DQ Landscape

The Data Quality Landscape - Q1 2013

The data quality market for the calendar year 2012 was worth around $994 million, of which software sales and maintenance accounted for around $825 million. The overall figure includes the professional services arms of data quality vendors, but excludes the (substantial) revenues of systems integrators and consultancies involved with data quality initiatives. This represents 5% growth over 2011, reflecting a slow recovery in the economy.

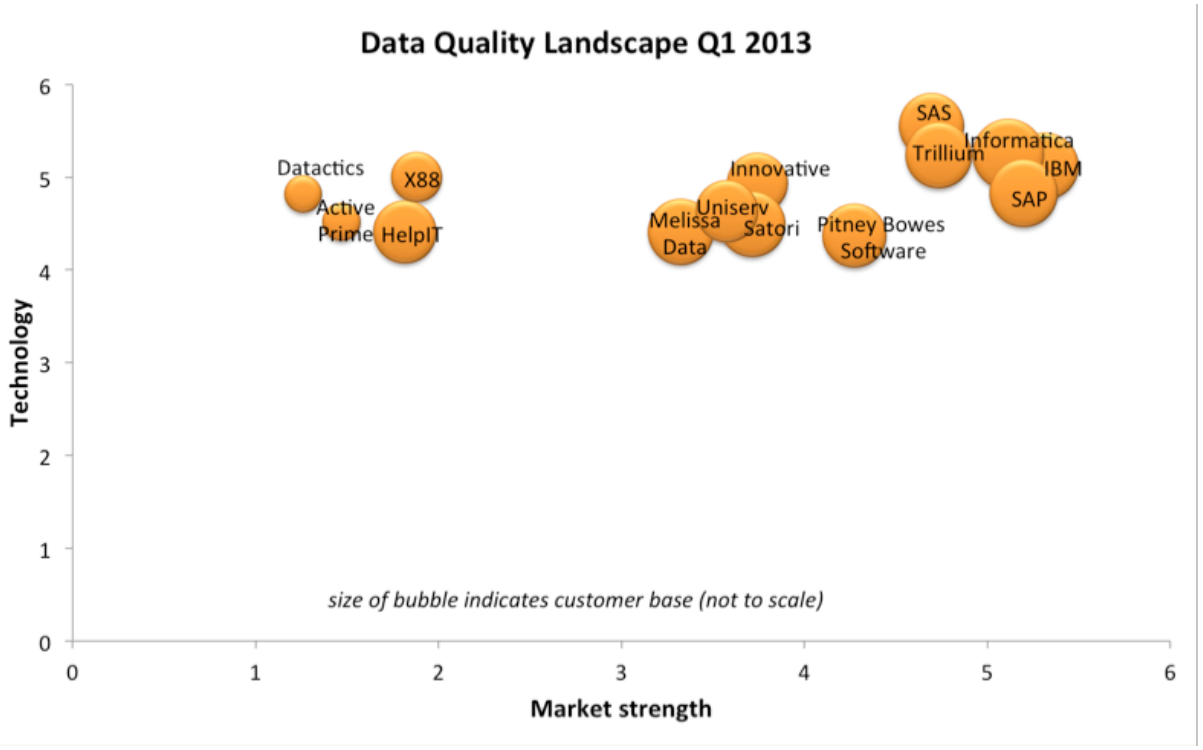
Data quality is a persistent problem in any large organization. At any point where human being are asked to type data into a computer system errors will occur, no matter how carefully the system is designed. Even data that was once correct can become out of date, and copied elsewhere into other systems. Large organizations store data in more than one system, increasing the scope for problems with the data. This duplication means it is easy to accidentally create a new account or product code when in fact one already exists. We are all familiar with errors in the spelling of our names on communications, but data quality can have much more serious consequences than mild irritation. Errors in data quality can (and have) caused contracts to be issued with the wrong specifications, orders to be delivered to the wrong address in the wrong quantities. More dramatically, poor data quality can caused drilling to be done in the wrong place, with hazardous and expensive consequences, or even cause a space mission to fail, as in the case of the Mars Orbiter.

The software industry has developed numerous software responses to this issue, from simple ones like validating data entry better, through to developing sophisticated algorithms to attempt to detect possible duplicate entries in systems. Software can be used to detect common misspellings and to realise that Bob, Rob and Robert may refer to the same person, despite not being identical. The data quality industry grew up around name and address validation, which even today can be a tricky thing, with many subtle variations by country, as anyone who has ever tried to find an address in Tokyo can verify. Many vendors can now add a great deal of added value to a simple address: showing not just the location on a map, but perhaps bringing in 3rd party data to show the number of employees at a particular office, or the revenue of a company based there. The rise of mobile devices has encouraged the development of more location awareness of addresses, for example showing all branches of a particular store in the vicinity of a mobile user.

Data quality software has become more sophisticated over the years, and can now be applied to other types of data, such as product codes, themselves a tricky and complex beast to tame. Over the last few years data quality suites have emerged that are quite broad in scope, from data profiling and discovery, through to merge/matching, data correction, workflow and reporting. As data governance has grown as a discipline the industry has responded, with several vendors now offering explicit support for data stewards. Recently a couple of vendors have even enabled features allowing business users to put a price on data quality, showing the monetary value at risk during a business process, and so potentially identifying the hard savings from data quality improvements, which are frequently elusive. Validating whether a person is really who they claim to be has become a fruitful area for data quality software, from simple email validation to trying to detect possible fraud.

As well as increasing in scope, data quality has started to be subsumed within broader data management initiatives. Master data management projects have data quality as a major component; indeed an Information Difference survey showed that master data projects consume on average 30% of their budget on issues related to data quality. Data quality is increasingly being regarded as just one facet of data governance, with ever more links between these areas. One emerging area is that of in-database data quality, since for very large datasets traditional data quality performance becomes challenging. Indeed the rise in interest in “Big Data” has been dramatic recently, and in a December 2012 survey The Information Difference found that 80% of survey respondents felt that data quality was of “key importance” to big data initiatives.

The diagram that follows shows the major data quality vendors, displayed on three dimensions. See later for definitions of these.



It is important to understand that this is a high-level representation of the market, with vendors represented on the chart specialising in different areas and at very

different price-points (HelpIT is an example of a vendor with a quite complete data quality toolset at a low price-point). If you are considering data quality software, it is important to tailor your selection process to the particular needs that you have rather than relying on high-level diagrams such as this. The Information Difference has various detailed models that can assist you in vendor selection and evaluation.

As part of the landscape process, each vendor was asked to provide at least ten reference customers (some provided over 50 references) which were surveyed to determine their satisfaction with the data quality software of the vendor. The happiest customers based on this survey were those of X88, Active Prime and Trillium, followed closely by the customers of Satori, Datactics, SAS, IBM and Informatica.

Main Vendors

Below is a list of the main data quality vendors.

Vendor	Brief Description	Website
Address Doctor	Vendor that specialises in providing wide coverage of name and address information; now owned by Informatica.	www.addressdoctor.com
Ataccama	Prague-based company with a modern data quality suite.	www.ataccama.czw
Active Prime	California-based vendor of data quality for CRM systems.	www.activeprime.com
Business Data Quality	UK-based data profiling vendor.	www.businessdataquality.com
Capscan	London-based provider of address management and data integrity services, now owned by GB Group.	www.capscan.com
Datactics	UK-based vendor specialising in product data quality.	www.datactics.com
DataQualityFirst	US start-up whose application lives on top of IBM Quality Stage.	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.dqglobal.com
Experian QAS	UK-based vendor specialising in customer name and address validation.	www.qas.co.uk
Google	The search engine giant now does data quality.	code.google.com/p/google-refine/
HelpIT	US/UK vendor of batch and real-time data quality solutions including address validation.	www.helpit.com
Human Inference	Dutch data quality vendor.	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 specialising in controls and compliance.	www.infogix.com
Infoglide	US vendor specialising in identity resolution.	www.infoglide.com
Infoshare	UK data quality specialising	www.infoshare.ltd.uk

	in the public sector market.	
Inquera	Israeli company with innovative approach to product data quality using machine-learning technology based on subject domain experts' knowledge.	www.inquera.com
Innovative Systems	Long established data management vendor that has leveraged their knowledge-based data quality solutions into Data Discovery, MDM, and Data Governance offerings.	www.innovativesystems.com
Intelligent Search	Identity management company now with a more general data quality capability.	www.intelligentsearch.com
Melissa Data	US/German global data quality vendor offering address verification, geocoding and matching solutions.	www.melissadata.com
Microsoft	DQS is the data quality offering of the Redmond software behemoth.	www.microsoft.com
Netrics	New Jersey vendor of matching software. Now owned by Tibco.	www.netrics.com
Oracle	The software giant's data quality offerings are based on the acquisitions of Datanomic and SilverCreek.	www.oracle.com
Pitney Bowes Software	The data quality vendor formerly known as Group 1, part of the Pitney Bowes group.	www.pbinsight.com
Postcode Anywhere	UK vendor of web-based addressing software.	www.postcodeanywhere.co.uk
SAP	The software giant is a major data quality player.	www.sap.com
SAS	One of the leading players in data quality.	www.dataflux.com
Satori Software	Seattle-based provider of address management solutions.	www.satorisoftware.com
Talend	Open source vendor with wide range of quality functions that are tied to data integration and MDM.	www.talend.com
Trillium	Part of Harte Hanks, one of the leading data quality vendors.	www.trilliumsoftware.com
Uniserv	Large German data quality vendor.	www.uniserv.com
X88	UK vendor specialising in data profiling.	www.x88software.com

Other vendors of data quality software include:
AMB DataMiners (www.ambpdm.com)
Ciant (www.ciant.com)
Data Lever (www.datalever.com)

Data Mentors (www.datamentors.com)
Infosolve (www.infosolvetech.com)
Intervera (www.intervera.com)
Irion (www.irion.com)
Ixsight (www.ixsight.com)
MSI (www.msi.com.au)
Rever (www.rever.eu)
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 (this has a high weighting), which we surveyed. In each case the scoring is on a scale of 0 (worst) to 6 (best).

Vendors were asked to submit answers to various questions via a questionnaire. Vendors were interviewed directly by an analyst and their software demonstrated and assessed. Reference customers were surveyed to give their experience of the software of each vendor. 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

Copyright © 2013 The Information Difference Ltd. [Updated: 6 May 2013]