



InComparison

InComparison Paper by Bloor
Author **Philip Howard**
Publish date **October 2017**

Policy-based Data Governance



If you buy into DATUM's approach – which we do – and need a product that can be deployed very rapidly (within days in some cases), and with a short-term (the company claims less than a month) time to value, then DATUM is an obvious potential choice to provide data governance.



Author **Philip Howard**

Executive summary

This paper compares DATUM Information Value Management with equivalent data governance capabilities from Collibra and Informatica. Our approach has been to research and write detailed technical evaluations – affirmed as correct by the respective vendors – of each of the relevant product sets; and then to use those evaluations as the basis for this comparison. These detailed evaluations are (or shortly will be) available on the Bloor Research website.

We should make it clear that this comparison is specifically related to elements of data governance which are primarily related to policy creation, monitoring and enforcement, but which also extends to constructs such as data catalogues. Thus, this report will be most relevant to users wishing to take a best of breed approach to data governance. We say this because neither DATUM nor Collibra provide data quality products, master data management or data-centric security products, all of which are provided by Informatica. In practice, this comparison is, more or less, a direct comparison between the two market leaders DATUM and Collibra but only a partial comparison with Informatica.

We start with a brief overview of each product set and what it does and then we will look specifically at different elements of the three products. Details provided in italics (and in quotation marks) are direct excerpts from the relevant technical evaluations or from Bloor Research's (forthcoming at the time of writing) Market Update on Data Lake Management where, again, each of the vendors in this paper has had the opportunity to agree a short description of their products and capabilities.



This comparison is specifically related to elements of data governance which are primarily related to policy creation, monitoring and enforcement, but which also extends to constructs such as data catalogues.



DATUM Information Value Management

DATUM started life leveraging services and consulting as a means to accelerate R&D while getting close to customer problems and, as a result, it has significant domain expertise that you would not normally expect from a vendor coming from a purely technical background. This business plan has paid off in product vision and intellectual property that can be used in practical implementations.

“DATUM Information Value Management is a data governance platform that aligns traditional governance with business objectives and business value.”

cataloguing, which help to keep up with changing data volumes, sources, types and formats, as well as structural changes. The capabilities provided extend to the discovery of sensitive data, for which the company has implemented machine learning.”

More generally, Information Value Management has a number of what the company calls “composers”, which relate to specific capabilities. These include (briefly):

- **Goals and Objectives**, whereby “users can document enterprise goals and then break them down into actionable objectives. DATUM provides guidance capabilities that help users collaborate with different stakeholders to agree on the most important objectives.”
- **Measures and Metrics**, “this is where you define relevant metrics and measures so that you can trace visibility from the metric down to the information assets that are used in the metric formulation.”
- **Business processes**, “what you are doing here is to define the roles of those tasked with creating and managing the details defined within the platform, together with the responsibilities and tasks that are attendant on those roles. In so far as roles are concerned, here you would define who is accountable: who owns the data, who manages the data, who is responsible for its remediation, and so on. And this would be aligned with relevant departments, such as finance, engineering, marketing and so forth.”
- **Data dictionary**, which is really “a comprehensive business data glossary that includes relevant business and technical attributes such as definitions and usage. It is not only used to support collaboration between the business and IT, it also underpins the rules, processes and metrics that are defined within other DATUM applications. Specifically, the data dictionary is pre-populated with some 1,700 pre-defined but

Figure 1: DATUM Information Value Management capabilities



Its approach is illustrated in **Figure 1**, and shows how Information Value Management’s capabilities align to enable customers to link their data to business goals and objectives. It also gives a good indication of DATUM’s target audiences for regulatory compliance, operational improvement and analytics. With respect to the last of these and specifically with regard to data lakes, “Information Value Management provides data discovery, a searchable data catalogue, creation of cross-lake metadata and tracking, and data lineage and metadata management, amongst other features.

“While DATUM’s platform is primarily a cloud-based application there are particular functions that may be implemented on-premises, namely discovery and

customisable definitions and, of course, you can add your own. In this context, it is worth noting that Information Value Management can automate the collection of metadata from applications such as SAP's Business Suite."

- **Governance rules**, this is "where you define and monitor business rules that apply to any particular dataset. Some of these will be quite simple: for example, defining the format of a customer ID, but others may be quite complex with multiple rules applying to the same data element." Additional features "include remediation strategy, approval workflow and organisational alignment. The last of these needs further description. It links to business impact but also includes information about relevance on a location basis or related to your organisation's structure. Most importantly this is where you assign roles and responsibilities to data stewards, data providers, data maintainers and so on. Remediation strategies directly link to this, as does approval workflow."

- **Standards Repository**, "this is similar to the previous application except that where you were defining rules in that module, here you are defining the standards that you want to apply, by data element, across your various systems. The aim here is to ensure accuracy, consistency, completeness and so forth: all the things that you would think of in terms of data quality and which are required by various regulations. In this context, it is worth noting that the company plans to introduce pre-built applications that are specific to particular compliance requirements. The first of these is the GDPR (General Data Protection Regulation) offering, which is currently available."



It is worth noting that the company plans to introduce pre-built applications that are specific to particular compliance requirements. The first of these is the GDPR (General Data Protection Regulation) offering, which is currently available.



Collibra

“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.”

Collibra Catalog, while not limited to data lake environments, provides “cataloguing capability that simplifies the process of data discovery by users. There are automated data discovery tools (including data profiling, which can also assist in finding sensitive data); and the ability to tag, document and annotate datasets with information such as the source of the data, the data owner, criticality, service level details with respect to data quality and business rules, and whether the data is sensitive and subject to security policies. The Collibra Catalog also includes facilities to support the ingestion of data into your data lake, with workflow capabilities and role-based approvals.”



Collibra's principle product is the Collibra Data Governance Center, an enterprise-wide data governance solution available both in-cloud and on-premises.



Informatica

While Informatica is not considered a pure-play Data Governance platform, it has a suite of products that when combined can provide the bulk of the capabilities one might require for data governance. These include Data Profiling and Data Quality, Master Data Management, both static and dynamic Data Masking, the company's Secure@Source product that complements Data Masking within a data-centric security environment. The directly competitive products to DATUM and Collibra are the Enterprise Information Catalog (EIG) and Axon. There is also the Informatica Intelligent Data Lake, which extends the cataloguing capabilities of EIG with the addition of data preparation capabilities.

The Enterprise Information Catalog, which includes a business glossary, is "available in the cloud, on-premises and for big data (the latter usually in conjunction with its sister product, Intelligent Data Lake). It automatically scans and indexes data assets from within your enterprise, then makes them available for discovery – via a data catalogue – by your users. This has all the usual benefits of a data catalogue, namely that it allows your users to get self-service access, thereby eliminating any bottlenecks that might otherwise occur. Its basic features are a global search of your data assets and data profiling of those same assets, but it also includes data lineage, impact analysis, the ability to trace how your data is moving, and a graph-based map of the relationships between data in your enterprise."

Axon, which Informatica acquired early in 2017 when it bought Diaku (which is the Greek for govern), "provides a collaborative business environment to support data governance: though organisational understanding might be a better way to put it. It provides a significant level of support for the creation and implementation of policies oriented around data quality and data privacy", as well as monitoring thereof. "Specifically, it allows you to design and build a policy hierarchy, beginning with an abstract description of your policy, and ending with concrete business rules that can be executed on your data assets. Policies will typically be organised by project although, of course, policies may be reused across projects. Projects may be purely internal or they may be related to regulations such as the General Data Protection Regulation (GDPR). A policy can be connected to its business use, value and objectives via the various business facets of Axon – including Glossary, Process, Capability, Quality, Project, Client Types and Regulation. The way policies are linked to these facts is customisable, so the link between policy and business can be described in a detailed, flexible way."

“
Axon provides a collaborative business environment to support data governance: though organisational understanding might be a better way to put it.
”

Comparison

Having outlined the capabilities of each of the solutions under consideration we will start by making some general observations about the three vendors:

- DATUM was built for the cloud and its solution is multi-tenanted. It supports top-down, bottom-up and middle-out approaches to data governance and its facilities are built on top of a graph database that allows easy exploration of data assets, policies and other artefacts, and their relationships. There is a clear emphasis on the business, as opposed to IT, but the focus is arguably more on business users rather than business analysts and more technical users.
- Unlike DATUM, Collibra was originally developed as an on-premises solution and while it has been ported into the cloud it is not “cloud native”. It also requires either an Oracle or SQL Server database, which is a limiting factor. While the company certainly has users across multiple industries its primary focus is on financial services and healthcare and one could get the impression that Collibra is more interested in regulatory compliance rather than data governance per se. Not that that is necessarily a bad thing. We would also comment that while Collibra has a clear focus on the business, it is nevertheless quite technical as a solution and, in places, is quite prescriptive about the way things should be done.
- While Informatica’s products were generally developed for on-premises environments we would say that the company has successfully transitioned to supporting cloud-based environments and Axon itself is cloud-based. Historically, the company has marketed its products to IT departments and has had a significant technical bias. The company is trying to move towards a more business-oriented approach but this is not an easy thing to do, either from a product or sales perspective.

As far as its data governance is concerned it has a similar underlying graph engine to DATUM, so exploration should be easy, and its ability to integrate with third-party tools surpasses those of either Collibra or DATUM. One weakness is that the Informatica products were never built as a single product suite and Axon, as a recent acquisition, is not as tightly integrated into the environment as one might like.

As far as specific capabilities are concerned, all the vendors offer the ability to define, manage and enforce policies; they all provide a data catalogue; they all support the discovery of sensitive or private information; they all provide specific features designed to support data stewards; and they all offer the ability to monitor data quality. We could discuss each of these elements in detail but the truth is that all these products are very capable. Any one product may have a bell or a whistle that is not in another offering but that is nit-picking and it would serve little purpose to discuss these differences in detail. We will therefore confine ourselves to areas where we think that there are significant differences, even if these are as much philosophical as technical.

Workflow

As one would expect, all the vendors offer workflow capabilities and support approval processes and the like. However, we find that Collibra’s approach can be excessively prescriptive. For example, *“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*



All the vendors offer the ability to define, manage and enforce policies; they all provide a data catalogue; they all support the discovery of sensitive or private information; they all provide specific features designed to support data stewards; and they all offer the ability to monitor data quality.



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." The danger here is that users finding this to be a bottleneck will game this approach and subvert these processes. It is not self-service if you have to wait for approval before you can start accessing the data you need. Governance in this environment needs to be a light touch. Both Informatica and DATUM have far less prescriptive approaches. In this context, note that all three vendors support role-based access security.

Approaching policies

DATUM believes – and we agree – that being able to link policies to business value, goals and objectives, is not only important for its own sake, but is essential if you want to prioritise how you approach data governance and the implementation of relevant policies. Consider the following description of Informatica's capabilities: *"the link between policy and business can be described in a detailed, flexible way"*. This is essentially also the approach taken by Collibra. However, while DATUM can be used in this manner, more usually you will use business value, goals and objectives to drive the development of policies as opposed to simply linking them. We believe that this is a superior approach for general purpose use.

Projects

All the vendors support the concept of projects as supersets of policies. In the case of Collibra, the Data Governance Center provides *"a selection of automated, out-of-the-box business management workflows and it ships with several (in excess of a dozen) of these"*, which are

designed to support common use cases such as GDPR or Solvency II regulatory requirements. DATUM provides pre-built capabilities, including over 120 best practice templates and tools to accelerate industry-specific initiatives and adoption of data governance solutions. Accelerator templates can include work products required to overhaul and optimise a new product introduction (NPI) process, respond to Life Sciences compliance regulations (such as the Sunshine Act and HIPAA) or implement/adopt a data governance operating model. For GDPR, it is offering what we would consider a more complete solution with dashboards (single view of privacy), templates with pre-loaded policies, and machine learning algorithms running in the background to identify personal information. DATUM has plans for further packaged solutions to be released in the future. As far as we are aware, Informatica has yet to introduce such capabilities.

Collaboration

All three vendors offer collaborative capabilities but Collibra, in our opinion, offers more advanced crowd-sourcing capabilities. Specifically, the Collibra Catalog provides *"automated data discovery tools (including data profiling, which can also assist in finding sensitive data); and the ability to tag, document and annotate datasets with information such as the source of the data, the data owner, criticality, service level details with respect to data quality and business rules, and whether the data is sensitive and subject to security policies."* With the exception of the crowd-sourcing facilities already mentioned, all three vendors have broadly comparable cataloguing functionality; all of them have embedded machine learning capabilities.

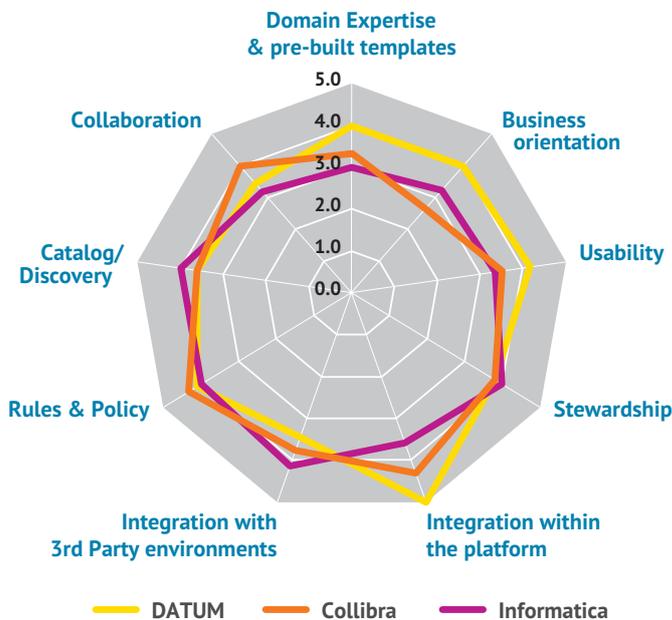
Conclusion

When DATUM commissioned Bloor Research to conduct the research that underlies this paper, it of course believed that its Information Value Management would come out ahead of both Collibra and Informatica in this head-to-head comparison. If we start with the premise that data governance should be business-driven rather than IT-driven

– which is generally agreed by data governance professionals – and if we also add into the mix that governance is best driven from business goals and objectives: then we agree with DATUM’s presumption. In our view, starting with business value makes obvious sense, though advocates of Informatica and Collibra would argue that you can do something similar with their respective products, even if not so cleanly and neatly. However, we would also add that DATUM outscores its competition in terms of its ease of use and suitability for non-technical personnel. While there are certainly aspects of both Collibra’s and Informatica’s offerings that are well designed for business use, we find Collibra’s approach to be inclined to be too prescriptive, while Informatica’s products can seem over-technical. In addition, it is worth commenting that Collibra’s editor-based pricing model is not conducive to a business-driven approach, because the more the business is involved the more editors (as opposed to just users) there are likely to be and the higher the cost potentially becomes.

The bottom line is that it depends on what is important to you. If you buy into DATUM’s approach – which we do – and need a product that can be deployed very rapidly (within days in some cases), and with a short-term (the company claims less than a month) time to value, then DATUM is an obvious potential choice to provide data governance. Even if you don’t require these things, then it remains worthy of your shortlist.

Capabilities	DATUM	Collibra	Informatica
Domain Expertise and pre-built templates	4	3.3	3
Business orientation	4	2.7	3.2
Usability	4.2	3.5	3.4
Stewardship	3.8	3.8	4
Integration within the platform	5	4.3	3.6
Integration with 3rd party environments	3.5	3.8	4.1
Rules and Policy Management	4.1	4.3	4
Catalog/Discovery	3.7	3.7	4
Collaboration	3.5	4	3.2
Pricing	4 Pricing Tiers including a Freemium option. No per user fees.	3 Pricing Tiers and a per editor fee (users free).	Pricing varies greatly and can depend on size of the organisation.



FURTHER INFORMATION

Further information about this subject is available from www.BloorResearch.com/update/2348



About the author

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**