

- Classify and Search for Hidden PII
- Extract and Report on Values and Metadata
- Mask Simultaneously, Later, or in Batch
- Combine Multiple Functions and Formats

Discover, Deliver, and Delete PII in Dark Data Sources



Product Summary

The Dark Data Conundrum

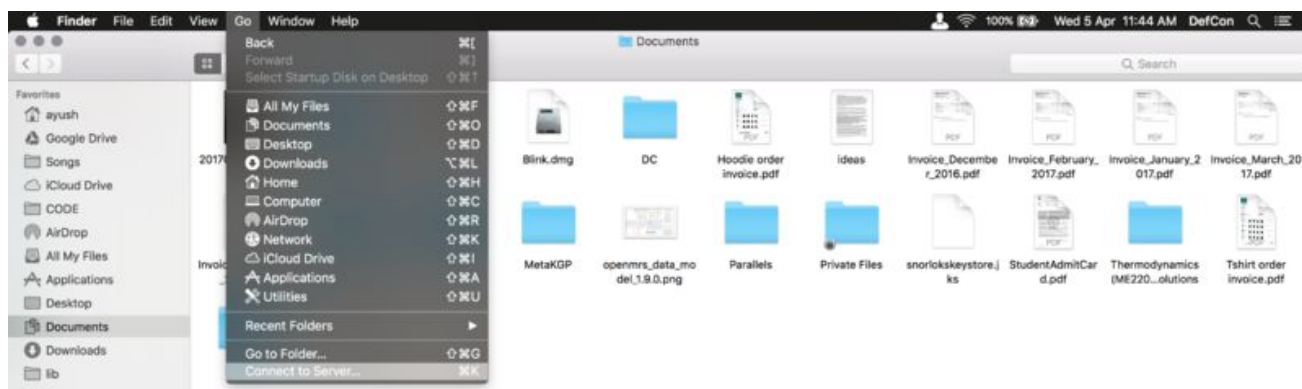
Gartner defines **dark data** as the "information assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes" like analytics or digital business ... Organizations often retain dark data for compliance or archive retrieval. Its storage and security "typically incurs more expense (and sometimes greater risk) than value."

Source: Gartner IT Glossary

Despite this observation, data loss prevention and the protection of personally identifiable information (PII) are critical elements of modern data governance and, in many cases, required by law. Unfortunately, safeguarding data at risk is a multifaceted problem -- especially when the data is hidden -- requires:

- 1) Knowledge of business and regulatory requirements,
- 2) Classification of sensitive data and its authorized recipients, and
- 3) The implementation of policies and techniques that support these requirements and protect PII.

PII search, remediation, and reporting techniques are particularly challenging to implement in dark data environments due to the volume, variety, and unstructured nature of the data sources in them.



Enter DarkShield Version 5

IRI DarkShield supports the risk and controls framework in enterprise IT environments by classifying, finding, extracting, masking, and reporting on PII and other data 'hidden' in unstructured sources.

DarkShield quickly and effectively scans structured, semi- and unstructured files, RDB C/BLOB columns, NoSQL DBs, Paquet, EDI and text files, plus images and PDF/MS documents. Local or SMB drives and folders, plus Azure, GCP, S3 and SharePoint Online buckets are supported. API users can source even further afield. DarkShield users find PII through data classes that match:

- 1) values stored in a lookup table (either exact or fuzzy matches)
- 2) named 'path' or column filters for JSON, XML, CSV, RDB, and Excel sources
- 3) stored or new Java Regex patterns, or elements in HL7, X21 and FHIR EDI files
- 4) drawn (bounding box) regions in images, or handwritten signatures in documents
- 5) machine-learned named entities in OpenNLP, PyTorch or Tensorflow (NER) models
- 6) OCR, OCR-A and MICR font recognition in images.

Whenever DarkShield finds a match, it applies the masking function you assigned to a "rule matcher" associated with data meeting specific search criteria. DarkShield can also detect and protect signatures and PII in audio files, and will eventually support handwriting and biometric identifiers like faces, too.

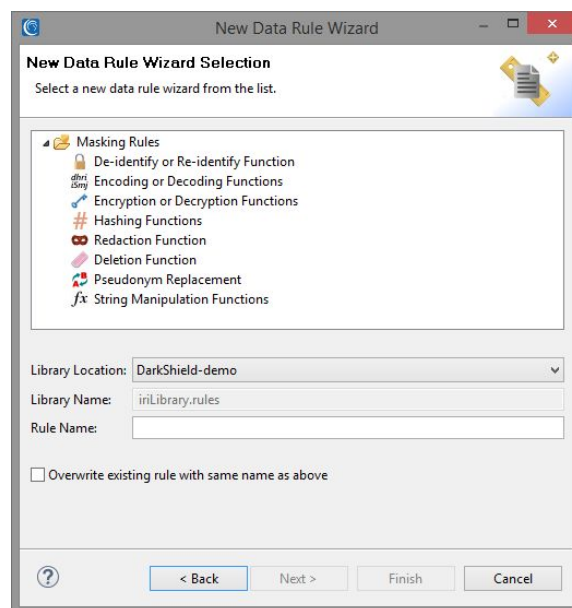
Searching and masking operations can be combined or performed separately, from a GUI wizard, batch job (CLI), API call, or web service. DarkShield can also extract and leverage search results and attendant metadata in HTML5 charts, delimited and JSON logs, DSARs, textual ETL, and SIEM/SOC tools.

DarkShield Masking Functions

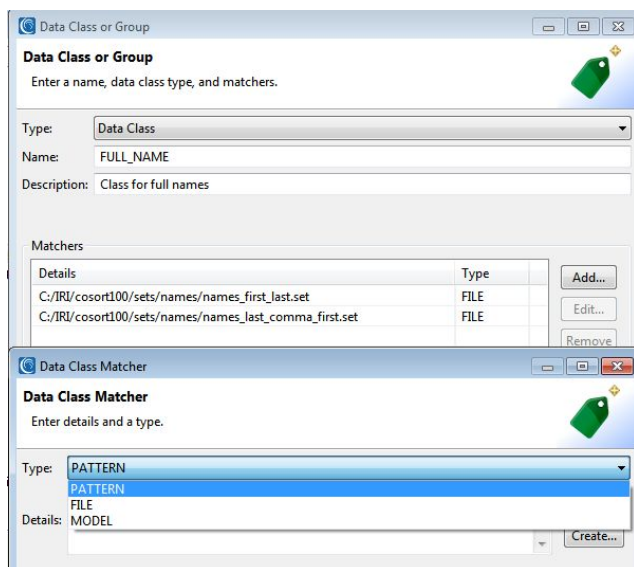
DarkShield users can employ many of the same [static data masking functions](#) that IRI FieldShield does to secure PII in DB and flat-file columns.

Commonly used DarkShield functions include:

- Format-preserving (or not) encryption
- Lookup pseudonymization
- Redaction / obfuscation
- String manipulation
- Deletion / erasure
- Scrambling
- Encoding
- Hashing



The masking rule you match to each data class should depend on the desired results for the ciphertext; i.e., whether they can be reversed, how they appear (conform to format constraints), and if they must be unique values. DarkShield can replace existing or create new files in current or cloned folder structures.



DarkShield Business Benefits

Only DarkShield supports the combination and automation of the difficult but necessary processes of [data classification](#), discovery, extraction, masking, and audit reporting across multiple data sources and silos.

Multiple search methods and threads are deployed in conjunction with multiple data masking functions. All this optimization and consolidation speeds compliance efforts, application deployment, systems testing, and document management activity.

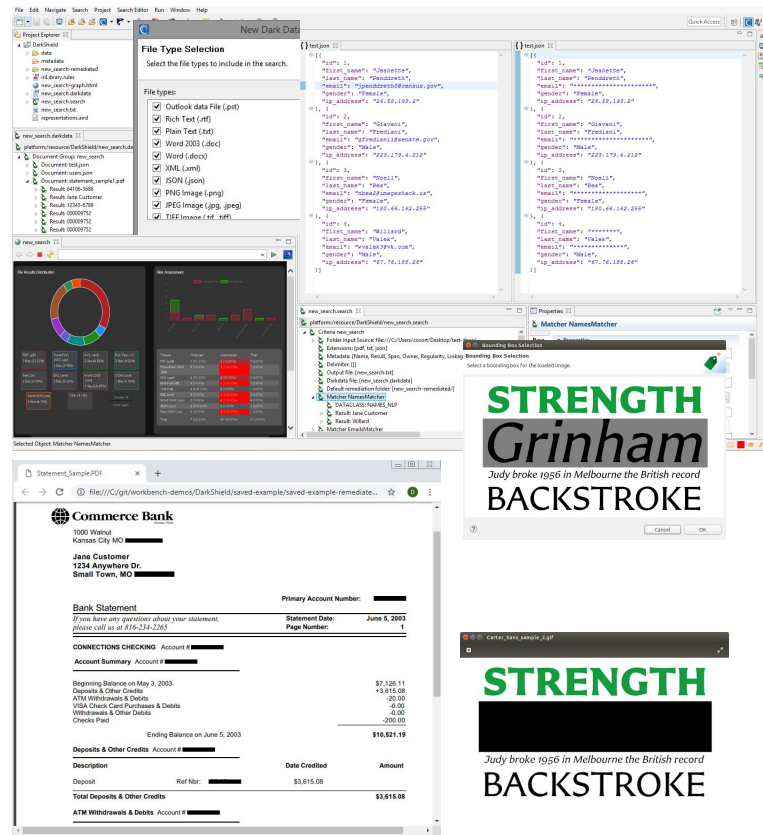
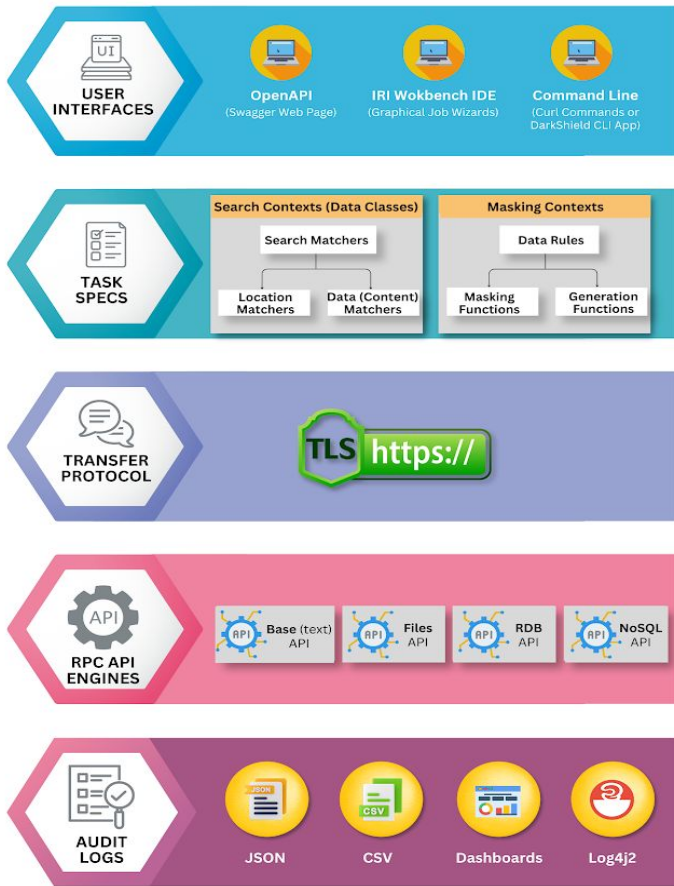
DarkShield also:

- helps GDPR data collectors and processors comply with data portability, rectification, and right-to-be-forgotten provisions.
- operates in the same free Eclipse™ job design and metadata environment, called IRI Workbench, with FieldShield and other data governance and management functions.
- is affordable for perpetual use standalone, in a bundle with other [IRI Data Protector](#) suite products, or for “free” inside IRI Voracity total data management platform subscriptions.

Supported Approaches and Data Sources



Version 5
Architecture



DarkShield API users can also find and mask PII in other sources through "glue code" in any language.

Compatible Platforms and Applications

DarkShield runs on Windows, Linux on-premise or in the cloud. It can reach databases and files in Amazon S3, Azure Blob, GCP, OneDrive, and SharePoint, plus LAN drives mounted through SMB.

DarkShield uses the same IRI Workbench front-end, data classes, and masking functions as:

- IRI FieldShield - DB and flat-file masking
- IRI CellShield - Excel spreadsheet masking
- IRI CoSort - Data transformation and reporting
- IRI Voracity - Big data integration, test data, etc.



© 2025 Innovative Routines International (IRI), Inc. All Rights Reserved. CoSort, Voracity, NextForm, FieldShield, CellShield, DarkShield, and RowGen are trademarks or registered trademarks of IRI. FACT is a trademark of DataStreams Corp. (CoSort Korea). Other product, brand, or company names may be (registered) trademarks of their respective holders.

2194 Highway A1A
Melbourne, FL 32937 USA
1.321.777.8889 * 1.800.333.SORT



iri.com/darkshield