



Data quality and compliance metrics

A business executives guide to the things you need to focus on to ensure you have fit-for-purpose data available for your data consuming initiatives such as AI and BI.

Data solutions. Data quality. Data compliance. Search & discovery.
Data Lakehouse. Custom data solution. Cloud migration.



Data quality and compliance metrics

A business executives guide to the things you need to focus on to ensure you have fit-for-purpose data available for your data consuming initiatives such as AI and BI.

info@infoboss.co.uk



**Business
Intelligence**



**Data
science**



**Artificial
Intelligence**



**Data
pipelines**



Search

Quality and compliant data

Metadata, Governance, Processing and Management



 **infoboss**



Structured, semi-structured and unstructured data

Introduction

We all know that data quality and compliance is important for our data initiatives. You know what I'm referring to, the projects that, seek to consume and leverage your data for insights or other purposes from the likes of BI (analytics), reporting, data science, artificial intelligence and other potentially value adding data consuming practices. They all rely, no, depend, on trusted quality & compliant data. So, what do we need to measure and monitor to ensure our data is fit for its intended purpose? and why?

Fortunately, a lot of the groundwork has been done for us by the Data Management Association (DAMA) who identified six dimensions that we should consider with regard to ensuring data suitability for its intended purpose. For each dimension I've listed the definition, why it matters, the metric and how infoboss helps. Sometimes people extend these dimensions adding ones of specific relevance to their sector. A common one you often see added is a regulatory dimension which for our purposes is covered in a section on its own for data compliance metrics.

Data quality metrics

The DAMA six dimensions are listed below and mainly cover data quality considerations.

1. [Data Accuracy](#):

- Definition: How closely data reflects the real-world entities or events it is supposed to represent.
- Why it matters: Inaccurate data can lead to poor business decisions, flawed reporting, and financial losses. It is critical for all those seeking to
- Metric: Percentage of accurate records (e.g., % of customer details that are up-to-date).

This dimension is perhaps the most difficult to measure as it will typically involve inspection of the real world entity the data represents to confirm that it is in fact accurate. Another method is to check your data against trusted data and highlight any discrepancies. These are things that infoboss tools can help you achieve. As a business executive you should ensure that a data accuracy assessment is undertaken at a frequency to suit the importance and potential use of the data.

2. [Completeness](#):

- Definition: Whether all required data is available and populated within the dataset.
- Why it matters: Missing data may result in incomplete insights, making it difficult to perform accurate analyses.
- Metric: Percentage of missing values (e.g., fields missing in a customer database or sales transactions).

This is a fairly simple measurement. I.e. have all the things we need to capture, been captured?

Filter preferred communication method

Values Export

Search for values

- (300) <None>
- (746) Text message
- (617) Face to Face
- (482) Mail
- (141) Phone
- (124) Email

Minimum matches

1

Apply

This is a simple validation rule that can be applied to any field in infoboss. There are however often subtleties to this dimension in that completeness may not be achieved on day 1, it might be that a new client for example may take some time in coming back with KYC information so completeness checks would need to be undertaken at a later date. Again, this is something the infoboss tooling can do for you.

3. [Consistency](#):

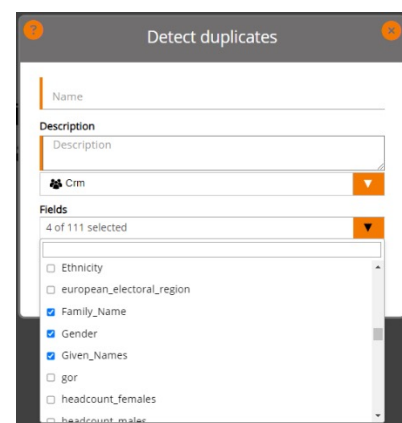
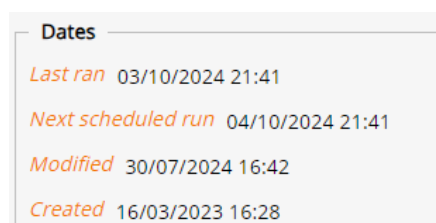
- Definition: Whether the same data is consistently presented across different systems or datasets.
- Why it matters: Inconsistent data (e.g., mismatched names or addresses) across systems can undermine trust in analytics or reporting.
- Metric: Percentage of inconsistent entries across systems.

This issue tends to arise in more complex organisations that have several systems being used to maintain data with integration between them. Fortunately, infoboss can do this for you by connecting to both sets of data and then running comparison rules across key data points e.g. are the mobile phone numbers or email addresses the same for this customer in both systems.

4. [Timeliness](#):

- Definition: Whether data is up-to-date and available when required.
- Why it matters: Delayed data can lead to out-of-date insights, impacting decision-making in real-time business processes.
- Metric: Average data update latency (e.g., how long it takes for transaction data to update in your reporting system).

Infoboss tracks the data collection cycle for every data set. You can see when it was run, when it's next scheduled to run, when it was changed and when it was first created. This timeliness data is available directly to the user when looking at the data.



5. [Uniqueness \(De-Duplication\)](#):

- Definition: The degree to which data is free from duplicates.
- Why it matters: Duplicate records can inflate metrics, skew reporting, and cause inefficiencies.
- Metric: Number or percentage of duplicate entries within datasets.

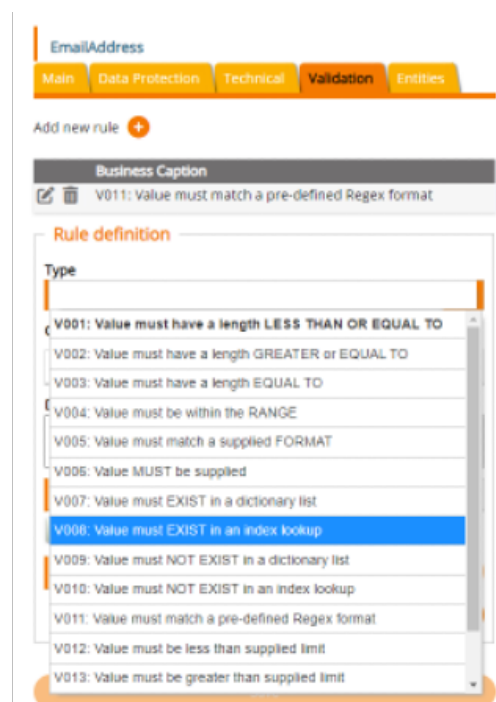
Infoboss has comprehensive duplicate detection functionality. Simple point and click during discovery will identify instances of duplicity in individual fields, but in the real world that doesn't tell the story. Two surnames of "Smith" aren't necessarily duplicates, however surnames, and date of birth when combined would indicate potential duplication, but what

if the data was twins or more then you may want to also include the person's first names. Essentially you can combine any combination of fields to help identify duplicity in your data.

6. Validity:

- Definition: Whether the data is stored in the correct format and meets the business rules or data types required.
- Why it matters: Invalid data (e.g., letters in a phone number field) can disrupt processes or cause system errors.
- Metric: Percentage of data fields meeting the format requirements.

This type of error along is perhaps the one that requires the greatest amount of checks or business rule consideration. Simple format checks like post codes, payment card numbers, bank sort codes can be relatively easily applied. But with infoboss it can do much more. For example, you may want to check that all surnames are correctly capitalised (including for double barrelled names). You may also want to check the length of a surname i.e. must be 2 or more characters. You may want to check the range, minimum or maximum values. For example, a date of birth that yields an age in excess of 110 years or that is in the future? You will also want to apply checks for categorised values e.g. religion, gender, title and so on.



Other data quality considerations

In addition to the dimensions there are other things you might want to measure and monitor with regard to your data quality management. Ideally you want the business to own the process. For this to work, there must be clearly defined roles and responsibilities usually established by your data governance team. But that alone does not ensure that things like actually fixing the data get done. Often the scale of the task is so big it needs to be tackled in incremental steps, perhaps focussing on high priority or high impact issues first (such as regulatory or non-compliance) and then move on to perhaps more voluminous lower priority issues later like missing or invalid data. Tracking progress towards acceptable levels of data quality requires analysis of trends in quality scores associated with data sets but also fix rates. You also need to bear in mind that new issues may come in because you haven't root cause analysed the reason for the error occurring and this is imperative to ensure you actually switch off the dirty data tap. You cannot reasonably expect to fix data if you are still generating and accepting poor quality data as it enters the business.

One of the challenges that data teams face is that they are not given reasonable priorities and with a need to serve many masters in the business often are exposed in terms of what to focus on first. Clearly, this is something the executive team need to take a lead on. If AI is your priority for example, then focus on the issues that will affect that, but if you don't resource the team effectively you may not be able to ensure that data is fit for other

purposes like management reports or operations. As you begin to embark on the data quality & compliance journey and you have achieved some success you need to continually ensure the process is managed in order to maintain the standards you are now setting for the business.

Data Compliance Metrics

1. Regulatory Compliance:

- Definition: The degree to which your company's data practices align with relevant regulations (e.g., GDPR, CCPA, HIPAA).
- Why it matters: Non-compliance can lead to heavy fines, legal risks, and reputational damage.
- Metric: Compliance audit results (e.g., % of compliant records, or number of compliance violations over time).

Infoboss' comprehensive rules engine allows for cross field, even cross data set checks to ensure compliance. For example, you may have a watch list for GDPR where the data subject has requested to be removed from your records. What you don't want to happen is that data that you believe you have removed suddenly re-appears (perhaps a backup restore from before the data was removed?) and the client continues to be mailed. Dissatisfied client and a breach of the regulations.

2. Data Retention and Deletion:

- Definition: Whether data is stored and deleted in accordance with legal and regulatory requirements.
- Why it matters: Keeping data longer than necessary can increase security risks and breach regulatory guidelines.
- Metric: Percentage of data meeting retention policies (or percentage of data flagged for deletion).

Infoboss works across not only structured data assets but also unstructured. You are able to classify and categorise all types of data and apply data retention rules to flag up data that breaches your business rules.

3. Data Access and Security:

- Definition: Whether sensitive data is protected by proper security measures and only accessible to authorised personnel.
- Why it matters: Data breaches can result in financial penalties, loss of customer trust, and reputational damage.
- Metric: Number of unauthorised data access incidents or percentage of sensitive data protected by encryption.

Knowing what data you have, whether it is classified correctly and where its stored is often the first steps in understanding who should have access to it? For example, finding PII data on a publicly accessible shared drive is probably not what you want to see! Infoboss helps

you classify (automatically) data and determine what rules should apply to monitor your data assets to ensure access and security considerations are in place.

4. Auditability:

- Definition: The ability to trace data changes, movements, and access patterns for auditing purposes.
- Why it matters: Audits are essential for verifying data compliance with internal policies and external regulations.
- Metric: Percentage of processes that have traceable logs and complete audit trails.

Infoboss has developed a special type of data set called a timeline index. This index tracks every change to the data and records the original data and the changed data so a complete audit trail is available. This helps when auditing and investigating data issues and even more so with the tools to search and analyse huge volumes of data are available out of the box.

5. Consent Management:

- Definition: Tracking whether customer data is collected, stored, and used based on their explicit consent (important for GDPR and similar regulations).
- Why it matters: Using data without consent can lead to hefty fines and legal challenges.
- Metric: Percentage of customer records with valid consent or number of non-compliant consent incidents.

As detailed in point (1) above the ability to check data against lists is crucial for determining any breach of compliance rules.

6. Breach Response Time:

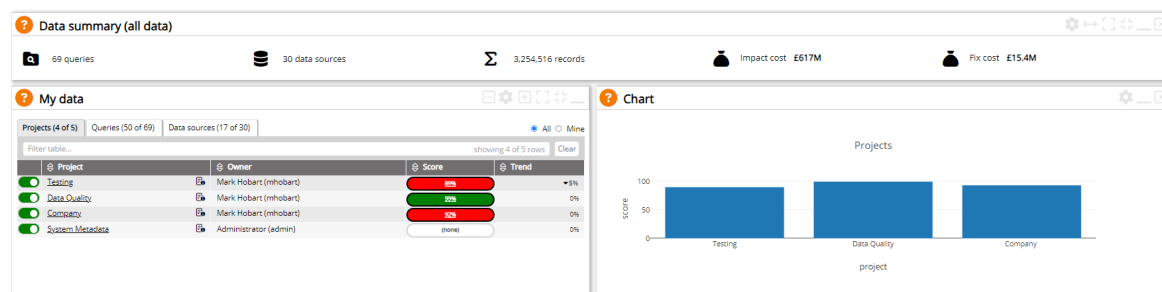
- Definition: The time it takes to detect, report, and mitigate a data breach.
- Why it matters: Quick responses to data breaches can reduce regulatory penalties and limit reputational damage.
- Metric: Average time to respond to a data breach or percentage of incidents resolved within a specified time frame (e.g., 72 hours for GDPR reporting).

Many of our clients have used infoboss to support their Data Subject Access Request processes. Providing the means to search data, documents and emails to find relevant texts to service the request.

Sadly, data breaches do occur and as said, they require a rapid response and the ability to not only identify what data was breached but also whether there are any GDPR or other regulatory compliance issues arising is crucial. Indeed one of our clients used infoboss after an email breach to identify all parties that had emails compromised as a consequence of a breach. They were able to do this within a matter of hours and keep within the 72 hour GDPR reporting target.

How to Monitor These Metrics:

- Dashboards and Reports: Are a common way, but you need to collect the data in the first place. Infoboss allows you to do this and implement automated dashboards that track data quality and compliance metrics, providing real-time updates for decision-makers.



- Periodic Audits: Conduct regular internal audits or leverage third-party audits to measure adherence to data quality and compliance standards. Particularly relevant for the accuracy dimension.
- Data Governance Frameworks: Establish a formal data governance framework that defines roles, processes, and responsibilities for ensuring data quality and compliance. If you don't have a data governance function and are endeavouring to capitalise on your data, then you simply must put in place the required governance to ensure your data is fit for purpose.
- Data Quality Tools: Use tools like infoboss for automated data quality & compliance monitoring.

Conclusion:

Infoboss provides not only the means of capturing the data required to support your business monitoring of your data quality & compliance requirements, but also affords the opportunity to put in place data management processes to support your data initiatives.

By closely monitoring data quality and compliance metrics, you can ensure that your organisation maintains high standards for data integrity, supports effective decision-making, and avoids regulatory penalties. Monitoring these metrics will also help build customer trust and operational efficiency and deliver a quality & compliant data set to your data initiatives that seek to add value to your business such as AI and BI.

To discover more, why not drop an email to info@infoboss.co.uk outlining what you are trying to achieve with your data and we'll setup a meeting to discuss.

infoboss.co.uk
info@infoboss.co.uk
0333 772 1963
[linkedin.com/company/infoboss-limited](https://www.linkedin.com/company/infoboss-limited)

