Data Quality 2022

Introduction

The data quality market is far from new, with organizations around the world struggling with their own data quality ever since the first customer's name and address was entered into a computer system. In a global organization, data about customers, products and assets is scattered amongst multiple systems, across business divisions and geographies, throughout multiple ERP implementations and amongst local marketing systems. Survey after survey shows that C level executives have limited trust in the reliability, accuracy, timeliness and completeness of their data, despite their own accounts depending on this data being correct. Despite the considerable costs associated with poor data quality, the issue has remained a thorny one over the years. Internal office politics, data ownership and control disputes and basic human nature makes it difficult to be sure that data is pristine and reliable across an enterprise.

Customer name and address data was the first focus of the data quality software industry, since just about every corporation and organization has such data in one form or another: hospitals may have patients rather than *"customers"*, and government departments may have *"citizens"*, but the problem is the same. Just as night follows day, customer data ends up being duplicated in multiple systems, with duplication rates of 20% or more being normal for customer databases, with similar duplication rates in other key systems such as materials master files. The author of this report recalls an 80% duplication rate of corporate customers in one global corporation's customer master file that he worked on. Data quality software applies clever algorithms to detect likely duplicate records and either automatically remove them or highlight them for expert human intervention.

These days, data quality suites do much more. Data can be profiled, duplicates detected and cleansed, business rules applied, missing data can be filled in from reliable sources, source systems scored for reliability and support given to data stewards in businesses to help them deal with issues and monitor ongoing data quality. Existing data can be substantially enriched with all manner of potentially useful additional data brought in from assorted third-party sources, from credit risk scores to precise geocoding and more. Many of the same techniques can be applied to other data domains, such as data about products, assets and locations.

This report covers both vendors that are pure play data quality vendors, and others that have data quality solutions embedded within broader software solutions, which may extend to the related areas of master data management, data governance or data transformation.

Figure 1: The highest scoring companies are nearest the centre. The analyst then defines a benchmark score for a domain leading company from their overall ratings and all those above that are in the champions segment. Those that remain are placed in the Innovator or Challenger segments, depending on their innovation score. The exact position in each segment is calculated based on their combined innovation and overall score. It is important to note that colour coded products have been scored relative to other products with the same colour coding.
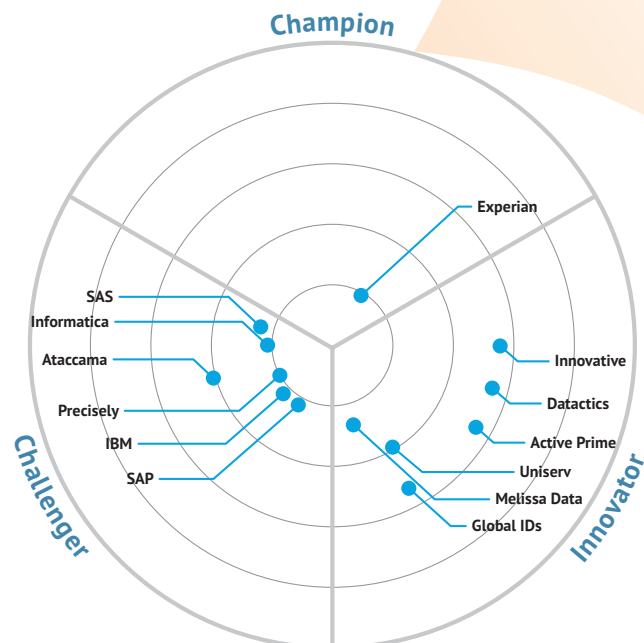
Market trends

In recent years data quality suites have generally expanded in scope to cover most common elements of functionality, and vendors seek to differentiate themselves on the quality of their algorithms or the performance and scalability of their products. Machine learning can be trained to observe human data stewards resolving potential data quality issues, and can suggest new business rules based on these observations. Artificial intelligence can be brought to bear to better detect data anomalies and outliers, which can be crucial in financial fraud detection and screening of airline passengers against terrorist watch lists, for example. Some vendors have developed industry specific datasets and business rules that are good at spotting common issues, and can be adapted and tuned over time as new patterns emerge. Vendors have provided support for cloud and hybrid solutions as well as on-premise solutions as customer themselves gradually migrate more of their core systems from on-premise to private and public cloud environments. There is an increasing emphasis on self-service tools that are aimed at business users, who are increasingly taking responsibility for data quality rather than relying on IT departments.

The Vendors

We approached a wide range of vendors in this field, and although many responded and participated in our research, not all did. Nonetheless, we have endeavoured to include the major players as well as those smaller vendors that we consider are providing particularly innovative solutions. Inevitably there are others that we did not include, and this is usually because of a matter of definition.

Metrics

This market update was based on interviews with vendors and additional research regarding their products and progress in the market. Criteria considered for each vendor and product offering were:

- Vendor stability and risk
- Vendor support and location
- How fit for purpose is the product in question
- Ease of use of product
- Product Performance
- Product Architecture
- Value for money
- Clarity of the marketing story
- Level of innovation
- Level of market disruption
- Market adoption
- Market results

Conclusion

The data quality market continues to evolve in response to market requirements, with the increasing move to digitisation in many industries meaning that data quality technologies need to be able to handle a wider range of data domains than just customer name and address. Data quality suites continue to broaden out in functionality, and are often now used in conjunction with related technologies such as master data management and data integration software. Just as some vendors are broadening their offerings in this way, there is still room for deeper specialization in specific industry verticals, and we see some vendors building industry-specific datasets and business rules that address common issues in industries such as financial services.

For customers, it can be difficult to fully understand and validate the claims of competing vendors within the blizzard of marketing material. Almost all will claim rapid implementation, high scalability and performance and the best and cleverest matching algorithms. It is important for customers to consider their own specific use cases and benchmark vendors on their own data rather than relying on vendor demonstration datasets. Careful testing of claims and taking up customer references is vital, and a consultation with an independent expert in the field to guide them through the myriad vendor claims can be a wise investment.

MarketUpdate

About the author

Andy Hayler
Senior Analyst

Andy is an established software industry authority, an independent strategy consultant advising corporations, venture capital firms and software companies. He is the founder of Kalido, which under his leadership was the fastest growing business intelligence vendor in the world in 2001. Kalido was recognised as an innovator in data warehousing, and then launched arguably the first true master data management product, a market which at the time did not exist but is now a well recognised and fast growing industry. Andy was the only European named in Red Herring's *"Top 10 Innovators of 2002"*. He was a pioneer in blogging with his award winning *"Andy On Enterprise Software"* blog.

Andy started his career with Esso, working in a number of technology roles before moving to Shell. He was Technology Planning Manager of Shell UK, then Principal Technology Consultant for Shell International. He later established a global information management consultancy, which under his leadership grew to 300 staff.

Bloor overview

Technology is enabling rapid business evolution. The opportunities are immense but if you do not adapt then you will not survive. So in the age of Mutable business Evolution is Essential to your success.

We'll show you the future and help you deliver it.

Bloor brings fresh technological thinking to help you navigate complex business situations, converting challenges into new opportunities for real growth, profitability and impact.

We provide actionable strategic insight through our innovative independent technology research, advisory and consulting services. We assist companies throughout their transformation journeys to stay relevant, bringing fresh thinking to complex business situations and turning challenges into new opportunities for real growth and profitability.

For over 25 years, Bloor has assisted companies to intelligently evolve: by embracing technology to adjust their strategies and achieve the best possible outcomes. At Bloor, we will help you challenge assumptions to consistently improve and succeed.

Copyright and disclaimer