# Data Governance

Market trends

The market for data governance is diffuse. Historically, there have been, and remain, several different sectors. In particular, the governance of data has typically been split between structured data—whose governance has typically been provided by data quality vendors—and spreadsheet and unstructured data governance, with both of these latter categories, especially the former, attracting its own corpus of suppliers. In addition, there are vendors, particularly in the structured data governance space, that have developed capabilities specifically to address the concerns of governance and stewardship. These are typically used to complement data quality suites that do not have the requisite capabilities when it comes to governance.

This disparity of offerings reflects the stances of both the IT industry and of the organisations that have or are implementing data governance. On the one hand companies tend to view governance on a piecemeal and siloed basis while, on the other, vendors do not have anything in the way of an integrated solution to offer. Thus, while governance should be viewed holistically, it is not.

The first of the actual trends we see in the market is that governance is gradually being more widely adopted, though not as fast as we would like. This is, in part, driven by new regulations, especially in financial markets, that mandate "the accuracy, timeliness and appropriateness" of data. This couldn't go much closer to requiring data governance without explicitly saying so. What is most interesting about this phraseology—which is from the Solvency II requirements—is that, along with Basel III, MiFID II and, to a certain extent, Dodd-Franks, these are the first regulations to focus on data quality. Previously, compliance was typically process-focused: you had to prove the lineage of data, for example, but not its accuracy. It is our belief that we will see more regulations that will require good data governance.

The second trend is that companies have been and are implementing governance over unstructured data. Indeed, there are probably more implementations of the single unstructured governance product that we are covering in this update than there are of all the structured data governance products put together. On the other hand, these have been, as we have stated, largely siloed implementations that divorce unstructured and structured implementations. We should also note that while unstructured products are capable of providing some level of governance of spreadsheets they do not provide the depth of capabilities that spreadsheet governance products do.
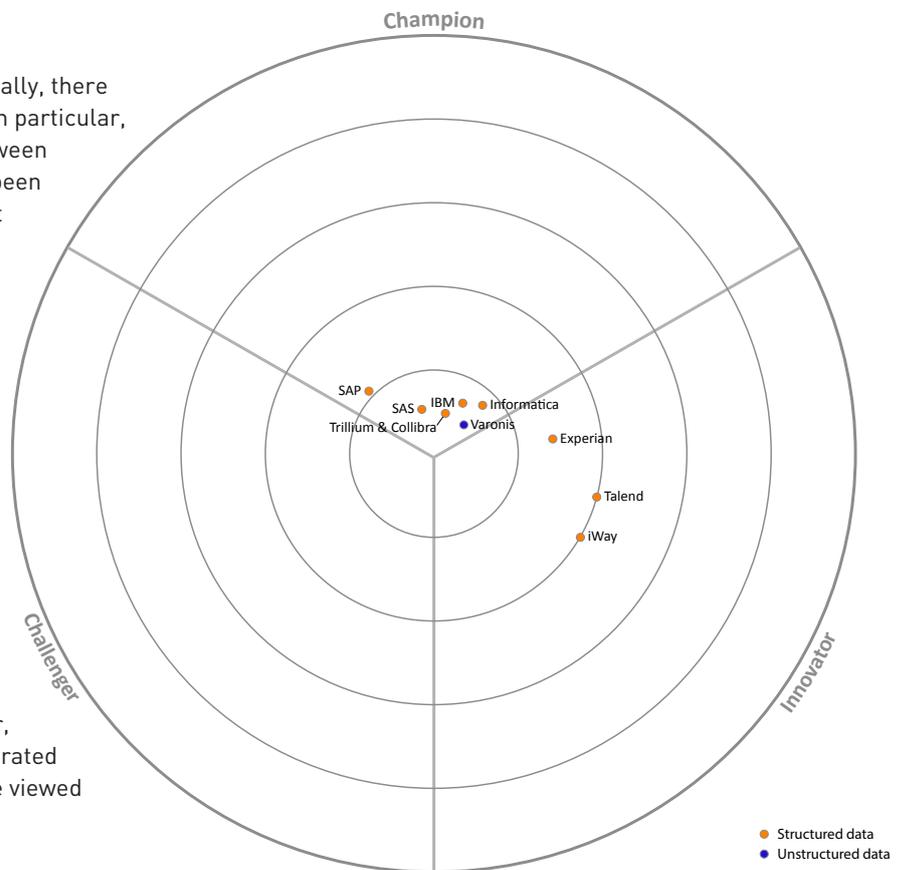
Figure 1: The highest scoring companies are nearest the centre. The analyst then defines a benchmark score for a domain leading company from their overall ratings and all those above that are in the champions segment. Those that remain are placed in the Innovator segment if their innovation rating is over 2.5 and Challenger if it is less than 2.5. The exact position in each segment is calculated based on their combined innovation and overall score.

One of the features of spreadsheet governance products is that they allow you to identify the importance of individual spreadsheets by, amongst other factors, the values embedded in the data. We are starting to see similar features appear within data quality/governance products that allow you to prioritise governance initiatives and remediation by value. We expect this trend to continue.

Finally there is, of course, Hadoop and all things NoSQL and big data: and, for that matter, the cloud. All the signs are that a significant proportion of corporate data is going to end up being stored in one or other of these environments, or in some sort of combination. Governance therefore needs to encompass these infrastructure options. Moreover, if you are going to archive, for example, both structured and unstructured data onto a Hadoop platform, then wouldn't it make sense to have the same management console for all of your data? We think so. We therefore hope that the trend towards NoSQL and cloud adoption will drive closer integration between governance vendors in different areas.
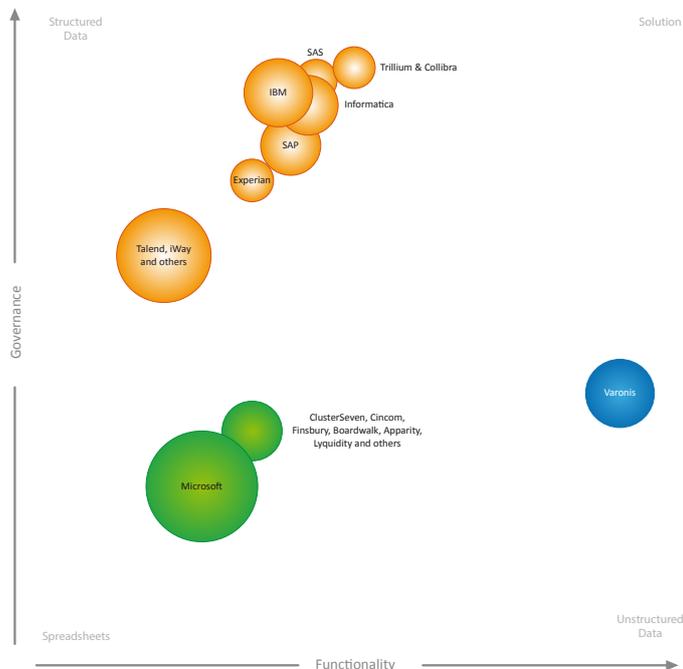
Figure 2: Market Map. The different coloured segments represent the different sectors of the market. No vendors appear in the solution segment since there is no product that caters to the governance of all types of data. Within their quadrants the further to the right the vendors are the richer their data quality functionality and the higher they are the more complete is the data governance aspect of their solution. The larger the circle for each vendor the larger that company's share of this market.

The vendors

The most notable change in the market has been the partnership between Trillium and Collibra. The latter is a specialist in providing data governance capabilities but is not a data quality vendor. In theory, Collibra could be used in conjunction with any of the other (structured) products that are available in the market but its partnership with Trillium, along with the tight integration between the two products, means that these two are likely to be preferred in combination and we have duly represented them in Figure 2 in conjunction.

Apart from this partnership it tends to be the established vendors (IBM, Informatica, SAS et al) that offer advanced data governance capabilities while stand-alone data quality suppliers are somewhat behind in this area. One company that we can imagine moving from the latter category to the former is Experian, which is now making a serious play for the data quality market and we can easily imagine the company expanding more forcibly into the data governance arena in due course.

In the unstructured space we have only included one vendor—Varonis, the clear market leader—so there is little to comment upon. However, for spreadsheet governance we had anticipated that Microsoft would wipe the floor with its competitors after the acquisition of Prodiance: that it would build on the capabilities it acquired to the extent that the likes of ClusterSeven and Finsbury Solutions would no longer be able to compete. That has turned out not to be the case. Indeed, rather than building on the features of Prodiance, Microsoft has deprecated some of the capabilities it inherited, so there is still a clear requirement for spreadsheet governance independent of Microsoft.

We have omitted Oracle from Figures 1 and 2 because a) the company did not respond to our requests for information and b) although the company offers its own data quality capabilities it also continues to resell Trillium Software products. It is therefore unclear as to the company's approach to this market. We can say that Oracle in its own right has particular capabilities with respect to product data cleansing. We have also omitted Microsoft DQS (Data Quality Services) for the primary reason that it is focused on SQL Server and is not a general-purpose tool. As DQS comes with SQL Server it is only likely that readers will be interested in other tools if they have found DQS insufficient for their purposes. We have also left out Global IDs, at the company's request, even though we know from experience that its product portfolio has significant capabilities in this area. The Bullseye diagram (Figure 1) ignores spreadsheet governance.

Summary

The governance of data remains disparate and we do not believe that this will change anytime soon. The advent of big data may cause this to change, at least as far as structured and unstructured (but not spreadsheets) are concerned. That said, few of the data quality vendors, with the exception of IBM (and to a lesser extent Informatica), are talking much about the importance of governance for big data so this is an area that has yet to fully develop.

Philip Howard
Research Director - Data
December 2013