Comparative costs and uses for Data Integration Platforms in Agile Enterprises
Data Integration should be at the heart of data architecture modernisation initiatives such as next generation analytics, application modernisation, total customer relationship, data governance and so on.

Author Philip Howard
In 2014 Bloor Research surveyed users of data integration platforms to discover how they were using that technology and what their three-year cost of ownership was. Our analysis, based on nearly 300 completed questionnaires, concluded that Informatica’s data integration platform generally offered the best cost of ownership figures and, on average, it was the most widely deployable across a range of use cases, providing extensive reuse of skills, tools, and code.

In this paper we are revisiting those results, not because we have any reason to believe that the responses we received in 2014 are any less valid today, but because we think that there has been a shift in the market. In our survey we identified a number of use cases, which organisations were using data integration platforms to support. However, over the last couple of years we have seen a shift towards a broader focus on data-driven enterprises, and we see data integration as a fundamental part of that. Data integration needs to be seen not merely as a tactical tool, which supports a single use case or implementation, but as a strategic foundation for today’s enterprise.

Data Integration should be at the heart of data architecture modernisation initiatives such as next generation analytics, application modernisation, total customer relationship, data governance and so on. As such, organisations should make sure that their data integration platform inherently provides the ability, scale and performance – not to mention flexibility and agility - to support multiple use cases across the organisation.

1.1 Supporting the agile enterprise

Enterprise agility (Bloor Research uses the term the “mutable enterprise”) is a far bigger topic than we have space to discuss in this paper. Briefly, an agile enterprise is one that is sufficiently adaptable and flexible so that it can embrace change as and when needed. This applies at all levels: people, processes and infrastructure. The fundamental driver for such change is information (data) because without information about its operations and performance (amongst other things) no organisation can know whether and how it should adapt to changing circumstances. Thus, the infrastructure that supports this data must incorporate some level of strategic capability to enable change, transformation and agility.

In practical terms, IT has a twofold mandate. On the one hand it needs to keep the lights on but on the other it needs to support the evolving data landscape in which companies now operate. With respect to the latter, this means leveraging far more sources of data than was previously the case (big data), in more locations and based on a wide range of different physical deployment architectures. From a data integration perspective, this means supporting a much greater range of capabilities than was previously the case. It is a significantly greater challenge to maintain timely, relevant, connected and high quality data in a big data world than it was previously.

On the other hand, the need to keep the lights on puts pressure on budgets, which in turn means that companies are focusing on application consolidation and rationalisation, thereby reducing costs in these areas in order to enable investment in the infrastructure that supports the enterprise agility. In particular, and in the context of this paper, this means data architecture modernisation. It is with these considerations in mind that we have extended our 2014 report with additional comments with respect to agility in the enterprise.

1.2 Augmenting our previous data

In order to gather the appropriate information to reach our conclusions, Bloor Research conducted a survey which produced 292 responses from companies using data integration tools (or, in some cases, using hand coding) for a variety of purposes. While there was a significant emphasis on cost this is, and should be, only one of several determining factors when selecting data integration solutions. Nevertheless, the highlight of our 2014 results was probably the following chart (Figure 1). This shows total cost of...
We believe that our original results are still valid with respect to both cost of ownership and individual use cases. We have, however, augmented our original results with some survey data obtained by TechValidate in late 2015.

In this update to our original paper we have not re-surveyed users. We believe that our original results are still valid with respect to both cost of ownership and individual use cases. We have, however, augmented our original results with some survey data obtained by TechValidate in late 2015. This company was asked to survey Informatica’s user base and we are grateful to Informatica for providing this information for us to reuse here. However, we should add the caveat that the inclusion of these details does not represent an endorsement of Informatica or its technology: the information provided is by way of being an exemplar for the sorts of facilities you should look for when considering data integration from a strategic perspective.

In terms of format, Section 2 discusses the sort of features that a data integration platform needs to include if it is to support data architecture modernisation or, more broadly, as a technology enabling the Mutable Enterprise. Where relevant, we have included details from our 2014 survey in this section. Other such details follow on in Section 3 where we drill down into the figures more deeply. For the sake of brevity, we have edited this in order to provide only the most salient details. The full text of the original 2014 report is available here.

**Microsoft figures**

For a number of reasons, we believe that the figures attributed to Microsoft in this report do not represent a true apples-to-apples comparison with the other products covered here. There are two main reasons for this. Firstly, Microsoft environments tend to be largely homogeneous, so that the number of sources and targets involved in projects does not accurately reflect the complexity of those projects. Secondly, Microsoft’s data integration capabilities are bundled into SQL Server. What this means is that respondents to our survey ascribed zero licence fees and zero maintenance fees to the use of Microsoft’s data integration capabilities. For obvious reasons this significantly skews the resulting calculations.
2. Data integration as an enabler for agile enterprises

There are a variety of features that a data integration platform needs to offer if it is going to be an enabler for change and modernisation. We will discuss these momentarily (in no particular order) but it will be useful to start with a business oriented question that TechValidate asked. As Figure 2 suggests, data integration isn’t simply about resolving technical issues. It can be about enabling business change. Note that although these responses apply specifically to Informatica PowerCenter we would expect similar responses from any comparably capable products.

2.1 Flexibility

The first, and perhaps most obvious characteristic of the sort of data integration platform we are looking for, is that it can support multiple use cases. This was an issue we addressed in 2014. We outlined six common scenarios in our survey, for which data integration tools might be used. In our experience, these scenarios, or use cases, represent the vast majority of project types for which integration products are deployed. The scenarios are:

1. Data migration and consolidation projects
2. MDM and associated solutions
3. Application to application integration
4. Data warehousing and business intelligence implementations
5. Synchronisation of data with SaaS applications

Figure 3, with data from our 2014 survey, shows how relatively important these different integration tasks are with, not surprisingly, more than 50% of integration work being focused on data warehousing and/or data migration while MDM and SaaS integration add up to less than 15% of the market between them.

As we shall discuss shortly, we are today seeing a number of additional use cases. However, as far as our 2014 figures are concerned we also asked respondents to identify the one, or more, integration products that were used in their organisation for these scenarios and to identify the single product with which they had the most experience. For that selected product, respondents recorded their view of how suitable they thought their tool was for each of the above scenarios. It was not a requirement that the chosen product or vendor was actually being used for each scenario, simply that respondents believed that the products were suitable.
Note that we did not receive sufficient responses from users of either open source tools (such as Talend or Pentaho) or other vendor products (SAS, Ab Initio and so forth) to include these in our analysis. **Figure 4** reflects this measure of suitability. In other words, more than half of Informatica users thought the product was suitable for use in almost all the use cases discussed above. Conversely, the uses of Actian and SAP were considered suitable for less than half of our use cases.

There are a number of new data integration use cases that are coming to the fore. This is reflected in the TechValidate data shown in **Figure 5**. For example, 26% of PowerCenter users are using it for real-time use cases and 6% are already using it for cloud/hybrid purposes. More generally, 88% of PowerCenter users deploy it for data warehousing and analytics, while figures for “data warehouse modernisation, migration or consolidation” and “application migration and consolidation” have figures of 44% and 31% respectively. Application to application is also significant at 48%.

In addition to asking users what they thought about the suitability of particular tools, we also examined the number of projects for which users were using or planning to use their data integration platform, by use case. The results are shown in **Figure 6**.

Informatica is the most suitable product overall and it is also the leading vendor in two categories: B2B and data warehousing. Oracle leads in application integration and SaaS integration, SAP in data migration and IBM leads for MDM. Of course, IBM, Oracle and SAP (not to mention Microsoft) all have a degree of lock-in with their client base so, to a certain extent, these figures probably flatter those vendors if we are considering general-purpose capabilities. This makes Informatica’s leading position even more impressive.

### 2.2 Connectivity

A major feature of today’s IT landscape is the proliferation of databases ([www.NosQldatabase.org](http://www.NosQldatabase.org) lists 225 NoSQL databases) and (cloud-based) applications, along with a variety of new data formats, such as JSON, that need to be supported. Therefore, connectivity is an increasingly important topic. This may be provided out-of-the-box in some instances but the sheer number involved means that you will also need development tools for building connectors and community sites where users can share connectors. This was not a subject that we addressed in our 2014 survey but, as noted, it is becoming increasingly significant. According to TechValidate (**Figure 7**), a significant majority of users recognise the importance of such capabilities. In addition, it is worth remembering that connectivity is no longer limited to relational sources and targets but also to many other environments that support unstructured and semi-structured as well as structured data.

It is worth noting that this increased connectivity requirement has led to a greater focus on the architecture that underpins data integration products. When the number of sources and targets that might need to be supported is small, then it doesn’t much matter how connectivity is implemented. However, when that number expands into, potentially, hundreds of endpoints, then tools have to be suitably agile to allow the development and implementation of connectors without impacting on integration logic. In practice, this will typically mean that data integration providers will need integration logic and connectivity to be separate tiers within the product architecture.
2.3 Productivity and agility
While we did not ask about connectors per se, we did ask survey respondents about the productivity of their chosen data integration platform on a per source/target basis, which is clearly related to connectivity. The results from this part of our survey show how long development takes, dependent on the number of end-points involved in the project. The results are presented in Figure 8.

As can be seen, Informatica and Microsoft, followed by IBM, appear to be significantly more efficient in this regard than their competitors. It is also notable that hand coding is the least efficient.

A second measure, related to the first, is presented in Figure 9, which shows the average project length, weighted by complexity (the number of end-points) for each type of integration task identified above.

Here we can see, for example, that hand coding is clearly not good for data warehousing while you wouldn’t recommend Oracle for B2B integration. Perhaps, more importantly, Informatica and Microsoft offer consistent performance across all integration scenarios, whereas other vendors and approaches are clearly more suitable for some tasks than others.

The results from the TechValidate survey confirm this result with 70% of respondents claiming improved productivity through "ease of use, automation, shorter development time, code reuse, skills reuse, metadata-driven visual tools and out-of-the-box templates and mappings." Another major factor not mentioned in this quote is collaboration and, in particular, collaboration between IT and the business. This is enabled by such factors as a business glossary (which enables a common understanding of terminology), workflow, role-based tools that are specific to certain classes of users (for example, analysts), and prototyping that enables an iterative approach to development. Facilities such as impact analysis (enabled through metadata management) are also useful in the sense that they reduce risk. While business/IT collaboration was not considered in our 2014 survey, nor by TechValidate, it has been addressed in previous research conducted by Bloor Research. For example, in our most recent research on data migration we found that lack of collaboration between the business and IT was cited as the number one reason for projects running over time and/or over budget.
2.4 Testing

Testing is a part of the development process in integration environments just as it is for conventional applications. In general, far too much testing is manual, resulting in wasted time and expense. However, it is rarely addressed within data integration environments even though it represents a major part of the cost of development. Features such as test data management and data masking (see next) will be welcome, as will conventional techniques such as unit and regression testing. With specific respect to data integration, it is also useful to be able to offer “before and after” testing that compares data before and after it is either moved or transformed in order to check that no data integrity errors have arisen.

TechValidate has addressed the use of such comparative tools which, in Informatica’s case, is called Data Validation, which also provides conventional testing techniques. Figure 10 highlights the reasons behind investing in Data Validation. It is interesting to note the number of respondents citing data-driven business decisions and improved business trust in the data, as these are fundamental characteristics of the agile enterprise. Similar results can be seen in Figure 11 with respect to monitoring of operational performance. It is important to recognise that this is not just a technical consideration but also about confidence (see next section) that your data is really up-to-date and complete. Given that some surveys suggest that as many as one in three business executives report that they do not trust the data that they use to make decisions, then supporting this trust is very important. In addition, proactive monitoring is also used as a part of the testing process to enable automated code reviews and helps to reinforce development best practices.

2.5 Trust

The foundation of the agile enterprise is the data, and the analysis thereof; that is used by the organisation to guide its path into the future. However, these analyses need to be trusted by the executives that rely on them for information and this, in turn, means that the data itself must be trusted. Recent research suggests that as many as one third of business leaders do not trust the information upon which they have to make decisions.
When it comes to data, trust is not an island unto itself: it specifically incorporates both the governance and security of data, as well as ensuring compliance accuracy and timeliness.

With respect to compliance in banking and financial services, data is expected to be accurate, consistent and complete and this represents best practice in other industries also. Detailed auditing and data lineage (visibility into where the data has come from and how it has been derived) is de rigueur. Moreover, it helps to build trust in the data when lineage is provided by built-in reports and dashboards. In addition, the use of a business glossary – discussed previously – can promote confidence in the data as well as collaboration. Timeliness of the data (supported by Proactive Monitoring) as well as performance and availability is another important consideration for building trust, as is the ability to meet service level agreements.

On the other hand, sensitive data should not be visible to those not authorised to view that data. While integration with data quality tools is a well-known characteristic of leading data integration platforms, the implications of protecting sensitive data may not be immediately obvious. In the context of data integration, developers of integration processes should not be able to see personally identifiable information but it will be necessary for them to test their processes against data that at least looks real. It will, therefore, be advantageous if data masking is supported by the integration platform. Note that this also applies to developers (data scientists) of analytics who are using, for example, a data preparation platform. Although not covered by our survey, Bloor Research published a Market Update on data masking in 2015.

We should further comment that data governance is evolving to support agility in the enterprise. Historically, this discipline was all too focused on the quality of data for its own sake rather than on what was important for the organisation. We are pleased to see that vendors of data governance solutions are moving towards a more business-driven approach. As with data masking, this was not addressed in our survey and there is a Market Update available on this subject. It is notable that in the Champions section for both data governance and data masking were IBM and Informatica.

### 2.6 Other factors

There are a number of other factors that we believe are required for a data integration platform to truly support an adaptable and agile enterprise. Some of these have been alluded to, or discussed, within other sections of this paper. However, we believe these deserve to be specifically referenced. In particular, we would call out collaboration on the one hand and an integrated platform (with data quality, data governance, data masking and so on) on the other. In addition, the one factor we have not mentioned, or discussed, is scalability – with multiple projects, increasing numbers of sources and targets, more diverse data, and huge amounts of it – this is hardly surprising.

When it comes to data, trust is not an island unto itself: it specifically incorporates both the governance and security of data, as well as ensuring compliance accuracy and timeliness.
In this section we summarise the results from our original 2014 report, which can be broken down into five major areas.

3.1 Support for multiple integration scenarios
In our 2014 report, we broke down the figures previously provided (see Figures 4 and 5) into suitability by use case. However, in the current context – that of a strategic platform to enable corporate agility – this seems less relevant. The bottom line is that the majority of large enterprises need a data integration platform that will support all of the use cases we have identified. This should maximise your return on investment.

3.2 Project timelines and resources
Specifically, we investigated the effort involved – in man weeks, for both internal and external (including consulting) resources – with respect to each of the identified use cases; and with respect to the very first project that was undertaken. On its own the results are limited in their utility: Not surprisingly, the products that tend to be used in less complex environments required less effort than those supporting more sophisticated environments. One point that is worth noting is that Informatica, and to a lesser extent Microsoft, provide much more consistent project lengths across use cases compared to other vendors. One might reasonably conclude, all other things being equal, that inconsistency in resource requirements would mean that relevant products were not particularly well-suited to some use cases; confirming what we already know. We might also deduce that the platforms provided by Informatica and Microsoft support transferrable skills across projects more flexibly than do their rivals. Finally, we compared first project resources to the resources required for subsequent projects. In the latter case, Informatica was the only vendor to have significantly reduced resource requirements for follow-on projects across both internal and external staff. This suggests a superior learning curve.

3.3 Scale and complexity
We noted in the previous paragraph that required resources on their own are not a particularly useful measure because that ignores the complexity of the projects you are fulfilling. Measuring the number of projects for which a particular platform is used, which we also did, suffers from the same problem. Of course, it is difficult to measure complexity, especially in a survey: Complexity involves the number and type of transformations, the volumes of data to be processed, the number of end-points (both sources and targets) and the heterogeneity of those end-points. In practice, we used a number of end-points as a proxy for complexity, though with hindsight, we should also have enquired about their homogeneity and heterogeneity. In practice, we found that hand coded solutions, on this measure, tended to involve very little complexity – not surprisingly – and this was also the case for Actian. Microsoft, on the other hand, scored highly here but we expect that this is largely because Microsoft environments tend to be more homogeneous than those catered to be the likes of IBM, Informatica, Oracle and SAP. We concluded that, apart from Microsoft, Informatica, followed by IBM, were the most suitable for supporting complex environments.

3.4 Maintenance and on-going support
We addressed two questions here; monthly maintenance costs (in minutes) and days taken to resolve critical issues. As far as the former goes, it would be misleading to simply present the raw data as this would take no account of the number of integrations that need to be maintained at any one time. This will be dependent both on the number of projects that are supported by the integration platform and the complexity of those projects. As it turned out, Microsoft is the most productive in this respect, followed by Informatica. If one looked at the raw figures, Actian would appear to be the least expensive in terms of man days per month but that is because it is used on fewer projects with fewer sources and targets. As far as the second question is concerned, the results are illustrated in

“ The bottom line is that the majority of large enterprises need a data integration platform that will support all of the use cases we have identified. This should maximise your return on investment. ”
A Bloor White Paper

Figure 12: Informatica performs nearly a day per issue better that IBM and Oracle. A day may not seem like a long time but when that means that you have no access to your data or, worse, that customers have no access to their data, that can be a very expensive 24 hours. Whilst the latter can mean losing customers, it can also be worse. As an example, consider that on June 20th 2012, customers of the Royal Bank of Scotland, Ulster Bank and National Westminster Bank could not gain access to banking services. As a result, the banking group was fined £42,000,000 by the Financial Conduct Authority.

3.5 Costs
Turning to costs there are two fundamental measures to consider; initial costs and costs over a three year (or longer) period. It is important to appreciate that running costs hugely outweigh initial costs. This is typical for all software products. Of course, it is often more difficult to get authorisation for capital expenditure, as opposed to operational expenditure, but looking solely at the initial cost will result in false economies.

However, if we looked only at annual costs these would make Oracle and Informatica (for example) look expensive and Actian great value for money. But this is without adjusting for either complexity or scale. Simply citing raw costs is not a useful comparative measure: You wouldn’t expect a product that is to be used for one or two projects per year to cost the same as a tool with which you implement twenty or more projects, you wouldn’t expect an offering that was only suitable for one or two uses to cost the same as a platform that supports a much broader range of integration scenarios, and you wouldn’t expect a product that is mostly only used with a handful of end-points (sources and targets) to be comparable in price to one that is commonly used with a dozen or more end-points. To put this more simply, a data integration technology that supports more complex integrations with greater scale is going to be more expensive than one that supports neither of these things. Thus the key metrics are not simply initial costs and costs over a three-year period but those costs weighted by the number of projects to be supported and the complexity of those projects as measured by the number of end-points involved within each project.

To cut a long story short, Figure 13 shows 3-year total cost of ownership, weighted by complexity and number of projects. Microsoft shows the lowest total cost of ownership, partly because of the homogeneity of its end-points (thereby reducing complexity) but also because its software is bundled with SQL Server licensing and therefore not seen as an independent cost. To be clear here, we do not believe that the Microsoft figures presented here provide a fair apples-to-apples comparison. Leaving this aside, perhaps the most notable results in this chart are that:

a) Actian no longer looks great value for money
b) Hand coding is actually more expensive than most vendor solutions, the exceptions being IBM and Oracle
c) Leaving Microsoft aside, Informatica offers the best TCO

These are particularly good arguments for not simply considering up-front costs: they can be misleading.

Figure 12: Days taken to resolve critical issues

Figure 13: Three year TCO per project per source/target ($k)
Organisations adopt integration products for a disparate range of integration scenarios: there is clearly no shortage of data integration challenges before we even begin to consider data integration as a strategic platform that enables corporate evolution. Reusability across multiple scenarios is potentially very valuable in terms of both ROI and an organisation’s agility in the face of change. Our numbers appear to show that Informatica and Oracle are considered by their users to be more reusable – suitable for multiple use cases – than their competitors though Oracle is considerably more expensive. Hand coding and, to a lesser extent Actian, are not perceived to be as flexible as other solutions.

In our view, the figures speak for themselves: Informatica is deemed to be the most suitable product across a wide range of scenarios and Microsoft is the clear winner from a cost perspective though, as discussed, this is not a fair apples-to-apples comparison: Microsoft tools tend to be very Microsoft-centric, so they are widely used by organisations that predominantly use a Microsoft platform but won’t be applicable elsewhere. For non-Microsoft purposes Informatica is clearly the value choice here and is well ahead of both IBM and Oracle, though Actian may be a good choice for smaller environments. Note that hand coding is relatively expensive.

We should add one final note on product functionality for specific tasks. The more frequently a vendor or product is used, in different scenarios, is certainly a reflection of the functionality of the product, but we must make the caveat that we have not attempted to distinguish between products on the basis of their functionality; so one product may be more suitable for supporting complex transformations than another, for example. Similarly, another product may offer better performance or have more real-time capabilities or have features that are not present elsewhere. Thus, cost is, and should be, only one of several determining factors when selecting data integration solutions.

4.1 Enabling the agile enterprise
Several points regarding support for the data aspects of the enterprise agility (or, what Bloor Research calls the mutable enterprise) were made in the preceding conclusions. For example, “reusability”, support for an “organisation’s agility in the face of change”, and flexibility. However, there is more than this that is required; a broad swathe of connectivity options, support for collaboration (both within IT – DevOps – and with the business), governance and security, scalability, and so on. We stated in the previous section that cost is only one of several determining factors in evaluating data integration platforms. We would go further than this: Features and functions and other technical considerations also represent only one element in such evaluations.

Bloor research would argue that possibly more important than either of these is the ability to support business adaptability and agility. It is often the case that the business impact of these initiatives can be an order of magnitude (or more) greater than the cost impact to IT. Businesses today are going through unprecedented change in order to meet the demands of the digital economy and to transition to data-driven enterprises. A requirement to enable this is a data integration platform that has the necessary characteristics to support such a transition, and this means one that supports all of the characteristics highlighted in this report.

FURTHER INFORMATION
Further information is available from www.BloorResearch.com/update/2277
Philip started in the computer industry way back in 1973 and has variously worked as a systems analyst, programmer and salesperson, as well as in marketing and product management, for a variety of companies including GEC Marconi, GPT, Philips Data Systems, Raytheon and NCR.

After a quarter of a century of not being his own boss Philip set up his own company in 1992 and his first client was Bloor Research (then ButlerBloor), with Philip working for the company as an associate analyst. His relationship with Bloor Research has continued since that time and he is now Research Director focused on Data Management.

Data management refers to the management, movement, governance and storage of data and involves diverse technologies that include (but are not limited to) databases and data warehousing, data integration (including ETL, data migration and data federation), data quality, master data management, metadata management and log and event management. Philip also tracks spreadsheet management and complex event processing.

In addition to the numerous reports Philip has written on behalf of Bloor Research, Philip also contributes regularly to IT-Director.com and IT-Analysis.com and was previously editor of both Application Development News and Operating System News on behalf of Cambridge Market Intelligence (CMI). He has also contributed to various magazines and written a number of reports published by companies such as CMI and The Financial Times.

Philip speaks regularly at conferences and other events throughout Europe and North America.

Away from work, Philip’s primary leisure activities are canal boats, skiing, playing Bridge (at which he is a Life Master), dining out and foreign travel.
Bloor overview

Bloor Research is one of Europe’s leading IT research, analysis and consultancy organisations, and in 2014 celebrated its 25th anniversary. We explain how to bring greater Agility to corporate IT systems through the effective governance, management and leverage of Information. We have built a reputation for ‘telling the right story’ with independent, intelligent, well-articulated communications content and publications on all aspects of the ICT industry. We believe the objective of telling the right story is to:

- Describe the technology in context to its business value and the other systems and processes it interacts with.
- Understand how new and innovative technologies fit in with existing ICT investments.
- Look at the whole market and explain all the solutions available and how they can be more effectively evaluated.
- Filter ‘noise’ and make it easier to find the additional information or news that supports both investment and implementation.
- Ensure all our content is available through the most appropriate channels.

Founded in 1989, we have spent 25 years distributing research and analysis to IT user and vendor organisations throughout the world via online subscriptions, tailored research services, events and consultancy projects. We are committed to turning our knowledge into business value for you.
Copyright and disclaimer
This document is copyright © 2016 Bloor. No part of this publication may be reproduced by any method whatsoever without the prior consent of Bloor Research. Due to the nature of this material, numerous hardware and software products have been mentioned by name. In the majority, if not all, of the cases, these product names are claimed as trademarks by the companies that manufacture the products. It is not Bloor Research’s intent to claim these names or trademarks as our own. Likewise, company logos, graphics or screen shots have been reproduced with the consent of the owner and are subject to that owner’s copyright.

Whilst every care has been taken in the preparation of this document to ensure that the information is correct, the publishers cannot accept responsibility for any errors or omissions.