

Data quality platforms

Introduction

This is the last of four Market Updates on data discovery, data profiling, data cleansing, and data quality platforms respectively. The question, of course, is what constitutes a data quality platform? At the lowest level this must include both a data profiling tool and a data cleansing/matching tool. We have regarded it as advantageous if the platform can also provide specific capabilities to support data stewards and data governance and, to a lesser extent, if there are particular facilities available to support data migration. We have taken only marginal note of the fact that other capabilities such as master data management and ETL [extract, transfer and load] may also be available.

Key market issues

The key issue, of course, is whether you should opt for a platform at all and, if so, how broad it should be? If this were a Market Update on data integration platforms we would necessarily have considered ETL along with data quality, but it is not. However, there remains the question of when and if it is better to source data profiling and data cleansing/matching from the same vendor?

The answer is that it depends. As our Market Update on data profiling tools discussed, there are a lot of things that you can do with a pure play data profiling tool that you will not usually be able to do with a product that has been tightly integrated into a data quality platform. This is because the suppliers of the latter have data quality as a focus rather than the broader capabilities of data profiling for supporting MDM, data migration, data archival, data modelling and so on. If you want to use your data profiling tool to support these sorts of capabilities as well as data quality then it may well make sense not to opt for a platform-based approach but to look for best-of-breed vendors as otherwise you will either end up with an inferior tool for discovering data relationships across data sources or you will need to license two different data profiling products.

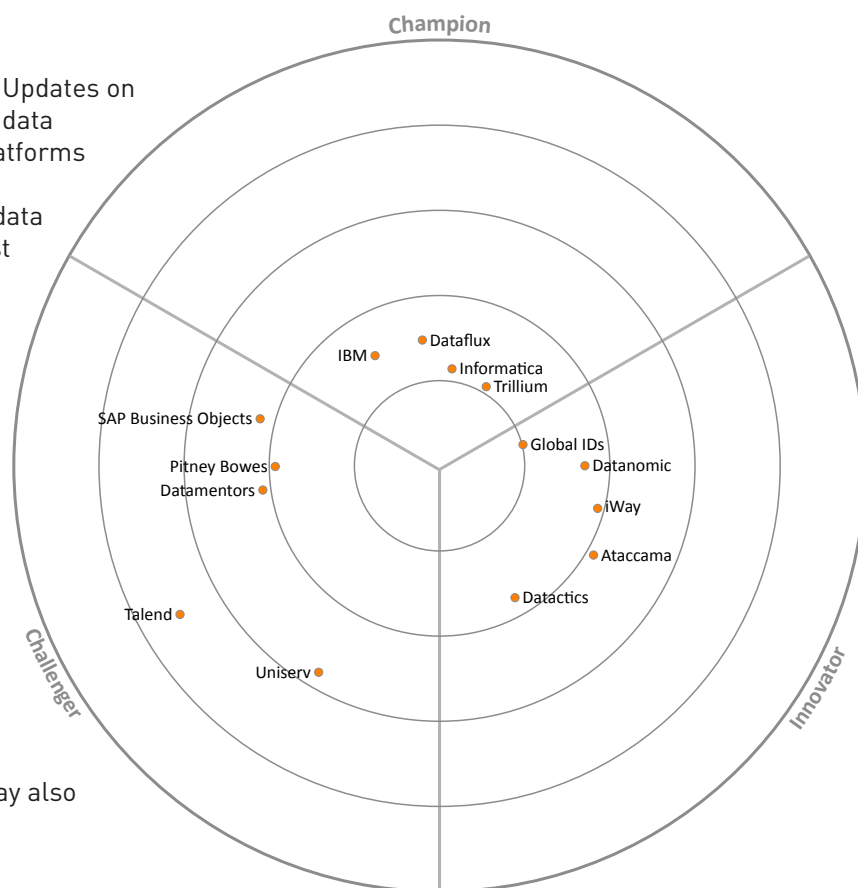


Figure 1: The highest scoring companies are nearest the centre. The analyst then defines a benchmark score for a domain leading company from their overall ratings and all those above that are in the champions segment. Those that remain are placed in the Innovator segment if their innovation rating is 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.

The downside of a best-of-breed approach is that the two products may not integrate well. Certainly, they are unlikely to share a metadata repository, which is a must-have feature in a platform. In addition, you would prefer not to have to extract the data once for profiling purposes and then have to do it again for cleansing, which is likely to be the case if the products do not tightly integrate. This would more or less preclude an iterative approach whereby you profile, cleanse, profile, match, do some profiling, then some cleansing and so on.

Note that care must be taken when considering platform-based products as some vendors resell data profiling products from other vendors, notably BDQ (Business Data Quality), which may or may not have tightly integrated with their own data cleansing/match products.

Vendor landscape

As may be imagined there are fewer vendors in this report than in the preceding ones, primarily because of the number of vendors who specialise in only one part of the market. We also have three

notable omissions that declined to participate in one or more parts of this series and therefore could not be included here: Microsoft, Oracle and Pervasive. As Oracle resells software from several other vendors (including Informatica and Trillium) this is no great loss, while the other two are not involved because of conflicts over product release cycles.

To a significant extent, choices about potential vendors may be dictated by geography: Uniserv is German and has only recently started to expand out from its heartland, Datanomic and Datactics are both British and, although they have overseas accounts, they are in similar position as is Pitney Bowes, except that it is American, while Global IDs is wholly confined to the US at present. Ataccama is Czech but is in a slightly different position as the product is resold by iWay, which has embedded the product into its real-time EIM offering.

If you are an international company you might therefore assume that you would be better off selecting your platform from one of the major vendors with a world-wide presence. However, the foregoing should not necessarily put you off: all the vendors discussed are very alive to the fact that they need to support Unicode, international addresses and nationalisation of interfaces, so the fact that a vendor is not itself global does not mean that it cannot support the global operations of your company.

A further factor is that while we have not paid much attention to the broader scope of a data quality platform that doesn't mean that it won't be relevant to you and there is a clear spread of capabilities across the various vendors. Thus Uniserv, Datamentors, Datanomic, Datactics, Trillium and Ataccama simply provide data quality and associated features but all the other vendors also provide data integration capabilities. Beyond that there are some vendors, notably IBM, Oracle, Dataflux and, to a certain extent, SAP Business Objects that provide MDM.

Summary and conclusions

Sticking strictly to our brief of limiting our considerations to combined data profiling and data cleansing/matching products, possibly with data governance as well, then we view Global IDs, Trillium and Informatica as the leading products in this category. While we particularly like Global IDs' product suite (which is extensive) the company is, as we have mentioned, geographically limited and a relatively small company to boot, so users may prefer a solution from Informatica or Trillium for general purpose use. In addition, for those of you who have particular concerns over product data quality (for which we refer you back to the third of these Market Updates) then Dataflux and Datactics are certainly worth consideration. For ease of use, Datanomic is probably the market leader.

Philip Howard
Data Management
May 2009



2nd Floor
145–157 St John Street
London, EC1V 4PY
United Kingdom

Tel: +44 (0)207 043 9750
Fax: +44 (0)207 043 9748

Web: www.BloorResearch.com
email: info@BloorResearch.com