data providers: The onus is on organisations to ensure due diligence is undertaken with their data providers. Can you trust your data vendor? There are some minimum requirements and assurances that any reputable provider should be able to provide. See 'Section 4: Questions to ask your data provider'.

Almost 18 months since GDPR came into force, we can see that the number of fines and actions are increasing, and consumers are increasingly exercising their rights:

- The ICO's helpline, chat and written advice services received 471,224 contacts in 2018-19, a 66% increase from 2017/18 (283,727 contacts).
- Data protection complaints received by the ICO almost doubled from 21,019 in 2017/18 to 41,661 in 2018/19.
- According to the ICO, 64% of organisations said they had noticed an increase in users exercising their information rights.
- 37% of GDPR decision-makers were obliged to report a breach to the ICO in the past 12 months. [14]
- A new study by IBM shows the cost of a data breach has risen 12% over the past 5 years and now costs $3.92 million on average, with increased regulation cited as one of the key reasons. [15]
- The past year has witnessed the largest ever publicly known fines associated with data compromise. British Airways were hit with a proposed fine of £183m ($230m) after customers were redirected to a fraudulent site [16]. While Marriott International were handed a proposed fine of £99m ($124m) following a large scale breach of customer data [17].

During its first year, enforcement bodies across Europe took a lenient approach to GDPR enforcement. It understood that organisations would require extra time to better understand the requirements of the new legislation and to ensure their processes were compliant. The French regulator CNIL reflected the views of other national enforcement bodies when it said the first year "should be considered a transition year" for GDPR, as the focus remained on finalising their rules and approaches, while tying up probes under the previous regime [18].

The fines for BA and Marriott demonstrate that the period of leniency is now over. At a recent 'Data Privacy Question Time' in central London, it was noted that regulators are also starting to target smaller businesses in order to signal that it's not just large multi-nationals who need to be concerned [19].

Most organisations have faced a steep learning curve when it comes to ensuring GDPR compliance, as we can reveal: *"At the tail-end of last year, a lot of brands were focused on making sure their data infrastructure and estate were secure. It was all about holding data securely and ensuring it wasn't going to be hacked. But there was also a lot of concern surrounding permissions in the immediate build-up and aftermath of the GDPR. 'Do I have the right permission, do I have the right lawful basis to process personal data?'."*

"Now that most organisations have got to grips with those aspects of compliance, brands are now having to consider whether the data they hold is clean, accurate and up-to-date. Those conversations have grown significantly this year because data maintenance represents the next stage of GDPR compliance for many companies."

"The solution lies in effective data management and processing tools embedded within an effective data governance framework."

Section 4: Staying clean - the future of data-cleansing

Serving a new proactive consumer

The ICO's helpline, chat and written advice services received 471,224 contacts in 2018-19, a 66% increase from 2017/18 (283,727 contacts) [14].

The data landscape will grow in complexity as new digital technologies create new data streams to add to the myriad which already exist. AI, VR and voice-recognition technologies are just three areas of innovation which are evolving at a rapid pace. Meanwhile regulations will have to adapt – not only to datatech innovations and the ethical questions they pose, but also to meet the demands of a more savvy and data literate consumer.

Recent Gartner research has found that organisations believe poor data quality to be responsible for an average of $15 million per year in losses. However, according to Gartner [8]: *"This is likely to worsen as information environments become increasingly complex — a challenge faced by organisations of all sizes. Those with multiple business units and operations in several geographic regions, and with many customers, employees, suppliers and products will inevitably face more severe data quality issues."*

In other words, effective data cleansing methodologies, will become ever more important. The rate of your own customer data decay will depend on many factors, from the consumers or industries you serve, to the countries you do business in.

Data decays at a faster rate than ever before, everyday

9,590

Households move

1,500

Pass away

3,000

Daily changes to PAF

1 in 5 (19%)

of businesses have lost a customer due to incomplete or inaccurate data

Protect your organisation against regulatory fines up to: **€20 million or 4% of global turnover**

It's perhaps unsurprising that according to CrowdFlower, data scientists spend 60% of their time organizing and cleansing data [20]. However, this time can be massively reduced by API-driven data cleansing technologies which update and clean data as near to real-time as possible.

We believe this auto-clean capability will play a fundamental role in ensuring businesses are able to draw on useful data in a safe and secure way.

Automation - taking the strain out of data quality

With GDPR Article 5.1 (d) specifying that personal data shall be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy'). Consequently, having an effective and efficient solution has never been more important.

The evolution of technology that automates data cleansing provides a solution that takes the heavy lifting out of keeping data clean and accurate – freeing up human resource and infrastructure as well as enhancing information security.

There are automated solutions currently on the market which provide close to real-time update capabilities. For example, Connect - our automation solution - seamlessly provides flags for updates as soon as changes in their comprehensive cleansing database are detected. Consequently, when a contact moves house or goes through another life event which undermines the accuracy of your data, Connect enables you to action an update to the data according to your business rules and processes.

A key element is the use of an API which provides access to a secure, scalable environment to detect and process changes in the end-users database – reducing the need for data movement and associated data security.

Unifying data governance and management

Data governance and data management are often talked of as separate concepts with contradictory objectives. This is illustrated by Ursula Cottone, Chief Data Officer (CDO) of Citizens Bank [21]:

"When the word Data Governance gets used, it automatically brings into play concepts like 'control', 'security', 'protection', and 'compliance.'" Yet data management brings in concepts such as 'usage', 'functionality' and 'innovation', which seemingly contradict what data governance is about.

Harmonising data management and governance should be a major objective of data-driven organisations today. Once again, powerful integrative data cleansing tools provide the solution because it supports both governance and management processes.

Effective data cleansing technology, for example, is a key weapon in the battle to become GDPR compliant. Article 5 (e) of the GDPR states 'personal data shall be kept for no longer than is necessary for the purposes for which it is being processed'. The period for which the personal data is stored should be limited to a strict minimum and time limits should be established by the data controller for deletion of the records or for a periodic review.

Organisations must therefore ensure personal data is securely disposed of when no longer needed. Technology which can flag up these potential issues will play an important role in an organisation's compliance capability and reduce the risk of exposure to financial penalties – now and in the future.

Prevention is the future

Data protection regulations are still maturing. Regulators are still striving to address the ethical questions posed by Big Data and the various (and advancing) technologies which are evolving to extract meaning and make use of it. But the direction of travel is clear: regulations will evolve to give ever-greater protections to individuals and their right to privacy as new technology continues to pose new ethical dilemmas.

Bart Willemsen, Senior Director Analyst, Gartner: *"Multiple countries are implementing regulations inspired by the GDPR principles, a movement that is likely to continue into the foreseeable future. These privacy requirements dramatically impact an organisation's strategy, purpose and methods for processing personal data. Furthermore, breaches of these requirements carry financial, reputational and regulatory implications."* [22]

For data-driven organisations, this puts increasing pressure on the ability to manage and govern data. The regulatory landscape has accelerated to a place where preventing 'age-related' data storage error is the requirement, and not remedial action.

The knock-on effect is that technology-driven data cleansing solutions need to be capable of acting swiftly to solve potential data storage issues before they happen. Hence, a rapid 'scan-and-flag' API-driven solution which can sift through all the relevant data points will not only drive better business results but also play a key role in ensuring compliance with evolving data protection regulations.

Section 5: Questions to ask your data provider

1

Source and provenance

How is the data collected and what is the source? You should also ask for confirmation of the collection methods and audit trails to ensure the principles of the regulation have been meet and the data is being processed lawfully, fairly and in a transparent manner, regardless of the data and purpose.

2

Permission

Your supplier should be able to provide you with the permission statement used at the point of collection.

3

Validation and Due Dilligence processes

Ask for confirmation of the validation process. A provider with nothing to hide should be able to provide on request an outline of their due dilligence process and the steps they take to ensure data fully satisfies legislative requirements.

4

Recency

When was the last engagement?

5

Quality

Is the data accurate and up to date? Has it been screened against a reliable suppression file to remove deceased and Gone Away contacts to meet GDPR data quality equirements under Article 5.1 (d)?

6

Reputation

Check out their creds and ask peers for a recommendation or ask to speak to an existing customer of the supplier for a candid view.

7

Results

Ask for some examples of the results and case studies - especially if you are using the data for acquisition campaigns.

8

Do they offer a trial?

If you are new to buying data or using a new supplier - ask if you can run a trail campaign to test the quality of the data.

References

1. International Data Corporation (IDC)/Seagate, 'Data Age2025' whitepaper, 2018. Accessible at: https://www.seagate.com/gb/en/our-story/data-age2025/
2. Royal Mail Data Services Insight Report, 'Dynamic Customer Data in a Digital World', 2018. Accessible at: https://www.royalmail.com/business/system/files/royalmail-data-services-insight-report-2018.pdf
3. Dun & Bradstreet, 'Past, Present and Future of Data' report, 2019. Accessible at: https://www.dnb.co.uk/perspectives/master-data/datamanagement-report.html
4. Data Protection Network (DPN), 'GDPR Data Retention Guide', 2017. Available at: https://www.dpnetwork.org.uk/gdpr-data-retentionguide/
5. Insights2020 – Driving Customer-Centric Growth, Millward Brown Vermeer, 2015. Accessible at: http://www.millwardbrown.com/globalnavigation/news/press-releases/fullrelease/2015/09/29/first-wave-of-insights2020-study-highlights-the-role-insights-and-analytics-play-in-drivingcustomer-centric-business-growth
6. Harvard Business Review, 'What's Your Data Strategy?', 2017. Accessible at: https://hbr.org/2017/05/whats-yourdata-strategy
7. DATAVERSITY, 'Data Architecture and Information Architecture: What's the Difference?', 2013. Accessible at: [https://www.dataversity.net/data-architecture-andinformation-architecture-whats-the-difference/]
8. Gartner, 'How to Create a Business Case for Data Quality Improvement', 2018. Accessible at: https://www.gartner.com/smarterwithgartner/how-tocreate-a-business-case-for-data-quality-improvement/
9. Ralph Kimball & Margy Ross, The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence, 2016. Accessible at: https://books.google.co.uk/books?id=ao8gCwAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
10. IBM, 'The Four V's of Big Data'. Accessible at: https://www.ibmbigdatahub.com/infographic/four-vsbig-data
11. Druva, 'Data Deduplication White Paper', 2015. Accessible at: https://pages2.druva.com/Whitepaper-DataDeduplication-and-DruvainSync.html?utm_source=social&utm_medium=Blog
12. ICO, 'Penalties' guidance. Accessible at: https://ico.org.uk/for-organisations/guide-to-dataprotection/guide-to-law-enforcementprocessing/penalties/
13. ICO, 'Time limits for responding to data protection rights requests'. Accessible at: [https://ico.org.uk/your-datamatters/time-limits-for-responding-to-data-protectionrights-requests/]
14. ICO, 'ICO Annual Report 2019'. Accessible at: https://ico.org.uk/about-the-ico/news-and-events/newsand-blogs/2019/07/ico-publishes-annual-report/
15. IBM, '2019 Cost of a Data Breach Report'. Accessible at: https://newsroom.ibm.com/2019-07-23-IBM-StudyShows-Data-Breach-Costs-on-the-Rise-FinancialImpact-Felt-for-Years
16. ICO, 'ICO announces intention to fine BA', 2019. Accessible at: https://ico.org.uk/about-the-ico/newsand-events/news-and-blogs/2019/07/ico-announcesintention-to-fine-british-airways/
17. ICO, 'Intention to Fine Marriott International Inc', 2019. Accessible at: https://ico.org.uk/about-the-ico/newsand-events/news-and-blogs/2019/07/intention-to-finemarriott-international-inc-more-than-99-million-under-gdpr-for-data-breach/

18. Data Protection Network, 'Why GDPR is not a fuss about nothing', 2019. Accessible at: https://www.dpnetwork.org.uk/opinion/why-gdpr-is-nota-fuss-over-nothing/
19. Global Marketing Alliance, 'Award for Responsible Marketing launches at Data Privacy Question Time', 2019. Accessible at: https://www.the-gma.com/award-forresponsible-marketing-launches-at-data-privacyquestion-time
20. CrowdFlower, '2015 Data Science Report'. Accessible at: https://visit.figure-eight.com/2015-data-scientist-report
21. DATAVERSITY, The Future of Data Governance: Balancing Data Governance and Data Management, 2018. Accessible at: https://www.dataversity.net/future-data-governancebalancing-data-governance-data-management/
22. Gartner, Future of Privacy 2019, 2019. Accessible at: https://www.gartner.com/smarterwithgartner/gartnerpredicts-2019-for-the-future-of-privacy/
23. Coleman Parkes, Big Data LDN and Cloudera; The Fourth Industrial Revolution Report 2019.

Sagacity

w sagacitysolutions.co.uk | e enquiries@sagacitysolutions.co.uk