

Global IDs Data Lineage

The company

Global IDs is a privately funded company that was established in 2001. It provides solutions for enterprise information management and governance, focusing on the largest, most complex environments, which are typically multi-domain and often span multiple geographies. What's more, the company's technology suite is extremely broad ranging and offers a variety of functions, including data profiling, data quality, data governance, data lineage, and master and reference data management.

What is it?

The Global IDs platform provides a variety of enterprise information management capabilities, with an evident emphasis on data privacy. This includes data lineage, data profiling, data cataloguing, data quality, data discovery, and data classification, among other things. Additional capabilities, such as data masking, are available via partners. For the purposes of this report, the company's data lineage solution is of particular interest, but we also examine parts of its (sensitive) data discovery and related capabilities, as they form an essential part of the lineage story.

“Global IDs Data Lineage was a lifesaver in helping us to meet audit and regulatory requirements. Their solution is highly automated and provided insight to data movement and transformations we were unable to determine before using their tool. Reports were easily exported enabling us to share proof and evidence and the ability to monitor and detect changes was instrumental in building trust in our data.”
Head of Data Governance,
Financial Institution

In general, the platform's data discovery and lineage capabilities allow you to find the personal and other critical data in your system, track its movements throughout said system at the level of both the individual record and the system as a whole, and then prove that you are doing so via visualisation and report generation. This is of significant benefit for both data privacy and regulatory compliance. Specific supported regulations include GDPR and CCPA, among others, but note that the platform is able to address privacy issues in a general sense and is not limited to any particular set of compliance mandates.

What's more, the product can do all of this on a massive scale with a high level of automation of accuracy – as is typical for Global IDs – and it supports a wide range of data sources and file formats, as you would expect from a product designed to support large and often highly varied ecosystems. In particular, it provides support for both relational and NoSQL databases. The latter most notably includes MongoDB and Cassandra, as well as – in principle – any data source that can be resolved into a columnar structure. Mainframes are also supported, as is Amazon S3, while unstructured support extends to text files and emails.

What does it do?

Global IDs uses machine learning (ML) driven classification and semantic tagging (with corresponding semantic domains) to automatically scan and categorise your data, including personal data as well as other critical data (perhaps relating to operational processes or decision making). Disambiguation and validation are provided, helping to eliminate false associations, and discovery results can be viewed through the platform's data catalogue and/or its data privacy dashboard. This process can be leveraged to retrieve all personal information related to a specific individual – and to generate a data privacy report to that effect, potentially in response to a Data Subject Access Request (DSAR) – but more pertinently, it forms the foundation of Global IDs' data lineage capability.

To wit, off the back of this discovery process the product provides full data lineage and traceability functionality, allowing you to see where a given

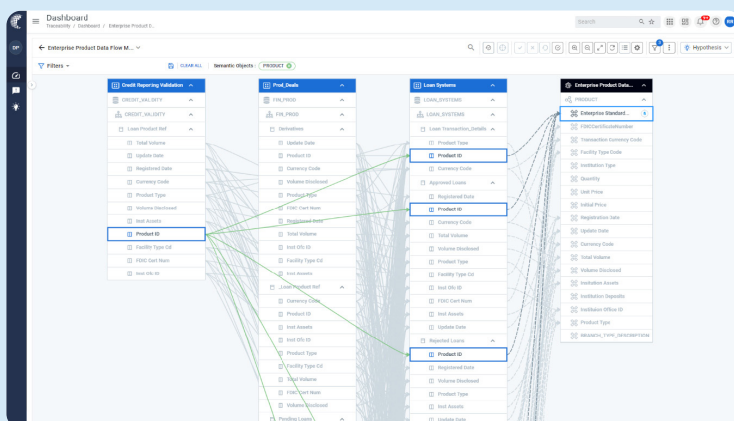


Figure 1 – Data lineage dashboard showing impact analysis

“
The auto-discovery
of data and its profiling
is a miracle. The Machine
Learning Algorithms for the data
profiling are well executed.
The overall workflow of the
tool is exceptional and
well thought out.”

C-Level,
Vendor Selection and
Purchasing

individual’s data is used within your system. It does this by generating hypotheses regarding where said data is being (or has been) used, based on both human and machine input. It is thus able to recommend probable flows that can be confirmed or rejected by your users by validating them against your actual system. You can also scan ETL flows and stored procedures in order to determine how they move data around, and even stitch together lineage flows manually if you so desire.

to regulatory compliance, as already discussed, and so on and so forth. Global IDs specifically can take a financial report and map all of the data represented in it back to its original source, track records back to individual consumers in order to engage in anti-money laundering monitoring, and more. Note that this is not a comprehensive list of Global IDs’ capabilities, or even its capabilities as they pertain to finance, but rather a brief elucidation of how the platform can be applied to this particular regulatory context. Other contexts can benefit in similar ways.



Figure 2 – Data lineage in virtual reality

The end result of this process is both a high-level visualisation of the flow of data within your system as a whole – shown in **Figure 1** – as well as a series of lower-level visualisations that show the movement of each (and all) of your individual records as they move between tables, applications, and any other part of your system. Both views are important aspects of data lineage, and of modern regulatory compliance. Moreover, the platform supports automated report generation for the lineage of a record in much the same way as it does for the personal data relating to a specific individual. Combined with the visualisations mentioned above, this should serve as adequate proof of your ability to trace personal data, and thus ensure compliance with GDPR, CCPA, and any other relevant regulations (at least as far as lineage is concerned). What’s more – and rather uniquely – all of this lineage information can be viewed in virtual reality (VR), as shown in **Figure 2**.

It is also worth demonstrating the real-world applicability of the lineage capability offered by Global IDs, and highlighting some of its features in that context. Consider, for example, the use of data lineage as part of financial operations - reporting, regulatory compliance, reconciliation, and so forth. Lineage in general is useful here for all of the reasons we’ve already described: it allows you to trace the flow of your data at all levels, which naturally feeds into effective reporting, it contributes

Why should you care?

The most obvious triumph of Global IDs’ data lineage (and indeed, Global IDs in general) is in its ability to operate at extremely large scales, encompassing huge numbers of data sources – which may be geographically distributed – and extracting comprehensive lineage information from even the largest systems. Moreover, the platform supports a broad range of data sources, including mainframes, relational databases, multiple types of NoSQL, S3, and others, making it eminently suitable for enterprise deployment.

That said, Global IDs’ lineage capability has more going for it than just scalability, although that alone makes for a substantial differentiator. The degree of automation offered is also a strong point, for instance, as is the ability to generate lineage information based (at least partially) on human input and feedback. Moreover, the aforementioned automation and scalability advantages combine to allow Global IDs to generate validated lineage information from even highly complex data flows with aplomb.

The end result is that by employing Global IDs you can learn how your data is flowing through your system and provide proof to that effect when regulators and auditors ask for it. The importance of being able to both understand and prove you understand your system really cannot be overstated when it comes to regulatory compliance, and these are some of the core values that Global IDs provides through its data lineage. Moreover, the increased understanding of data flow that its data lineage offers can also be beneficial in general: for generating robust impact analyses, or for resolving data quality issues, for example.

The Bottom Line

Global IDs offers useful, compliant data lineage that is suitable for even the largest and most complex enterprise systems. If that’s you – and perhaps even if it’s not – there is every reason to check it out.

[FOR FURTHER INFORMATION AND RESEARCH CLICK HERE](#)