



Total Data Management

www.iri.com

2194 Highway A1A, 3rd Floor, Melbourne, FL 32937 USA

Tel: +1 321 777 8889 Email: info@iri.com

IRI FieldShield now at version 6

The company

IRI is a privately-owned independent software vendor founded in 1978. Its offices are in Florida and it relies on a partner network of resellers for international coverage (in 40 locations throughout the world).

The company's oldest product, CoSort, is a high-performance data transformation utility that was first designed to offload JCL sort/merge steps to CP/M. Needless to say, this has been extended and ported to other environments, but it remains at the heart of IRI's offerings. This includes Voracity, IRI's data management platform, that is designed to perform and consolidate common work in data discovery, integration, migration, governance, and analytics. In addition, IRI products can be operated using a common interface: the IRI Workbench, a bespoke, wizard-driven Eclipse IDE backed by graphical modelling.

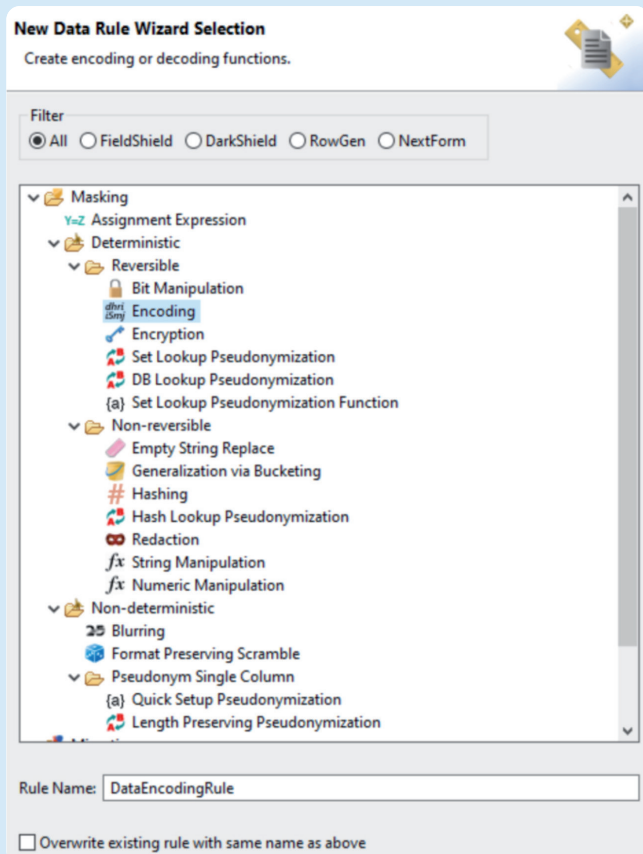
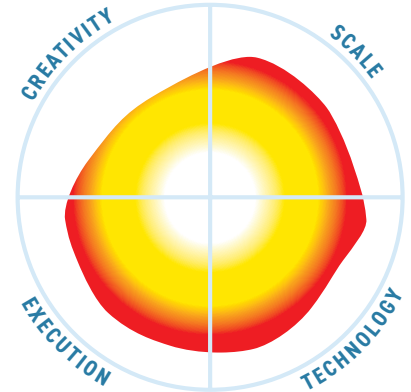


Figure 1 – Masking method selection in IRI Workbench

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?

IRI offers several products that are capable of data masking, namely FieldShield, DarkShield, and CellShield. Of these, CellShield is the most specific: it is designed to operate exclusively on Excel spreadsheets. DarkShield, conversely, is the most general, capable of searching for and masking PII and other sensitive data in a variety of different formats, including structured, unstructured, and semi-structured. FieldShield sits somewhere in the middle, focusing on structured data (primarily relational databases and flat-files, but also including spreadsheets, though it is less specialised than CellShield in regards to the latter). All three products are available individually, as a service, or as part of the IRI Data Protector Suite within the Voracity platform. They each offer data discovery capabilities in addition to data masking, and this can be used for a variety of purposes, not the least of which is to identify sensitive data to mask.

FieldShield, in particular, was first released in 2011 as a data masking solution that would "shield" sensitive data at the field level (hence the name). Historically, one of its biggest use cases has been anonymising data sets for use in test environments (which is to say, test data management), and this continues today. Moreover, FieldShield is built on the SortCL engine that also drives several other IRI products, such as CoSort. This provides FieldShield with various advanced functionality, including input phase filtering and complex field logic that can combine data masking with data cleansing, joining, reformatting, transforming, and more.

FieldShield can also be leveraged in real-time database replication scenarios via IRI Ripcurrent, a CDC (Change Data Capture) facility included in Voracity that refreshes target table rows whenever there are changes in the source table. By applying FieldShield's masking functionality in Ripcurrent, sensitive data classes being refreshed or replicated to can be updated with masked data automatically and immediately.

FieldShield works on RDBs and MongoDB, ASN.1 CDRs, Excel sheets, and fixed and delimited files – typically on-premise, but increasingly often in AWS, Azure, and GCP environments.

The product itself can be deployed on-prem or in the cloud, and can be consumed using the IRI Workbench or an API. It will readily integrate with other IRI products, including DarkShield for masking unstructured data and RowGen for generating synthetic data. In addition, FieldShield metadata is interoperable with other SortCL-driven products, notably Voracity for ETL et al., plus its RowGen, NextForm, and CoSort components. The latest FieldShield version, version 6, was announced in December 2024 and includes several new or updated features.

What does it do?

Masking in FieldShield is rule-based and powered by the aforementioned SortCL engine. A variety of out-of-the-box masking methods are available, including several dozen static masking functions that range between deterministic, non-deterministic, reversible, non-reversible, format-preserving, and so on (as shown in **Figure 1**). String and numeric manipulations are both available, and nested functions can be used to combine string manipulations and/or value lookups with other masking rules. Moreover, you can invoke DarkShield on free text fields from within FieldShield, scanning the contents of those fields and masking any sensitive text contained therein. Masked data is always kept consistent across multiple data sources, meaning that structural and referential integrity are maintained. As mentioned above, you can also leverage SortCL to combine masking with other data manipulations. For example, you can use it to action ETL processes simultaneously with data masking, which could be useful for, say, creating protected data sets for analytics.

FieldShield also provides a substantial data discovery and profiling capability, although unlike in DarkShield this is a separate process from its data masking. This enables you to search through and classify your data against a centralised library of either pre-configured or bespoke data classes shared between all of your IRI masking jobs and products, which can in turn be married to masking rules when they correspond to sensitive data. These rules are acted on at execution time, ensuring that the associated sensitive data is protected.

Considerations are made for performance: for instance, tables that have already been scanned will be skipped during repeated discovery phases, and you can choose to exclude specific tables or data classes from the process entirely. In addition, data classes can be grouped together at either a global or a project level, and these data class groups can be categorised by assigning them one or more sensitivity levels and/or applicable compliance regulations.

A range of discovery methods can be used as part of this process, including lookup value or pattern matching, column name matching, and dictionary matching. Moreover, any number of these methods can be used in concert with each other to improve the accuracy of your results, albeit at the cost of performance. There is a configurable matching threshold for discovery, allowing you specify how sure you want to be before settling on a result. Predefined methods for finding data protected by GDPR and HIPAA are available out-of-the-box.

FieldShield also comes equipped for handling quasi- or indirectly identifiable demographic data, which refers to

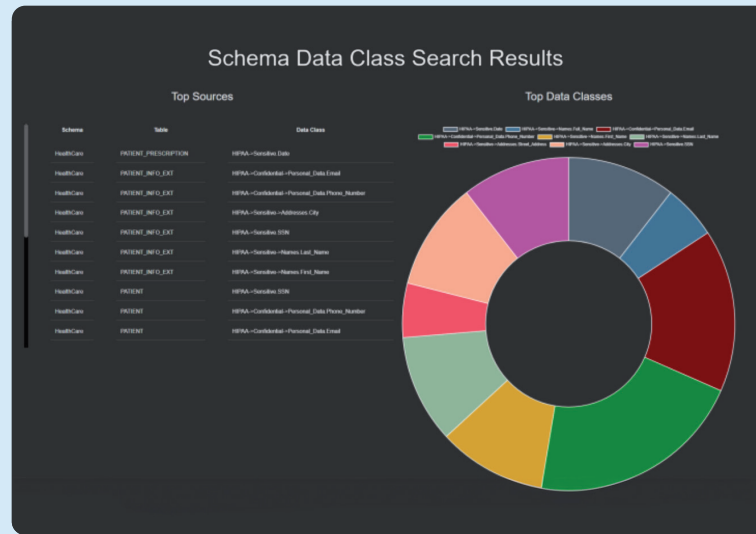


Figure 2 – HTML report produced by FieldShield

data that does not identify an individual by itself, but can be combined with other, similar data in order to do so. The product can score the re-ID risk from, and anonymise, this kind of data in order to keep it compliant but accurate enough for analytic or marketing purposes.

The results from FieldShield's discovery process are summarised within an HTML report (a small sample of which is shown in **Figure 2**). This report provides a range of visualised information, including (but far from limited to) job performance (such as how long it took to run) and configuration (such as which data sources and classes of sensitive information were used). This can be used to copy the configuration and constraints present during a previous job in order to quickly reproduce it.

Why should you care?

FieldShield is a robust solution for data masking. It is one of the most mature products currently available on the market, and while its focus on structured data is somewhat old-fashioned, the sophisticated capabilities that this focus – along with SortCL and Ripcurrent – provides are difficult to ignore. Indeed, SortCL is a significant differentiator for FieldShield, both within and without the IRI product catalogue. In particular, it is one of the primary reasons to use it over DarkShield, which can mask both structured and unstructured data but cannot leverage SortCL in the same way. That said, given the abundance of unstructured data available in many enterprise environments, we would expect that most organisations will want to leverage both products together rather than either individually. Indeed, the fact that you can easily do this is a major strength of IRI's data masking offering.

The bottom line

Whether for test data management or more general data masking applications, FieldShield is a formidable product for finding and protecting the sensitive data that is hidden in your structured sources.