

The Data Quality Market - Spring 2009

The data quality market is very fragmented. There are a number of reasons for this, but one is that many vendors have their origins in the customer name and address matching business, and this varies a great deal from country to country. For example the UK Royal Mail postal code database covers pretty much every UK address, as does the US Postal Service Zip+4 database for the USA, but the equivalent from the Italian national postal service covers barely half the country, while the Republic of Ireland does not have a national post code system at all.

Consequently a network of different information providers have sprung up, and software companies have also adapted to serve the specialist needs of local markets, with different ways of enriching the data found.

By no means all data quality is about customer name and address, and a few companies specialize in product data, which is much less structured and tends to be inherently more complex than address data (a product may have hundreds of attributes, for example). Here Silver Creek Systems, Inquera and Datactics each take different approaches to this thorny problem area. Other vendors provide broader, general purpose data quality suites. Another more recent trend is towards identity management, with some companies providing solutions that aim to ensure that the person at the other end of a business transaction is really who they claim to be. Despite there being literally dozens of data quality vendors, new ones still pop up. Ataccama is a recent entrant making solid commercial progress, while DataQuality First launched just as this research was being completed.

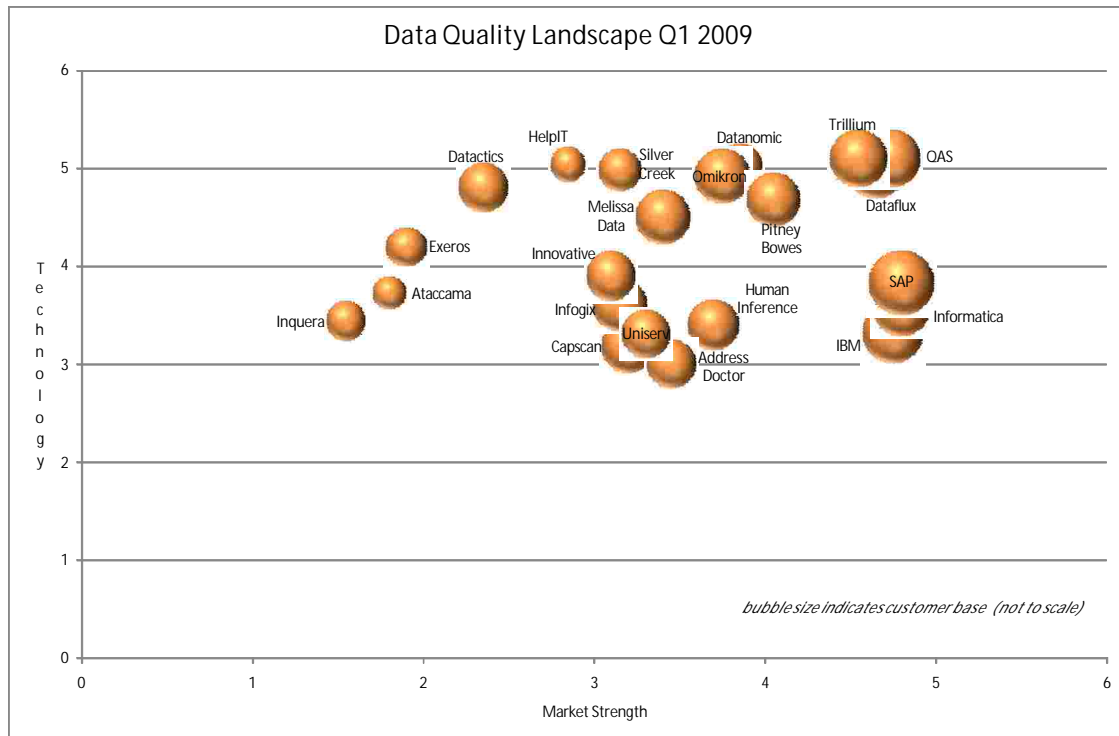
The Information Difference believes that the data quality market overall is worth around USD 740M, and according to our estimates grew at 17% over the previous twelve month period (this figure is the total associated with software and related services from data quality vendors, and does not include systems integration revenues that are associated with data quality). The pure software license market is around USD 400M.

A few years ago there was a clear distinction between vendors offering data profiling, merge/matching, enrichment or data quality monitoring. Customers increasingly expect an integrated solution, and in response these functional lines are blurring as many vendors have expanded their product coverage to provide broader functionality across the various elements of data quality. They have constructed a more complete suite of offerings either by building out their own software, through acquisitions, or via OEM of other software. There are some notable exceptions to this trend; for example Exeros have specialized in data discovery, and Netrics in matching algorithms with software that can be embedded in other applications, while Address Doctor partners with other data quality vendors to provide expanded international address coverage.

One technology trend has been for data quality vendors to embrace SOA standards, which has enabled data quality applications to be embedded into broader business processes, and for this to happen on-line rather than just in batch, which was usually the case in the past. Trillium, for example, report that over half their deployments are real-time rather than batch.

The major vendors in the data quality market are described using the Landscape diagram that follows (see later for more on how this is derived; note that due to a modification in the

methodology this time around by including customer satisfaction feedback, like-for-like comparisons between this chart and previous versions need to be treated with caution).



The largest players in terms of market strength (see below) in the data quality arena are SAP, Informatica, Trillium, QAS Experian, IBM and Dataflux. These have broad data quality offerings with large numbers of customer deployments. The data quality market is by no means entirely US-centric, with companies such as QAS Experian in the UK, Omikron and Uniserv in Germany and Human Inference in the Netherlands prospering in Europe.

A 2008 Information Difference study found that 30% of large companies had not deployed any kind of data quality tool, and the take-up even within companies that have some experience is very uneven, so despite the abundance of vendors this is far from a saturated market.

The profile of data quality has increased due to the rapidly escalating interest in master data management, since the quality of data is an inherent part of any master data management initiative. Companies that have set up data governance teams to help with master data issues have logically taken an interest in data quality tools, and this has raised the profile of data quality significantly within many companies. We see more partnerships developing between MDM platform vendors and data quality vendors. Indeed, as some data quality vendors build out data governance offerings within their own software, they are starting to overlap more with the MDM players. We believe that further inter-locking of these inherently related markets is certain, both through further extended partnerships and through acquisitions.

Main Vendors

In addition to the vendors shown on the landscape diagram, the following lists vendors which provide data quality software but did not qualify for inclusion on the landscape diagram on the basis of size in terms of their revenue.

Vendor	Brief Description	Website
AMB DataMiners	Chicago based vendor specializing in just-in-time data quality with in-stream profiling and outlier detection.	http://ambpdm.com/
Address Doctor	Vendor that specialises in providing wide coverage of name and address information; it is used by many other data quality vendors.	http://www.addressdoctor.com
Ataccama	Prague-based start-up with a modern data quality suite	http://www.ataccama.cz/
Business Data Quality	UK-based vendor with good government customer references.	www.businessdataquality.com
Capscan	London-based provider of address management and data integrity services.	http://www.capscan.com
Datactics	UK-based vendor specialising in product data quality.	http://www.datactics.com
Datanomic	Cambridge-based vendor of data quality solutions.	http://www.datanomic.com/
Dataflux	Part of SAS, one of the leading players in data quality.	http://www.dataflux.com
DataQualityFirst	US start-up whose application lives on top of IBM Quality Stage.	http://www.dataqualityfirst.com/
Datiris	Colorado vendor of data profiling technology.	www.datiris.com
Datras	Munich-based vendor with wide ranging data quality functionality.	www.datras.de
DQ Global	UK data quality and address verification software.	www.dgglobal.com
Exeros	California-based vendor specialising in data discovery.	http://www.exeros.com/
HelpIT	UK vendor of data cleansing technology.	http://helpit.com/
Human Inference	Dutch data quality vendor.	http://www.humaninference.com/
IBM	Data quality software from the industry giant.	www.ibm.com
Informatica	California based vendor, a major player in data quality.	www.informatica.com
Infogix	Illinois-based vendor	http://www.infogix.com

	specialising in controls and compliance.	
Inquera	Israeli company with innovative approach to product data quality using machine-learning technology based on subject domain experts knowledge	http://www.inquera.com/
Innovative Systems	Long established Pittsburgh-based vendor whose software uses an extensive knowledgebase.	http://www.innovativesystems.com/
Intelligent Search	Identity management company now with a more general data quality capability.	http://www.intelligentsearch.com/
Melissa Data	Data quality US vendor with a focus on the Microsoft software environment.	http://www.melissadata.com/
Netrics	New Jersey vendor of impressively accurate matching software.	www.netrics.com
Pitney Bowes Software	The data quality vendor formerly known as Group 1, part of the Pitney Bowes group.	http://www.g1.com/
Postcode Anywhere	UK vendor of web-based addressing software.	http://www.postcodeanywhere.co.uk/
QAS Experian	UK-based vendor specialising in customer name and address.	Qas.co.uk
SAP	The major software vendor has a significant data quality presence.	www.sap.com
Silver Creek Systems	Colorado-based vendor of product data mastering software.	http://www.silvercreeksystems.com/
Talend	Paris-based open source data quality software vendor.	www.talend.com
Trillium	Part of Harte Hanks, one of the leading data quality vendors.	http://www.trilliumsoftware.com
Uniserv	Large German data quality vendor.	http://www.uniserv.com/
X88	Recent UK market entrant specialising in data profiling.	http://www.x88software.com/

Other vendors of data quality software include:

Ciant	www.ciant.com
Data Lever	www.datalever.com
Data Mentors	www.datamentors.com
Infosolve	www.infosolvetechnology.com
Intervera	www.intervera.com
Irion	www.irion.com (an Italian vendor)
Ixsight	http://www.ixsight.com
MSI	www.msi.com.au
Stalworth	www.stalworth.com
TIQ Solutions	www.tiq-solutions.com
Winpure	www.winpure.com
Wizsoft	www.wizsoft.com

Research Methodology

The Information Difference Landscape diagram shows three dimensions of a vendor:

- Market strength
- Technology
- Customer base.




“Market strength” is made up of a weighted set of five factors: revenues, growth, financial strength, geographic scope and partner network. Each of these individual elements is scored, the total producing the “market strength” figure. Similarly “technology” is made up of four factors: “technology breadth” (the coverage of the vendors in various data quality areas as illustrated below), the longevity of the software in the market, analyst perception of the product via briefings, and customer feedback from reference customers, which we surveyed. In each case the scoring is on a scale of 0 (worst) to 6 (best). The highest customer feedback scores in this research were for Melissa Data, Datactics and Omikron.

Vendors were asked to submit answers to various questions via a questionnaire. Those which were larger than a particular size threshold were then interviewed directly by an analyst and reference customers were surveyed to give their experience of the software and vendors.

The technology functions which the vendors were asked about are as shown below. These are drawn from the Information Difference vendor functionality model; if you are interested in more detail on this then please contact The Information Difference.

Functional Areas

Data Quality Functionality Areas

Data profiling			Cleansing rule configuration
data type & format validation			Merging & survivorship
Data dependency validation			Data enrichment
Matching algorithm tunability			Language standardisation
Fuzzy matching			External source integration
Matching accuracy			Data type support
Match rule configuration			Data quality monitoring
Parsing			Industry-specific templates
Context-sensitive data cleansing			