IRI DarkShield

Unstructured Data Search & Security

Version 5

Product Overview

IRI

Total Data Management

Executive Summary

In an information technology era when both big data opportunities and privacy laws exist and converge, there is a pressing need to discover, work with, and protect data hidden in unstructured sources. Data in these silos is often referred to as dark data:

> *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 (for example, analytics, business relationships and direct monetizing). Similar to dark matter in physics, dark data often comprises most organizations' universe of information assets. Thus, organizations often retain dark data for compliance purposes only. Storing and securing data typically incurs more expense (and sometimes greater risk) than value.*

Every company and government agency collects and stores such data in logs, emails and other free text, plus documents, images, and audio/video files. Like transactional data in structured sources, the information contained in semi- and unstructured data sources carries both analytic value and business risk.

Innovative Routines International (IRI), Inc., founded 1978 and best known worldwide as The CoSort Company, expanded its high-volume, high-performance data transformation capabilities into the world of sensitive data discovery and masking in 2007.

The addition of encryption, redaction, pseudonymization, and other anonymization functions was a natural evolution of the field-level manipulations IRI software was already performing in CoSort-driven mainframe sort and data migrations, big data integration and wrangling, test data generation, custom reporting, and so on.

IRI has created fit-for-purpose data masking tools from this foundation, and has enjoyed both commercial success from them, and recognition from the data security analyst community; e.g., Gartner, which now features five IRI products in its Market Guide for Data Masking Technologies. IRI DarkShield® Version 5 is designed to reduce the cost and risk involved in finding and securing information in dark data repositories, and to help you nullify the risk of data breaches and comply with data privacy laws.

For its innovations in PII security for semi- and unstructured data in relational and NoSQL DBs as well as files, DarkShield is recognized as a trend-setting product by DBTA Magazine.

Contact Information

Innovative Routines International, Inc.
2194 Highway A1A, Suite 303
Melbourne, FL 32937 USA
Tel. +1.321.777.8889
darkshield@iri.com

IRI DarkShield
Unstructured Data Search & Security

database
TRENDS AND APPLICATIONS
TREND-SETTING
PRODUCTS

Product Introduction

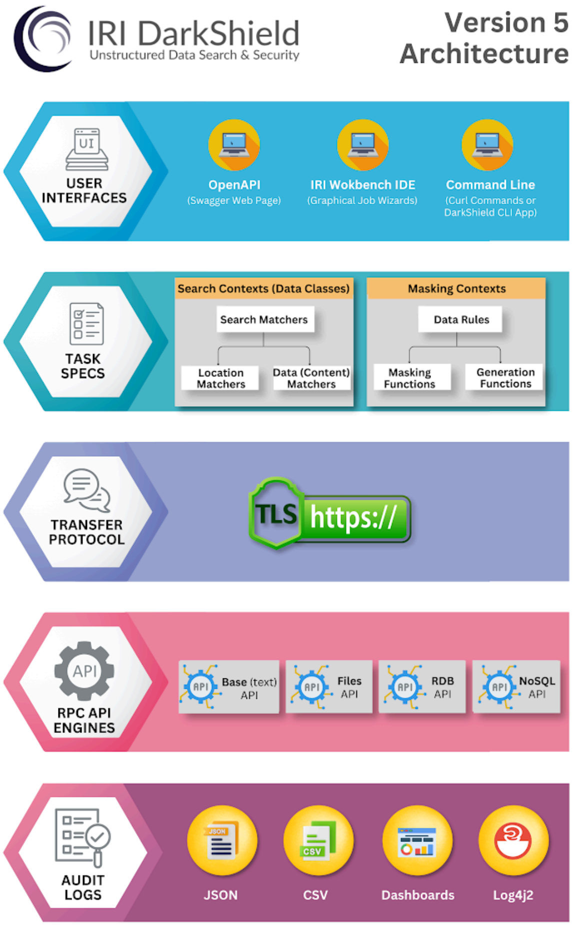
IRI DarkShield Version 5 is a software package for finding and masking Personally Identifiable Information (PII) and other sensitive data hidden in semi-structured and unstructured files, databases, and streams. It can be licensed and used standalone, or within the IRI Voracity data management platform.

DarkShield can use regular expressions, value lookups, path filters, signature detection, and Named Entity Recognition (NER) models to find and mask PII floating in: relational and NoSQL database collections; Microsoft Office documents; PDF, JSON, XML, and many EDI, log and plain-text files; plus, many image file formats through Optical Character Recognition (OCR).

In the same or a separate pass from the search operation, DarkShield can extract for delivery (data portability), mask with industry-standard protection functions, and report on the found values and their associated locational data. Supported masking functions include: redaction, encryption, pseudonymization, hashing, encoding, bit and string manipulation, random noise (blurring), scrambling, and deletion.

DarkShield jobs created in Workbench, are serialized in XMI files, and API specs in JSON to ease modification, res-use and sharing in repositories like Git. Search and masking results are in JSON, which can be audited in several ways, including: ad hoc via built-in interactive dashboards, CoSort/SortCL programs, Datadog, and Splunk ES.

DarkShield Architecture



DarkShield search and mask operations are powered by one or more RPC APIs running on Windows or Linux , on either on-premise or cloud servers in infrastructure *that you (not IRI) own and control.*

DarkShield jobs are configured through:

1) the IRI Workbench Graphical User Interface (GUI) for job design and management, built on Eclipse™

2) Swagger UI to make direct API calls via a web app

3) Custom calling program logic.

DarkShield jobs designed in Workbench also rely on the DarkShield APIs for search and masking work; calls are made to the DarkShield API based on the configurations of the job. Workbench is also where IRI FieldShield, RowGen and other SortCL-driven jobs are managed.

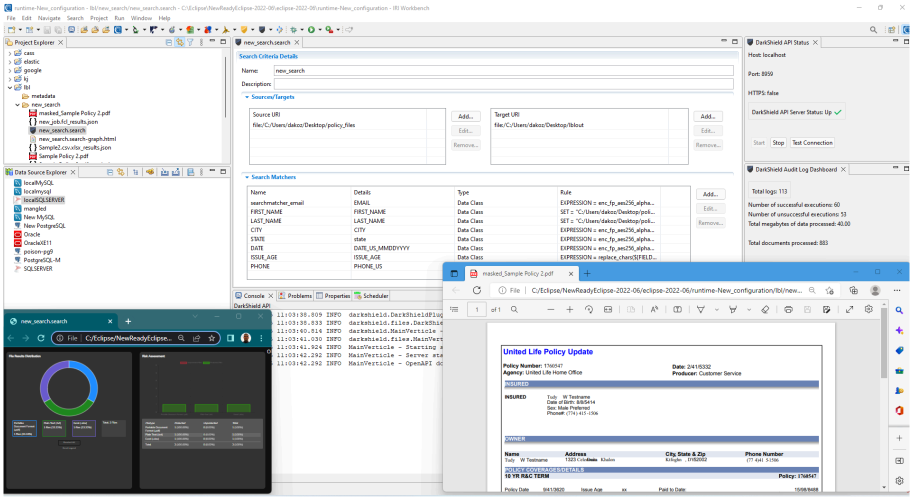
The flexibility of the DarkShield API also makes it possible to integrate with document management and other software systems. For example, DarkShield can leverage image pre-processing tools, and the NGINX reverse proxy for load balancing and user authentication.

For DarkShield users in particular, the IRI Workbench GUI for DarkShield includes:

1. The *New File Search and Masking Job* wizard for creating DarkShield jobs for unstructured text, document, and image sources
2. The *New NoSQL Search and Masking Job* wizard for creating DarkShield jobs for the NoSQL databases MongoDB, Cassandra, and Elasticsearch
3. The *New Relational Database Search and Masking Job* wizard for creating DarkShield jobs for any relational database that can be connected via JDBC
4. The *export .search files to Contexts* wizard for exporting information serialized in DarkShield jobs – such as data matchers and masking rule definitions and pairings, file configuration options, and file-type-specific filters to a set of DarkShield API contexts
5. Form editors for viewing and editing data classes, and DarkShield jobs
6. Click-to-run, run configuration dialog, and built-in task scheduler options to launch and automate DarkShield jobs that search, mask, or do both search and mask at once
7. The ability to manage (start/stop) a DarkShield API server
8. A view that displays the status of a DarkShield API server
9. A graphical view of DarkShield execution logs containing items such as job names, success status, runtimes, etc.
10. Named Entity Recognition (NER) Model wizards to leverage the power of Natural Language Processing (NLP) and Machine Learning (ML) to train and use custom NER models to identify persons, organizations, locations, etc.
11. Textual and graphical views of DarkShield search and remediation results
12. Offline and online technical documentation, learning articles and videos, and support from IRI engineers and IRI partners located in 40+ cities worldwide.

IRI Workbench and DarkShield run under Windows or Linux, on physical or virtual nodes, as well as in containers hosted on-premise or in your cloud.

DarkShield operations should be staged on a system with at least 4GB of RAM.



runtime-New_configuration - lbl/new_search/new_search.search - C:\Eclipse\NewReadyEclipse-2022-06\eclipse-2022-06\runtime-New_configuration - IRI Workbench
File Edit Navigate Search Project Run Window Help
Project Explorer
Data Source Explorer
new_search.search
Console
Problems
Properties
Scheduler
masked_Sample Policy 2.pdf

Search Criteria Details

Name: new_search

Description:

Sources/Targets

Source URI: file:/C:/Users/dakoz/Desktop/policy_files

Target URI: file:/C:/Users/dakoz/Desktop/lblout

Search Matchers

Name	Details	Type	Rule
searchmatcher_email	EMAIL	Data Class	EXPRESSION = enc_fp_aes256_alpha...
FIRST_NAME	FIRST_NAME	Data Class	SET = "C:/Users/dakoz/Desktop/poli...
LAST_NAME	LAST_NAME	Data Class	SET = "C:/Users/dakoz/Desktop/poli...
CITY	CITY	Data Class	EXPRESSION = enc_fp_aes256_alpha...
STATE	state	Data Class	EXPRESSION = enc_fp_aes256_alpha...
DATE	DATE_US_MMDDYYYY	Data Class	EXPRESSION = enc_fp_aes256_alpha...
ISSUE_AGE	ISSUE_AGE	Data Class	EXPRESSION = replace_chars(${FIELD...
PHONE	PHONE_US		

DarkShield API Status

Host: localhost

Port: 8959

HTTPS: false

DarkShield API Server Status: Up ✓

Start Stop Test Connection

DarkShield Audit Log Dashboard

Total logs: 113

Number of successful executions: 60

Number of unsuccessful executions: 53

Total megabytes of data processed: 40.00

Total documents processed: 883

DarkShield API

```
11:03:38.809 INFO  darkshield.DarkShieldPlug
11:03:38.833 INFO  darkshield.files.DarkShie
11:03:40.814 INFO  darkshield.MainVerticle -
11:03:41.030 INFO  darkshield.files.MainVert
11:03:41.924 INFO  MainVerticle - Starting s
11:03:42.292 INFO  MainVerticle - Server sta
11:03:42.292 INFO  MainVerticle - OpenAPI do
```

C:/Eclipse/NewReadyEclipse-2022-06/eclipse-2022-06/runtime-New_configuration/lbl/new...

United Life Policy Update

Policy Number: 1760547
Agency: United Life Home Office

Date: 2/41/5332
Producer: Customer Service

The use of DarkShield features are outlined and explained in further detail below.

DarkShield Workflow (GUI)

These steps describe the most common (but not the only) way DarkShield is used:

Download & Installation

1. ***Install and Configure***. Obtain and open IRI Workbench, and license the back-end data masking executable(s) per the IRI installation guide. Activate the DarkShield API from *Window > Preferences > IRI > DarkShield*. Paste *C:\IRI\DarkShield\API\plankton-1.5.1* into the folder location for the DarkShield API distribution. Click Start Server, then Apply and Close. Gather details on your sources and targets and JDBC drivers for any RDBs to mask.

Data Classification

2. ***Classify Your Data***. Define Data Classes (e.g. names, phone numbers, PINs) and Groups (e.g., PHI) which require masking from the Data Discovery menu; see this article. You can associate each class or group with a search method or methods (pattern, lookup value, NER model matches, etc.), and if desired, run a Search-only operation for the PII.
.

Masking Rules

3. ***Define Masking & Sensitivity Rules***. Accept default rules in the library or assign masking functions to each data class and/or group per the above. You can also associate sensitivity levels to those classes and groups where masking rules can be different.

Data Sources

4. ***Specify Your Sources/Targets****.* In either your DarkShield API calling program (a/k/a "glue code"), *New Job* wizard, or job configuration file in Workbench, provide the location of the source or target file folders or DBs in your LAN or cloud store(s).

Rule Matchers

5. ***Create or Use Search Matchers***. The mapping between Data Rules (masking functions) and Data Classes/Groups in Workbench happens through Search Matchers. To create these matchers, browse to an existing Data Rule created in Step 3 above, or create a new one and associate it with a Data Class or Group in the Dark Data Discovery Wizard. If you are using the DarkShield API, see how Search and Mask Contexts build and co-relate.

Job Execution

6. ***Run the Job****.* When you click Finish in one of the New DarkShield Job wizards, your specifications serialize into a *.dsc* (DarkShield Configuration) file in your project folder. You can run the job from that .dsc file directly, the Run Configuration menu, the built-in task scheduler, or later from the DarkShield CLI or API outside Workbench. If you specify a search-only job, the search results (including PII found) get logged in JSON annotation files which can be re-run as needed, analyzed or used in a separate mask-only job later. Masking jobs affect all associated PII found in the search, remove PII from the annotation file, and write masked data to new target folders or DB collections with the same formats and names.

Job Review

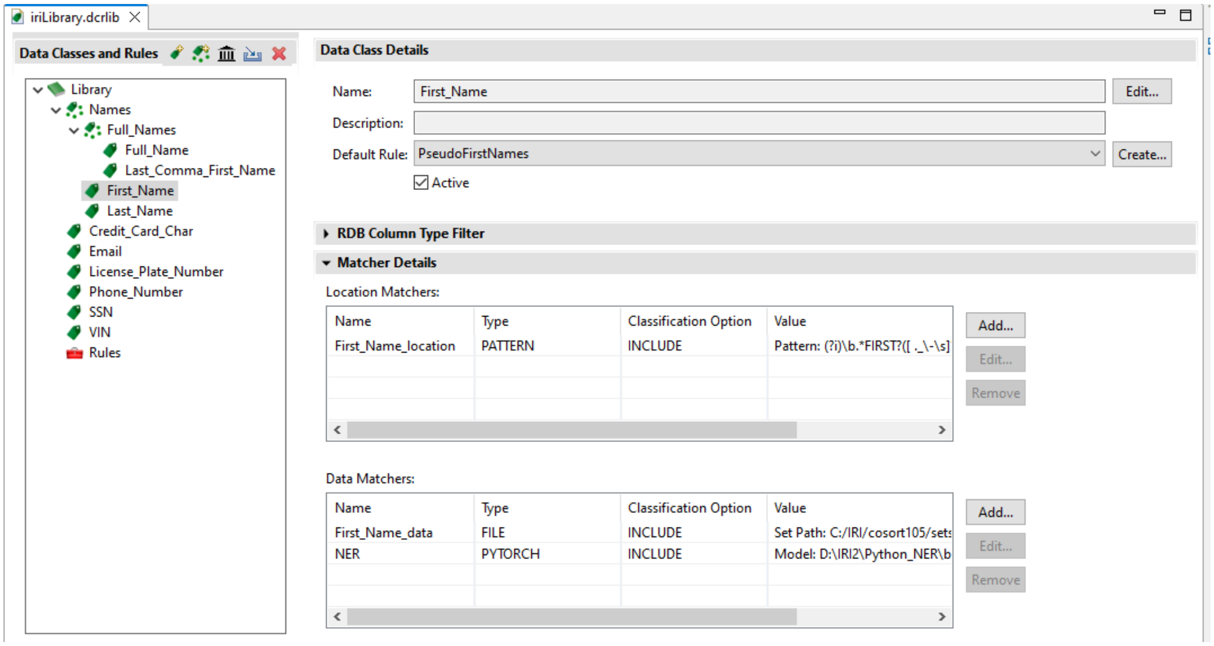
7. ***Review the Results***. DarkShield jobs in Workbench are logged for immediate review or export to analytic tools: a delimited file containing matched data in files and the metadata of those files, plus JSON annotation files with search results and locations from all sources. DarkShield also builds customizable HTML5 charts ranking data by silo and class and whether it was masked. You can also verify what was masked by examining the job targets.

Job Scheduling

8. ***Automate the Job****.* Once you are familiar with running DarkShield and comfortable with the results it produces, you can rerun jobs in the Workbench or other scheduler, in a DevOps pipeline or via CLI. Each time a job runs, it will do the same searching and masking on new data in your sources, or it will re-scan and mask updates since the last search.

Data Classification

DarkShield uses the built-in data classification facilities in IRI Workbench to define and catalog one or more items of PII or other data which you classify as sensitive, typically just once.



Name: First_Name

Description:

Default Rule: PseudoFirstNames

☑ Active

Location Matchers:

Name	Type	Classification Option	Value
First_Name_location	PATTERN	INCLUDE	Pattern: (?i)\b.*FIRST?([._\-\s]

Data Matchers:

Name	Type	Classification Option	Value
First_Name_data	FILE	INCLUDE	Set Path: C:/IRI/cosort105/sets
NER	PYTORCH	INCLUDE	Model: D:\IRI2\Python_NER\b

These items are organized into **Data Classes** or **Data Class Groups**, which will be discovered through Data content and/or Location (structure) *Search Matchers* – and then masked per a *Rule* (i.e., data masking or generation function) – that you assign to each class or group.

The search matchers you can use (one or more in combination) to find classified data include:

1. **Pattern** matchers. Strings conforming to IRI-supplied or custom-defined Java Regular Expression (Regex) patterns, which are ideal for ID and phone numbers, email addresses and credit cards. These Regex searches can also be computationally verified with IRI-supplied or custom-written JavaScript validation routines to avoid false positives.

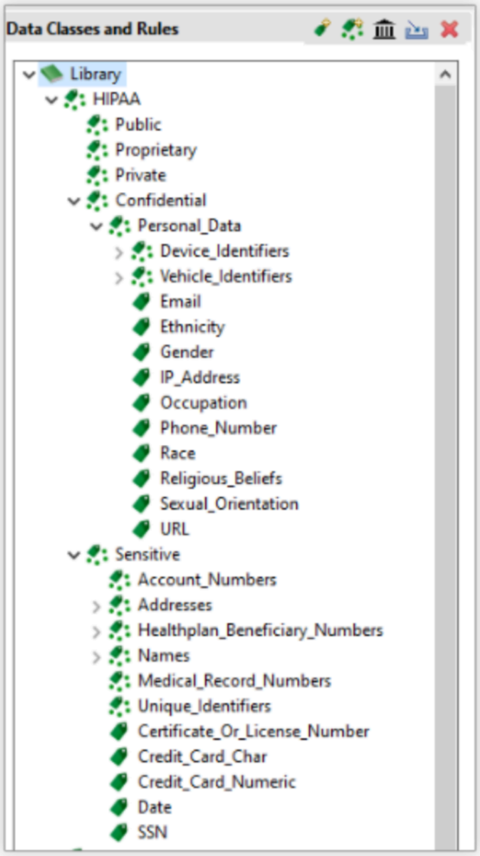
2. **Dictionary** matchers. Exact or fuzzy matches to strings in a lookup (set) file or table.

3. **Location** matchers; i.e., specified 'path' or column filters for JSON, XML, CSV, Excel

4. Named-Entity Recognition (**NER**) matchers. Use or train machine-learning Natural Language Processing (NLP) or TensorFlow/PyTorch models to words in context.

5. Bounding boxes which define specific, repeated **regions** within images to mask.

Note that the same Data Classes and Groups, as well as most of the masking rules you assign to them, are also used in IRI FieldShield and IRI CellShield EE. However, only DarkShield supports NER matchers.



Data Classes and Rules

- Library
  - HIPAA
    - Public
    - Proprietary
    - Private
    - Confidential
      - Personal_Data
        - Device_Identifiers
        - Vehicle_Identifiers
        - Email
        - Ethnicity
        - Gender
        - IP_Address
        - Occupation
        - Phone_Number
        - Race
        - Religious_Beliefs
        - Sexual_Orientation
        - URL
    - Sensitive
      - Account_Numbers
      - Addresses
      - Healthplan_Beneficiary_Numbers
      - Names
      - Medical_Record_Numbers
      - Unique_Identifiers
      - Certificate_Or_License_Number
      - Credit_Card_Char
      - Credit_Card_Numeric
      - Date
      - SSN

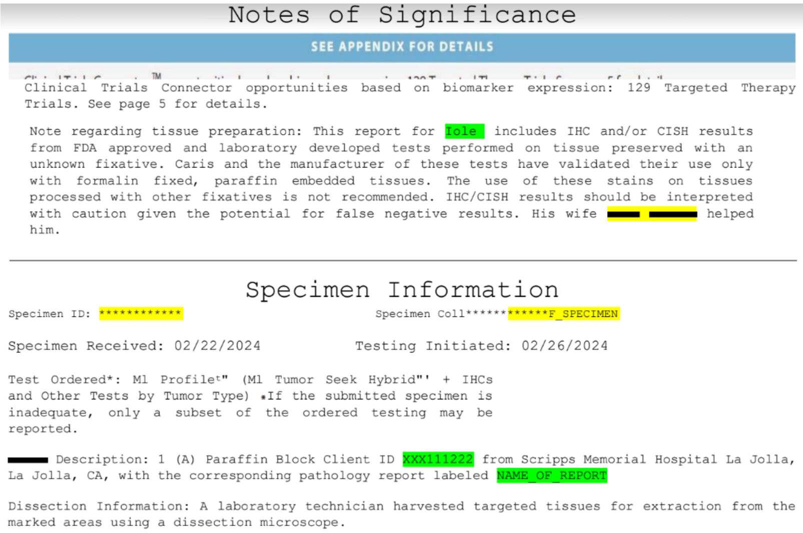
An optional feature of Data Class Groups is the ability to further categorize Data Classes according to their level of sensitivity. *Sensitivity Level Groups* are Data Class Groups with an assigned priority level. Higher sensitivity (priority) groups would typically have more restrictive masking functions assigned to them.

DarkShield also ships with some default Data Class Groups categorized as *Privacy Law Groups*. Privacy Law Groups are pre-populated Data Class Groups that provide a launching board for business rules to adhere to different privacy law requirements. These privacy law groups have pre-populated data classes, search matchers, and masking functions.

For example, to facilitate compliance with the GDPR and laws based on the GDPR, DarkShield provides a number of international patterns for direct identifiers, as well as set files containing first and last names – as well as labor unions and ethnicities – across many different countries.

Note however that IRI provides such out-of-the-box configurations for convenience; i.e., they may not identify every element, or conform to your specific data masking requirements. IRI thus recommends that you review and modify these settings to make sure they address your needs.

In general, DarkShield (and FieldShield) can facilitate GDPR compliance by finding, classifying, extracting, correcting, erasing, and otherwise anonymizing PII through multiple functions like pseudonymization, deletion, encryption, and redaction. It can find, extract, and replace PII to meet portability and rectification requirements, plus score re-identification risk and anonymize quasi-identifiers to comply with Article 29 provisions and the HIPAA EDM security rule.



Notes of Significance

SEE APPENDIX FOR DETAILS

Clinical Trials Connector opportunities based on biomarker expression: 129 Targeted Therapy Trials. See page 5 for details.

Note regarding tissue preparation: This report for Iole includes IHC and/or CISH results from FDA approved and laboratory developed tests performed on tissue preserved with an unknown fixative. Caris and the manufacturer of these tests have validated their use only with formalin fixed, paraffin embedded tissues. The use of these stains on tissues processed with other fixatives is not recommended. IHC/CISH results should be interpreted with caution given the potential for false negative results. His wife [redacted] [redacted] helped him.

Specimen Information

Specimen ID: ************   Specimen Coll************F_SPECIMEN

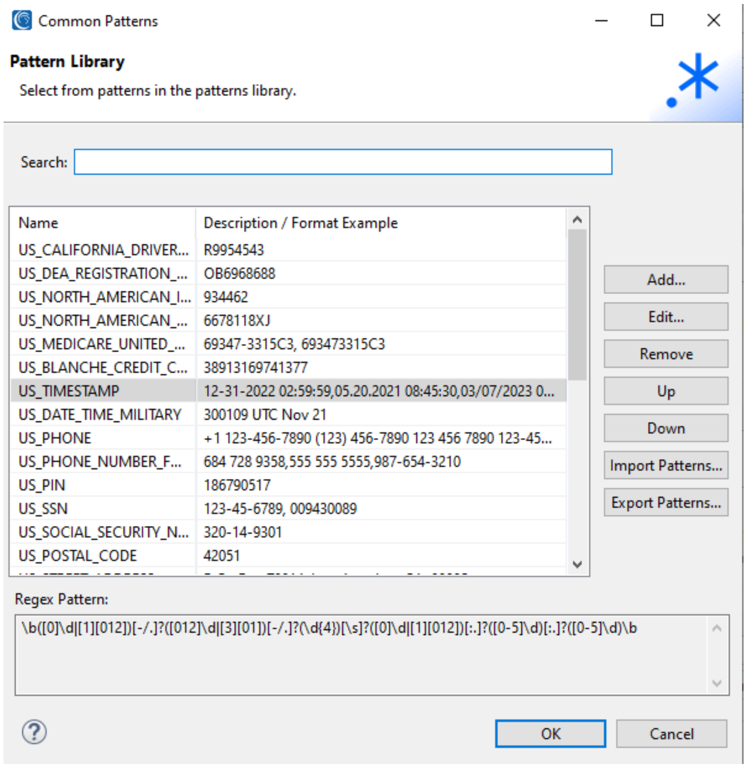
Specimen Received: 02/22/2024   Testing Initiated: 02/26/2024

Test Ordered*: Ml Profileᵗ" (Ml Tumor Seek Hybrid"' + IHCs and Other Tests by Tumor Type) *If the submitted specimen is inadequate, only a subset of the ordered testing may be reported.

[redacted] Description: 1 (A) Paraffin Block Client ID XXX111222 from Scripps Memorial Hospital La Jolla, La Jolla, CA, with the corresponding pathology report labeled NAME_OF_REPORT

Dissection Information: A laboratory technician harvested targeted tissues for extraction from the marked areas using a dissection microscope.

Data Class Search Matchers



Pattern Library

Select from patterns in the patterns library.

Search:

Name	Description / Format Example
US_CALIFORNIA_DRIVER...	R9954543
US_DEA_REGISTRATION_...	OB6968688
US_NORTH_AMERICAN_I...	934462
US_NORTH_AMERICAN_...	6678118XJ
US_MEDICARE_UNITED_...	69347-3315C3, 693473315C3
US_BLANCHE_CREDIT_C...	38913169741377
US_TIMESTAMP	12-31-2022 02:59:59,05.20.2021 08:45:30,03/07/2023 0...
US_DATE_TIME_MILITARY	300109 UTC Nov 21
US_PHONE	+1 123-456-7890 (123) 456-7890 123 456 7890 123-45...
US_PHONE_NUMBER_F...	684 728 9358,555 555 5555,987-654-3210
US_PIN	186790517
US_SSN	123-45-6789, 009430089
US_SOCIAL_SECURITY_N...	320-14-9301
US_POSTAL_CODE	42051

Regex Pattern:

```
\b([0]\d|[1][012])[-/.]?([012]\d|[3][01])[-/.]?(\d{4})[\s]?([0]\d|[1][012])[:.]?([0-5]\d)[:.]?([0-5]\d)\b
```

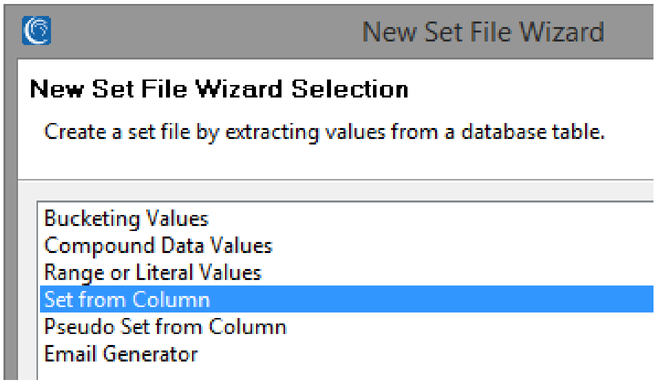
Regular Expression Patterns

DarkShield can use any Java regular expression (RegEx) to find PII data that conforms to a well-defined format (email address, credit card number, etc.), along with further validation logic in Javascript to prevent false positives in search/mask jobs.

IRI Workbench ships with many common patterns, and allows you to create and save your own patterns for re-use in other IRI data classification, searching and masking wizards and projects, too.

Set File Lookups

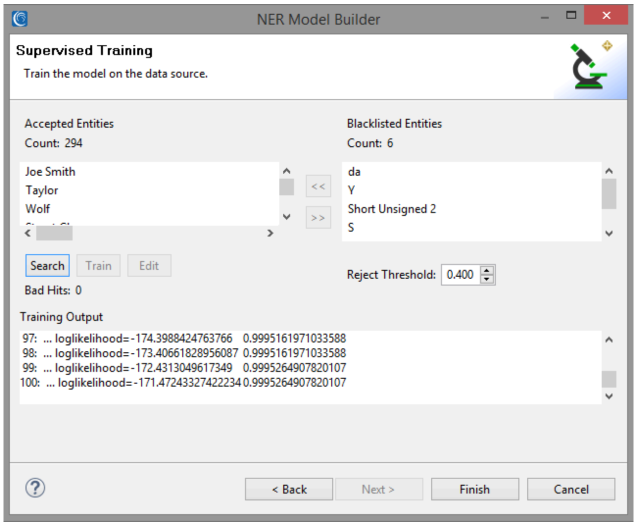
DarkShield supports the use of exact or fuzzy matches to values in external, tab-delimited files. IRI provides many lookup international files for you, including names and demographic indicators. You can also provide your own set files, create them in IRI Workbench set file wizards, or build them from database column values via JDBC.



New Set File Wizard Selection

Create a set file by extracting values from a database table.

- Bucketing Values
- Compound Data Values
- Range or Literal Values
- Set from Column
- Pseudo Set from Column
- Email Generator



Supervised Training

Train the model on the data source.

Accepted Entities
Count: 294

Joe Smith
Taylor
Wolf

Blacklisted Entities
Count: 6

da
Y
Short Unsigned 2
S

Bad Hits: 0

Reject Threshold: 0.400

Training Output

```
97:  ... loglikelihood=-174.3988424763766   0.9995161971033588
98:  ... loglikelihood=-173.40661828956087 0.9995161971033588
99:  ... loglikelihood=-172.4313049617349   0.9995264907820107
100:  ... loglikelihood=-171.47243327422234 0.9995264907820107
```

NER Models

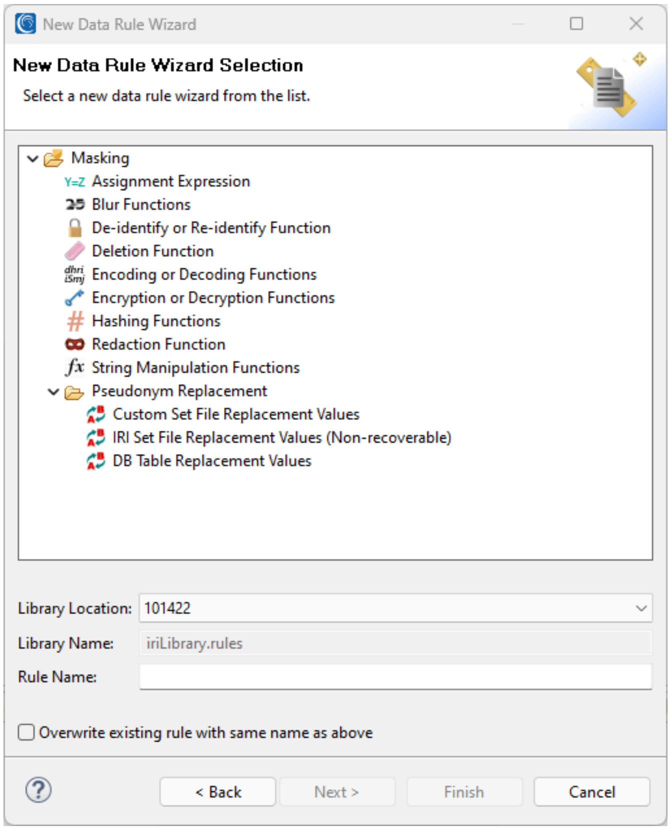
DarkShield also supports the use of any OpenNLP, Tensorflow and PyTorch Named Entity Recognition (NER) models to find (and mask) names and other words using sentence grammar and structure. In cases where the available models do not provide accurate results after searching your documents, semi-supervised machine-learning modules in IRI Workbench can help train your models.

DarkShield Masking Rules

DarkShield applies masking functions by using data rules. Data rules are created and stored for future use and modification in an IRI Data Class Rule (.dcrlib) library stored in an IRI Workbench project folder.

These Data Rules can be paired with Data Classes when creating a DarkShield job. The Data Class to Rule Mapping are then used to consistently mask the discovered PII via:

1. SHA-1 and SHA-2 hashing
2. multiple, NSA Suite B and FIPS-compliant encryption
3. ASCII de-ID (bit scrambling)
4. binary encoding
5. deletion (erasure / removal)
6. redaction (full or partial string masking)
7. lookup value pseudonymization
8. byte shifting and (sub)string functions



New Data Rule Wizard

New Data Rule Wizard Selection

Select a new data rule wizard from the list.

- Masking
  - Assignment Expression
  - Blur Functions
  - De-identify or Re-identify Function
  - Deletion Function
  - Encoding or Decoding Functions
  - Encryption or Decryption Functions
  - Hashing Functions
  - Redaction Function
  - String Manipulation Functions
  - Pseudonym Replacement
    - Custom Set File Replacement Values
    - IRI Set File Replacement Values (Non-recoverable)
    - DB Table Replacement Values

Library Location: 101422

Library Name: iriLibrary.rules

Rule Name:

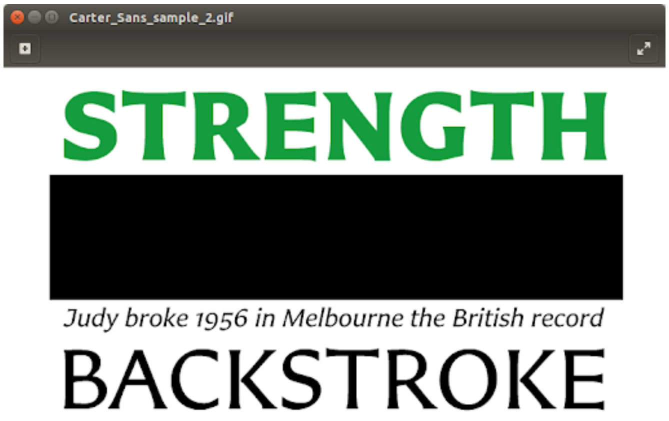
☐ Overwrite existing rule with same name as above

Restoring masked data to its original state is possible under certain circumstances. The data must have been masked using a reversible function like encryption or two-column (restore-set) pseudonymization, and the data must be discoverable using location (i.e., not data) matchers. See the API example provided in the article, "Restoring Masked Values with IRI DarkShield."

Bounding Boxes

DarkShield supports the definition of regions within image files to mask. This is especially useful if other PII discovery (search) methods do not work, and the area in which the PII exists in one or more files with a common layout (e.g. image location) is known.

A user-friendly area drawing tool allows the definition of "bounding boxes" around the content you want redacted in each file.



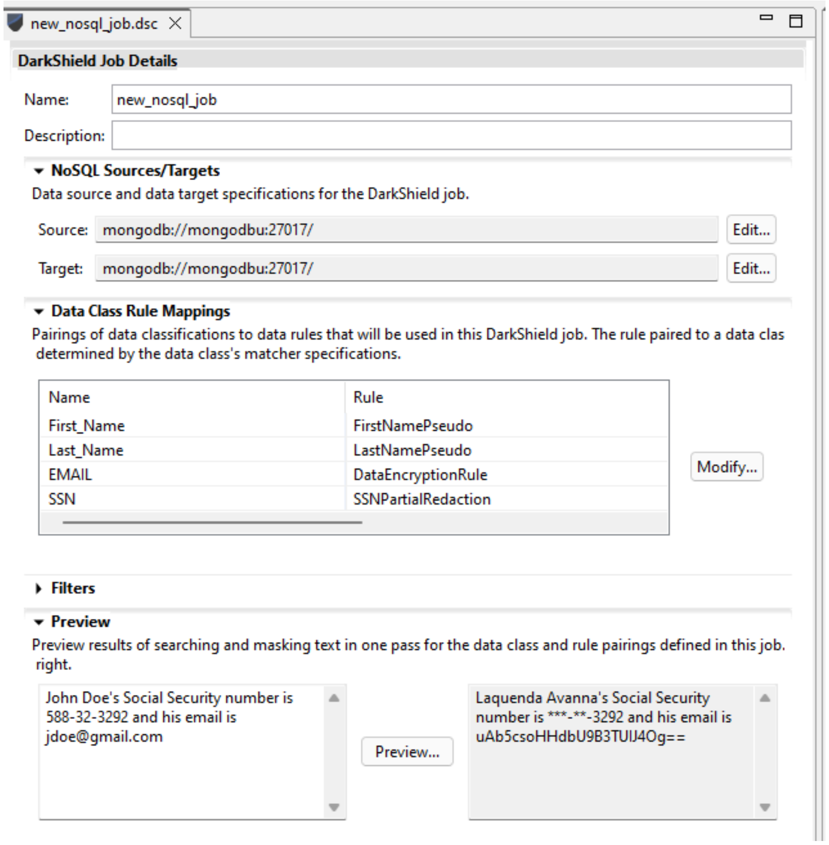
Running DarkShield Jobs in the Graphical User Interface (GUI)

PII searching and masking jobs can be designed, managed, and run from IRI Workbench. Jobs can perform searching and masking in the same pass by running a “search & mask” job from the *.search* configuration file, or separately by first running a “search” job and then running a “mask” job generated from the search.

You can save the configuration for use in either ad hoc or scheduled executions. Repeated DarkShield runs can detect changes in files that were previously searched on subsequent runs, and repeat the search.

Read the New File Search/Masking Job … wizard article to learn how to find and mask the PII within different file formats – including MS Office, PDF and image files – and silos all at once.

If you have PII in a MongoDB, Cassandra or Elasticsearch NoSQL database, read about the NoSQL wizard in Workbench. Note that it is also possible to use the DarkShield API with glue code to search and mask PII in at least 7 more NoSQL DBs; find the sample projects here.



DarkShield Job Details

Name: new_nosql_job

Description:

NoSQL Sources/Targets

Data source and data target specifications for the DarkShield job.

Source: mongodb://mongodbu:27017/ Edit...

Target: mongodb://mongodbu:27017/ Edit...

Data Class Rule Mappings

Pairings of data classifications to data rules that will be used in this DarkShield job. The rule paired to a data clas determined by the data class's matcher specifications.

Name	Rule
First_Name	FirstNamePseudo
Last_Name	LastNamePseudo
EMAIL	DataEncryptionRule
SSN	SSNPartialRedaction

Filters

Preview

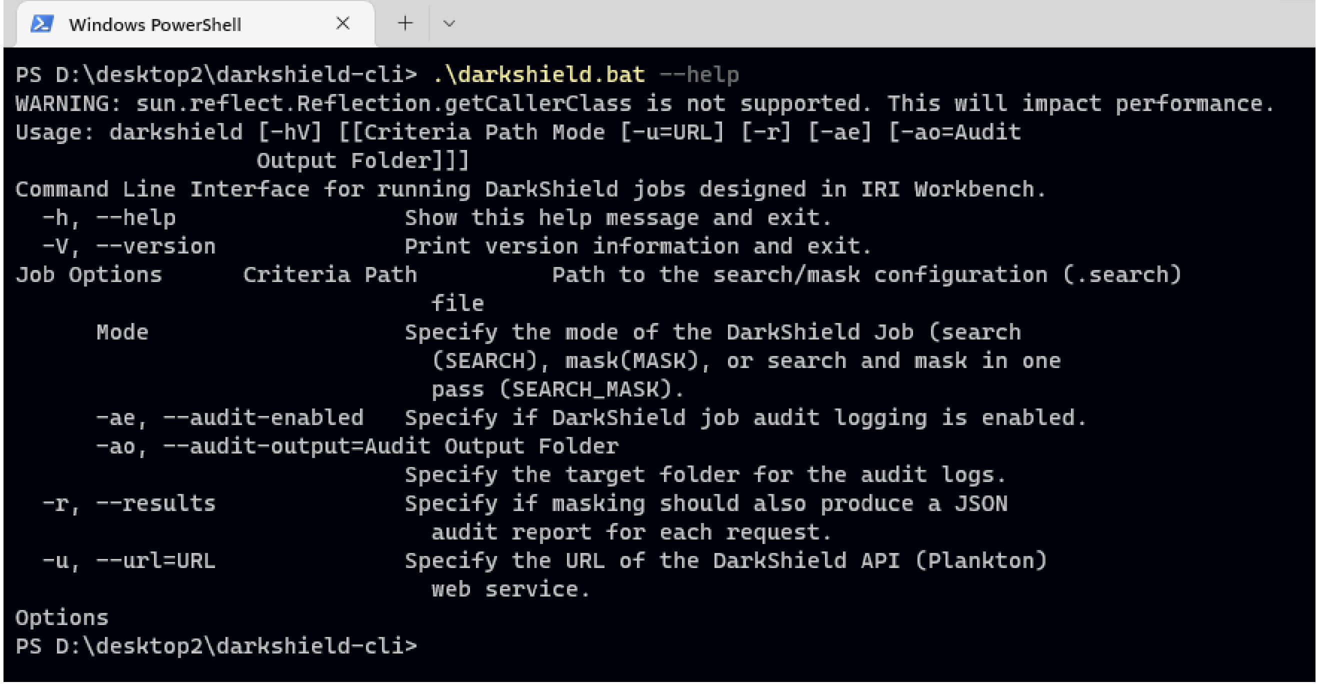
Preview results of searching and masking text in one pass for the data class and rule pairings defined in this job. right.

John Doe's Social Security number is 588-32-3292 and his email is jdoe@gmail.com

Laquenda Avanna's Social Security number is ***-**-3292 and his email is uAb5csoHHdbU9B3TUIJ4Og==

For relational database sources, DarkShield will support any instance that can be accessed via a JDBC connection. See the JDBC sections of these articles for DB-specific connection advice, and the DarkShield RDB wizard article to learn more about working with RDB sources.

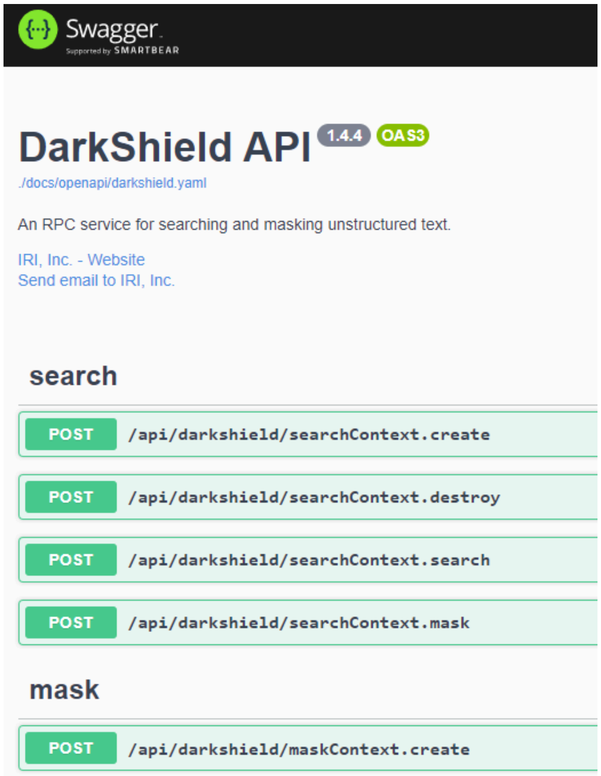
Command Line Interface (CLI)



```
PS D:\desktop2\darkshield-cli> .\darkshield.bat --help
WARNING: sun.reflect.Reflection.getCallerClass is not supported. This will impact performance.
Usage: darkshield [-hV] [[Criteria Path Mode [-u=URL] [-r] [-ae] [-ao=Audit
                  Output Folder]]]
Command Line Interface for running DarkShield jobs designed in IRI Workbench.
  -h, --help                Show this help message and exit.
  -V, --version             Print version information and exit.
Job Options      Criteria Path          Path to the search/mask configuration (.search)
                              file
      Mode                  Specify the mode of the DarkShield Job (search
                              (SEARCH), mask(MASK), or search and mask in one
                              pass (SEARCH_MASK).
      -ae, --audit-enabled   Specify if DarkShield job audit logging is enabled.
      -ao, --audit-output=Audit Output Folder
                            Specify the target folder for the audit logs.
  -r, --results             Specify if masking should also produce a JSON
                              audit report for each request.
  -u, --url=URL             Specify the URL of the DarkShield API (Plankton)
                              web service.
Options
PS D:\desktop2\darkshield-cli>
```

The DarkShield CLI runs DarkShield jobs from outside IRI Workbench, via other programs in server environments with a Java runtime.

Remote Procedure Call (RPC) APIS



DarkShield API 1.4.4 OAS3

./docs/openapi/darkshield.yaml

An RPC service for searching and masking unstructured text.

IRI, Inc. - Website
Send email to IRI, Inc.

search

POST /api/darkshield/searchContext.create

POST /api/darkshield/searchContext.destroy

POST /api/darkshield/searchContext.search

POST /api/darkshield/searchContext.mask

mask

POST /api/darkshield/maskContext.create

The base DarkShield API, as well as with the DarkShield-Files API, DarkShield-RDB API and DarkShield-NoSQL API, allow your application programs and web forms to leverage DarkShield's many search methods and masking functions for streaming and static sources in a virtually unlimited range of formats and systems (subject to "glue code" customizations).

Embedding this functionality allows you to bypass IRI Workbench and deploy DarkShield in more automated, and orchestrated, environments on-premise or in the cloud.

Note that you can also export these API contexts directly from IRI Workbench configurations as well.

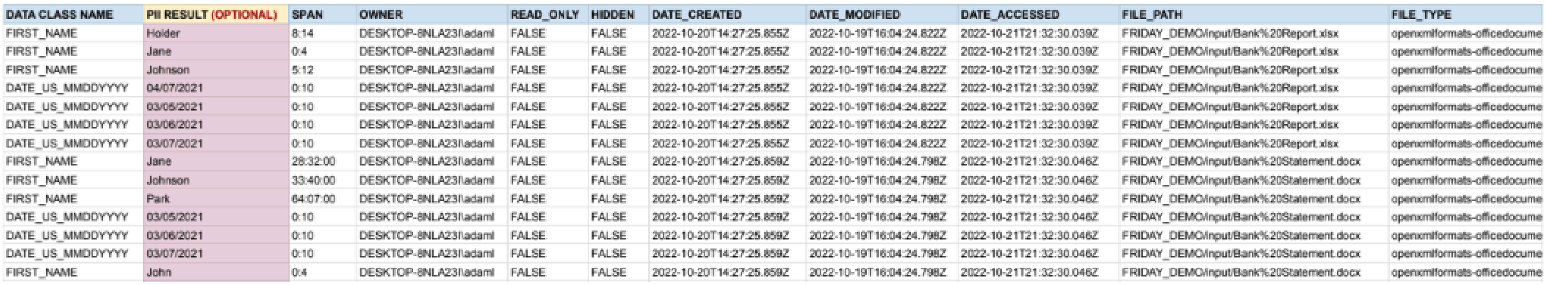
Log Reporting and Using Results

When DarkShield runs, it produces several logs which can be reviewed for audit purposes, and compliance with GDPR SAR provisions. One such log is a delimited file containing search results and any pre-selected metadata information on source *files* containing PII within.

DarkShield can also create a Data Definition File (DDF), or metadata repository defining the fields you picked for that search file. In the same Workbench GUI environment, you can build IRI CoSort (SortCL) jobs – using the DDF field layouts in the job scripts – to create custom reports.

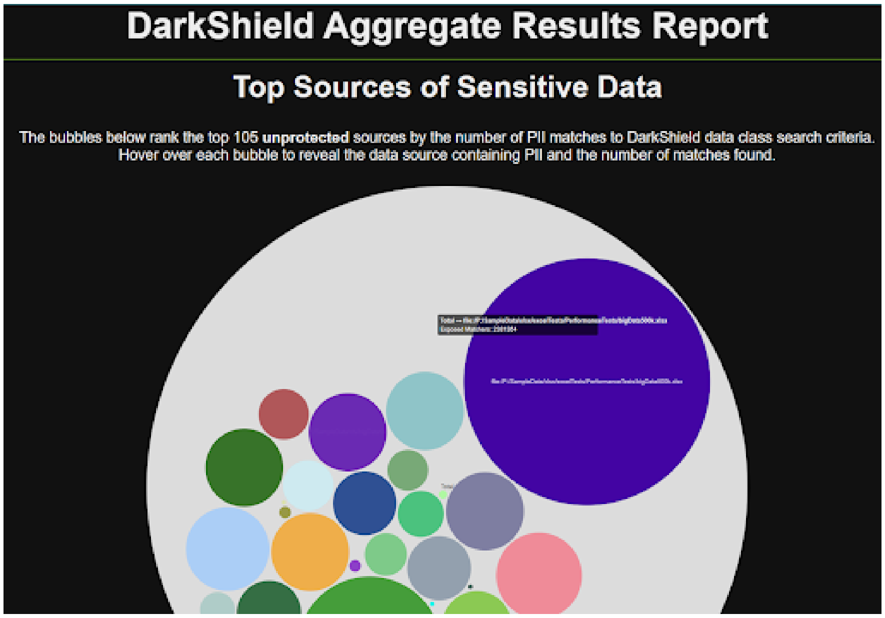
With or without a query tool like CoSort, you can extract, delete, and/or furnish specific PII values to auditors. For GDPR compliance, you can also provide the results of individual name searches to those requesting "data portability" and "the right to be forgotten." You will be able to show them what data about them was found, and what data was deleted.

Here is an example of a DDF-defined report from DarkShield with source metadata:



DATA CLASS NAME	PII RESULT (OPTIONAL)	SPAN	OWNER	READ_ONLY	HIDDEN	DATE_CREATED	DATE_MODIFIED	DATE_ACCESSED	FILE_PATH	FILE_TYPE
FIRST_NAME	Holder	8:14	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.855Z	2022-10-19T16:04:24.822Z	2022-10-21T21:32:30.039Z	FRIDAY_DEMO/input/Bank%20Report.xlsx	openxmlformats-officedocume
FIRST_NAME	Jane	0:4	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.855Z	2022-10-19T16:04:24.822Z	2022-10-21T21:32:30.039Z	FRIDAY_DEMO/input/Bank%20Report.xlsx	openxmlformats-officedocume
FIRST_NAME	Johnson	5:12	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.855Z	2022-10-19T16:04:24.822Z	2022-10-21T21:32:30.039Z	FRIDAY_DEMO/input/Bank%20Report.xlsx	openxmlformats-officedocume
DATE_US_MMDDYYYY	04/07/2021	0:10	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.855Z	2022-10-19T16:04:24.822Z	2022-10-21T21:32:30.039Z	FRIDAY_DEMO/input/Bank%20Report.xlsx	openxmlformats-officedocume
DATE_US_MMDDYYYY	03/05/2021	0:10	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.855Z	2022-10-19T16:04:24.822Z	2022-10-21T21:32:30.039Z	FRIDAY_DEMO/input/Bank%20Report.xlsx	openxmlformats-officedocume
DATE_US_MMDDYYYY	03/06/2021	0:10	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.855Z	2022-10-19T16:04:24.822Z	2022-10-21T21:32:30.039Z	FRIDAY_DEMO/input/Bank%20Report.xlsx	openxmlformats-officedocume
DATE_US_MMDDYYYY	03/07/2021	0:10	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.855Z	2022-10-19T16:04:24.822Z	2022-10-21T21:32:30.039Z	FRIDAY_DEMO/input/Bank%20Report.xlsx	openxmlformats-officedocume
FIRST_NAME	Jane	28:32:00	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.859Z	2022-10-19T16:04:24.798Z	2022-10-21T21:32:30.046Z	FRIDAY_DEMO/input/Bank%20Statement.docx	openxmlformats-officedocume
FIRST_NAME	Johnson	33:40:00	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.859Z	2022-10-19T16:04:24.798Z	2022-10-21T21:32:30.046Z	FRIDAY_DEMO/input/Bank%20Statement.docx	openxmlformats-officedocume
FIRST_NAME	Park	64:07:00	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.859Z	2022-10-19T16:04:24.798Z	2022-10-21T21:32:30.046Z	FRIDAY_DEMO/input/Bank%20Statement.docx	openxmlformats-officedocume
DATE_US_MMDDYYYY	03/05/2021	0:10	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.859Z	2022-10-19T16:04:24.798Z	2022-10-21T21:32:30.046Z	FRIDAY_DEMO/input/Bank%20Statement.docx	openxmlformats-officedocume
DATE_US_MMDDYYYY	03/06/2021	0:10	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.859Z	2022-10-19T16:04:24.798Z	2022-10-21T21:32:30.046Z	FRIDAY_DEMO/input/Bank%20Statement.docx	openxmlformats-officedocume
DATE_US_MMDDYYYY	03/07/2021	0:10	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.859Z	2022-10-19T16:04:24.798Z	2022-10-21T21:32:30.046Z	FRIDAY_DEMO/input/Bank%20Statement.docx	openxmlformats-officedocume
FIRST_NAME	John	0:4	DESKTOP-8NLA23I\adaml	FALSE	FALSE	2022-10-20T14:27:25.859Z	2022-10-19T16:04:24.798Z	2022-10-21T21:32:30.046Z	FRIDAY_DEMO/input/Bank%20Statement.docx	openxmlformats-officedocume

You can also determine what data was found and masked through JSON (search) annotations and (masking) results files. The search annotation files contain a list of *files, documents or database collections* that were searched, along with the search results found within each. DarkShield also produces interactive HTML5 charts from its JSON logs like this one:



DarkShield Aggregate Results Report

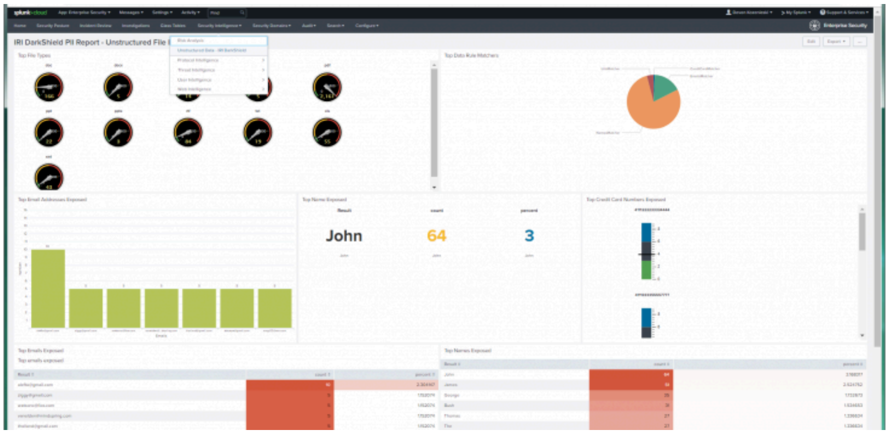
Top Sources of Sensitive Data

The bubbles below rank the top 105 **unprotected** sources by the number of PII matches to DarkShield data class search criteria. Hover over each bubble to reveal the data source containing PII and the number of matches found.

SIEM Tool Integration

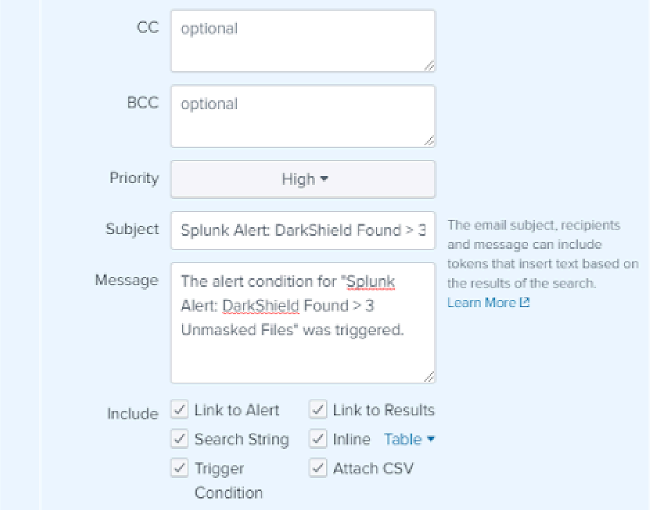
Security Information Event Management (SIEM) tools like Splunk Enterprise Security (ES) and modern analytic platforms like Datadog designed to create insights and enable actions from machine log data. As such, they can be used to categorize, graphically reveal, and report on security incidents from data in log files.

DarkShield produces a high volume and quality of log file data from its PII search and mask operations. The flat-file logs produced by [DarkShield can feed Splunk ES]() directly. This supports insight into PII-related vulnerabilities in the files searched on the network, as well as those in which DarkShield has fully masked the PII it found.



For example, you can design graphical widgets that use the results of discrete log queries in Splunk, and arrange them inside a dashboard accessed from a URL. Custom views can thus reveal fine details about the PII DarkShield finds.

With that data indexed in Splunk, it is also possible to leverage the Adaptive Response Framework in the ES version to [send alerts]() or run Phantom Playbooks [based on conditions]() detected in DarkShield logs. For example, an email can be sent when a certain number of files with unmasked PII was recorded, thus telling DarkShield or its user to run or re-run a data masking job against the current search results. As DarkShield searching and masking lobs are generated, they can also be [sent to Datadog]() or [forwarded to Splunk automatically](). That updates the DarkShield log data indexed in Splunk, and can thus trigger new response actions -- like a dashboard refresh or email.



CC	optional
BCC	optional
Priority	High ▾
Subject	Splunk Alert: DarkShield Found > 3
Message	The alert condition for "Splunk Alert: DarkShield Found > 3 Unmasked Files" was triggered.
Include	☑ Link to Alert ☑ Link to Results ☑ Search String ☑ Inline Table ▾ ☑ Trigger Condition ☑ Attach CSV

The email subject, recipients and message can include tokens that insert text based on the results of the search. Learn More

File Formats & Databases Supported

DarkShield v5 can find and mask PII in these file types:

Text	MS Documents	Images
Fixed Width	.doc/x	.bmp
FHIR, HL7, X12	.ppt/x	.gif
CSV, TSV	.xls/x	.jpg
JSON	**Other Documents**	.png
TXT	.pdf	.tiff
XML	Parquet	DICOM

as well as in audio files, plus PII in these:

Data Silos

File Stores	RDBs	NoSQL	Additional Sources
Local & SMB	MS SQL, DB2	MongoDB	Kafka*
Amazon S3	MySQL, Oracle	Elasticsearch	HTTP/S
Azure Blob, OneDrive SharePoint Online	Salesforce, Snowflake	Cassandra	SFTP*
GCP Storage	Any other RDB connected via a JDBC driver	Couchbase*, Redis*, Solr* OpenSearch* Google BigTable* CosmosDB*	More via Custom API Calls *

If your file format or silo is not on the list above, please contact darkshield@iri.com to ask if it has been added since the publication of this booklet, or when it could be added. Sources marked with a * require "glue code" we offer or you write to work with the DarkShield API.

Compatible Applications

DarkShield uses the same IRI Workbench IDE, data classes, and masking engines as:

- IRI FieldShield - DB and flat-file masking
- IRI CellShield EE - Excel spreadsheet masking
- IRI RowGen - Test data synthesis
- IRI Voracity - Big data management, ETL, etc.

IRI Data Manager Suite

IRI
Total Data Management

www.iri.com
info@iri.com
+1.321.777.8889

IRI Data Protector Suite

IRI CoSort
Sort, Transform & Report

Speed or replace legacy sorts, batch/ETL/SQL transforms

- Filter, join, aggregate, pivot, cleanse, lookup, calc, etc.
- Map, migrate, federate, and replicate data from 150 sources
- Segment data, capture changes, report details / summaries
- Analyze changing dimensions, support complex transforms

IRI FACT
Fast Extract for DBs

Speed RDBMS unloads for archival, migration, reorg, and ETL

- Extract tables to flat files in parallel using SQL queries
- Convert and re-format to change data types and layouts
- Create the data definitions for IRI software and DB loads
- Pipe to CoSort and DB loaders for faster reorg and ETL

IRI NextForm
Data, File & Database Migration

Unlock data and move between apps, DBs, and platforms

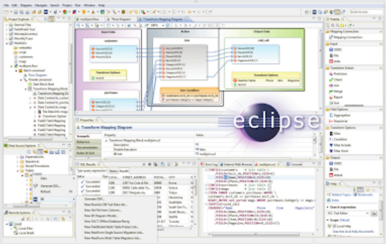
- Convert, federate, remap, and replicate legacy data
- Migrate data between databases and create new tables
- Change file formats, data types, and endian conditions
- Find, extract, and structure data in unstructured sources

IRI RowGen
Smart Test Data Generation

Prototype DBs and ETL, stress-test, outsource, benchmark

- Use real data models and formats, not production data
- Combine generation and selection, create new formats
- Preserve referential integrity and frequency distributions
- Feed test DBs, files, reports, and DevOps simultaneously

IRI Voracity
An Insatiable Appetite for Data



Consolidate tools and tasks to process, protect, prototype, present

- Discover, define, and manage data in legacy and new sources
- Combine data integration, migration, governance, and analytics
- Use IRI Ripcurrent to replicate or mask changed data in real-time
- Leverage the familiarity of Eclipse and the power of CoSort

IRI FieldShield
PII / PHI Classification & Masking

Static and dynamic masking of structured data sources

- Search, profile, and classify sensitive data in DBs and files
- Encrypt, hash, redact, pseudonymize, randomize, tokenize
- Apply cross-table rules to save time and referential integrity
- Score re-ID risk and audit your jobs to verify compliance

IRI CellShield
PII / PHI Search & Mask in Excel

Discover and de-identify PAN/PHI/PII in Excel spreadsheets

- Define or use patterns to search for sensitive data
- Locate, report, and open all found ranges in the LAN
- Click to encrypt, mask, or pseudonymize data directly
- Auto-log protections to verify privacy law compliance

IRI DarkShield
Unstructured Data Search & Security

Discover, deliver, and delete sensitive information everywhere

- Find PII in LAN and cloud souces using multiple methods
- Simultaneously de-identify, remove, or report those values
- Mask text, MS, PDF, Parquet & image files + LOBs & NoSQL
- Comply with the right to erasure, portability, or rectification

IRI DMaaS
Data Masking as a Service

Leverage expert data privacy engineers to find and mask PII

- Avoid learning curves, software expenses and staff diversion
- Reduce risk by agreement, monitored VPN, or secure cloud
- Use operational logs for reporting and compliance audits
- Select from competitive hourly, daily or project rates

DESIGN

Wizards with Rules | Graphical Dialogs
Scripts with Outlines | Form Editors
Workflow & Mapping Diagrams
Erwin Mapping Manager
DataSwitch No-Code

SOURCES

- Hadoop & Streams
- ASN.1 CDRs
- Flat & EDI Files
- Cloud & SaaS
- Relational DBs
- NoSQL DBs
- Text & Images
- Mainframe
- Logs, Excel, etc.

DISCOVER

Data Classification
Dark Data Search
DB & File Search
DB & File Profiling
ER Diagramming
Multi-Source Metadata

INTEGRATE

Slowly Changing Dimensions
Public/Private Mashups
Change Data Capture
Fast DB Un/Load
Data Federation
One-Pass ETL

MIGRATE

Incremental Replication
Database Platforms
Data & File Types
Legacy Sorts
Endianness
ETL Tools

IRI Voracity
An Insatiable Appetite for Data

GOVERN

Data Quality
Data Masking
DB Subsetting
Re-ID Risk Scoring
Test Data Synthesis
Data & Metadata Lineage

ANALYZE

IoT Feeds
In Datadog
Embedded BI
Data Wrangling
KNIME & Splunk
Predictive Analytics

TARGETS

- Kafka & MQTT
- BI & Analytic Tools
- Cloud Stores
- Relational DBs
- NoSQL DBs
- Custom Reports
- DevOps
- Flat & EDI Files
- Logs, Excel, Images

DEPLOY

GUI, CLI, API | MapReduce 2 (Grid)
Spark (In-Memory) | Storm (Streaming)
Tez (Batch) | CI/CD | Java | SQL | YARN
Eclipse or Any Scheduler

intel Software Partner · IBM Business Partner · hp Business Partner · DataSwitch Connecting Data · erwin by Quest · eclipse · msdn · MICRO FOCUS · melissa · Red Hat · ORACLE PartnerNetwork

IRI DarkShield
Unstructured Data Search & Security

INNOVATIVE ROUTINES INTERNATIONAL (IRI), INC.

Innovative Routines International, Inc.
2194 Highway A1A, Third Floor
Melbourne, Florida 32937 USA
Tel. +1.321.777.8889
iri.com/darkshield