

# What's Driving the Demand for Self-Service Data Quality in Insurance?

## **The drivers and benefits of a holistic, self-service data quality platform**

In this paper we describe trends and technologies bringing data quality functions closer to the data and moving responsibility and control from central IT functions to data stewards and business users, thereby achieving greater operational efficiency and higher value data assets.

**Published on 27<sup>th</sup> July 2021**

# Contents

1. Data Quality fundamentals driving valuable Data Insights in Insurance .....	2
2. Key Features a Self-Service DQ Platform Should Have .....	4
3. Conclusion .....	10
4. Next steps .....	10
5. References .....	10
6. About Dataactics .....	10
7. General Enquiries: .....	11

## 1. Data Quality fundamentals driving valuable Data Insights in Insurance

### Data in a Changing World

The Insurance industry traditionally uses data to inform decision-making and manage growth and profitability across marketing, underwriting, pricing and policy servicing processes. However, like most established financial institutions, insurance companies have many data repositories and different teams managing analytics functions. Traditionally, they also struggle to share this information or communicate with one another, with many organisations having their own processes for capturing data. These factors combine to cause poor quality and inconsistent data, creating barriers toward seamless integration.

The Insurance industry recognises the importance of maintaining a competitive edge, with many companies looking to adopt a '**single platform**' approach using Cloud Services from AWS, Azure or Google in the short to medium term. Such a platform needs to be flexible to support different skill sets, react to changing market conditions and able to integrate alternative sources of data. Fundamental to this is the quality of data across different data sources, ensuring it is trusted, of a high degree of integrity, and complete for business decisioning purposes.

### Challenges

Customer insights are isolated to silos and scattered across lines of business, functional areas and even channels. As a result, much of the work that surrounds the handling of data becomes manual and time consuming, with no common keys or even set definitions of key terms, i.e., 'customer'. It is estimated that as much as 70% of a highly qualified analyst's time is spent locating and fixing the data.

The challenge for Insurance companies is being able to recognize the same customer across product lines and/or at different stages of the policy lifecycle. Direct and agency channels may compete for the same customer or attract a high-risk prospect that was turned down previously by underwriting. Since the claims department data is not available to pricing and marketing to inform their decisions, the result is often extra expenditures and a larger than necessary marketing budget that could easily be streamlined

should these inefficiencies be addressed. It also causes poor customer experiences, which harm the brand.

There is, however, a significant demand for customer-centric solutions which allow insurance companies to link different pieces of data about a customer. These solutions use Data Quality tools to match, merge and link records, creating a holistic view across product lines and throughout the policy lifecycle.

Customer-centric solutions help insurance companies realise important business goals, including more accurate targeting, longer retention and better profitability.



## Opportunity

Generating valuable insights from expanding data sets is becoming significantly harder. On top of this, leveraging the right technology, people and process to analyse data remains a key challenge for Executives. Prepping the data is often where the real heavy lifting is done and using Data Quality automation and a Self-Service approach can really benefit a company in terms of significantly reducing costs and accelerating decision making.

While the Insurance industry faces a plethora of challenges with data and analytics, it's imperative that executives recognize that the quality of the data is fundamental to capitalising on market opportunities. By overcoming these barriers, the industry will be better prepared to embark on the next frontier of Data and Analytics (D&A).

## 2. Key Features a Self-Service DQ Platform Should Have

To enable the evolution towards actionable insight from data, D&A platforms and processes must evolve too. At the core of this evolution is the establishment of ‘self-service’ data quality – whereby data owners and organisations have ready access to robust tools and processes, to measure and maintain data quality themselves, in accordance with data governance policies.

From a business perspective such a self-service data quality platform must be:

---

***“Self-Service Data Quality empowers Business Users to implement Data Quality automation, reduce heavy lifting and associated cost, and drive valuable Data Insights. Data Owners can remediate measure and maintain data quality themselves in accordance with governance policies”***

---

- ❖ Powerful enough to enable business users and organisations to perform complex data operations without highly skilled technical assistance from IT
- ❖ Transparent, accountable and consistent enough to comply with firm wide data governance policies
- ❖ Agile enough to quickly onboard new data sets and changing data quality demands of end consumers such as AI and Machine learning algorithms
- ❖ Flexible and open so it integrates easily with existing data infrastructure investment without requiring changes to architecture or strategy
- ❖ Advanced to make pragmatic use of AI and machine learning to minimize manual intervention

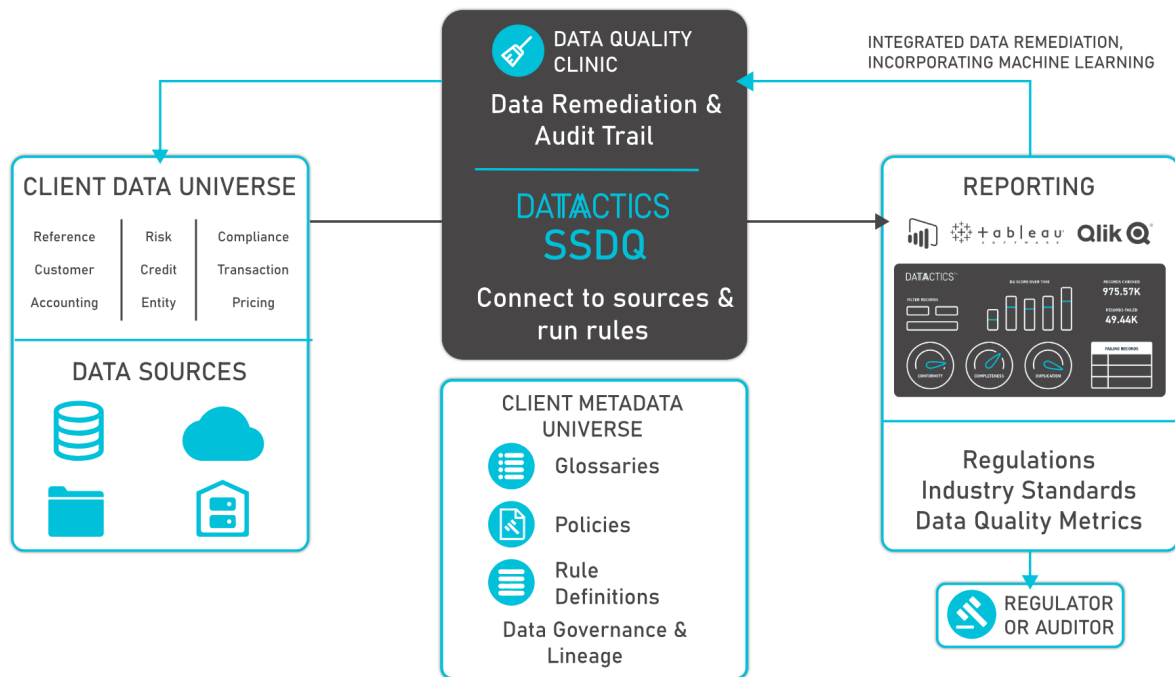


Figure 1 Self-Service Data Quality Platform integrations, outputs and remediation loop.

This goes way beyond the scope of most stand-alone data prep tools and ‘home grown’ solutions that are often used as a tactical one-off measure for a particular data problem. Furthermore, for the self-service data quality platform to truly enable actionable data across the enterprise, it will need to provide some key technical functionality *built-in*:

- **Transparent & Continuous Data Quality Measurement**

Not only should it be easy for data stewards and business users to implement large numbers of data domain specific data quality rules, but also those rules should be simple to audit, and easily explainable, so that ‘DQ breaks’ can be easily explored and the root cause of the break established.



Figure 2  
Example DQ Dashboard, built in PowerBI, taken from Datactics SSDQ Platform

In addition to data around the actual breaks, a DQ platform should be able to produce DQ dashboards enabling drill-down from high level statistics down to actual failing data points and publish high level statistics into data governance systems.

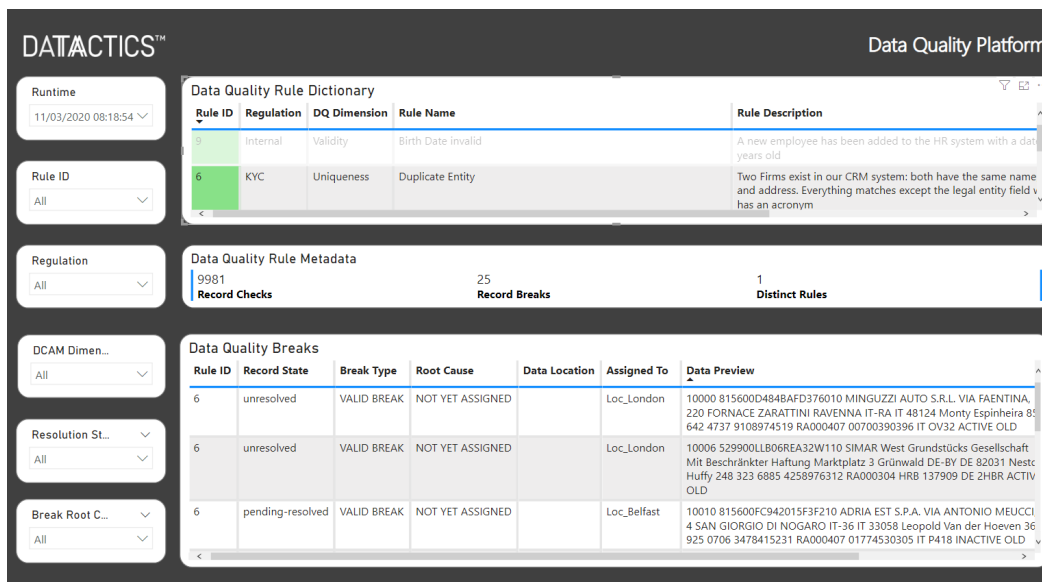


Figure 3  
Detail portion of data breaks

- **Powerful Data Matching - Entity Resolution for Single View and Data Enrichment**

Finding hidden value in data or complying with regulation very often involves joining together several disparate data sets. For example, enhancing a Legal Entity Master Database with an LEI, screening customer accounts against sanctions and PEP lists for KYC, creating a single view of client from multiple



data silos for GDPR or IFRS17 compliance. This goes further than simple deduplication of records or SQL joins – most data sets are messy and don't have unique identifiers and so fuzzy matching of numerous string fields must be implemented to join one data set with another. Furthermore, efficient clustering algorithms are required to sniff out similar records from other disparate data sets in order to provide a single consolidated view across all silos.

Source	SourceID	Prefix	FirstName	LastName	DOB	Gender	StreetAdd	State	Postcode	Country	Company
114825											
CRM	10120738	Mr.	Alexander	Brown	2017-07-08	Male	39554 Laura Centers	Michigan	29980-7918	United States	Warner PLC
CRM	10560119	MR	Alexander	BROWN	2017/07/08	M		MI		USA	WARNER
INV	20161358	Mr.	Alexander	Brown	2017-07-08		39554 Laura Centers	Michigan	29980-7918	United States	
115531											
CRM	10121462	Mr.	Alexander	Browning	1941-01-18	Male	778 Bradley Hills	South Carolina	36199	United States	Lewis-Whitney
CRM	10560502	MR	Alexander	BROWNING	1941/01/18	M		SC		USA	LEWISWHITNEY
123691											
CRM	10129974	Mr.	Alexander	Brown	2005-12-12	Male	171 Beasley Cove	Montana	65707	United States	Hobbs, Brooks and Phillips
CRM	10564766	MR	Alexander	BROWN	2005/12/12	M		MT		USA	HOBBS BROOKS & PHILLI
MAR	30032436	Mr.	Alexander	Brown			171 Beasley Cove	Montana	65707	United States	
MAR	30127585	MR	Alexander	Brown			171 Beasley Cove	MT	65707	united states	

Figure 4 Clustering algorithms find similar data across multiple data sets

## • Integrated Data Remediation Incorporating Machine Learning

It's not enough just to flag up broken data, you also need a process and technology for fixing the breaks. Data quality platforms should have this built in so that after data quality measurement, broken data can be quarantined, data owners alerted and breaks automatically assigned to the relevant SMEs for remediation. Interestingly, the manual remediation process lends itself very well to machine learning. The process of manually remediating data captures domain specific knowledge about the data – information that can be readily used by machine learning algorithms to streamline the resolution of similar breaks in the future and thus greatly reduce the overall time and effort spent on manual remediation.

***“The process of manually remediating data captures domain specific knowledge about the data – information that can be readily used by machine learning algorithms to streamline the resolution of similar breaks in the future”***



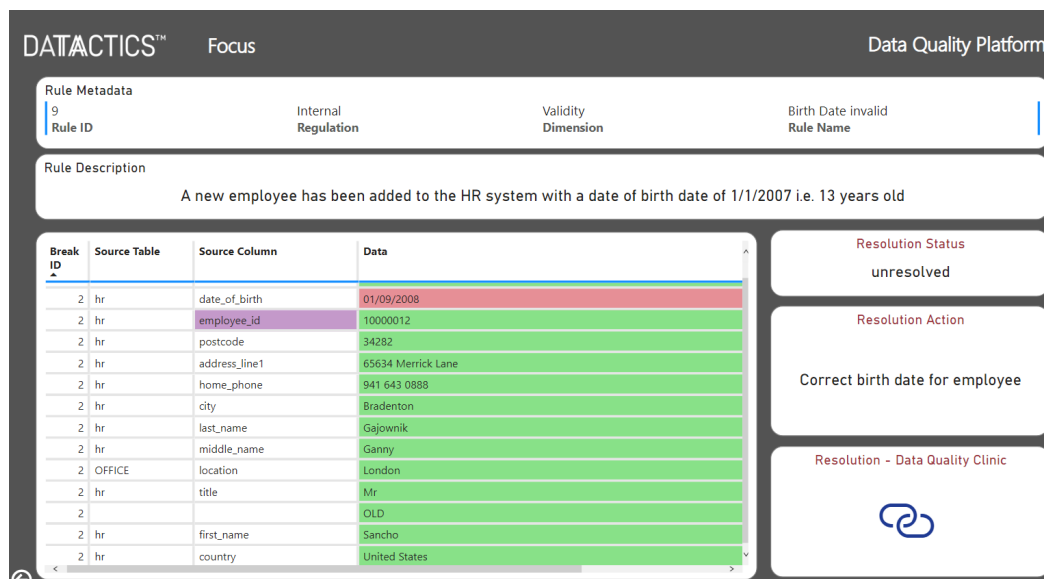


Figure 5  
Specific break highlighted with link through to resolution tool.

## • Data Access Controls Across Teams and Datasets

Almost any medium to large sized organisation will have various forms of sensitive data, and policies for sharing that data within the organization e.g., ‘Chinese walls’ between one department and another. In order to enable integration across teams and disparate silos of data, granular access controls are required – especially inside the data remediation technology where sensitive data may be displayed to users. Data access permissions should be set automatically where possible (e.g. inheriting Active Directory attributes) and enforced when displaying data, for example by row- and field-level access control, and using data masking or obfuscation where appropriate.

DATA-SET-COLLECTION A0100 / DATA-SET ea3ddf38 / CLUSTER 0df06607 / RECORDS

Next	Previous	Save	Copy Field	Copy All Fields	Save Resize	Cancel	RULE ID	LOCATION	BREAK TYPE	X ID	X TITLE	X FIRST NAME	X MIDDLE NAME	X LAST NAME
							A0055	london	1	GA6439	MS	MAUD	LORRAINE	WYLIEN
							A0055	belfast	1	GA6438	MISS	MAUD	LORRAINE	WYLIE
							A0055	edinburgh	1	GA6431	MRS	MAUD	LORRAINE	WYLLIE

Winning Record: A0055 | edinburgh | 1 | GA6439 | MISS | MAUD | LORRAINE | WYLIEN

State: resolved Assigned To: london

Figure 6  
Example manual remediation screen in self-service data quality

## • Audit Trails, Assigning and Tracking Performance

Providing business users with tools to fix data could cause additional headaches when it comes to being able to understand who did what, when, why and whether or not it was the right thing to do. It stands to

reason, therefore, that any remediation tool should have built-in capability to do just that with the associated performance of data break remediation measured, tracked and managed.

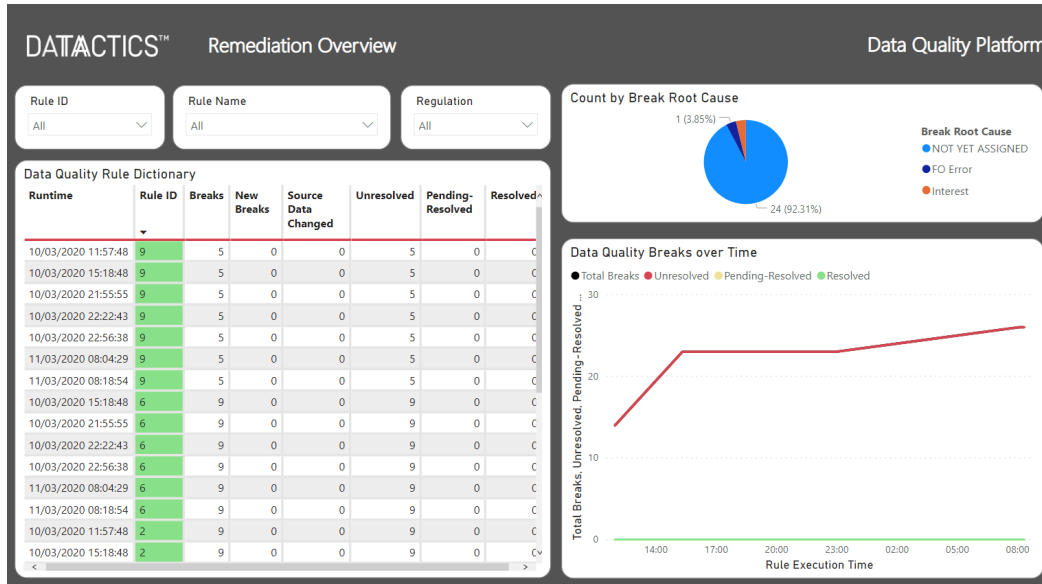


Figure 7  
Tracking and measuring DQ break remediation activity

## • AI-Ready

There can be no doubt that one of the biggest drivers of data quality is AI. AI data scientists can spend up to 80% of their time just preparing input data for machine learning algorithms, which is a huge waste of their expertise. A self-service data quality platform can address many of the data quality issues by providing ready access to tools and processes that can ensure a base level of quality and identify anomalies in data that may skew machine learning models. Furthermore, the same self-service data quality tools can assist data scientists to generate metadata that can be used to inform machine

***“AI data scientists can spend up to 80% of their time just preparing input data for machine learning algorithms, which is a huge waste of their expertise”***

learning models – such ‘Feature Engineering’ can be of real value when the data set is largely textual as it can generate numerical indicators which are more readily consumed by ML algorithms.

### 3. Conclusion

The move towards a self-service oriented model for data quality is a logical way to keep up with the expanding volumes and varieties of data. According to Corinium Intelligence in the [Future of Insurance Data \(coriniumintelligence.com\)](https://coriniumintelligence.com/future-of-insurance-data) "Data leaders at these companies must make tough decisions about which investments provide the most accurate view of risk, for the best value." Organisations that look to deploy a Data Quality Framework approach using a self-service data quality platform are more likely to benefit from actionable data, resulting in deeper insight, in more effortless compliance, and in significant competitive advantage.

### 4. Next steps

To discuss Self-Service Data Quality, book a demonstration or obtain case studies, please contact Julian Wavell, Head of Business Development. [Julian.Wavell@Datactics.com](mailto:Julian.Wavell@Datactics.com)

### 5. References

1. EDMC's DCAM: <https://edmcouncil.org/page/aboutdcamreview>

### 6. About Datactics

[Datactics](#) helps Insurance companies drive valuable Data Insights, supports Operational Data needs and process, including Data Governance and Compliance & Regulation by removing roadblocks common in data management. We specialise in class-leading, [self-service data quality](#) and [fuzzy matching](#) software solutions, designed to empower business users who know the data to visualise and fix the data.

## 7. General Enquiries:

Website: [www.dataactics.com](http://www.dataactics.com)

Email address: [info@dataactics.com](mailto:info@dataactics.com)

Phone number: +44 (0)28 9023 3900

Dataactics Ltd.  
1 Lanyon Quay  
BT1 3LG  
Belfast  
+44 (0)28 9023 3900  
[dataactics.com](http://dataactics.com)

# DATACTICS™

