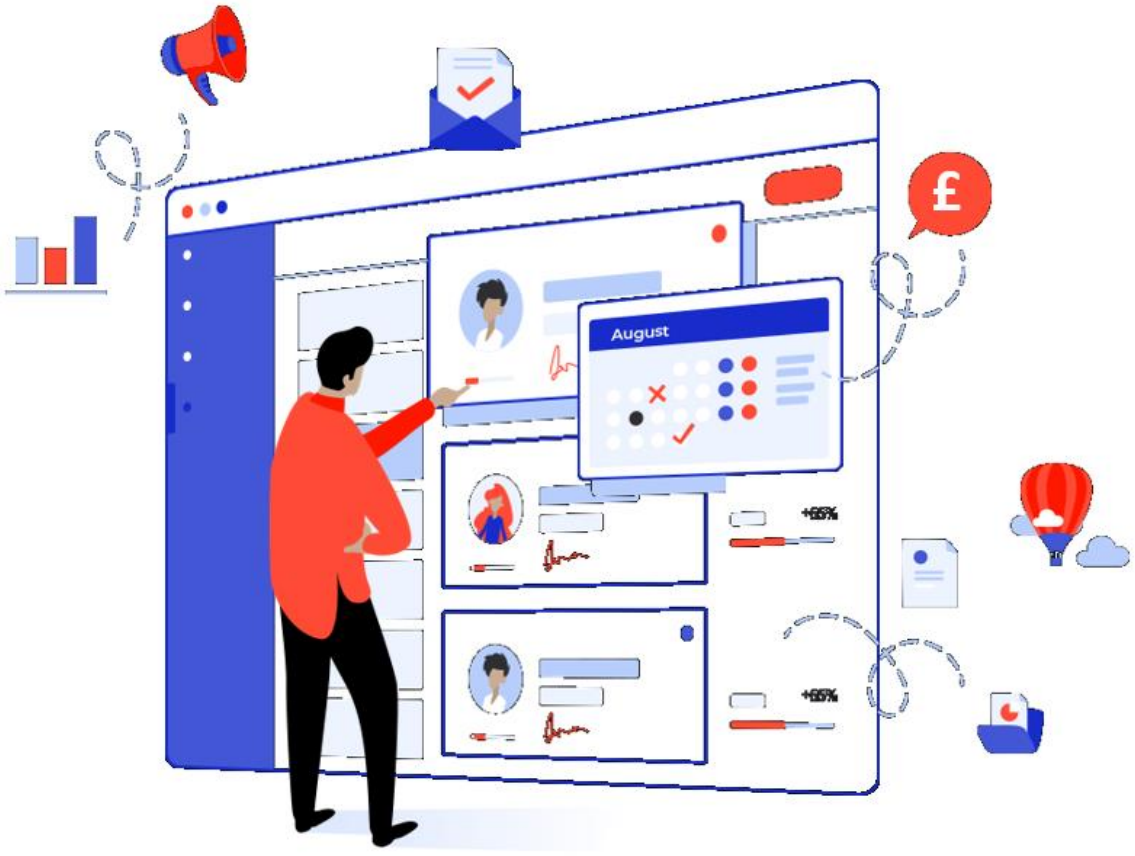
Using automation to manage data in the information age

Author: Mark Hobart – Co-Founder and CEO



Contents

1 Introduction

We are living in the Information Age. The idea is that ***access to and the control of information*** is the defining characteristic of this current era in human civilization. The emboldened words are worth noting as things are about to get much more difficult for businesses of all shapes and sizes.

You don't need to be a fortune teller to predict the future were data is concerned. There are three important factors to consider:

- Data growth
- The rise and prevalence of unstructured data
- Increasing levels of data related regulatory compliance

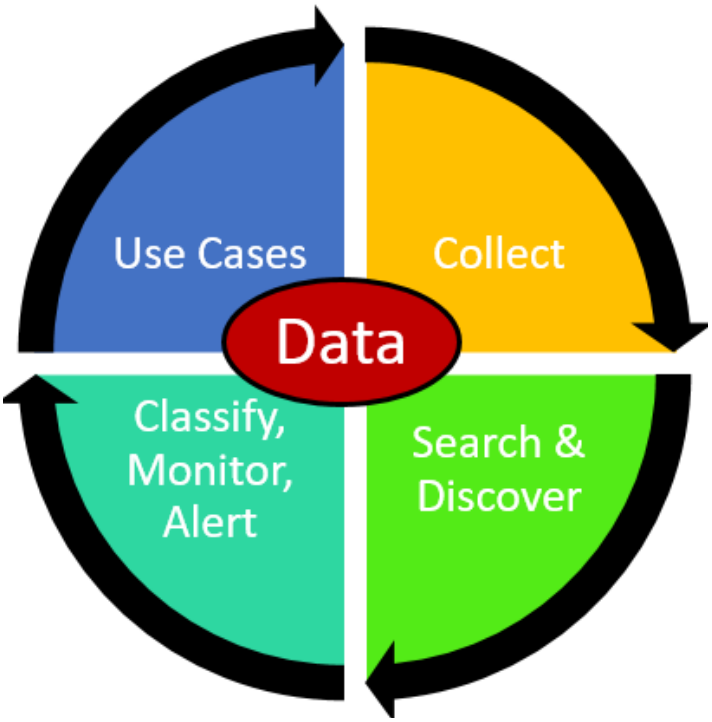


Figure 1Information life-cycle management

This paper will explore these topics, drawing on recent research and articles to illustrate their consequential impact on business. We'll also look at solutions such as Infoboss' automated data discovery and management platform and how it can be used to prepare an organisation for the tsunami of data and regulatory compliance coming their way by utilising and embedding data automation processes for data classification, data quality improvement, compliance and general monitoring of data for practical and effective information life-cycle management.

2 Data growth

I've heard people say that "Data is the new oil". The Economist recently described it as, "The world's most precious resource". Whether you agree or not, there is no doubting the increasing importance data has for business. It is the raw material that business decisions are made upon to deliver a competitive advantage or enhance the product and service offerings of the future.

2.1 New people around the office

This significance can be seen in the job roles now being advertised that relate to data and information. These include roles such as Chief Information Officer, Data Scientist, Data Analyst, Data Quality Manager, Data Protection Officer, Data and Compliance Manager and more. Indeed, the World Economic Forum recently described the role of Data Analyst as "The most in demand job of the 21st Century". Another economist described it as "The sexiest job of the 21st Century". IBM have indicated a requirement for 700,000 new data analysts worldwide by 2020! But don't think this is a big business thing, it's not! A recent survey in Europe discovered that 5 out of ten SME businesses are intent on hiring a data analyst over the next three years. Businesses are beginning to be populated by people who understand and can exploit data.

The International Data Corporation (IDC) conducted research in 2016 projecting a ten-fold increase in data created and replicated by 2025. Rising from 16.1 ZettaBytes [1]in 2016 to a colossal 163 ZettaBytes by 2025!

What is significant in this research is that they also predicted that 80% of the data will be generated by business, in contrast to just 20% today.

3 Unstructured data

Worryingly for almost all business is that IDC also predicted that 80% of that data will be of the unstructured variety. Notorious in that it's the most difficult to access, manage and control. One compliance consultant recently described it as, "the mother of all headaches".

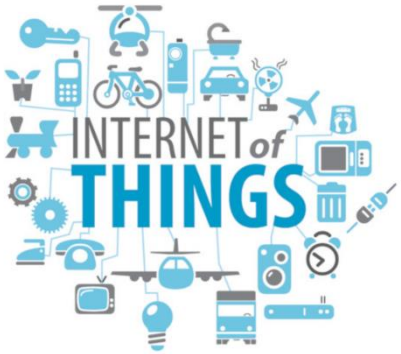
If you consider that many organisations are sitting on tens if not hundreds of thousands of document files of varying types that have perhaps accumulated over many years, it's perhaps no surprise to discover that businesses feel somewhat overwhelmed by the task ahead.

3.1 Sources of unstructured data

The most common sources of unstructured data today are:

- Document files
- Social media
- Voice and video
- Scanned files
- Note and comment fields in application databases

However, the future looks a little different with the lions share of new unstructured data being generated by Internet connected devices, transmitting status and other information to ever ready recipient devices for monitoring and other types of processing.



Thought for the day: **If IDC are correct, then what are you going to do with ten times the amount of data you currently store today?**

[1] A zettabyte is expressed as 10^{21} (1,000,000,000,000,000,000,000 bytes) or 1 sextillion bytes

4 Data protection and related compliances

GDPR has rightly changed the dynamic and business focus around data and its management. It has rightly, imposed a new way of thinking about it.

But it's not just GDPR in Europe, companies the world over are coming to terms with a society that is getting increasingly concerned about the way organisations handle their personal and sensitive data. CCPA in California, PDPA in Singapore, summits in Asia, Africa on data protection are now the new norm.

The threat of regulatory fines which can be significant is one thing, but perhaps the greatest concern is reputational damage. Recent analysis of data breaches by BitGlass revealed an average 7.5% drop in stock value after a breach is not untypical. The threat of brand erosion and negative impact on the business is very real and one that senior executives simply cannot ignore.

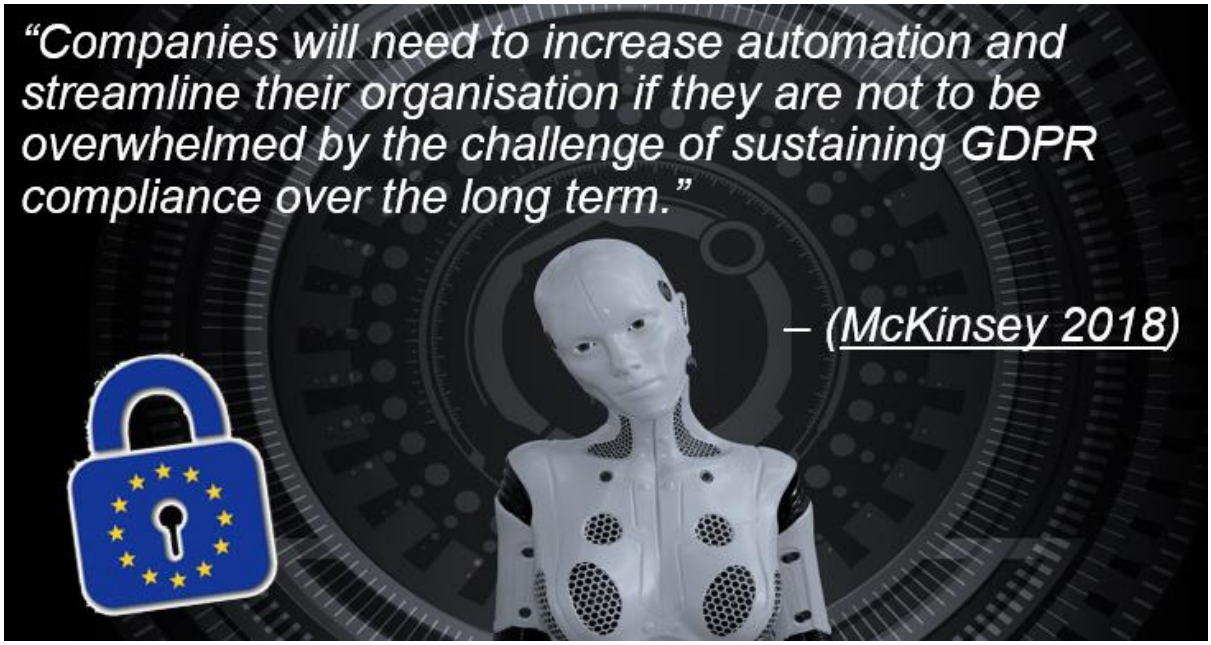
4.1 The challenges of GDPR compliance

Back in May 2018, there was a frenzy of activity within almost every business that processed the data of EU citizens. This included some or all of the following…

- Establishing a data inventory
- Perhaps a little data discovery
- Updating privacy notices
- Understanding and mapping data handling processes and the reason for processing
- Updating policies and procedures
- Contract reviews
- Understanding data roles - data processors, data controllers
- Appointing a Data Protection Officer
- Getting consent from customers for processing
- Undertaking data protection impact assessments
- Training staff on GDPR
- Preparing the way to service data subject rights such as access requests
- And more…

This was recently described to me as "the easy part"! The holy grail must however be "Data protection by design and default" for every business.

Sadly, many organisations have defaulted back into old ways of working and have begun to rely heavily on people to maintain compliance. Given the increase in data, the complexity of it (unstructured) and a heavy dependency on people, it is perhaps not unsurprising that leading consultancy McKinsey summarised the situation that many now find themselves in.



With over 50% of data breaches occurring as a consequence of "insider" activity, the growth of data and the largely manual approach to tackling the problem falling on people based processes within many organisations data leaks and the subsequent fines and reputational damage are likely outcomes for many.

4.2 Data quality

Almost every year we hear business leaders proclaim that poor quality data is holding them back. Fixing data quality is a daunting task and too often protagonists pay lip service to the job in hand and elect to adopt short term strategies to resolve specific issues rather than take the brave steps to build a data quality management strategy that delivers sustainable improvements, forever.

Research from Gartner has shown a marked increase in the cost of poor-quality data on an organisation.

"The average cost to an organisation of poor-quality data in 2018 is $15 million per annum!" – (Gartner 2018)

This is up from $9.7 million dollars in 2017, and is a trend that is unlikely to improve due to the reliance of an increasing number of business initiatives, on quality data.

"Two-thirds of all business leaders believe that their companies must pick up the pace of digitalization to remain competitive." (Gartner 2017)

The consequence of poor-quality data is therefore that initiatives that rely on data, such as digital transformation, and business analytics, will not deliver the business outcomes expected. In extremis they may fail completely if both the data quality is not improved, and the required processes and procedures for effective data input, collection and use are also not put in place.

However, we believe this is all starting to change. With so many contemporary business drivers, like digital transformation, being underpinned by a deep reliance on quality data along with GDPR compliance concerns, business leaders are being forced to confront the problem head on and take the necessary steps to embed data quality management (DQM) within the DNA of the business operating model.

5 Automated data management explained

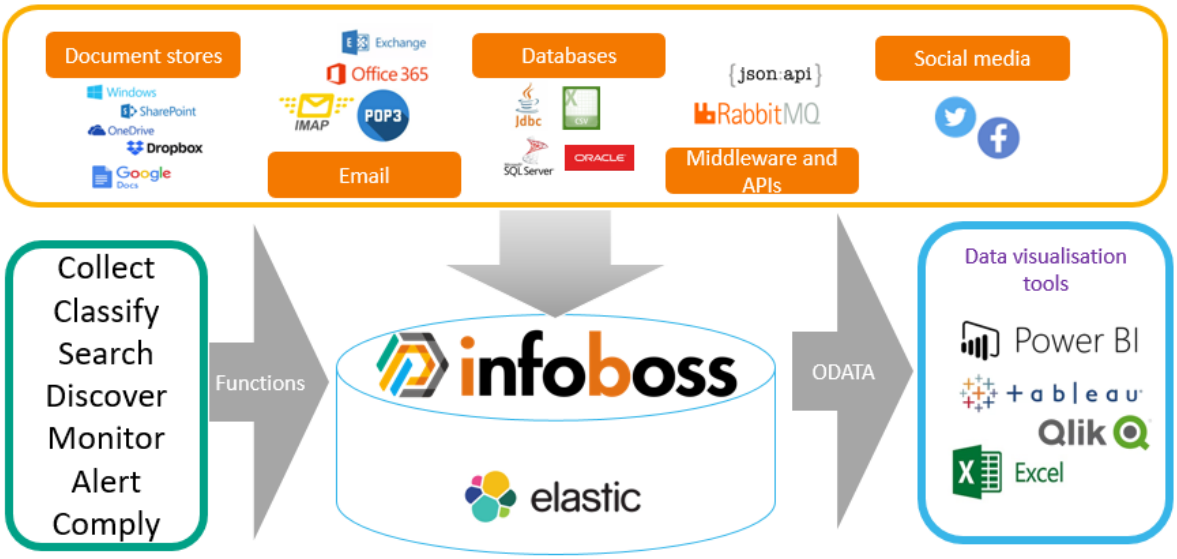
Automated data management (or governance) is a relatively new concept. The principal is that all enterprise data is collected, indexed and classified (catalogued) into one central location where it can be more easily accessed and controlled for data quality, compliance, retention, analysis and other business use case requirements.

Buried deep on the ICO.gov.uk website is advice that endorses the consolidation of data in this way to help support the GDPR compliance use case.

"Know what information your organisation holds, who it is about and where it is stored. Auditing and indexing your information properly will make it easier and more efficient to deal with…"

5.1 How it works…

Our Infoboss automated data discovery and management platform is one such solution. It collects data from any electronic data source and stores it in an indexed (searchable format) in an Enterprise Data Store. Data can be consumed from structured sources like application databases, data warehouses, CSV files and unstructured sources like email systems, document repositories, social media feed and IoT sources. It can even collect data from other applications within the enterprise or tap into an Enterprise Service Bus architecture (if you have one).



5.1.1 Unstructured data

Infoboss is particularly powerful at processing unstructured data. Any document containing text including images and photographs that can be processed by our OCR scanner and the textual data therein can be analysed. Infoboss can be used to add structure to a data set by enhancing the meta data associated with the file. For example, imagine a document folder that contains copies of letters sent to clients. Infoboss can be used to derive fields of data. Such as dates, customer references, names and even how the letter greeting and salutation were phrased.

5.1.2 Discovery and classification

As data is ingested, organisation specific rules based on regular expression pattern matching are applied to identify data and automatically classify it.

For example, identify records or documents that contain: payment card numbers, postal codes, phone numbers, bank accounts, currency values and so on. As data is discovered it is automatically labelled (classified) to enable simple and efficient access.

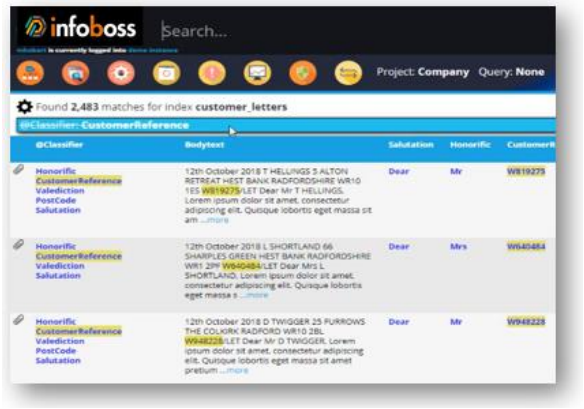


Figure 2 Automatic document labelling

In the screen shot all documents that contain a customer reference have been automatically labelled with the classifier “CustomerReference”

5.1.3 Set based classification (Whitelists and blacklists)

In addition to regular expression pattern matching, Infoboss also enables you to use whitelists (or blacklists) to further label data sets. Essentially if an item matches one or more fields in a list then it classifies it appropriately. For example, you might have a list of customer names, reference identifiers and addresses. This could be used to interrogate a document repository looking for any documents that match any record in the list. Answering questions like “Find all documents that relate to customers in active contracts?”

“Infoboss has enabled us to systematically classify and discover personal and sensitive data across our entire data estate and crucially who has access to it.”

5.1.4 Enterprise search and democratisation of data

Infoboss is very easy to use. Empowering staff to engage with data from across the organisation that is under their control. Staff can carry out Google™ like searching and simple ‘point and click’ style filtering to drill down and identify data of interest to them. Essentially providing an enterprise search facility for all to use to quickly find information to support them in their work.

This democratisation of data through a practical and engaging interface is particularly useful for enabling direct access to data by staff that can make a difference to the quality and compliance of the data under their control.

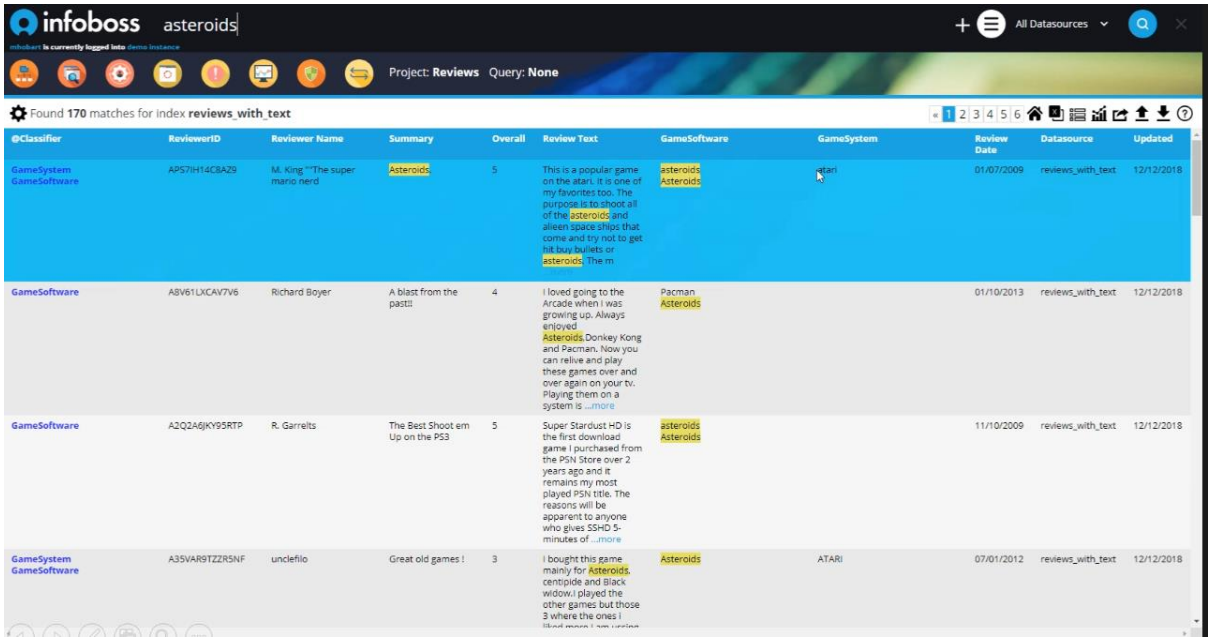


Figure 3Search and discovery

5.1.5 Automated data monitoring and auditing

Discovering data and resolving issues associated with it, is not a one-off event. Infoboss enables you to monitor data as it enters the data estate, looking for data items that are not compliant or meet the quality standards required.

For example, you might want to monitor the notes field of a CRM system for the appearance of payment card details or adverse comments about a customer. Perhaps use it to monitor a public document folder for the appearance of data relating to expired contracts or the sudden appearance of employee salary details. You could monitor a document store and any files that are of a certain classification and that should no longer be retained, can be automatically alerted to the data owner to enable action to be taken.

The uses of this functionality are endless and provide your quality and compliance managers with the ability to undertake ‘perpetual audits’ of the data estate, rather than rely on manual processes to check random samples of data if and when the data audit actually takes place.

6 The GDPR use-case – operationalising compliance

The GDPR has moved data and its management up the to-do list of every organisation. It's not just the threat of fines and the reputational impact of a data breach, but the realisation that good data governance and high-quality data is crucial to the success of many business transformation and change initiatives.

Many organisations ticked the GDPR compliant box back in May 2018. Although not as many as you might think…

AIIM research conducted in 2018 suggested that only 30% of organisations surveyed believed they were compliant by 25th May 2018. Gartner research conducted in 2018 also revealed that "75% of organisations were not GDPR compliant and a data protection strategy was needed".

Pragmatic organisations accept that they are still on the journey to achieve the overarching objective of the regulation namely, to embed **data protection by design and default** into the DNA of the organisation. Effectively operationalising data protection into the way the organisation conducts its business.

What organisations are now beginning to recognise is that the day to day effort to maintain GDPR compliance is a labour-intensive exercise, difficult to manage and fraught with 'data breach' risk.

The regulation isn't going away – it's a necessity that all businesses must adhere to. Moreover, the GDPR affects every department in the organisation that holds data from sales and marketing through to finance, HR, procurement and IT. Compliance can't be left to the legal experts alone.

Achieving GDPR readiness requires everyone to reliably streamline all personal data held in various documents, emails and databases, held across disparate systems, network folders, and even those still in paper-based storage. One problem is that organisations hold their data in multiple locations and, worryingly, some don't even know with a high degree of confidence as to where or how their data is stored and who has access to it.

However, it's not all bad news. The GDPR presents a major opportunity for organisations to transform their approach to privacy, harness the value of data, and ensure they are fit for the digital economy.

6.1 How does Infoboss help me comply with the GDPR?

6.1.1 Data audit and classification

A first step towards achieving GDPR compliance, and mitigating the risk of a potential data breach, is to conduct a data protection audit and in so doing classify your data to discover where personal and sensitive data is, and who has access to it.

- √ **Infoboss automatically collects, classifies and stores data in a searchable enterprise data store.**
- √ **Infoboss provides powerful search-based data discovery tools to find data of interest.**

"Infoboss has enabled us to systematically classify and discover personal and sensitive data across our entire data estate and crucially who has access to it."

– Information Compliance Manager

6.1.2 Enforce data retention policies

With so many systems in use within an organisation, and a variety of business rules surrounding what retention periods should apply for various types of data, it is not unsurprising that many data protection officers struggle to efficiently monitor and enforce their data retention policies.

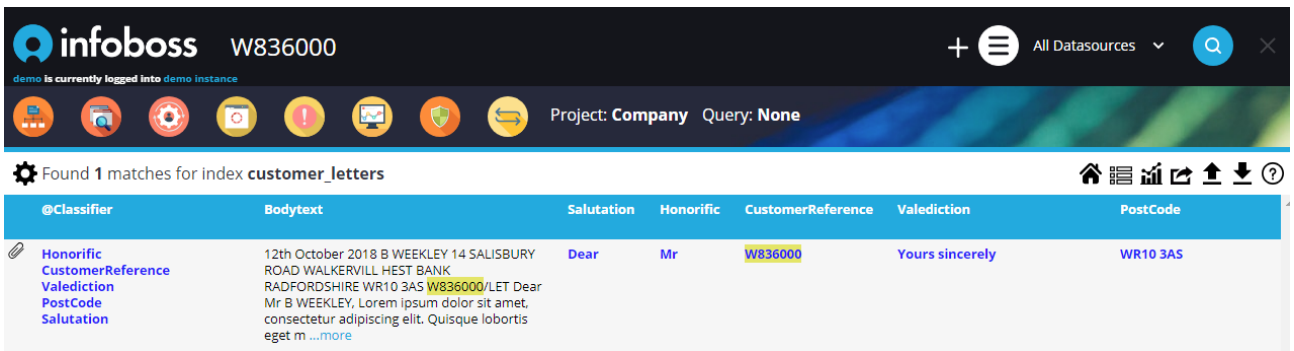
- √ **Infoboss can automatically monitor application databases and unstructured data sources to identify data that has passed its valid retention date. For example, board minutes that are over five years old.**

"We use infoboss to effectively police our data retention policy, ensuring it's enforced consistently across all systems." ***– Information Compliance Manager***

6.1.3 Service data subject rights

- √ **Data Subject Access Requests: Infoboss searches through your data and finds all related data in seconds. It can then export the data and related documents as a collection of files in a zip folder. The data can then be redacted if appropriate to do so, all in one place.**

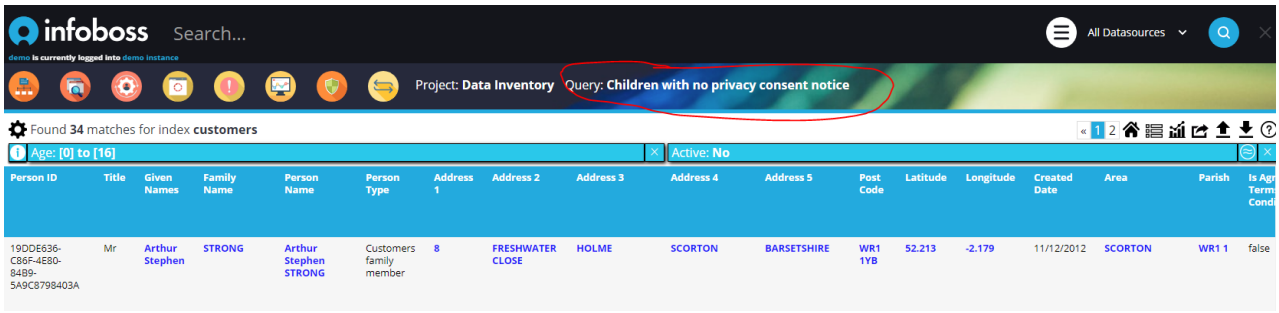
- √ **Right to erasure (right to be forgotten): Infoboss uses control lists which contain references to individuals whose data shouldn't be stored. Any individual requesting erasure can be added to the control list. If any data then appears in the data estate that matches the individual, an alert is triggered to the data controller highlighting that action needs to be taken.**



6.1.4 Consent management

Do you have consent to undertake action on a data subject's data in accordance with the GDPR? Are you processing child data with the appropriate level of consent?

- √ **Infoboss enables you to configure rules to check for consent and report on cases where consent has not been given (or has not been recorded).**



6.1.5 Data portability

Under the GDPR individuals can request that personal data is moved, copied or transferred easily and securely from one IT environment to another.

- √ **Infoboss ensures information can be easily located, retrieved and sent on within the set timescale in an approved format.**
- √ **Infoboss can anonymise personal and sensitive data automatically as part of the export process**

6.1.6 Data minimisation

GDPR recommends minimising the data held about data subjects to the minimum needed to fulfil contractual obligations.

- √ **Infoboss, through its data classification process and search-based rules, enables you to identify all types of personal information held. For example, if you are not entitled to hold payment card details, or process child data without consent, then Infoboss can be configured to find this data within the estate and then automatically alert the data controller.**
- √ **Infoboss can also be used to identify redundant, obsolete and trivial (ROT) data.**

"We used the results of our data audit to inform our data minimisation activities ahead of a significant data migration project." – Information Compliance Manager

6.1.7 Breach notification

The GDPR introduces a duty on all organisations to report certain types of data breach to the relevant authority, and in some cases to the individuals affected, within 72 hours of becoming aware of it.

- √ **Infoboss can be used to monitor system logs for suspicious activity and alert if an incident occurs. Affected data from source systems can be identified and used to inform the data breach reporting process to those affected and the necessary authorities.**

7 The business case for automated information governance

There is a compelling business case for automated information governance, improving the quality of your organisation's data and adopting more effective and efficient data protection and compliance processes. It's not just the threat of fines and the reputational impact for the business, but the very real costs associated with quality improvement and compliance activities in general.

To justify the investment in technology and process change to best prepare your organisation for the tsunami of data and regulatory change will require C-suite approval for budget and business transformation. This section of the document provides some insight into the business case for centralising and automating your organisation's information governance approach.

The Ponemon Institute is a well-regarded organisation that conducts research on compliance and data related matters globally. Here are some figures from their recent research…

- The average cost of compliance per employee is £2,601 per annum. This figure can be multiplied by 2.71 if the organisation is non-compliant!
- On average 14.3% of IT budgets are spent on compliance related projects
- The cost of a data breach in the UK is £99 per record. This includes fines and work to remedy the problem. Therefore a 10,000 record data leak could result in over a £1m of direct cost.
- GDPR compliance is *the most difficult* compliance to achieve

Further research by BitGlass recently revealed the stock value of companies that have had a data breach typically saw their company valuation drop by 7.5% in the aftermath.

The Ponemon Institute in their research observed that **organisations that centralised their information governance and undertook greater than five audits per annum saved on average 29%** of their compliance budget. As explained earlier, the opportunity to automate significant parts of GDPR compliance enables continuous monitoring of data for compliance purposes, potentially achieving even greater savings than this.

7.1 The data quality dimension

Research from Gartner has shown a marked increase in the cost of poor-quality data on an organisation.

"The average cost to an organisation of poor-quality data in 2018 is $15 million per annum!" (Gartner 2018)

This is up from $9.7 million dollars in 2017 and is a trend that is unlikely to improve due to the reliance of an increasing number of business initiatives, on quality data.

What is significant about this finding is this…

"Two-thirds of all business leaders believe that their companies must pick up the pace of digitalization to remain competitive." (Gartner 2017)

The consequence of poor-quality data is therefore that initiatives that rely on data, such as digital transformation, and business analytics, will not deliver the business outcomes expected. In extremis they may fail completely if the data quality is not improved, and the required processes and procedures for effective data input, collection and use are also not put in place. Information governance is crucial to the success of these initiatives.

8 Conclusion

We are in an age, the information age, that is characterised by our need to access and control information. Yet achieving this is severely challenged as we see huge growth in data volumes, unprecedented rise in unstructured data (the most difficult type of data to manage and control) and an ever increasing need to satisfy data related regulatory compliance.

There is a compelling business case both in terms of efficiency gains and tangible financial savings to adopt a strategy for centralising information governance and automating the management of data. Embedding data protection by design and default into the DNA of the organisation and future proofing the organisation's ability to tackle compliance challenges ahead.

8.1 Contact Infoboss

Infoboss is a software and services company that have developed an innovative technology platform for automating information management. Why not join the growing number of organisations that are Infoboss*ing* their data using the technology.

For further information, demonstration or to discuss your information governance challenges please contact us via email info@infoboss.co.uk or via the contact form on our web site www.infoboss.co.uk