



www.iri.com

2194 Highway A1A, 3rd Floor
Melbourne, FL 32937-4932, USA
Tel: +1 321 777 8889
Email: info@iri.com

Test Data Management in IRI Voracity

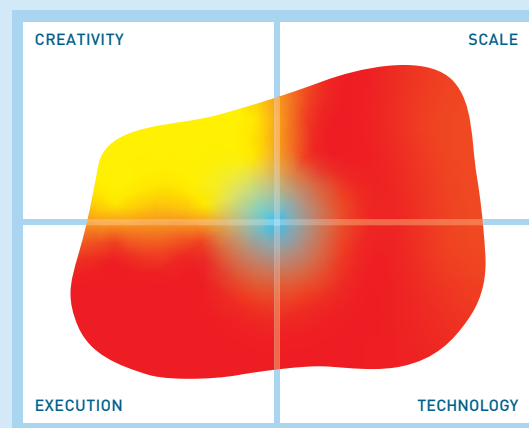
“We use a wide range of data masking and prototyping capabilities in Voracity to discover and protect individual identities, and to quickly produce quality test data for our heterogeneous technology ecosystem’s DevOps program and data virtualization scenarios.”
WD, Healthcare Enterprise Architect

The company

IRI is a privately-owned ISV. It was founded in 1978, and has international coverage. Its first product, CoSort, is a high-performance data transformation utility that remains at the heart of the company’s offerings today, including IRI Voracity, a “total data management” platform that spans data discovery, integration, migration, governance, and analytics.

What is it?

IRI Voracity contains two suites of products that provide functionality relevant to test data management (TDM): IRI Data Manager Suite, and IRI Data Protector Suite. The latter provides a selection of masking products (namely IRI FieldShield, CellShield EE, and DarkShield, plus a services option that leverages



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.

available for Voracity as a whole as well as individual products (and even APIs). Database virtualisation is not offered directly, but is provided through integration with partner vendors Windocks and Actifio. Other partnerships support integration with provisioning and CI/CD pipelines.

What does it do?

IRI’s approach to TDM consists of data masking (and subsetting) and/or synthetic data generation, either of which can run as standalone operations or in combination. Masking is rule-based and powered by the CoSort engine. FieldShield masks structured data and flat files, CellShield masks Microsoft Excel data, and DarkShield can mask structured, semi-structured and unstructured data simultaneously and consistently (shown in **Figure 1**). Several dozen static masking functions are available for FieldShield and DarkShield, and about half of those are available in CellShield as well. Dynamic data masking is also available.

In static operations, masked data is kept consistent across multiple data sources, so that referential integrity is always maintained. That consistency does not require formal constraints, but instead relies on masking rules applied to centrally defined data classes shared between all the shield products. Data classification, discovery, and masking is also offered as a service via DMaaS.

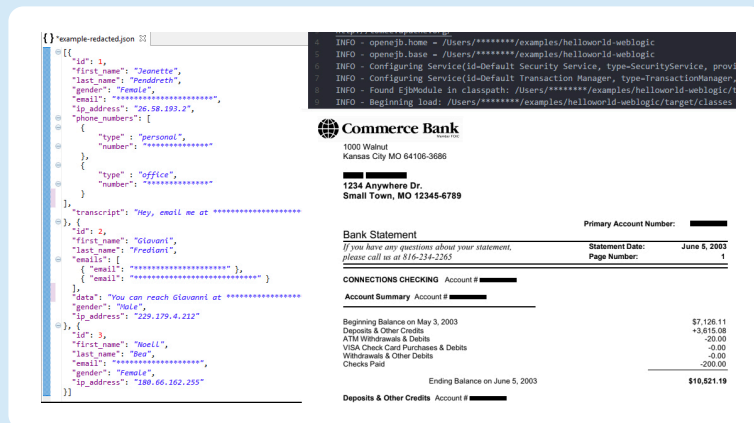


Figure 1 – Multi-source masking using IRI DarkShield

them called DMaaS) suitable for various use cases, including TDM. The former suite contains IRI RowGen, which can be used to generate synthetic data. In principle it also provides data subsetting, but in practice this is more typically delivered as part of the platform’s broader data integration capabilities.

The Voracity platform, including the above products, is accessed through IRI Workbench, a largely wizard-driven Eclipse interface backed by graphical modelling. Licensing is flexible, with options

Data subsetting	★★★★★
Synthetic data generation	★★★★★
Database virtualisation	N/A
Data discovery	★★★★★

Data masking	★★★★★
Automation	★★★★★
Ease of use	★★★★★
Integration	★★★★★

“
Test data management (TDM) is a critical part of our agile SDLC, and is subject to data privacy regulations. Integrated data classification, discovery, anonymization, subsetting, and synthesis functions in Voracity improve our time-to-market delivery strategy, and help us comply with GDPR and similar laws.”
RAK, Capgemini Technology Services

The various Data Protector Suite products provide multiple data discovery facilities that support both data profiling and masking operations. They allow you to categorise your data against an extensible library of pre-configured or bespoke data classes which can then be married to masking or test data generation rules acted on at execution time. Various discovery methods are associated with the data classes, including lookup value and pattern matching, with validation scripts to avoid false positives. DarkShield in particular can also leverage named entity recognition models (with semi-supervised machine learning to enable more sophisticated and effective language analysis of highly unstructured data), fuzzy matching, and facial recognition.

direction. The former is faster, but the latter is more comprehensive. Ultimately, you're handed the top 100 (or some other specified number) rows of the resulting subset under one of a variety of sorting methods.

All of the above functionality can be executed as individual scripts or batched jobs from within IRI Workbench, the command line, or a partnered database virtualisation environment (say, Windocks). You have several options for creating these scripts, including wizards, form editors, and mapping diagrams (see **Figure 2**). APIs are provided, meaning that Voracity TDM functions are also operable as part of an external pipeline, and the test data created by its processes can be exported to many databases and file formats, including spreadsheets, PDFs and images.

Why should you care?

IRI's subsetting, masking and synthetic data generation capabilities are all highly competent. For the latter in particular, the ability to create representative synthetic data sets via analysis is particularly notable and useful. Moreover, in this paper we've only really been able to scratch the surface of the product's capabilities. There is a significant depth of functionality here: IRI has been organically growing its technology for over 40 years, and it shows.

That said, TDM is only one aspect of Voracity. It is billed as a total data management platform, and to that end it offers a wealth of other capabilities – data integration, governance, quality, and so on – that stretch beyond just TDM. Moreover, these capabilities (including TDM) are offered through a unified and user-friendly interface, complete with wizards, visual programming and so on. This makes it easy to use each individual product and to shift your attention from one product to another.

The company's recently announced partnership (and integration) with containerised database virtualisation vendor Windocks is particularly promising, and its other relevant partners, including Actifio, CommVault and ValueLabs, should not be dismissed either.

The Bottom Line

IRI touts Voracity as a total data management platform, and in many ways, it lives up to that claim. As a solution for TDM specifically, it is both competent and – thanks to the platform's overall breadth – surprisingly versatile.

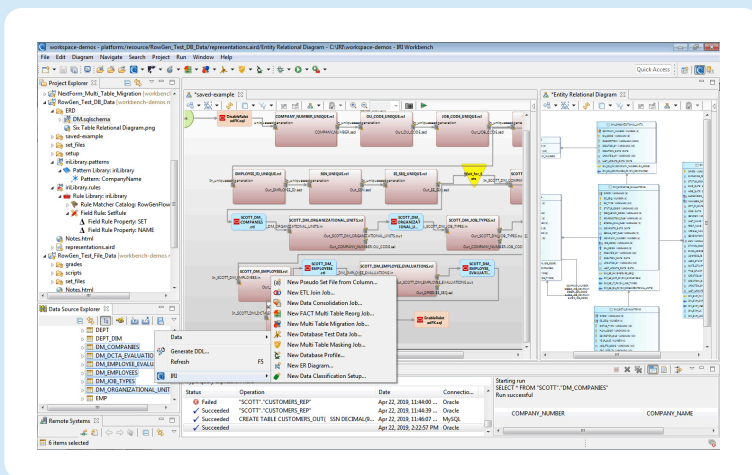


Figure 2 – Test data synthesis of a relational schema using RowGen in IRI Workbench

RowGen provides synthetic data generation. You can either use it to analyse the attributes of your production data and create synthetic data sets that resemble it, or take a more manual approach and apply generator rules yourself on a per-column basis. This can include drawing from an external set of values, possibly with a configurable distribution according to some secondary factor (gender, say).

Subsetting is delivered via either RowGen or Voracity's data integration capabilities. In either case, you can specify a driver table and trace its foreign key relationships to create a self-contained subset. Voracity gives you the option to follow these relationships "downhill" – only moving from parent to child – or to move through them in either

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