

InDetail Paper by Bloor

Authors **Daniel Howard and Philip Howard**

Publish date **November 2017**

Collibra Data Governance Center



Collibra Data Governance Center is a data governance platform containing a suite of data governance applications, with an emphasis on the automation of policy creation, management and enforcement via the use of workflows. These applications also include a data catalogue data stewardship manager and data privacy, each of which has been built with additional capabilities to enable and support collaboration between users in your organisation.



Executive summary

Data governance, as a term, can have a lot of different meanings. For a great many vendors, data governance capability consists of little more than some basic data quality enforcement – cleansing, matching, standardisation and so on – along with data profiling (again, oriented towards data quality) to go with it. That isn't to say that data quality isn't important: far from it. But there is more to data governance, and suitable capabilities are offered by only a small number of vendors. These suppliers go beyond quality to include sophisticated data discovery and cataloguing tools, data privacy and security, regulatory compliance (for GDPR as well as industry specific regulations) and still, of course, providing or integrating with data quality tools. However, at least in terms of data governance, notice that all of these capabilities centre on enforcement: making sure your data complies with the policies you or your organisation have already put in place. There is thus an additional step to be taken, that of assisting in the creation and maintenance of the policies themselves. The number of vendors that have implemented such a step is limited. What's more, many of these companies have only done so in a limited capacity. For instance, they might only offer policy creation and management in support of a single area of governance, such as analytics. Collibra, with Collibra Data Governance Center, is one of the few vendors that supports policy creation and management across every area of data governance.

Fast Facts

Collibra Data Governance Center is a data governance platform containing a suite of data governance applications, with an emphasis on the automation of policy creation, management and enforcement via the use of workflows. These applications also include a data catalogue, data stewardship manager and data privacy, each of which has been built with additional capabilities to enable and support collaboration between users in your organisation.

Key Findings

In the opinion of Bloor Research, the following represent the key features of the Collibra Data Governance Center:

- From the perspective of policy creation, management and enforcement in the data governance space, Collibra Data Governance Center is the clear market leader. Very few other products in the space have as sophisticated policy management capabilities, and even then, they rarely emphasise it to the same extent as Collibra.
- Unlike some other policy oriented data governance solutions, Collibra allows you to create policies for any area of your organisation. This allows you to have a common governance environment across your entire infrastructure.
- The product suite has many tools that enable or promote collaboration. This includes a full-fledged data catalogue, a business glossary, a ticketing system, and an application for managing data stewards and their relationships, as well as the policy management itself, which places a great deal of focus on making sure that everyone with a stake in the policy has the opportunity to provide input during its creation.
- The data catalogue leverages machine learning, both to make recommendations for data you might be interested in, as well as suggestions for which business terms might apply to a particular dataset.
- Policy management excepted, the products contained in Collibra Data Governance Center, in general, do what they do well, and are handily competitive with other products in their respective areas. However, they can be more limited in scope than competing products: for instance, for data lake management the Collibra Catalog (data cataloguing) does not provide any particular tools to support data preparation.



From the perspective of policy creation, management and enforcement in the data governance space, Collibra Data Governance Center is the clear market leader.





**It is worth
emphasising this:
Collibra is very
much focused on
the business.**



- The included data profiling (assisted by machine learning) provides both an overall score for an asset, as well as sub scores for individual criteria. This provides both a quick overview of an asset's quality and a detailed breakdown of it. This is nice to see, as many data governance offerings provide only one or the other. In fact, it is worth commenting that we particularly like this aspect of the product.

The bottom line

Collibra Data Governance Center aims, first and foremost, to do one thing well, and that is policy management and its automation. Not only does it achieve this in spades, it provides a great deal of other functionality as well, including data cataloguing, data stewardship and data privacy, each of which is capable in its own right, and all focused on supporting collaboration within your organisation. It is worth emphasising this: Collibra is very much focused on the business. All of these things come together to make Collibra Data Governance Center a data governance solution that is difficult to beat.

The products

Collibra's principle product is the Collibra Data Governance Center, an enterprise wide data governance solution available both in-cloud and on-premises. In fact, Collibra Data Governance Center is a suite of products that work together to support data governance. It runs on any platform that supports Java and an Oracle or SQL Server database, with a browser-based front-end. It consists of *Business Glossary*, a collaboratively created 'single source of truth' for business terms and definitions; *Data Dictionary*, a searchable repository of technical metadata; *Data Helpdesk*, a ticketing system that allows any user to flag data as incorrect; *Policy Manager*, a central location to create, review and update policies using automated workflows; *Stewardship*, a system for managing data stewards; and *Reference Data*, a tool that helps users to understand and use reference data. These products may be extended via *Collibra Catalog*, a data catalogue that spans databases and data lakes. Also available are *Collibra Connect* and *Collibra On-the-Go*, which provide integration with third party products, and mobile support, respectively.

It is possible to license and implement the Data Governance Center without the data catalogue provided by Collibra Catalog. Historically, the latter was not available. Nevertheless, we can think of no good reason, now that you can deploy it, why you would not do so. While we start this review with a discussion of Collibra's policy management capabilities, the bulk of this report assumes that you will be implementing the Collibra Catalog.

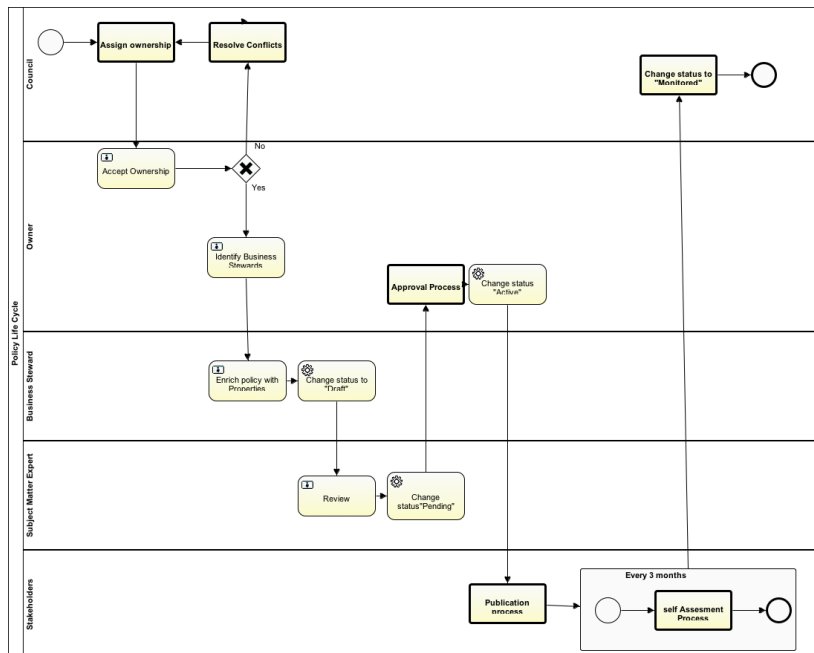
Policy Management

The Collibra Data Governance Center has many capabilities which are worthy of note. However, the area in which it most prominently stands out is its capacity to support policy creation, monitoring and enforcement. This is made most obvious by Policy Manager, a centralised location for viewing and managing all of your policies and standards, as well as linking them to business rules in order to make them enforceable.

Policy Manager's most potent capabilities come in the form of creating, reviewing and updating your policies, as illustrated in **Figure 1**. Each of these functions is facilitated by a selection of automated, out-of-the-box business management workflows. It ships with several (in excess of a dozen) of these. These workflows move repeatedly between each and every potential stakeholder for the proposed policy, allowing everyone to have their say and the policy to be altered to accommodate their wishes. Once a policy has been agreed upon and finalised, the workflow will also contact the appropriate users to have it implemented. A notable feature is that users can approve via email without actually opening the application. If the pre-built workflows are not sufficient, there is also the option to either edit any of the existing workflows or create entirely new ones.

“
The Collibra Data Governance Center has many capabilities which are worthy of note.
”

Figure 1: the policy publication process in Collibra



In addition, Collibra Data Governance Center contains Stewardship, a platform for managing your data stewards and their relationship with the other users in your organisation, including each other.

It allows you to establish an organisational structure, thereby empowering your data stewards to make data governance decisions collaboratively by engaging with the right users. In particular, it supports the creation of data stewardship processes and policies, for instance, to govern the use of sensitive data, by helping your data stewards to involve the appropriate stakeholders in the creation and management of these policies.

Data Cataloguing

Information catalogues are becoming increasingly useful tools to support organisational evolution in data-driven enterprises. This is because, essentially, they provide a registry of all the data available to you. This applies not just to conventional relational databases but also to data lakes, file systems and other types of data store. Moreover, it isn't simply a question of having a catalogue for each of these but of having one catalogue that spans all of them. In other words, catalogues – which are indexed and searchable – provide an integrating mechanism that enables both self-service and collaboration.

As far as Collibra Catalog is concerned, the first screen you are likely to see is the catalogue recommendation screen, as shown in **Figure 2**. This leverages machine learning to display assets that are likely to be useful, based on you or your teams' prior data usage, role in the organisation, and the activity of others with similar roles. You can search and filter this information, and once you have found something that looks interesting, you can then explore its relationships, usage, and data profile in more detail. The green flags in **Figure 2** indicate that these have been approved as authoritative sources of data.

Also worth discussing is the Catalog Dashboard, as shown in **Figure 3**. This screen is primarily composed of widgets that display data profiling information. These widgets are both personalised and customisable, allowing each user to display the information most useful to them. Further, you can create more dashboards if desired. This means that you could, for instance, create additional dashboards that display information pertaining only to a particular topic, such as data quality, data privacy or a particular regulation. From any of these screens you can run a global search on your catalogue, with filters for data type, context and category; create a new asset, including datasets, business terms and policies; or register a new data source.

Upon navigating to an asset, you will be able to view its properties in detail. For a dataset, this will generally mean a description, generated metadata, sample data, and data profiling information such as data distribution. Users are able to

Figure 2: the Collibra Catalog recommendation page

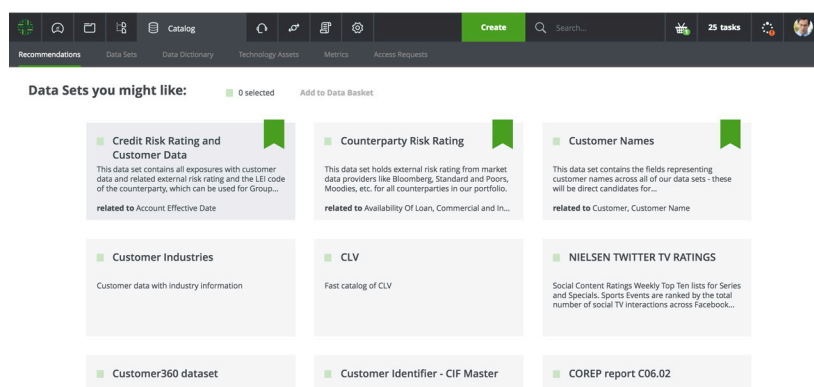
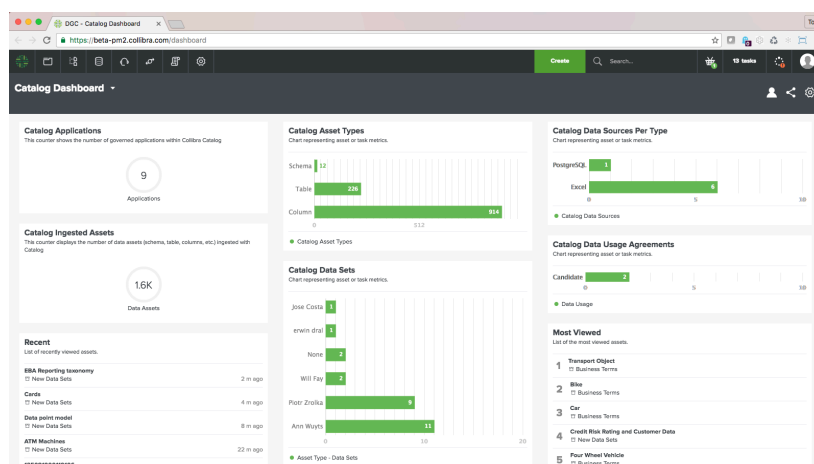


Figure 3: the catalogue dashboard in Collibra Catalog



comment on assets, as well as suggest customised tags that are searchable once added. You will also be able to see any relationships your dataset has to other assets. These relationships have several possible forms, depending on the type of asset being related to. For instance, your dataset might contain a data element, be related to a business term, or comply with a policy. Similarly, you can track the relationships between data and how it is used in reports, models or processes. Each category of relationship is listed separately and in natural language, making it plain to see which type of relationship is which. Furthermore, links specifically between data and business assets – that is, datasets and business terms – will be suggested automatically by the software, based on the existing associations and relationships within the catalogue. These suggestions are, again, backed by machine learning. Lastly, there are several additional views available for a given dataset, including data quality information and lineage. In this context, it is worth commenting that you can graphically explore – in an interactive manner – business traceability and policy hierarchies as well as lineage.

There are two other products in Collibra Data Governance Center that significantly aid in governing your data. The first of these is the Business Glossary. At its most basic, the glossary is a place to store and organise the definitions of business terms. In reality, it is much more than that, providing a system which enables your users to create those business definitions collaboratively. It also provides universal access to those definitions, as well as allowing you to associate business terms with business rules, policies and processes. These associations can, in turn, be leveraged by your catalogue. The product also integrates with IBM's Business Glossary and Informatica Metadata Manager, in case you are already using these.

The second product worth mentioning here is the Data Dictionary. Just as the glossary is the single source of truth for business assets, the dictionary performs a similar job for data assets and technical metadata, such as data structure, origin, format and use. It allows you to visualise data flows and traceability, it is searchable,

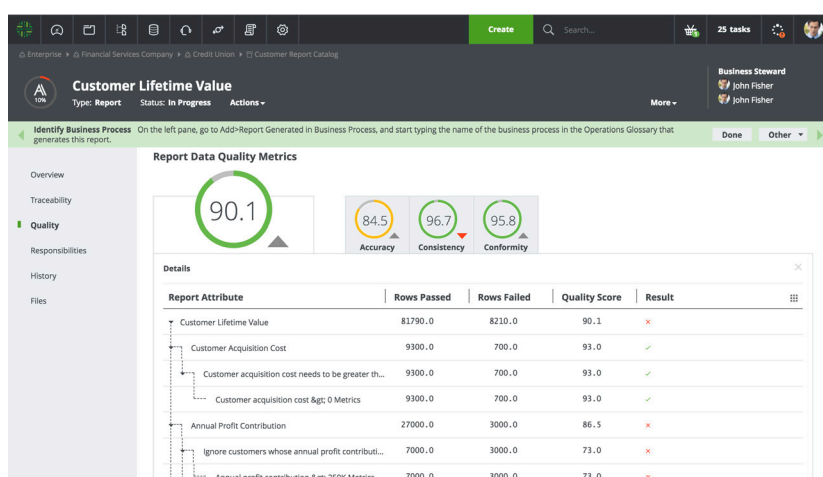
and, like the glossary, can be accessed by anyone. These two tools, along with the catalogue itself, highlight an important advantage of the data catalogue: making data readily available to anyone in your organisation, at any time.

Data quality monitoring

This is an especially strong area for Collibra. Every asset in the Collibra Catalog has a dedicated page for monitoring its data quality. Here, you can find all of the business rules pertaining to both data quality and the asset in question, the rules themselves being imported from whatever third-party data quality software you are using (and Collibra also integrates with third-party master data management solutions). You will also be able to see whether the asset passes or fails each of these rules on a row-by-row basis. The results from each rule are assigned a quality score, based on the percentage of passing rows. These scores are then aggregated to generate the asset's overall data quality score, which is displayed prominently on the page as shown in **Figure 4**.

“Users are able to comment on assets, as well as suggest customised tags that are searchable once added. You will also be able to see any relationships your dataset has to other assets.”

Figure 4: the data quality view for an asset in Collibra Catalog





The graph can also be actively navigated, so users can expand and collapse different relationships on the fly as well. Using the exploration mode gives complete access to the information about the data.



The overall score is colour-coded (green for good, red for bad, orange for middling) and features an arrow indicating the trend compared to the previous quality assessment (a detailed history is also available on a separate page). Additional, 'sub' scores are available that score the asset in a particular area of data quality, for example, accuracy, consistency or completeness. These can be very useful when addressing a low data quality score: you need to know why the score is low before you can take appropriate action to increase it. Finally, the overall scores are themselves aggregated in order to generate several different types of data quality assessment for your catalogue that are available on the dashboard.

The Data Helpdesk is a ticketing system that allows any user to issue a ticket to flag that something is wrong with an asset. Although it is not strictly a data quality tool – a user could, in principle, ticket a privacy or policy issue – it is integrated with Collibra's data quality dashboard and you can set quality thresholds that will automatically

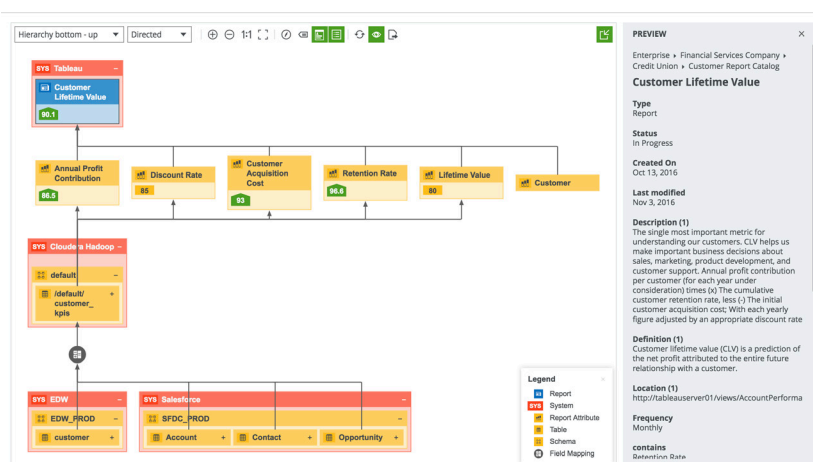
generate a ticket if the quality score (or a sub-score) falls below that threshold. More generally, enabling crowdsourcing and collaboration to support data quality is a very good thing, following the principle that many eyes will find more errors than a few. Moreover, the helpdesk intelligently routes tickets to the correct stakeholder for resolution, based on their relationship with the problem asset. This means that tickets can be resolved swiftly, and in so doing create a positive feedback loop: if a user flags an issue and it's resolved quickly, they are much more likely to flag more issues, rather than ignore them. The end result is that the quality of your data should improve over time.

Data Privacy

Assets in Collibra Catalog are equipped with views for both traceability (lineage) and responsibilities. The latter displays parties or roles responsible for the asset, such as the owner or owners, the chief data officer, or the stakeholders collectively, all of which could be individuals or groups. It should also be noted that the catalogue features role-based permissions. The former view indicates who is using the data and how they are using it, visualised in the form of a traceability diagram that includes all linked assets, regardless of whether they are datasets, business terms or governance policies. An example of this diagram can be seen in **Figure 5**. This view of a report asset is designed to show specifically what physical artifacts the data in the report comes from. This is the type of report that would be used by a developer or someone who wishes to understand how the data was derived at a technical level. Views can be built and stored, and there are a number of default views provided out of the box.

Other views could show the relationship with business elements such as business terms and uses. Users can design these views on the fly, and store them for future use, or sharing with their colleagues. The graph can also be actively navigated, so users can expand and collapse different relationships on the fly as well. Using the exploration mode gives complete access to the information about the data.

Figure 5: Collibra lineage view for a report asset



More generally, the entire structure of the catalogue is built around making sure data is accessed appropriately, as determined by a data owner or data steward. To this end, the default workflow is that in order for any user without special permissions to get full access to data assets found in the catalogue, they must go through a 'checkout' process, similar in style to a web store such as Amazon, in which they gather the data assets they would like to work with and request permission to access them collectively, providing both a reason for the request and, optionally, start and end dates. This workflow process is customisable so that you can ask for additional or different details such as the project for which the data is required or who will use the data. Once these details are completed this will start the workflow that forwards the request to the data owner or owners, allowing them to review the request and either approve or reject

it. While we are happy to see these sorts of features, there is a danger that the process for approving access described here is too onerous and lengthy. There will be occasions when access to data is needed urgently and the data owner isn't available. There therefore needs to be a workaround for occasions when access needs to be expedited.

Additionally, users can tag data as sensitive (and therefore subject to security policies) although Collibra relies on third party products to subsequently mask or encrypt such data. Automated data discovery and data profiling tools are also available, and can assist in finding sensitive data that can then be tagged.

“
The entire structure of the catalogue is built around making sure data is accessed appropriately, as determined by a data owner or data steward.
”



The Data Governance Center includes semantic modelling capabilities that support constructs such as synonyms, homonyms and words in different languages.



The vendor



ollibra is a small but high growth Data Governance software company. Its research and development is done in Europe (the company's founders are Belgian), but deploys most of its sales and marketing activities through its US headquarters in New York. The company's technology originally derived from research into semantics, which was first commercialised in 2008. With respect to this, it is important to note that the Data Governance Center includes semantic modelling capabilities that support constructs such as synonyms, homonyms and words in different languages.

In addition to building and selling its products, Collibra manages Collibra University, a free data governance training program; hosts the Collibra Community, a

support centre for its customers; holds the annual Data Citizens Conference; and runs on-going Data Citizen City Tours in cities around the world.

The company partners with Trillium (now part of Syncsort, itself to be acquired by Centerbridge Partners in Q3 2017) for data quality. Collibra also integrates with IBM, Informatica and others for data quality, via an open API. The company does not, at present, have any formal relationships with data masking vendors.

Website: www.collibra.com

Conclusion

In a sense, Collibra Data Governance Center has little real competition. Of course, there are other data governance products and solutions, and several of them even target the entire organisation in a similar way to Collibra Data Governance Center. The difference is that a lot of the competition seems to see policy management as an auxiliary capability where Collibra sees it – as do we – as a primary function of data governance. The end result is that few of the competition can even hope to match Collibra when it comes to its

area of expertise. On the other hand, the company is likely to face increased competition from data management vendors that provide data quality and cleansing solutions who are moving into the broader data governance space: these may not have the strength of policy and related support that Collibra can offer but they otherwise have a greater breadth of capability. Leaving this aside, Collibra Data Governance Center is a highly capable solution even if you don't rate policy management as highly as Collibra; if you do, it should absolutely be on your shortlist.

FURTHER INFORMATION

Further information about this subject is available from www.BloorResearch.com/update/2350



About the authors

DANIEL HOWARD
Information Management

Daniel started in the IT industry relatively recently, in only 2014. Following the completion of his Masters in Mathematics at the University of Bath, he started working as a developer and tester at IPL (now part of Civica Group). His work there included all manner of software and web development and testing, usually in an Agile environment and usually to a high standard, including a stint working at an 'innovation lab' at Nationwide.

In the summer of 2016, Daniel's father, Philip Howard, approached him with a piece of work that he thought would be enriched by the development and testing experience that Daniel could bring to the

table. Shortly afterward, Daniel left IPL to work for Bloor Research as a researcher and the rest (so far, at least) is history.

Daniel primarily (although by no means exclusively) works alongside his father, providing technical expertise, insight and the 'on-the-ground' perspective of a (former) developer, in the form of both verbal explanation and written articles. His area of research is principally DevOps, where his previous experience can be put to the most use, but he is increasingly branching into related areas.

Outside of work, Daniel enjoys latin and ballroom dancing, skiing, cooking and playing the guitar.



PHILIP HOWARD
Research Director/Information Management

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 Information Management.

Information management includes anything that refers to the management, movement, governance and storage of data, as well as access to and analysis of that data. It involves diverse technologies that include (but are not limited to)

databases and data warehousing, data integration, data quality, master data management, data governance, data migration, metadata management, and data preparation and analytics.

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), and dining out.

Bloor overview

Technology is enabling rapid business evolution. The opportunities are immense but if you do not adapt then you will not survive. So in the age of Mutable business Evolution is Essential to your success.

We'll show you the future and help you deliver it.

Bloor brings fresh technological thinking to help you navigate complex business situations, converting challenges into new opportunities for real growth, profitability and impact.

We provide actionable strategic insight through our innovative independent technology research, advisory and consulting services. We assist companies throughout their transformation journeys to stay relevant, bringing fresh thinking to complex business situations and turning challenges into new opportunities for real growth and profitability.

For over 25 years, Bloor has assisted companies to intelligently evolve: by embracing technology to adjust their strategies and achieve the best possible outcomes. At Bloor, we will help you challenge assumptions to consistently improve and succeed.

Copyright and disclaimer

This document is copyright © 2017 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.



Bloor Research International Ltd
20-22 Wenlock Road
LONDON N1 7GU
United Kingdom

Tel: **+44 (0)20 7043 9750**
Web: **www.Bloor.eu**
Email: **info@Bloor.eu**