



www.globalids.com

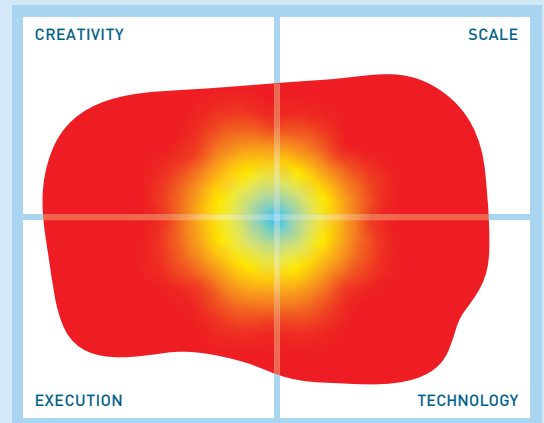
182 Nassau Street, Suite 202
 Princeton, NJ 08542, USA
 Tel: +1 (888) 514 0192
 Email: info@globalids.com

Global IDs Enterprise Data Automation

The company

Global IDs is a privately funded company established at the beginning of the century (2001). It is headquartered in New Jersey, although software development takes place primarily in India. Throughout its lifespan, the company has specialised in creating software for profiling, managing, and governing data within extremely large – and therefore extremely complex – data ecosystems. Its products are purpose built to work with the very largest enterprise systems containing tens or even hundreds of thousands of (potentially geographically distributed) data sources.

Global IDs’ product portfolio consists of a number of individual tools within the Data Ecosystem Evolution Platform. In this particular review we are focused on data quality, which is used by around thirty percent of clients. The software has been containerised and makes use of microservices while the company is currently working on elastic scaling and other cloud-native capabilities.



The image in this Mutable Quadrant is derived from 13 high level metrics, the more the image covers a section the better. Execution metrics relate to the company, Technology to the product, Creativity to both technical and business innovation and Scale covers the potential business and market impact.

What is it?

Global IDs does not see data quality as an isolated set of functions. In fact, the company’s methodology is illustrated in **Figure 1**, and as can be seen, data quality is actually step nine of ten. Specifically, it leverages earlier steps and particularly the discovery, profiling, classification, lineage, and catalog (the platform includes a data catalog) functions to establish effective quality controls, though it can be deployed without all of these previous steps. These all do much as you would expect them to do – or perhaps more – in terms of functionality but what really sets the whole Global IDs environment apart is its emphasis on data management at scale or, as in this case, data quality at scale. As an example of this, one of the company’s clients has two thousand applications, ten thousand databases, ninety thousand schemas, twenty million tables, and two hundred and thirteen million columns under management using Global IDs’ software. And this isn’t even its largest customer, which has five hundred million columns under management.

What does it do?

As a precursor to data cleansing operations, the Global IDs data profiling capability provides the sort of histograms and analysis that are commonplace for these sorts of tools. As far as data quality itself

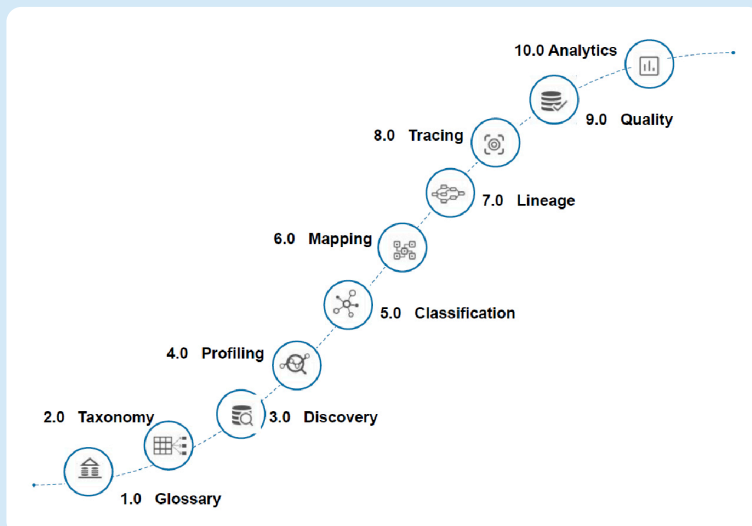


Figure 1 – The Global IDs methodology

Architecture	★★★★★	Extensions	★★★★★
Automation	★★★★★	Matching	★★★★★
Cleansing	★★★★★	Profiling	★★★★★
Ease of use	★★★★		

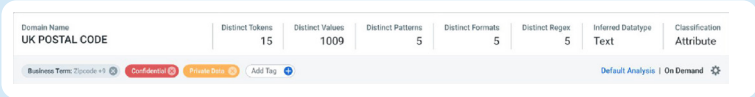


Figure 2 – Profiled UK postal code results based on the domain rule generator

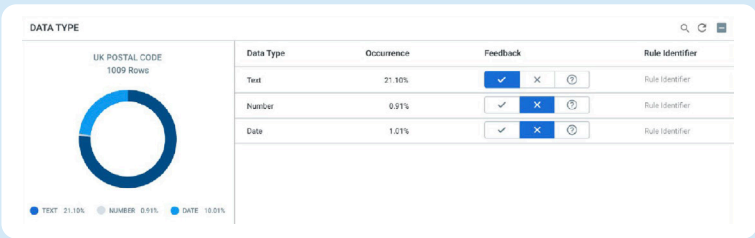


Figure 3 – Selecting datatypes in the domain rule generator

is concerned the company takes two approaches: rules-based data quality and machine learning based data quality, with the former applied to single and/or multiple columns and/or rows, while the latter is only used against single columns. In the case of multiple columns these are categorised by semantic domain (for example, checking the format and consistency of email addresses across multiple physical columns). More generally, data quality checks are used for completeness, conformity, consistency, integrity, and timeliness. While

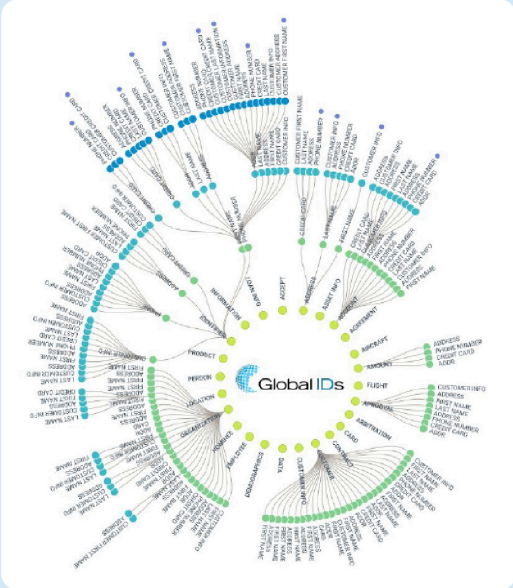


Figure 4 – Graph visualisation of data relationships

there is no match engine per se, rules can be used to check the uniqueness of any particular physical row to identify duplicates.

Global IDs provides a domain rule generator. For example, Figure 2 shows an example of profiled results for UK postal codes.

In the domain rule generator, you select the datatype – see Figure 3 – then the format (length plus pattern) using a similar clicking of the appropriate box and finally a word analysis, again based on the profiled results. The rule is then generated for you.

As far as machine learning is concerned, although its data quality use is confined to single columns it can be used in conjunction with classification to, for example, determine that a particular domain contains sensitive data. When running machine learning for data quality the software will make predictions about expected values along with confidence levels.

Other notable facilities include reconciliation analyses that compare source and target records. Needless to say, the company provides a data quality dashboard so that you can monitor data quality over time.

Why should you care?

We have made much of the fact that Global IDs is focused on data management at scale. However, we should note that what the company means by “at scale” is an order of magnitude greater than what most other vendors mean when they talk about scalability. For instance, Global IDs will currently admit that its row-based data quality is not “at scale” to the extent that it would like. But it’s probably already comparable to other tools in the market. However, running at scale causes other problems because it makes visualisation, in particular, difficult. The company therefore partners with Neo4j so that it can represent relationships in the data as a graph, as illustrated in Figure 4.

Aside from that, the major point in Global IDs favour is that it sees, and supports, data quality as a part of a chain of data management requirements rather than something that is a stand-alone function. We concur with this view but it does mean that you are unlikely to select Global IDs purely for data quality purposes but only as a part of a more holistic solution.

The Bottom Line

Global IDs focuses on the most intractable and complex data management problems. Typically, in the largest enterprises with huge volumes of data. We are not aware of any other vendor that has this singular focus. If your company falls into this category Global IDs should at least be on your short list of potential providers.

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