



Data Protector Suite

Data Protector Suite 2026



IRI FieldShield
PII / PHI Classification & Masking



IRI DarkShield
Unstructured Data Search & Security



IRI CellShield
PII / PHI Search & Mask in Excel



IRI DMaaS
Data Masking as a Service



IRI RowGen
Smart Test Data Generation



IRI, The CoSort Company

Vendor Background

- Specializing in fast data management and data-centric security
- Privately owned and profitable since 1978
- Sales and support in more than 40 cities worldwide
- Organically grown, shared metadata and Eclipse IP stack
- Featured in: CIO Review (top GRC and Compliance vendors);

DBTA; Bloor Research, SoftwareTestingHelp.com, the Gartner Market Guide to Data Masking Tools; and in QY, Markets & Markets, and Research & Markets *forecast reports* on Data Masking, DB Security, Data Classification, Data Governance

DESIGN

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

TARGETS

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



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 Databodog
Embedded BI
Data Wrangling
KNIME & Splunk
Predictive Analytics

DEPLOY

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

SOURCES

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



Selected IRI Data Masking References



Recent Recognitions of IRI in the Data Governance and Security Industry

- BARC Technology Map for Data Governance (IRI Voracity)
- [Bloor Research Data Searching and Masking InBrief](#)
- CIO Applications: Top 25 GRC Technology Providers
- CIO Review: 20 Most Promising Compliance Technologies
- [Computerworld Germany](#) (IRI FieldShield)
- Data Bridge Global Data Governance Market Size & Analysis
- Data Bridge Global Data Masking Market Size & Forecast
- [DBTA: Trendsetting Products](#) (IRI DarkShield, IRI Ripcurrent)
- Forrester Research Now Tech: Data Masking
- [Gartner Market Guide for Data Masking](#)
- Healthcare Tech Magazine - Top 10 Healthcare Security Solutions Providers
- Insight Partners - Data Classification Market
- Insight Partners - Test Data Management Market Outlook
- [Markets & Markets - Data Governance Market](#) - Visionary Leader
- [Markets & Markets - Data Masking Market Forecast](#)
- [Outlook Series: The Case for Data Masking](#)
- [QY Research - Global Test Data Management Market](#)
- [Privacy & Data Security Law Journal](#)
- [Research & Markets - DB Security Market](#)

IRI Data Manager Suite



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

IRI Data Protector Suite



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



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



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



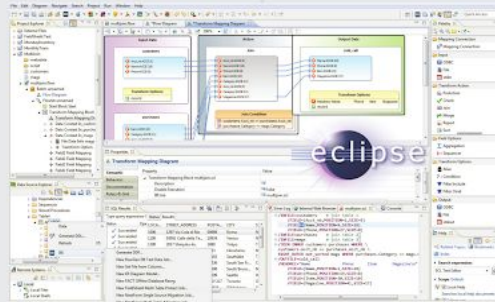
Smart Test Data Generation

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



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



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



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



Unstructured Data Search & Security

Discover, deliver, and delete sensitive information everywhere

- Find PII in LAN and cloud sources 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



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

IRI Data Masking Tool Differentiators

IRI FieldShield

- Introduced 2011 after CoSort began masking fields in files in 2007. Designed for flat-files and RDBs.
- Profiles, classifies, re-ID risk scores, masks, anonymizes, and audits. Preserves referential integrity (RI).
- Integrates with IRI Ripcurrent CDC, Voracity ETL, BI, etc. Supports advanced/custom business logic.

IRI DarkShield

- Uses same data classes, masking rules, and GUI as FieldShield, but on semi- & unstructured data, too
- Offers multiple AI search matchers, plus multi-threading and load balancing to scale performance
- Supports special use cases like EDI loops, conditions, signatures, test documents, and textual ETL

IRI CellShield

- Designed for and runs directly inside MS Excel to protect spreadsheets ergonomically
- Enterprise edition leverages DarkShield search methods and most common masking functions
- Supports bulk masking for hundreds of sheets at once. Produces audit logs and graphs in Excel.

IRI RowGen

- Subsets and masks RDBs, or synthesizes data from DDL, using rules. This preserves RI in test schemas.
- Generates realistic data values, ranges, frequencies, nulls, etc. via randomization or set value selection
- Supports transforms and reporting. Can feed data to DarkShield, CI/CD, DB cloning and TDM platforms.



All these jobs can be designed, modified, shared, and run graphically in a rich design client, called IRI Workbench, built on Eclipse. This front-end also serves as a multi-DB administration hub and IDE for IRI and 3GL jobs, and hosts many free third party plugins like Git and Apache DS.

Notably, no server framework is required to orchestrate jobs. This fosters more granular allocation and tuning of resources, and the sharing of metadata artifacts.

**Managing
Data & Risk
in One Place**

**Define
& Discover**

Connect & Classify
Profile & Diagram
Search & Report

**Prototype
& Provide**

Parse & Synthesize
Subset & Mask
Snapshot & Virtualize



IRI Voracity
An Insatiable Appetite for Data

**Process
& Protect**

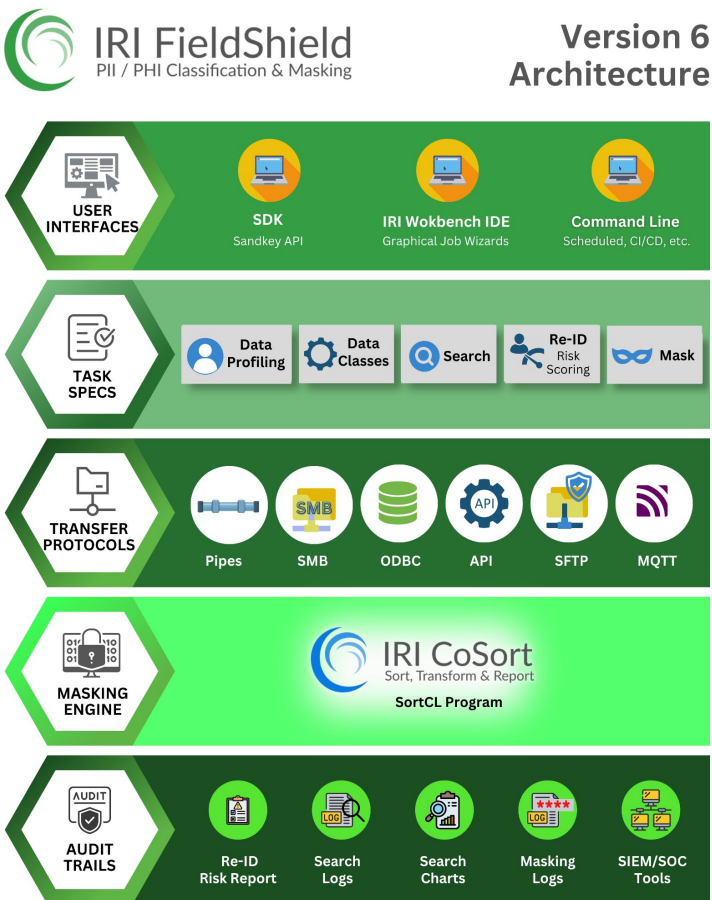
Integrate & Migrate
Cleanse & Mask
Wrangle & Report

**Audit
& Analyze**

Risk-Score & Log
Comply & Verify
Trace & Optimize

Multi-source platform for leveraging and securing data at the same time

IRI Data Masking Tool Architectures

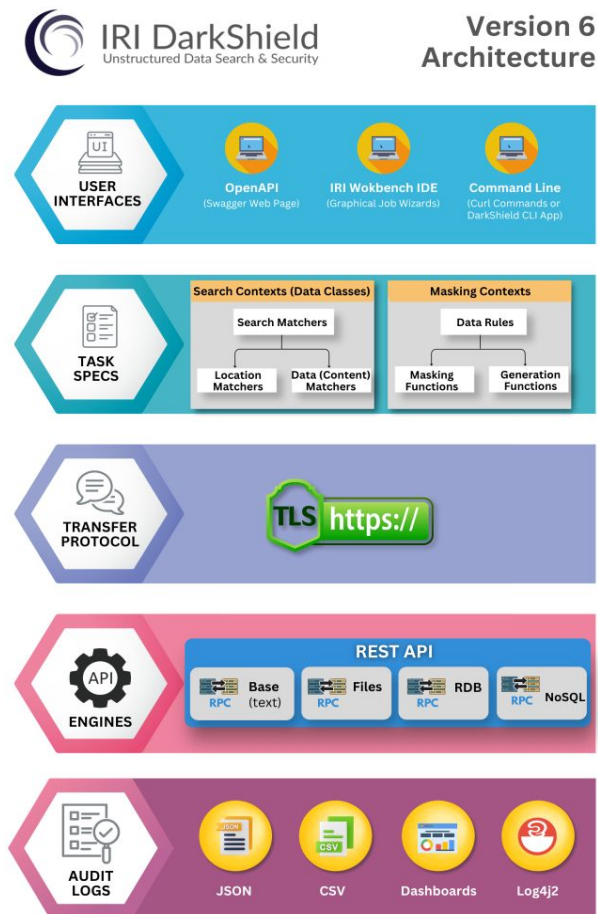


Both structured and unstructured data discovery functions -- including classification, search, and metadata creation -- are performed in IRI Workbench data discovery wizards.

Static data masking (SDM) jobs are usually built in IRI Workbench, while user-specific dynamic data masking (DDM) is or will be available via options on slide #32.

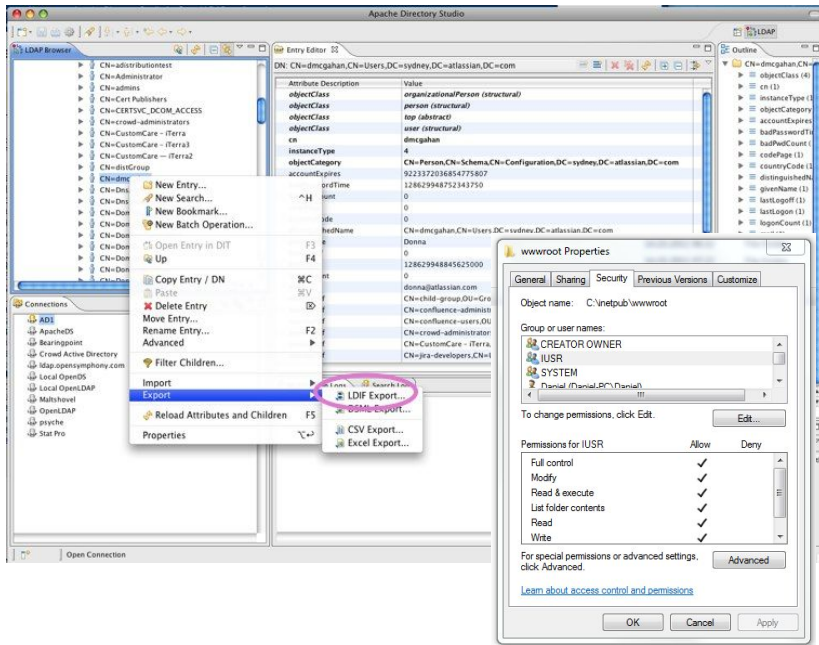
Voracity data manipulation and masking jobs use the IRI CoSort (SortCL) engine on commodity LUW hardware, on premise or in the cloud. No database or cloud API is needed. This reduces runtime overhead, administrative complexity, and risk. The executable is also compatible with the SortCL metadata that can mask PII within data integration, cleansing, and reporting jobs, too.

Finally, no server framework is required to orchestrate jobs. This fosters more granular allocation and tuning of resources, and the sharing of metadata artifacts.

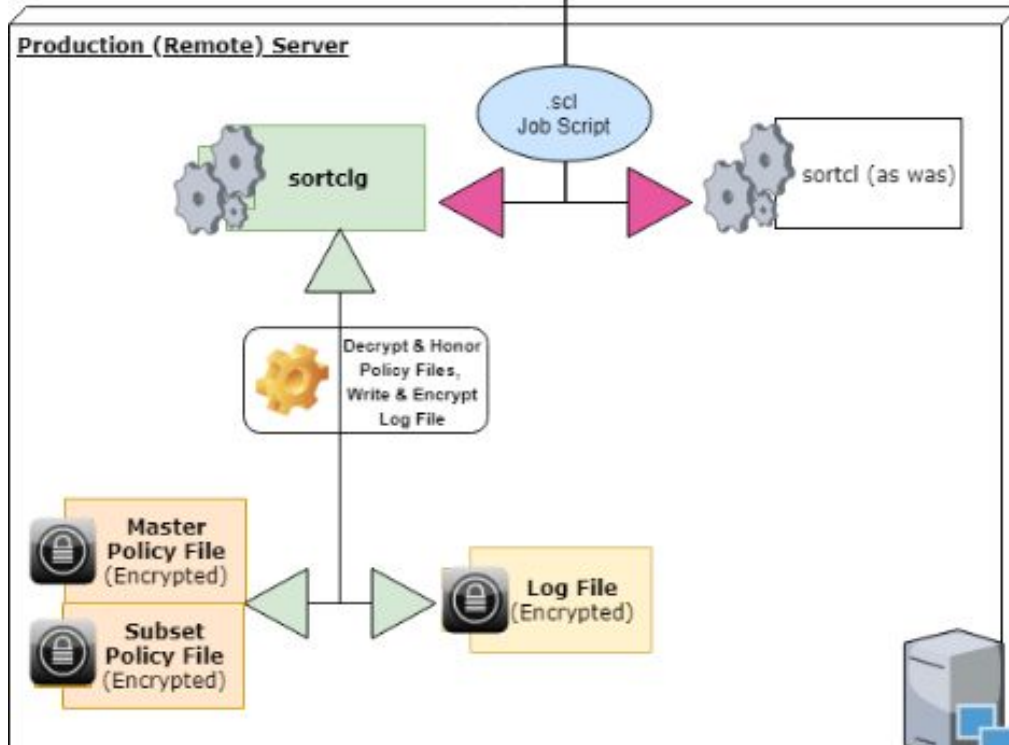
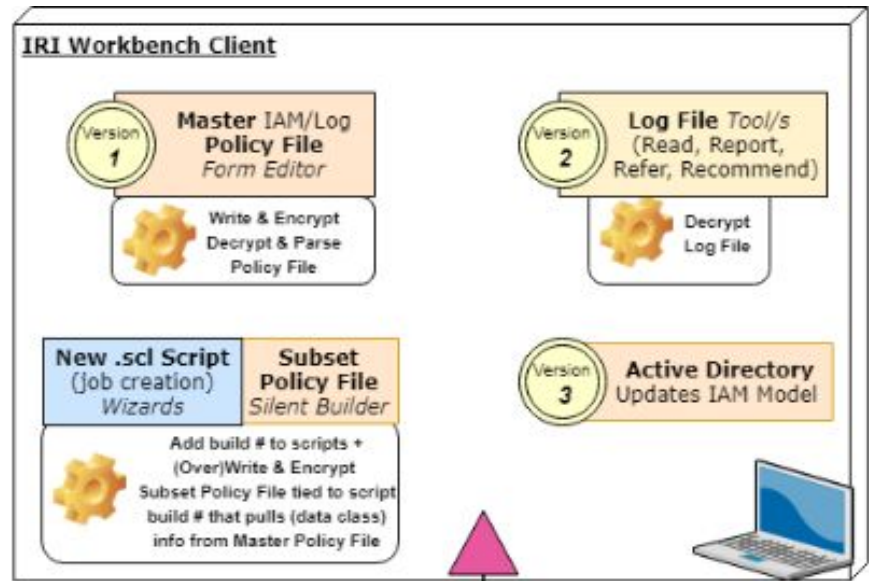


IAM/RBAC Now & Later

Today, you can assign permissions via [encryption keys](#) for select decryption, and: to data (file) sources, IRI masking programs (sortcl.exe), and the scripts they run (spec.fcl) in LUW file systems using central LDAP/AD settings. You can optionally control them via Apache Directory Studio in IRI Workbench:

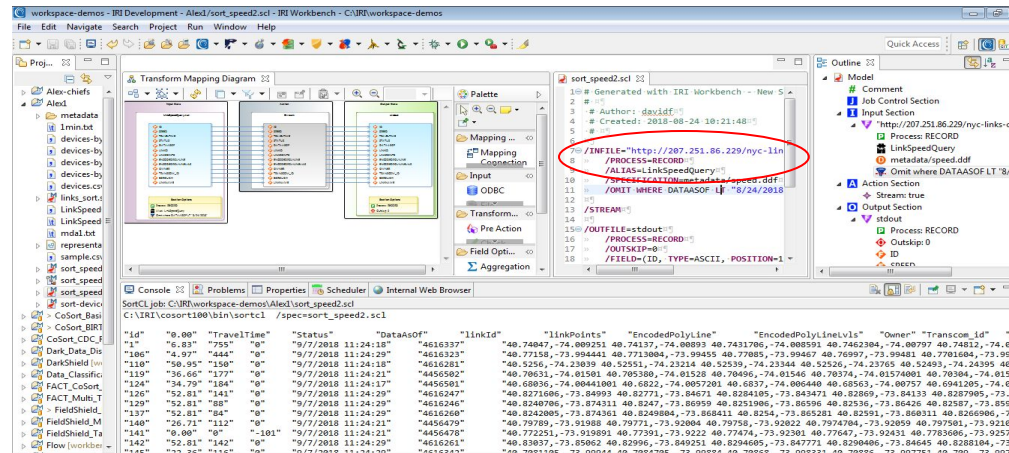
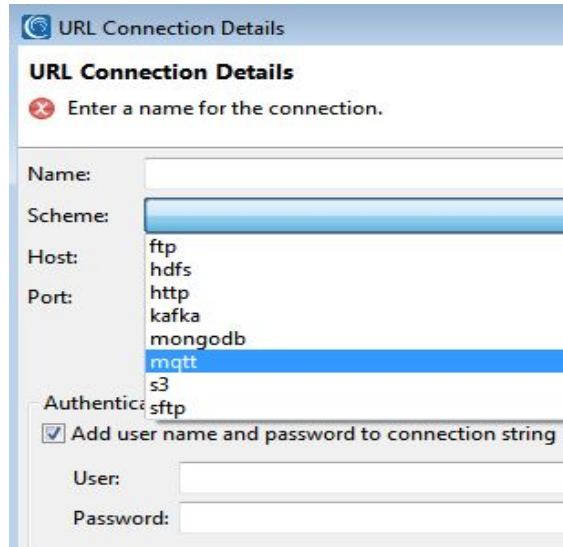
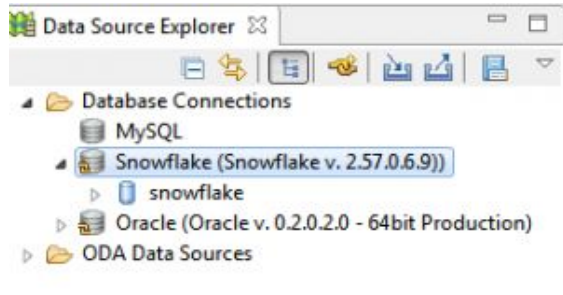
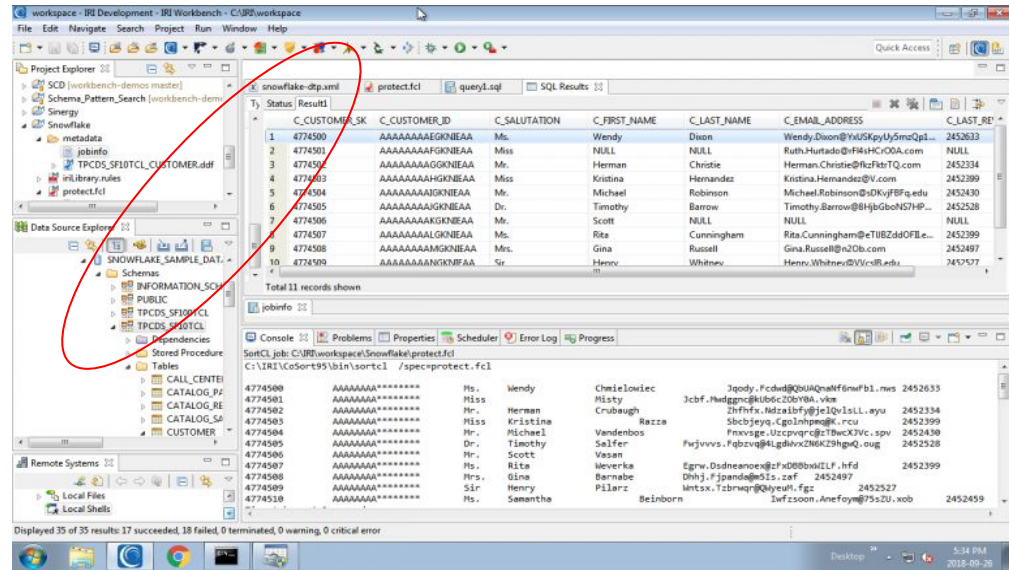


In late 2025, the IRI client/server governance system illustrated on the right will all you to assign and enforce RBACs to the same elements above, **and** to more granular elements like field names (mapped from data classes), functions, and perhaps specific data values (or ranges of values).



Cloud Data & Systems Support

RowGen can synthesize and FieldShield can mask data in cloud stores and DBs like Oracle 19c, Snowflake, MS SQL in Azure, AWS Redshift, etc. via J/ODBC, *plus* URLs & message queues. DarkShield supports files in S3, GCP and Azure Storage, plus any RDB, 10 NoSQL DBs, and SMB-tied cloud drives. All run on Linux, Unix or Windows on-premise, or in cloud shapes, VMs or containers.

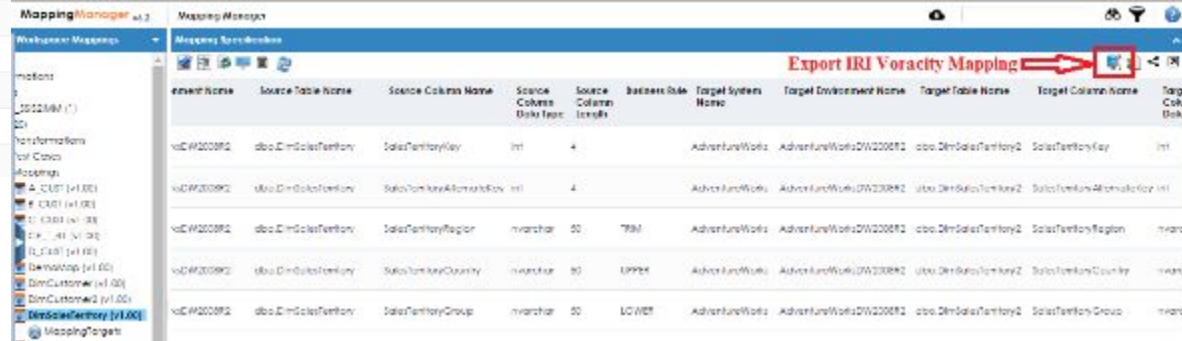


Metadata Integrations

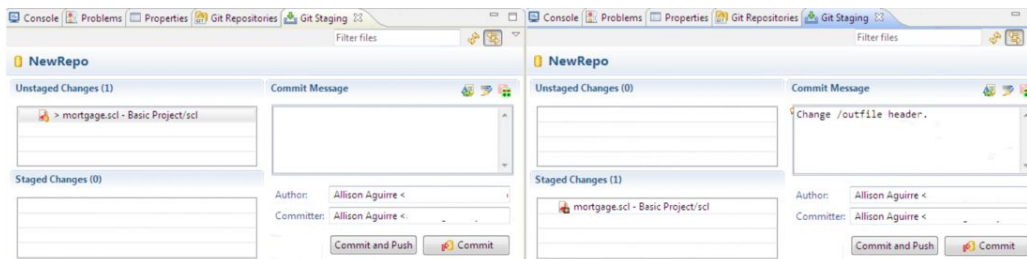
1. Voracity tooling *consumes* metadata from any structured source for data classification, profiling, search, de-ID, ETL etc.
2. FieldShield & RowGen job scripts also *produce* metadata for several **DB load utilities** in multi-DB masking & test data jobs.
3. Their data definition file metadata can also be *exported* (e.g. target field layouts) in CSV for catalog tools like **Collibra**.
4. DarkShield reads **attribute** metadata about source files, and produces artifactual metadata from its search and mask ops and it can auto-forward or populate Splunk ES with that information for analysis, dashboarding, or adaptive responses.
5. **MIMB, erwin, DataSwitch** and **ValueLabs TDH** hub and feed FieldShield and RowGen specs from external metadata:

Bridge Mapping

Meta Integration Repository (MIR) Metamodel (based on the OMG CWM standard)	"IRI CoSORT SortCL Data Definition File" Metamodel IriCoSort	Mapping Comments
Attribute	/FIELD	
Comment		# Comments in the DDF file
Description		# Comments in the DDF file
Name	field name (if physical name is not specified)	
PhysicalName	field name	
Position		
DataType	Data Type name	
Length	Field length	
Scale	Field scale	



6. All IRI metadata -- including data source/target layouts, job/task speci and batch files, workflows and metamodels, discovery configurations, search matchers and masking rules -- can also be team shared, secured and version controlled in **Git** et al



Data Sources (Standard)

Acucobol (MF) Vision	ESDS	MF- & RM-ISAM	Tibero (FACT)
Altibase (FACT)	Excel XLS/X	MF Var. Length	Teradata
ASN.1 CDRs	HL7 (DS)	MySQL / Aurora	Text
C-ISAM	HSQLDB (WB)	Oracle	TSV
CLF web logs	IDX 3, 4 & 8	PDF (DS)	MQTT
CSV	Informix	PostgreSQL / Redshift	Variable Block
DB2 (UDB)	Ingres	Record Sequential	Variable Sequential
DB2 for i5/OS	LDIF	RTF (WB)	VSAM MVS (UniKix)
DB2 for z/OS	JSON	SQL Anywhere	Web Services
Delimited	Line Sequential	SQL Server	Word (DS)
Derby (WB)	MariaDB	SQLite	X12 (DS)
ELF web logs	MaxDB	Sybase ASA/E & IQ	XML

*FACT: requires IRI Fast Extract (FACT) DS: requires IRI DarkShield
WB: requires IRI Workbench, the free Eclipse GUI for FieldShield, etc.*

Data Sources (Legacy)

Access	D3	GA-Power 95, R91	K-ISAM	Pathway	RMS
Adabas	Datacom	Gemstone	Knowledgeman	PDS	Reality/X
Advanced Pick	Dataflex	GENESIS	KSDS	PervasiveSQL	RRDS
ALLBASE	Db4o	Gigabase	Lotus	Pick/Pick64+	Sequoia
Alpha5	dBase	H2	Manman	PI-Open	SFS (VS*)
C-ISAM	RMS	IDMS	Mentor / pro	Powerflex	Sharebase
Azure	DL/1	IDS	MO	Powerhouse	Supra
BizTalk	DSM	QSAM	Model 204	Progress	Terracotta
Cache	Enscribe	IMS	Mumps	QueryObject	Total
Clipper	Enterprise Adapter	Interbase	MyBase	rBase	Ultimate
Codasyl	FileMaker	Intersystems	UDB	R83	UltPlus
CorVision	Firebird	ISM	NonStop SQL	Rdb	Unidata
ConceptBase	Focus	Jasmine	ObjectStore	REALITY	Universe
D-ISAM	FoxPro	JBase	Paradox	Red Brick	VSAM VSE

These sources are typically only accessible via IRI partner (SoftwareAG-CONNx) J/ODBC drivers.

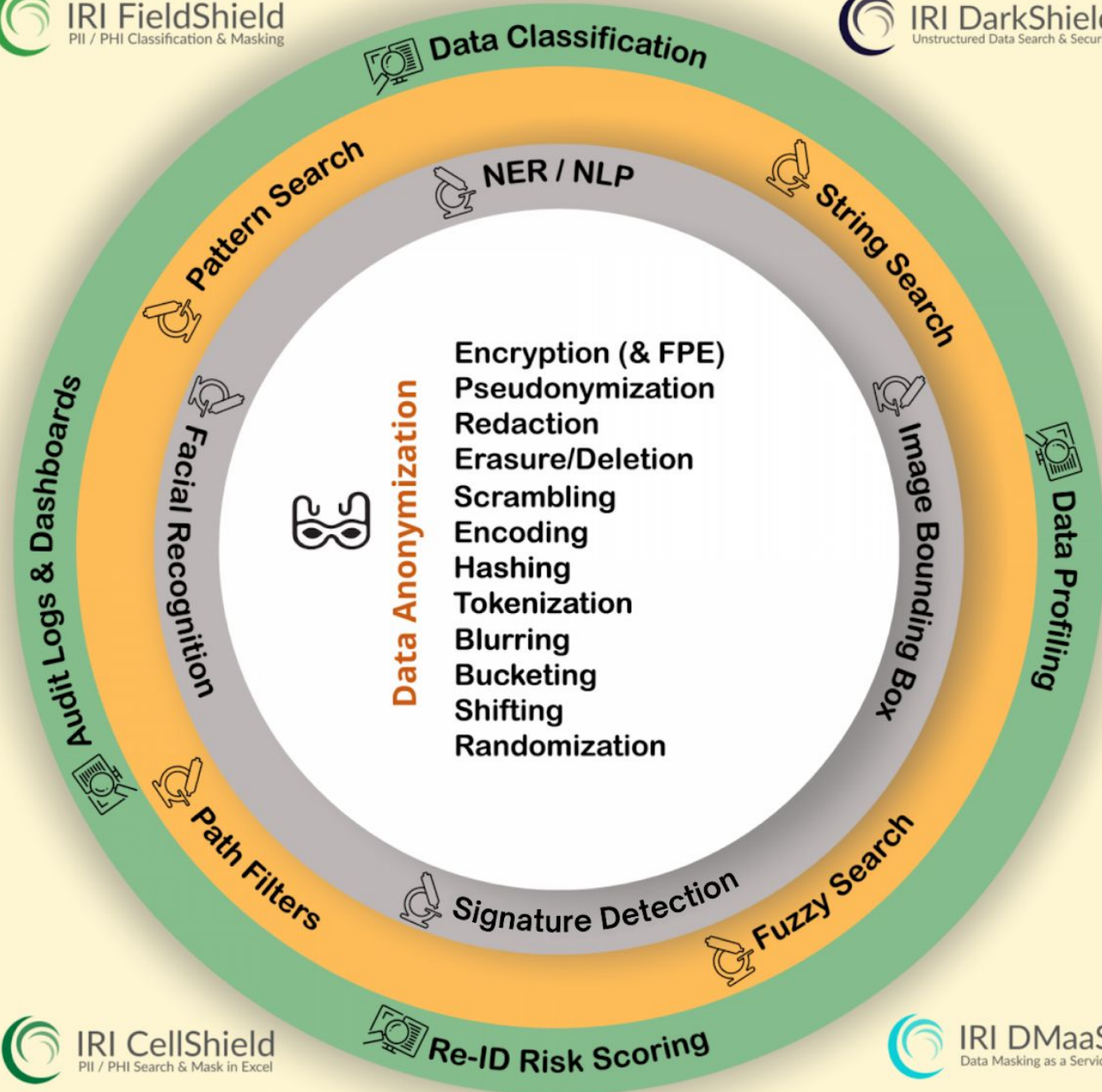
**IBM/Encina SFS files should be supported when written in COBOL using RECORDING MODE IS VARIABLE*

Data Sources (Modern)

Amazon EMR Hive	DynamoDB	Redis & Solr	Parquet files
Amazon RDS	ElasticSearch	MarkLogic (XML)	Pivotal Greenplum
Apache Cassandra	Google BigQuery	MongoDB	Pivotal HD Hive
Apache Hadoop Hive	Google BigTable	MS Dynamics CRM	SAP HANA
Azure CosmosDB	Hortonworks Hive	MS SQL Azure	Salesforce.com
Cloudera CDH Hive	Hubspot	Oracle Eloqua	Snowflake DB
Cloudera Impala	Kafka Connect	Oracle Cloud DB	Spark SQL
Database.com	MapR Hive	Netezza	Vertica DB

IRI FieldShield finds and masks structured RDB and flat-file data on-premise, or in HDFS, Sharepoint, AWS, Azure, GCP or OCI, plus files in ASN.1-encoded CDR formats, MF-ISAM or Vision (COBOL index), and XLS/X (Excel spreadsheet) formats.

IRI DarkShield supports RDB and flat file data, too, **plus:** semi- and unstructured data in static or streaming text, log and EDI formats like JSON, HL7, X12, FHIR, and XML; B/CLOB columns in RDBs; PDF, Parquet, and .SQL files, Excel, Word and PowerPoint (Office) docs including signature and PII in embedded images; NoSQL DBs; audio, and, image files in BMP, DICOM, GIF, JPG, PNG and TIFF. DarkShield runs on premise or in the cloud, and can read/mask/write PII from/to files in AWS S3 buckets, Azure Blobs, GCP storage, or SharePoint Online. **IRI CellShield** supports Excel XLS/X files on-premise or in Office 365.





IRI FieldShield
PII / PHI Classification & Masking

IRI Data Protector Suite

What FieldShield Does

- Finds and masks data in scalar RDBs plus flat, Excel, ACH, Fedwire and ASN.1-encoded CDR files
- Supports complex business logic for custom masking needs and complex output targeting
- Performs ETL while masking to push masked data into lower environments
- Automatically discovers and parses DDL and file metadata to speed job script production
- Uses built-in data classification infrastructure to assign sensitivity levels and masking rules
- Profiles, models, and searches data sources on premise or in the cloud
- Produces search reports in human and machine-readable log and dashboard formats
- Applies deterministic masking rules consistently to preserve referential integrity
- Works in combination with IRI data cleansing, transformation and report job specs
- Writes loader metadata and performs direct path loads for test DB populations
- Simultaneously creates flat-file and custom/structured detail and summary report targets
- Works with Voracity subsetting wizard to mask parent and child subset tables or files
- Runs from the Voracity Ripcurrent module to mask changed rows incrementally in real-time
- Scores the risk of re-identification based on unmasked quasi-identifiers
- Generates multiple runtime logs and diagrams for masking-related audit trails
- Pushes and pulls data classes, rules, and other job artifacts through Git for shared work

Sensitive Data Classification and Search Wizards

To facilitate data masking, IRI FieldShield includes: PII definition (cataloging through data classes); discovery through string (literal or dictionary), pattern, and fuzzy-logic searches; statistical reporting; and, automatic metadata creation.

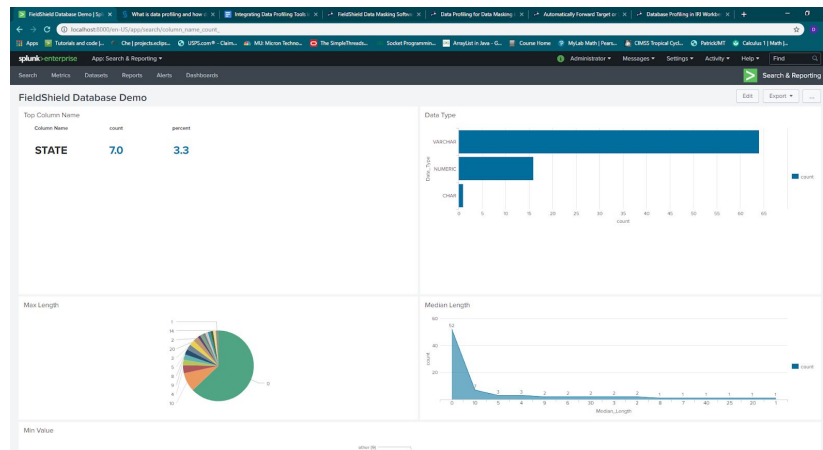
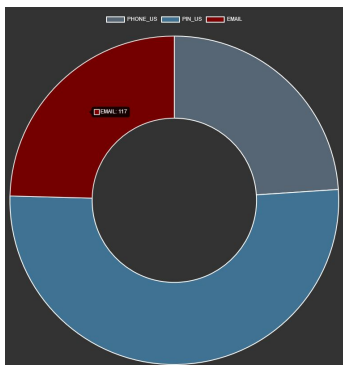
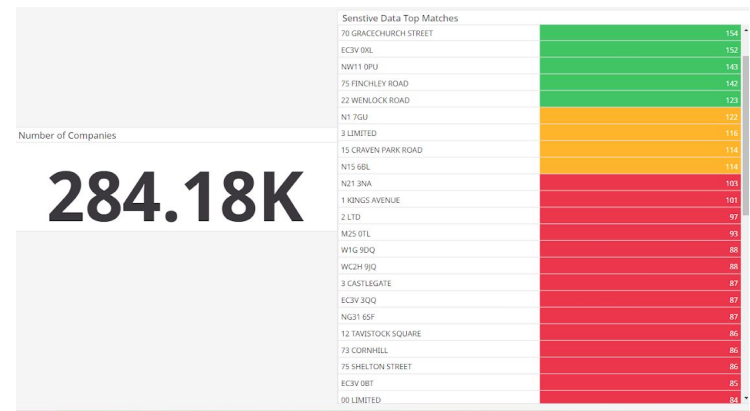
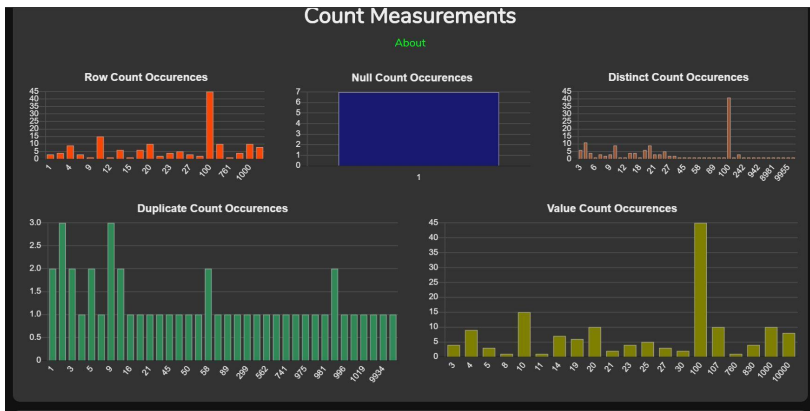
Fit-for-purpose GUI wizards deliver:

- DB and file data classification, with search and masking rule matchers
- DB profiling, ERDs, and table searches
- Flat-file profiling and value searches
- Data class searches through schema and directories for bulk discovery
- Metadata discovery and definition
- Dark data search and structuring, with metadata reporting (see DarkShield)

The image displays several screenshots of the IRI FieldShield software interface. The top row shows the 'New Database Profile' wizard, including 'Table Selection' and 'Pattern and Fuzzy Search' options. The middle row shows a 'Table Profile' window displaying a list of columns and their statistics, and a 'Dark Data Discovery' window showing search criteria. The bottom row shows a 'Project Explorer' with a tree view of data classes and a 'Dark Data Discovery' window showing a complex ERD diagram of data relationships.

Search Result Reports, Dashboards & Exports

In addition to report-formatted and machine-readable outputs from PII search operations in IRI Workbench, FieldShield users can also see details at a glance in digital displays, or feed that data to tools like Splunk Enterprise Security and Datadog for analytics and action-taking:



Multiple Masking Job Design Options

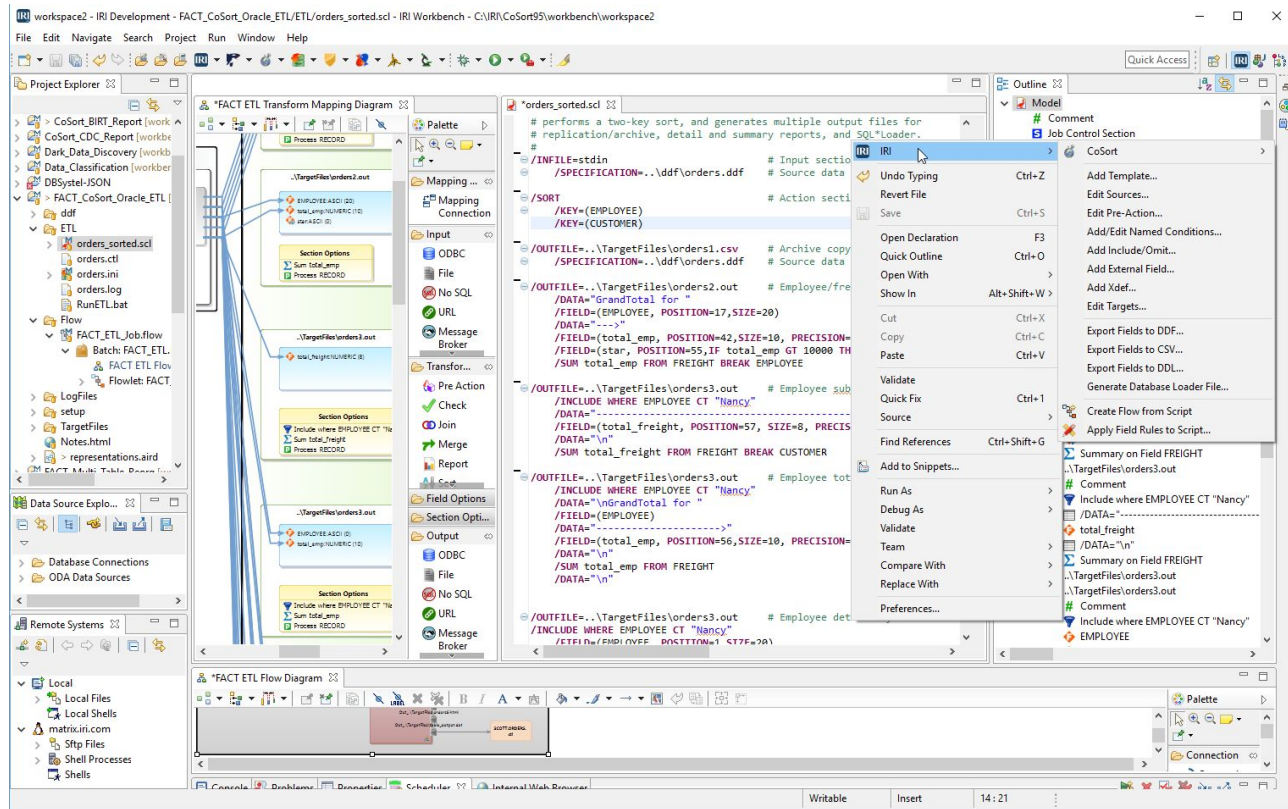
IRI FieldShield and other Voracity data masking, cleansing, transformation, migration, reporting, and wrangling jobs can be created and run *inside or outside* of IRI Workbench.

Job design methods supported *inside*:

- 1) Job creation wizards
- 2) Color-coded syntax-aware job script editor with outline
- 3) Form Editors
- 4) Graphical parameters Dialogs
- 5) Mapping Diagrams

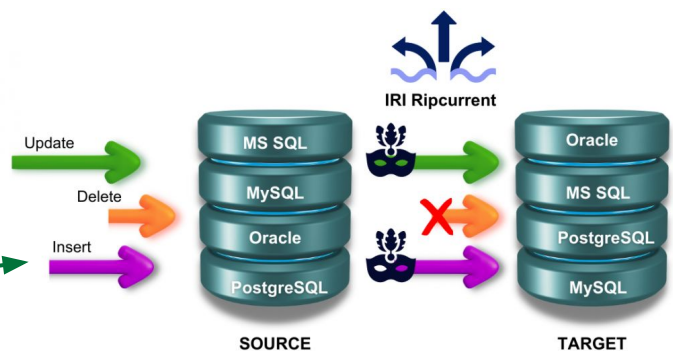
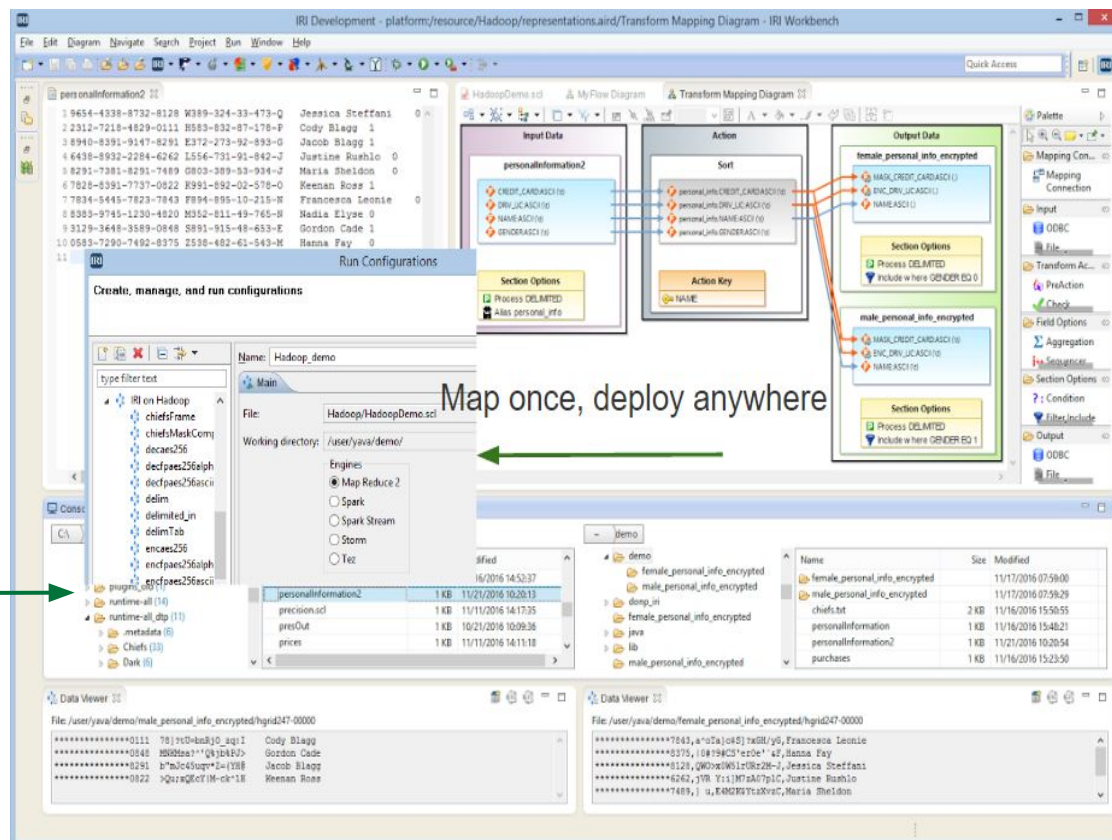
Job design methods supported *outside*:

- 6) DataSwitch no-code web app
- 7) erwin Mapping Manager
- 8) Value Labs Test Data Hub
- 9) Any external text editor
- 10) 3GL app (system or API calls)



and Multiple Job Deployment Options

- 1) 4GL scripts on the command line or in batch
- 2) From 3rd party automation tools like Stonebranch UAC, cron, etc.
- 3) Directly from KNIME in Eclipse, or a Splunk add-on app, as you report or index
- 4) Some jobs run without code changes in Hadoop via MR2, Spark, Spark Stream, Storm or Tez
- 5) Use graphical run configuration dialogs or the built-in task scheduler to launch local, remote, or HDFS jobs from IRI Workbench
- 6) System or API calls from 3GL programs or web services
- 7) Value Labs Test Data Hub, Cigniti Blueswan TDM, GitLab, Azure DevOps, Amazon CodePipeline, and Jenkins CI/CD
- 8) Actifio, Commvault & Windocks cloning tools for virtualized DB images / containers
- 9) DataSwitch no-code data engineering app
- 10) IRI Ripcurrent facility in Voracity for real-time, incremental DB mask/refresh ops upon source table inserts, updates, or deletes



Step 1: **Connect** to RDBs or **Create** Other Source Metadata



Step 2: **Define** Data Classes & Masking Rules (.dcr1 Library File)



Step 3: **Search & Classify** PII in RDB Schema or File Folder(s)



Step 4: **Review** the Data Class Map to Validate / Edit Rules



Step 5: **Mask** the RDB Schema or Files with Referential Integrity



IRI FieldShield
Workflow

Static Data Masking Function Categories (1-4 of 15)

New Data Rule Wizard

New Data Rule Wizard Selection

Create encoding or decoding functions.

Filter

All FieldShield DarkShield RowGen N

Masking

Y=z Assignment Expression

Deterministic

Reversible

Bit Manipulation

Encoding

Encryption

Set Lookup Pseudonymization

DB Lookup Pseudonymization

(a) Set Lookup Pseudonymization Function

Non-reversible

Empty String Replace

Generalization via Bucketing

Hashing

Hash Lookup Pseudonymization

Redaction

String Manipulation

Numeric Manipulation

Non-deterministic

Blurring

Format Preserving Scramble

Pseudonym Single Column

(a) Quick Setup Pseudonymization

Length Preserving Pseudonymization

Rule Name: DataEncodingRule

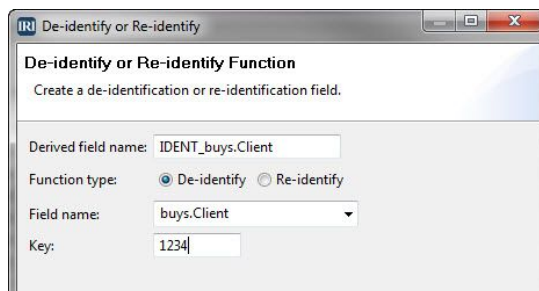
Overwrite existing rule with same name as above



< Back

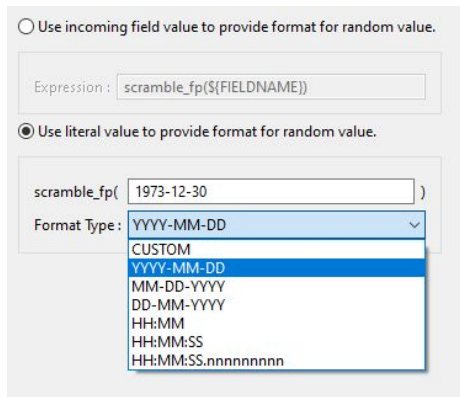
Next >

Bit Twiddling

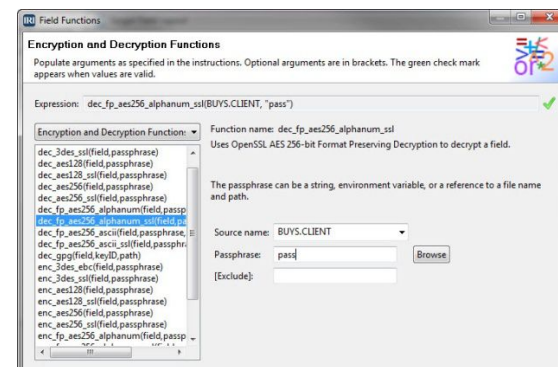


- For ASCII data
- Less secure
- Reversible

Format Preserving Scrambling

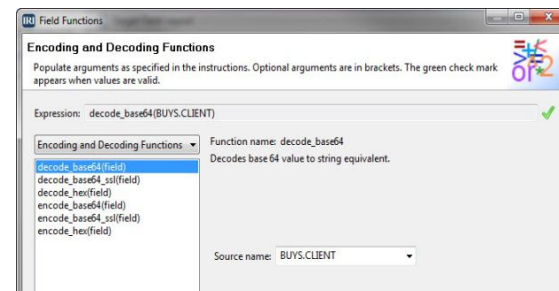


Encryption / Decryption



- Multiple algorithms
- AES-256 format-preserving
- Multiple key mgmt options

Encoding / Decoding



- Changes binary to ASCII
- Supports base64 & hex

Static Data Masking Function Categories (5-7 of 15)

Pseudonymization

Pseudonym Replacement

Create a pseudonym field that will use values in a set file as substitutes for the original field's values.

Pseudonymize field:

Use provided pseudonym list (non-recoverable)

Name options: PseudoSetPage_grpPseudoFiles=Pseudonym Files

Name type:

Sex:

Order:

Default pseudonym list file: C:\IRI\CoSort95\sets\names\names_male_first.set

Use only unique names from pseudonym list
(Blanks inserted when # of records is greater than # of unique names)

Use your own pseudonym list (non-recoverable)

User pseudonym list

Pseudonym list file:

Use original field as a look-up into pseudonym list

Use random draw from pseudonym list

Use only unique names from pseudonym list
(Blanks inserted when # of records is greater than # of unique names)

- Provides realistic names
- Reversible lookup values
- or Non-reversible selection

Redaction / Obfuscation

Masking Function

Replaces a range of characters in the required source field with a replacement character.

Source field:

Use predefined masks

Predefined Masks

Mask Example: (0)123456789 => *****

Mask:

Define mask

Arguments

Mask character (* by default):

Start position:

Length:

Type	Value

- Partial/full-field masking
- Conditional omission
- Non-reversible

Randomization

Random Value Generation

Randomly generate a new value for this field.

Derived field name:

Generate random value

Random value options

Type:

Random Min Size:

Random Max Size:

Random selection from a set file

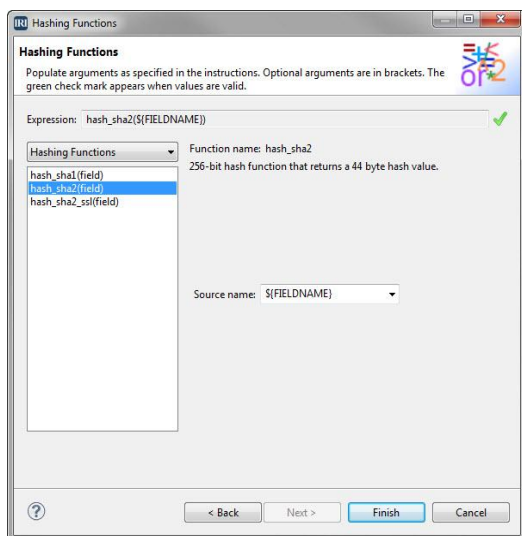
Set file

File:

- Random data generation
- Random data selection
- Non-reversible

Static Data Masking Function Categories (8-15 of 15)

Hashing



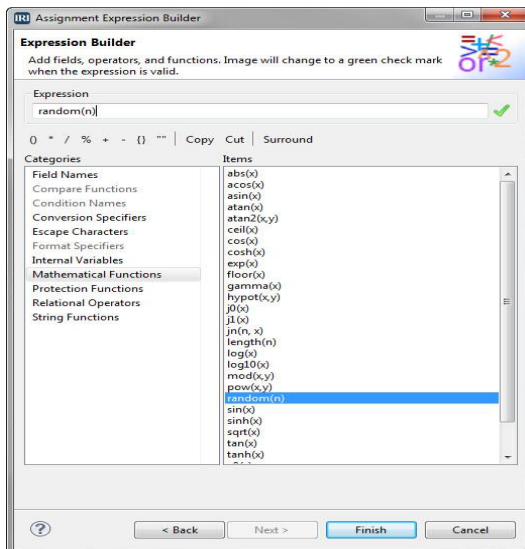
- SHA-1 & 2 cryptographic
- Returns hash of fieldstring
- Use for integrity checking

Blurring: Add random “noise” (perturbate) to ages and dates

Bucketing / Binning:

Generalize quasi-identifiers to to anonymize demographic traits

Expression Logic

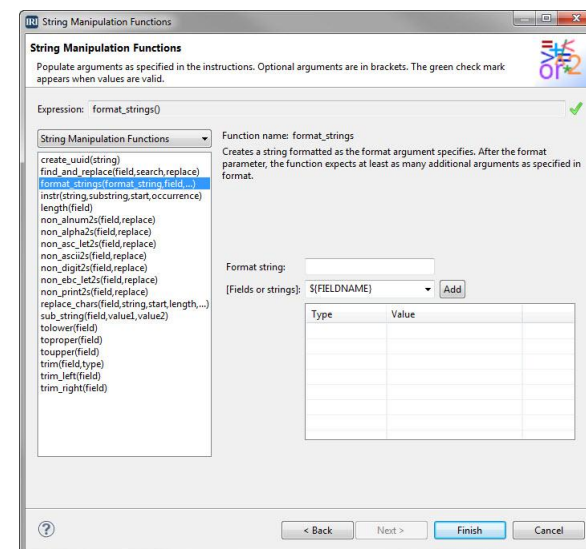


- Mathematical operations
- PCRE logic
- Concatenations, etc.

Tokenization: Substitute PANs with DB values for PCI-DSS

Suppression: Deletion (record erasure) to comply with the GDPR [Right to be Forgotten](#)

String Manipulations



- Find, replace, and add
- Reposition and trim
- Use INSTR information

Custom Function: Nested, or user-written field, transforms.

Re-ID Risk Measurement

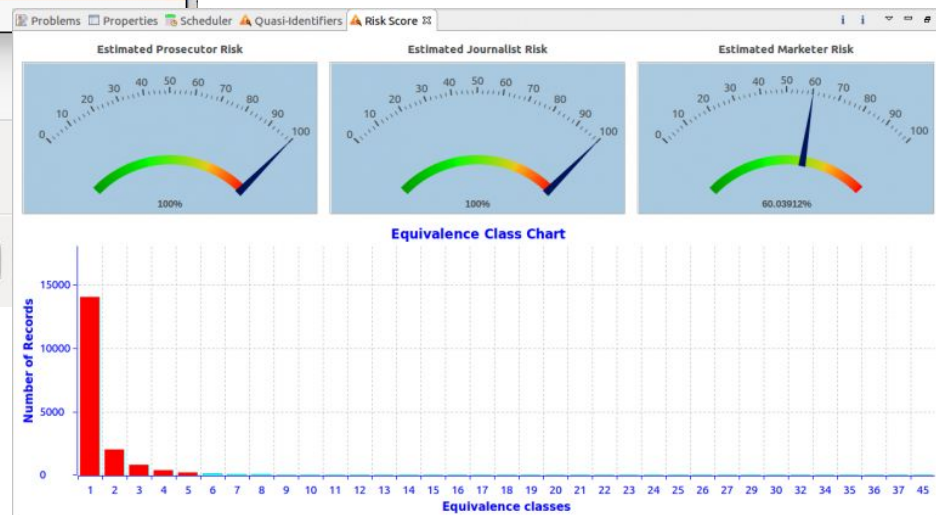
New Re-ID Risk Scoring

Attributes
Select the type of each attribute.

Name	Type
sex	Quasi-Identifying
age	Quasi-Identifying
race	InSENSITIVE
marital-status	Sensitive
education	Quasi-Identifying
native-country	Identifying
workclass	InSENSITIVE
occupation	InSENSITIVE
salary-class	Sensitive

< Back Next > Cancel

US HIPAA and FERPA regulations require that patient and student data sets used in research or marketing have a statistically certified “very small” chance of being re-identifiable.



- The risk scoring wizard produces re-ID probability scores in 3 modes
- Analyzes quasi-identifiers with multiple, peer-reviewed functions
- Detail and graphed scoring reports

Query-Ready XML Audit Log (JSON Log in Development)

The screenshot displays the IRI Workbench interface with the following components:

- Project Explorer:** Lists project files such as `map_name1.set`, `name_match.scl`, `password`, `patient_dec.scl`, `patient_enc.scl`, `patient_record.ddf`, `FS-Tables.SQL`, `Notes.html`, `patient_record.data`, `pr_metadata.ddf`, and various repository folders like `Flow 66 repo`, `JCL_SORT_Convert 55 repo`, etc.
- Data Source Explorer:** Shows database connections including `BIRT Classic Models Sample Databases`, `Oracle (Oracle v. 0.2.0.2.0 - Productic)`, and `ODA Data Sources`.
- FSlog.xml (Source View):**

```

1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <AuditTrail>
3 <AuditRecord>
4 <Product>CoSort</Product>
5 <Version>9.5.3</Version>
6 <VersionTag>R95140122-1600</VersionTag>
7 <Serial>99999.demo</Serial>
8 <OperatingSystem>Windows 7</OperatingSystem>
9 <User>IRIDEMO</User>
10 <ProcessId>1396</ProcessId>
11 <Terminal>console</Terminal>
12 <Program>sortcl</Program>
13 <Command>/spec=patient_enc.scl </Command>
14 <StartTime>2014-02-20 13:31:49</StartTime>
15 <EndTime>2014-02-20 13:31:54</EndTime>
16 <RunTime>00:00:05</RunTime>
17 <ReturnCode>0</ReturnCode>
18 <ErrorMessage>normal return</ErrorMessage>
19 <RecordsProcessed>10</RecordsProcessed>
20 </Script>

```
- FSlog.xml (Design View):** A table showing the structure of the XML log:

Node	Content
VersionTag	R95140416-1006
Serial	99999.demo
OperatingSystem	Windows 7
User	IRIDEMO
ProcessId	3732
Terminal	console
Program	sortcl
Command	/spec=patient_enc.scl
StartTime	2014-09-09 05:22:16
EndTime	2014-09-09 05:22:17
RunTime	00:00:01
ReturnCode	0
ErrorMessage	normal return
- Properties View:** Shows metadata for the `FSlog.xml` file:

Resource	Property	Value
FSlog.xml - FieldShield_Table_File\scl	Info	
	derived	false
	editable	true
	last modified	December 21, 2015 at 9:27:50 AM
	linked	false
	location	C:\IRI\CoSort95\workbench\workspace\FieldShield_Table_File\scl\FSlog.xml
	name	FSlog.xml
	path	/FieldShield_Table_File\scl\FSlog.xml
	size	45,058 bytes

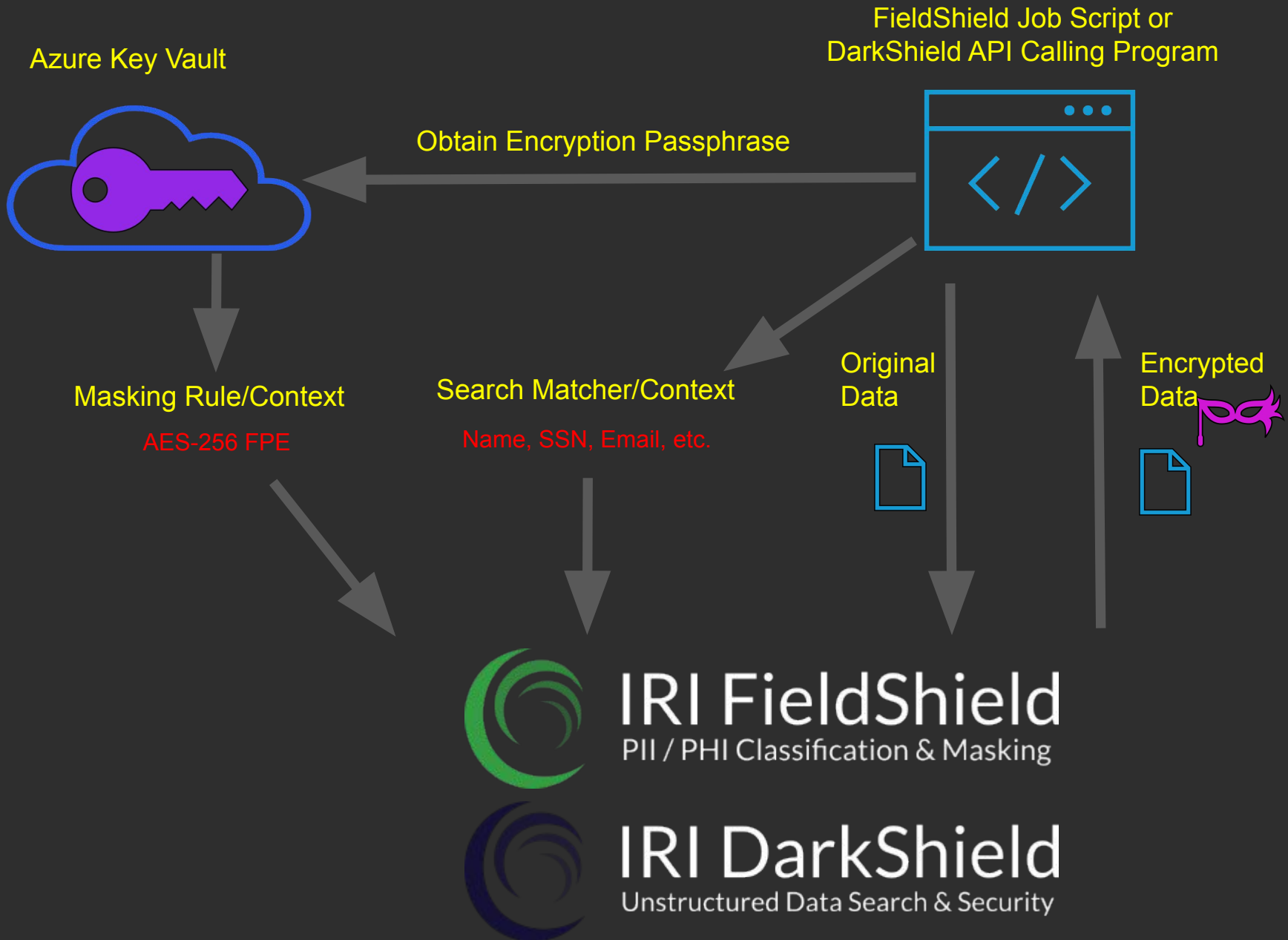
Dynamic & Real-Time Data Masking Options

Method	Operation
ODBC Select / Update	Apply FieldShield column masks to target (view) tables for specific users/rows
DB App Invocation	Use callable API library functions or system-call job scripts on the fly
In-Situ Redaction	User and SQL-specific full and partial column masking on query
Custom I/O Procedures	Drive real-time application data directly to/from FieldShield jobs in memory
Real-Time Processing	Database replication and masking with FieldShield rules via IRI Ripcurrent
Proxy-based	New development initiative for RDB, NoSQL and file dynamic data masking
Governance Mode	New development initiative for runtime facility tied to RBAC/IAM infrastructure

Encryption Key Management Options

1. Passphrase (key string) embedded in script, in clear or encrypted text
2. Passphrase string as environment variable
3. Passphrase string in (securable) key file
4. MFA, HSM/VM etc. via [Azure Key Vault](#)
5. Townsend Security [Alliance Key Manager](#)

PII Encryption through Azure Key Vault



User Profiles

- Vertical industries – especially BFSI and Healthcare – and government agencies storing, processing, or outsourcing applications with sensitive data, such as:
 - Banks
 - Hospitals
 - Census / Tax
 - Pharmas
 - Defense
 - Schools
- Application, DB, and DW users handling PII, PI, CUI / NPI
- CISOs, compliance teams, consultants, IT managers, and solution architects
- Test data (DevOps) architects

Use Cases

Tesco Bank/RBS UK

- Decrypt and re-encrypt fields in credit card migration and test files
- Generate and manage encryption and user ID keys
- Other projects protect 38,265 records per minute on Windows

Accenture Singapore

- Design and run encryption and masking jobs on Linux servers
- Secure PHI for the Ministry of Health Holdings (MOHH)'s Oracle DB
- Row sequencing and job audits

Medicx Media Solutions USA

- Encryption and hashing functions to PII and PHI in geo-medical consumer health databases
- Exceeds HIPAA requirements in provisioning mScores™ data to digital and direct marketers

Key Differentiators

Developer Support

- Metadata, rule, and job version control
- Master data definition
- 5 encryption key management options
- Git project management (teaming)
- SDK supports .NET and Java calls
- Data profiling and metadata discovery
- XML (and soon JSON) job logs, IAM

Price Performance

- The data-centric security tool with:
 - ➔ The most sources
 - ➔ The most protection functions
 - ➔ The most target file formats
- Fastest standalone protection software

One-Stop-Shop

- Integrated data classification & search
- Includes re-ID risk scoring for HIPAA
- Use w/Voracity ETL, migrate, cleanse
- Metadata-compatible with RowGen TDM
- Used in DB subsetting & replication
- Also works in Voracity BI & KNIME jobs
- Runs w/Actifio, Commvault, and Windocks DB cloning tools

Ease-of-Use

- Familiar Eclipse GUI
- Self-documenting 4GL syntax
- Easy management and modification of jobs/metadata

Competitive Advantages

vs. Oracle ([click](#))

vs. IBM

- FieldShield scripts simpler than Optim interoperability model and Javascript
- Seamless integration with more sources
- Same metadata as subset & synthesize More functions
- Lower cost

<https://www.iri.com/support/data-education-center/top-data-masking-solutions-of-2025>

vs. CA (Grid Tools)

- Built-in CoSort engine makes FieldShield faster than GT Fast Data Masking
- Tight integration with data profiling, ETL, data quality, and BI operations
- Multi-target/format options
- Lower cost

<https://www.iri.com/products/fieldshield/why-is-fieldshield-better>

vs. Informatica

- FieldShield DDM inclusive with product (compared to Informatica's upgrade)
- More SDM protection functions
- Integration with Eclipse and Excel
- Access to 4GL scripts
- Lower cost

vs. Imperva (Camouflage)

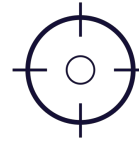
- FieldShield has more masking and encryption functions, hashing, etc.
- Re-ID risk scoring wizard
- Faster and more extensible job scripts in the IRI Workbench IDE for Voracity
- Lower cost



IRI DarkShield
Unstructured Data Search & Security

IRI Data Protector Suite

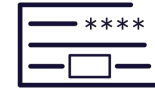
What DarkShield Does



Search



Extract



Redact



Audit

1. Simultaneously scans, extracts, and de-IDs or deletes PII and other sensitive data in multiple silos
2. Finds PII in structured, semi/unstructured and unstructured files, DB, images, Parquet, and audio
3. Finds defined data classes tied to RegEx patterns, lookup sets, NER models, signatures and regions
4. Builds, saves, and re-uses semi-supervised, machine learning models in project or cloud repositories
5. Redacts or replaces PII with encrypted (including FPE), pseudonymized, or other ciphertext values
6. Writes masked data atop originals, or to different targets with the same file/table names and formats
7. Shows on-screen progress of search, remediation, and model training activity
8. Generates logs and charts for values found or masked, plus IRI-compatible metadata for textual ETL
9. Creates interactive dashboards with search and mask results, or hand-offs log files to Splunk et al
10. Runs in IRI Workbench with other IRI and Eclipse tools, from the command line, or via RPC API
11. Works with reverse proxy, image preprocessing, CI/CD, AI models, and file conversion technologies

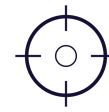
DarkShield Benefits

1. Combines PII discovery, delivery, deletion, and reporting in multiple structured, semi-structured, and/or unstructured source formats into one or few ergonomic operation(s)
2. Uses RegEx pattern matching, NER and other search methods for more accurate results
3. Supports multiple [right to be forgotten and data portability requests](#) into the same search/mask operation through literal name or lookup-file matches
4. Leverages multiple drives, nodes, and threads for searching and masking work
5. Operates in the same Eclipse job design and metadata environment – IRI Workbench – with related data classification, masking, test data synthesis, and data management activities
6. Also runs in CLI or RPC (Open)API mode, from a job scheduler or inside a CI/CD pipeline
7. Features affordable licensing options (standalone, bundled, or free in Voracity)
8. Works with FieldShield and CellShield EE data classes and masking [functions](#)
9. Serializes and models parameters to simplify job modification, batch execution and auditing
10. Integrates with IRI RowGen to [synthesize and insert test data into images, docs, etc.](#)

Development Roadmap

1. Detection and redaction support for PII in handwriting, video files, Splunk indexes, etc.
2. Support for biometric identifiers
3. Front-ending search/mask job support for on additional DBs in the Workbench NoSQL wizard
4. Plug-in integration with more SIEM tools *beyond* Datadog, Splunk ES, and Phantom Playbooks (which are now supported), like IBM QRadar or SolarWinds

Granular Sourcing/Targeting



Search



Extract



Redact



Audit

Data Source
Specify the URI of the input data source.

Type: File System

Directory: C:/Users/cosort/Desktop/demos/Da

[Include]:
[Exclude]:

Recursive

File Types:

- Email Message (.eml)
- Excel 2003 (.xls)
- Excel (.xlsx)
- Offline Outlook Data File (.ost)
- Portable Document Format (.pdf)

Data Class Details

Name: Emails Edit...

Description:

Default Rule: Create...

Active

▼ Matcher Details

Location Matchers:

Name	Type	Classification Option	Value
EmailLocation	PATTERN	INCLUDE	Pattern: .

Data Matchers:

Name	Type	Classification Option	Value
EMAIL	PATTERN	INCLUDE	Pattern: \

Data Target
Specify the URI of the target data source.

Type: SMB

URL: DarkShield-Demo New...

Path: /public/ Browse...

File Exists: Override

Flatten

Keep last modified date

OK Cancel

Use the DarkShield dark data discovery wizard to find sensitive data in unstructured data in LAN and cloud stores, mask it, and target the results.



Extract

workspace - IRI Development - DarkShield/DarkShieldOutput.sql - IRI Workbench - C:\IRI\workspace

File Edit Navigate Search Project Run Window Help

DDDReport.scl DarkShieldOutput.sql search_output.txt

Connection profile
Type: Oracle_11 Name: Oracle Database: XE Status: Connected, Auto Commit

Console Properties Scheduler SQL Results

Type query expression	NAME	RESULT	SPAN	OWNER	REGULARITY	LINKAGE	READONLY	HIDDEN	FILESIZE	DATECREATED	DATEMODIFIED	DATEACCESSED	FILEPATH	FILETYPE
Status Operatic	1	PHONE_INTE...	727	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
✓ Succes SELECT	2	PHONE_INTE...	444	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	3	PHONE_INTE...	6593	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	4	PIN_US	194-67-09...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	5	PHONE_INTE...	321	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	6	PHONE_INTE...	3333	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	7	PIN_US	367-54-23...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	8	PIN_US	123-45-67...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	9	PHONE_INTE...	205	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	10	PHONE_INTE...	3433	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	11	PIN_US	987-65-43...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	12	PHONE_INTE...	142	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	13	PHONE_INTE...	456	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	14	PHONE_INTE...	7777	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	15	PIN_US	103-81-23...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	16	PHONE_INTE...	321	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	17	PHONE_INTE...	555	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	18	PHONE_INTE...	1212	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	19	PHONE_INTE...	12345678...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	20	PHONE_INTE...	14725836...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	21	PHONE_INTE...	78945612...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	22	PIN_KOREA	040392-5...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf
	23	PIN_KOREA	120486-7...	Pag...	dmity	true	false	false	61611	2018-04-05T...	2018-04-05T1...	2018-06-07T1...	/home/...	pdf

Some records are hidden

Displayed 1 of 1 results: 1 succeeded, 0 failed, 0 terminated, 0 warning, 0 critical error

193 PHONE_INTERNATIONAL|3742|slideShow 1, slide 2, slide-content 2|dmity|
194 PHONE_INTERNATIONAL|3056|slideShow 1, slide 3, slide-content 3|dmity|
195 PIN_US|723-64-2357|slideShow 1, slide 8, slide-content 8|dmity|
196 PHONE_INTERNATIONAL|7779|slideShow 1, slide 5, slide-content 5|dmity|

Optionally and automatically extract all of the values you searched for (think GDPR data portability or CCPA/DPDPA [DSARs](#) or [textual ETL](#)), plus metadata associated with the files containing those values.



Redact

```
{ "id": 1,
  "first_name": "Jeanette",
  "last_name": "Penddreth",
  "gender": "Female",
  "email": "*****",
  "ip_address": "26.58.19...",
  "phone_numbers": [
    {
      "type": "personal",
      "number": "*****"
    },
    {
      "type": "office",
      "number": "*****"
    }
  ],
  "transcript": "Hey, email me at *****"
}, {
  "id": 2,
  "first_name": "Giavani",
  "last_name": "Frediani",
  "emails": [
    { "email": "*****" },
    { "em"
  ],
  "data": {
    <?xml version="1.0" encoding="utf-8"?>
    <root>
      <Customers>
        <Customer CustomerID="GREAL">
          <CompanyName>Great Lakes Food Market</CompanyName>
          <NATID>*****</NATID>
          <ContactName>Michael Johns</ContactName>
          <ContactTitle>Marketing Manager</ContactTitle>
          <Phone>(503) 555-7555</Phone>
```

```
1 Running org.superbiz.hello.HelloTest
2 Apache OpenEJB 4.0.0-beta-1 build: 20111002-04:06
3 http://tomee.apache.org/
4 INFO - openejb.home = /Users/*****/examples/helloworld-weblogic
5 INFO - openejb.base = /Users/*****/examples/helloworld-weblogic
6 INFO - Configuring Service(id=Default Security Service, type=SecurityService, provider-id=Default Security Service)
7 INFO - Configuring Service(id=Default Transaction Manager, type=TransactionManager, provider-id=Default Transaction Manager)
8 INFO - Found EjbModule in classpath: /Users/*****/examples/helloworld-weblogic/target/classes
9 INFO - Beginning load: /Users/*****/examples/helloworld-weblogic/target/classes
10 INFO - Configuring enterprise application: /Users/*****/examples/helloworld-weblogic/classpath.ear
11 INFO - Configuring Service(id=Default Stateless Container, type=Container, provider-id=Default Stateless Container)
12 INFO - Auto-creating a container for bean HelloBean: Container(type=STATELESS, id=Default Stateless Container)
13 INFO - Enterprise application "/Users/*****/examples/helloworld-weblogic/classpath.ear" loaded.
14 INFO - Assembling app: /Users/*****/examples/helloworld-weblogic/c
15 INFO - Jndi(name=MyHello) --> Ejb(deployment-id=HelloBean)
16 INFO - Jndi(name=global/classpath.ear/helloworld-weblogic/HelloBean)
17 INFO - Jndi(name=global/classpath.ear/helloworld-weblogic/HelloBean)
18 INFO - Created Ejb(deployment-id=HelloBean, ejb-name=HelloBean, conta
19 INFO - Started Ejb(deployment-id=HelloBean, ejb-name=HelloBean, conta
```



1000 Walnut
Kansas City MO 64106-3686

0701 Ogdsohii Kg.
Xlots Ltet, UW 89090-8827

Primary Account Number: 489671137

Bank Statement

If you have any questions about your statement,
please call us at 816-234-2265

Statement Date: June 5, 2003
Page Number: 1

CONNECTIONS CHECKING Account # 489671137

Account Summary Account # 489671137

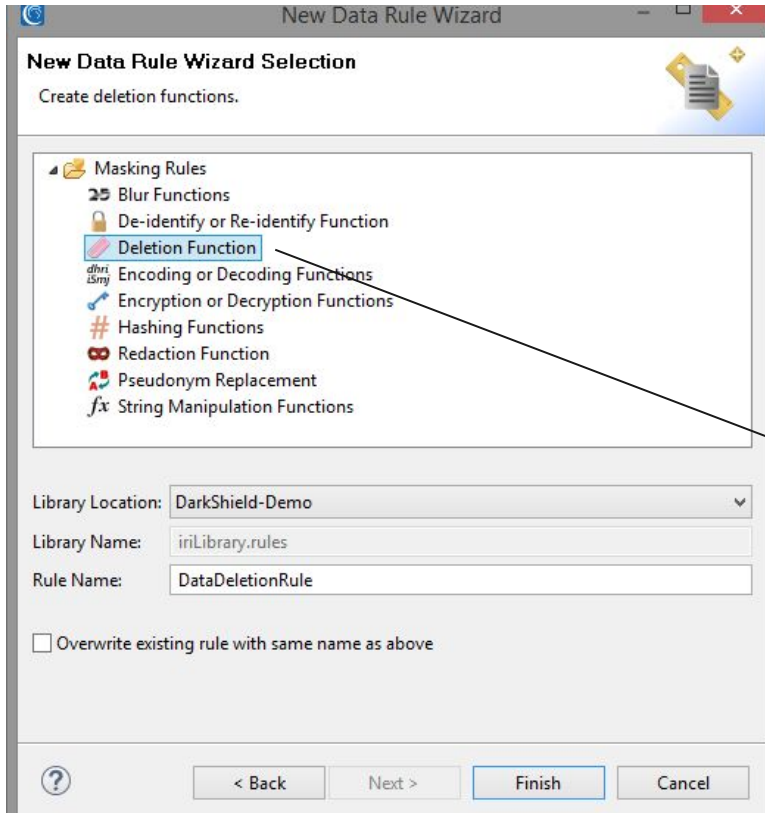
Beginning Balance on May 3, 2003	\$7,126.11
Deposits & Other Credits	+3,615.08
ATM Withdrawals & Debits	-20.00
VISA Check Card Purchases & Debits	-0.00
Withdrawals & Other Debits	-0.00
Checks Paid	-200.00

Ending Balance on June 5, 2003

\$10,521.19

Apply width-preserving redaction, blackout, deletion, encryption, pseudonymization, and other data masking functions to protect PII and comply with data privacy laws like the [GDPR](#).

Deletion Function



IRI FieldShield, DarkShield & CellShield and other features in Voracity combine to comply with [GDPR](#) (and thus CCPA, KVKK, etc.) provisions like:

- Discovery and **De-Identification** of PII and PI
- [The right to be Forgotten \(via erasure like this\)](#)
- Data **Portability** (via extraction and reformatting)
- Data **Rectification** (via discovery and cleansing)

Search & Mask via Column & Path Filters



Search



Extract



Redact



Audit

Search Matcher
Match Data Classes or Groups to a Data Rule.

Name:

Description:

Data Class Name: Details:

Rule Name: Details:

Filters:

Details	Type
\$..clients[*].email	JsonFilter

Filter
Create a filter for filtering content that will be searched.

Type:

Range from to
 Including this value Including this value

Literal entry

Preview:

Ignore First
 Filter by Row

```
MSH|^~\&|VA StarLIMsv10 Prod^2.16.840.1.114222.4.3.3.2.2.4^ISO|VA PHL Richmond^2.16.840.1.114222.4.1.9977^ISO|F
PID|1||987654321^^^HospitalSystem&2.16.840.1.114222.XXX&ISO^MR||SLK0Z^FZPXDVH|MOMMAIDENONE|798684589276-7989|F|
NK1|1|SLK0Z^JZNN|MTH^Mother^HL70063|0047 Lmzu Aeniqi^^Yhvozgzg^UT^96981|^^^^^804^5693861||N^Next-of-Kin^HL7013|
ORC|RE|XXXXX^HospitalSystem^2.16.840.1.114222.XXX^ISO|555550001^VA PHL Richmond^2.16.840.1.114222.4.1.9977^ISO|
OBR|1|XXXXX^HospitalSystem^2.16.840.1.114222.XXX^ISO|555550001^VA PHL Richmond^2.16.840.1.114222.4.1.9977^ISO|5
SPM|1|XXXXX&HospitalSystem&2.16.840.1.114222.XXX&ISO^555550001&VA PHL Richmond&2.16.840.1.114222.4.1.9977&ISO|
```

Allows the user to take the column, segment, or key-value pair structure of an Excel sheet, CSV, HL7, X12, FHIR, JSON or XML files -- or RDB column names -- into account in searches to:

- Ignore fields that do not match the filter
- Increase search speed, and narrow the scope of the search results

Relational and NoSQL Database Supports



Search



Extract



Redact



Audit

Data classes and filters can also be used to find and mask the PII within unstructured text columns in relational tables via JDBC drivers, or in CosmosDB, DynamoDB, Google BigTable, MongoDB, Cassandra, Elasticsearch, Redis, Solr, and Couchbase collections/clusters.

Combine table filters with XML or JSON path filters to pinpoint and mask PII in unstructured XML or JSON text within RDB columns.

DarkShield can also automatically detect, search and mask binary data (images and MS/PDF documents) embedded within BLOB columns of JDBC-connected RDBs.

The screenshot displays the IRI DarkShield interface with several key components:

- URL Connection Details:** A dialog box for configuring a connection. Fields include Name (Customers), Scheme (ELASTICSEARCH (DarkShield only)), Host (localhost), Port (9200), and Cluster (elasticsearch). An authentication checkbox is present.
- Search Criteria Details:** A configuration window showing search criteria for 'new_schema_search'. It includes a description, database schema sources/targets (Source: dtp-jdbc/DarkShield?schemaName=public&include=clients; Target: dtp-jdbc/DarkShield?schemaName=target), and search matchers like NameColumnMatcher, EmailMatcher, AddressMatcher, and NamesNerMatcher.
- Search Results:** Two tables are shown: 'clients' and 'clients_masked'. The 'clients' table lists records with columns for id, name, email, address_xml, and notes. The 'clients_masked' table shows the same records with the PII masked.
- Data Source Explorer:** A tree view on the left showing the project structure, including 'DarkShield (PostgreSQL v. 12.2.0)' and various schemas like 'information_schema', 'pg_catalog', 'public', and 'target'.

Image File Redaction or Value Replacement ...



Search



Extract



Redact



Audit

... or Test Value (RowGen) Synthesis into Images or Documents ...



at his touch of a certain icy pang along my blood. "Come, sir," said I. "You forget that I have not yet the pleasure of your acquaintance. Be seated, if you please." And I showed him an example, and sat down myself in my customary seat and with as fair an imitation of my ordinary manner to a patient, as the lateness of the hour, the nature of my preoccupations, and the horror I had of my visitor, would suffer me to muster.

"I beg your pardon, Dr. [REDACTED]" he replied civilly enough. "What you say is very well founded; and my impatience has shown its heels to my politeness. I come here at the instance of your colleague, Dr. [REDACTED] on a piece of business of some moment; and I understood..." He paused and put his hand to his throat, and I could see, in spite of his collected manner, that he was wrestling against the approaches of the hysteria—"I understood, a drawer..."

But here I took pity on my visitor's suspense, and some perhaps on my own growing curiosity.

"There it is, sir," said I, pointing to the drawer, where it lay on the floor behind a table and still covered with the sheet.

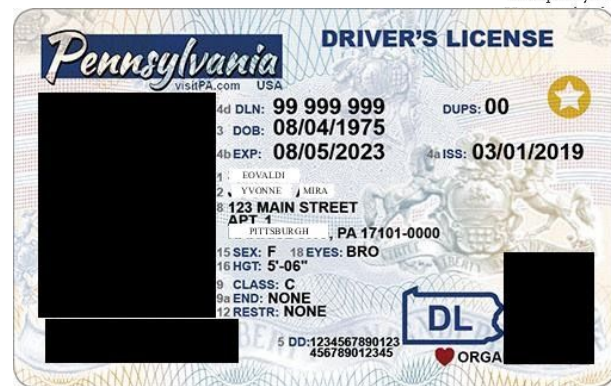
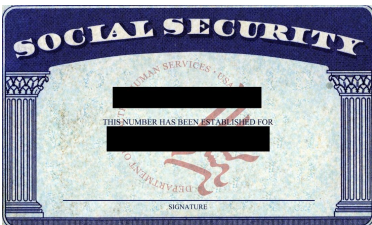
He sprang to it, and then paused, and laid his hand upon his heart: I could hear his teeth grate with the convulsive action of his jaws; and his face was so ghastly to see that I grew alarmed both for his life and reason.

"Compose yourself," said I.

careful smile to me, away the sheet. At such immense relief voice that was alighted glass?" he asked place with something

with a smiling nature and added c first of a reddish hue

BMP, DICOM, GIF, JPx, PNG, and TIFF, alone or in docs like PDFs and Word!



DICOM Medical Image De-ID / Anonymization



Search



Extract

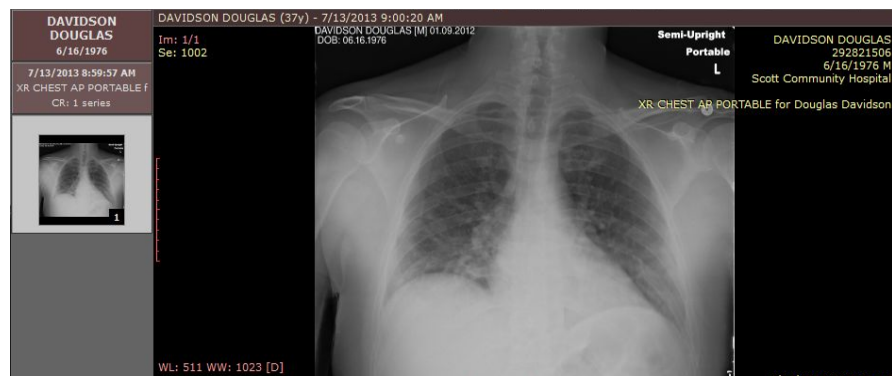


Redact



Audit

Before DarkShield:



	E	F	G	
1	Subject ID	Study UID	Study Description	St
2		292821506 2.25.10646195478329164104825442366895644619	XR CHEST AP PORTABLE for Douglas Davidson	:
3		292821506 2.25.10646195478329164104825442366895644619	XR CHEST AP PORTABLE for Douglas Davidson	:
4		339833062 2.25.541629416691863029175036346640219638	NA	

After DarkShield:



	C	D	E	F	G	H
1	3rd Party Analysis	Data Description URI	Subject ID	Study UID	Study Description	Study Date
2	NO	https://doi.org/10.7937s17z-r072	703523712	2.25.106461990375634587137034423668956446198	XR CHEST AP PORTABLE for Axodduq Wmhpzjcx	7/13/201
3	NO	https://doi.org/10.7937s17z-r072	703523712	2.25.106461990375634587137034423668956446198	XR CHEST AP PORTABLE for Axodduq Wmhpzjcx	7/13/201
4	NO	https://doi.org/10.7937s17z-r072	204498801	2.25.541629416691863029175036346640219638	NA	7/5/200

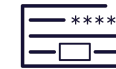
Healthcare EDI Files: Masking PHI in HL7, X12 & FHIR



Search



Extract



Redact



Audit



- New File Search/Masking Job...
- New NoSQL DB Search/Masking Job...
- New Relational DB Search/Masking Job...
- New Healthcare Search/Masking Job...**
- Export Job Specs for API Use...
- DarkShield API Configuration...
- DarkShield Audit Log Filter Configuration...
- NLP Tasks...

New Healthcare DarkShield Job

Healthcare Wizard Selection

Select which Healthcare job you would like to create

Select a Healthcare wizard.

Source Type:

- FHIR
- HL7
- X12

New HL7 DarkShield Job

PII General Selection and Advanced Selection

Select PII general groupings or precise PII selection

NOTE: These are general selections that may NOT cover all PII segments as labeled: review your documents for conformance to these specifications prior to runtime.

In the selections below, please select the groupings of PII you would like to identify to be masked, and/or use the advanced button for more precise control

Select All General Segments

General Selection For Segments

Ethnicity Gender Occupation Phone Number Race
 Account Numbers City County Street Address Zip Address
 ID Numbers/Code/Identifier Name Unique Identifiers (General) Date SSN

Advanced

Segment: PID Field: 5 Sub-Field(Optional): 2

Segment Name	Field	Sub-Field
PID	5	1
PID	5	2

Use the same masking functions and data classes across all sources.

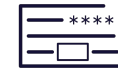
.SQL & Audio File Masking



Search



Extract



Redact



Audit

```
INSERT INTO DARKSHIELD.CUSTOMERS VALUES (1, 'Haszc', 'ixwnm@sgtttkv.qkd', '(013) 865-4046');
INSERT INTO DARKSHIELD.CUSTOMERS VALUES (2, 'Ikoso', 'dxt@hfdmztp.hga', '(276) 220-9396');
```

```
INSERT INTO DARKSHIELD.ACCOUNTS VALUES (101, 1, 'SAVINGS', 1500);
INSERT INTO DARKSHIELD.ACCOUNTS VALUES (102, 2, 'CHECKING', 500);
```

```
INSERT INTO DARKSHIELD.ALL_CONSTRAINTS_TEST (department_id, employee_id, email, phone_number) VALUES
(10, 101, 'ixwnm@sgtttkv.qkd', '025-9364'),
(20, 102, 'dxt@hfdmztp.hga', '187-3365'),
(10, 103, 'shyqe@bjbzyzc.waz', '284-3425'),
(30, 104, 'tnxb@xvdmjnw.ewy', '397-1824'),
(20, 105, 'aqp@kybiyvr.xma', '428-2929');
```

```
INSERT INTO DARKSHIELD.CATEGORY (CATEGORY_ID, CATEGORY_NAME) VALUES (100
INSERT INTO DARKSHIELD.CATEGORY (CATEGORY_ID, CATEGORY_NAME) VALUES (100
INSERT INTO DARKSHIELD.CATEGORY (CATEGORY_ID, CATEGORY_NAME) VALUES (100
INSERT INTO DARKSHIELD.CATEGORY (CATEGORY_ID, CATEGORY_NAME) VALUES (100
INSERT INTO DARKSHIELD.CATEGORY (CATEGORY_ID, CATEGORY_NAME) VALUES (100
```

Audio Manipulation Function

Audio Manipulation Function

Audio manipulation through either deletion or splicing.

Expression:

Delete Replacement Audio

BLEEP
BLEEP
BIRDS
WHITE_NOISE
SILENCE



< Back

Next >

Finish

Cancel

```
new_job.dsc {} *Trump_WEF_2018.mp3_results.json X
{
  "metadata": {
    "name": "file",
    "Content-Encoding": "UTF-8",
    "resourceName": "file /C /IRI/runtime-New_configur",
    "Content-Type": "audio/mpeg"
  },
  "document": {
    "type": "audio/mpeg",
    "audioSegments": [
      {
        "failedResults": [],
        "results": [
          {
            "annotation": {
              "additionalInfo": {
                "startTimestamp": 0.52,
                "endTimestamp": 0.64
              },
              "matcher": "Klaus_k_1",
              "type": "pattern",
              "dataClass": "Klaus",
              "dataClassGroup": "",
              "start": 12,
              "end": 17,
              "searchMatcherPriority": 1
            },
            "maskedResult": "",
            "ruleName": "DataAudioRule",
            "start": 12,
            "end": 17
          }
        ]
      }
    ]
  }
}
```

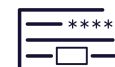
Multilingual NER via Machine Learning



Search



Extract



Redact



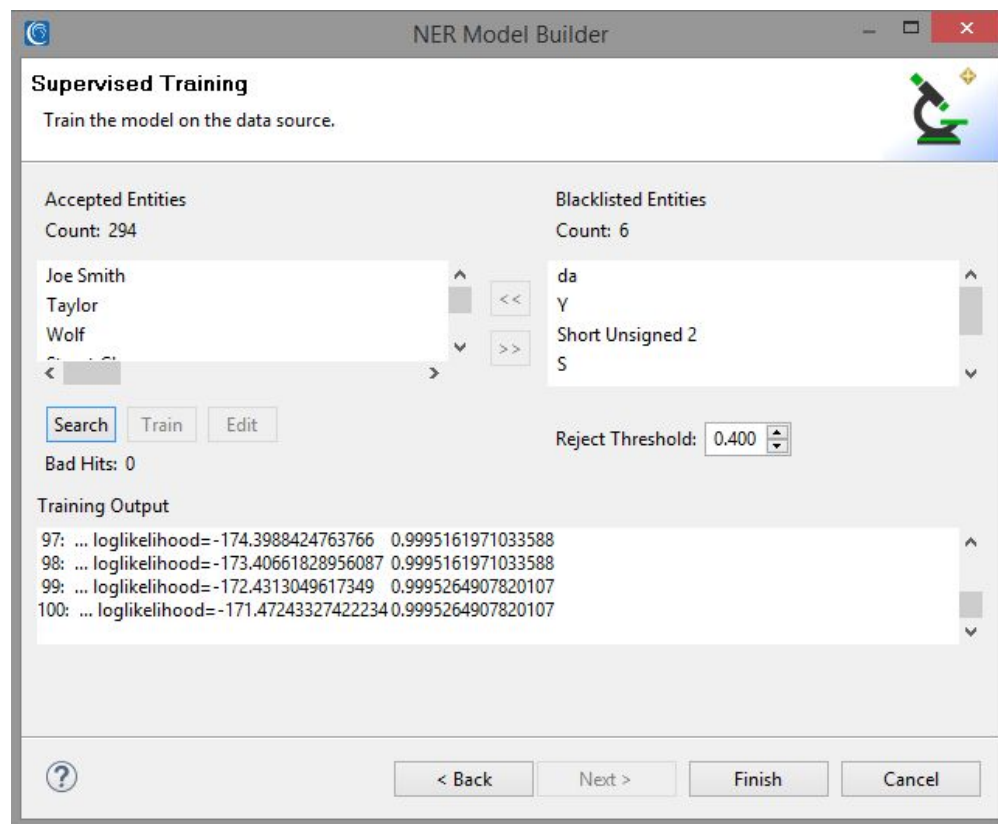
Audit

DarkShield supports both pre-trained OpenNLP Name Finder models or new Named Entity Recognition (NER) models that you can build and train inside its semi-supervised machine learning dialog.

This iterative process improves the accuracy of searches for names and other nouns based on their natural language (grammatical) context in sentences.

Also supported are Tensorflow and PyTorch NER models for richer, multi-language models.

Compare this search method to other DarkShield search methods, like pattern and lookup matches, path filters, or bounding-box areas (for images).



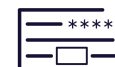
Signature Detection and Redaction



Search



Extract



Redact



Audit



File Specific Configurations

Image Configuration Page

Additional options for searching and masking image files.

OCR Configs | Synthesis Configs | Masking Configs | Bounding Box

Masking method: Black boxes

Match background

Use Signature Detection:

Signature Detection API Key: 19dedb1d2d8c5760b1fc1cf70437

Signature Detection API URL: http://localhost:8080

Signature Detection Minimum Confidence Threshold: 0.85

Search for masked status. Enter either masked, unmasked, or failed.

Finish Cancel



Aggregate Data Source Report

DATUM NAME	DATA CLASS	NUMBER OF MATCHES	MASKED STATUS
file:/D:/eclipse/ReadyEclipse-2023-12/Release/runtime-Workbench/signature-d/darkshield_file_jobs2/input/Cheque 083660_1.jpg	Signature	1	UNMASKED

Callable APIs



Search



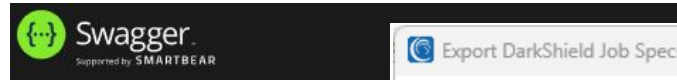
Extract



Redact



Audit



Export DarkShield Job Specs

Export Job Specs to API Context Files

Export job specifications that are serialized in a .dsc file for use with direct DarkShield API calls.

Select .dsc file:

Select target directory:

Enter Prefix:

needs to be performed. The Context contains a list of search matchers that will be used to be created.

application/json

DarkShield API

[/docs/openapi/darkshield.yaml](#)

An RPC service for searching and masking

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
- POST /api/darkshield/maskContext.destroy
- POST /api/darkshield/maskContext.mask
- POST /api/darkshield/searchContext.mask

```

{
  "name": "MixedMatcherContext",
  "matchers": [
    {
      "name": "EmailMatcher",
      "type": "pattern",
      "pattern": "\\b\\w_?+@[\\w.-]+\\.([\\w]{2,4})\\b"
    },
    {
      "name": "NameMatcher",
      "type": "ner",
      "modelUrl": "http://opennlp.sourceforge.net/models-1.5/en-ner-person.bin"
    }
  ]
}
    
```

```

curl -X POST http://localhost:8080/api/darkshield/searchContext.mask -H "accept: application/json" -H "Content-Type: application/json" -d '{"searchContextName":"MixedMatcherContext","maskContextName":"MixedMaskingContext","text":"Hello, my name is John Doe and my email address is johndoe@gmail.com."}'
    
```

Request URL

http://localhost:8080/api/darkshield/searchContext.mask

Server response

Code	Details
200	<p>Response body</p> <pre> { "failedResults": [], "maskedText": "Hello, my name is Wazz Edk and my email address is BqJA0RzCAMdtqXb3tJNBGB/RgNo3y+QKd0MsCjZsgMM=.", "results": [{ "end": 26, "maskedResult": "Wazz Edk", "ruleName": "FpeNameRule", "start": 18 }, { "end": 95, "maskedResult": "BqJA0RzCAMdtqXb3tJNBGB/RgNo3y+QKd0MsCjZsgMM=", "ruleName": "HashEmailRule", "start": 51 }], "unmatchedAnnotations": [] } </pre> <p style="text-align: right;"><input type="button" value="Download"/></p>

Multi- and single (restful) endpoint search/mask context APIs for text, files, RDBs and NoSQL DBs.

Masking PII in Cloud Stores

AWS S3, Azure, GCP, SharePoint, OneDrive



Search



Extract



Redact



Audit

masked-bucket

- OBJECTS
- CONFIGURATION
- PERMISSIONS
- RETENTION
- LIFECYCLE

Buckets > masked-bucket

- UPLOAD FILES
- UPLOAD FOLDER
- CREATE FOLDER
- MANAGE HOLDS
- DOWNLOAD
- DELETE



masked-container
Container

Search (Ctrl+)

- Upload
- Change access level
- Refresh
- Delete
- Change tie

Microsoft Azure
Blob Storage



Authentication method: Access key (Switch to Azure AD User Account)
Location: masked-container

Search blobs by prefix (case-sensitive)

Show deleted blobs

Copy S3 URI

masked/

- Objects
- Properties



Objects (3)
Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

- Copy S3 URI
- Copy URL
- Download
- Open
- Delete
- Actions
- Create folder
- Upload

Find objects by prefix

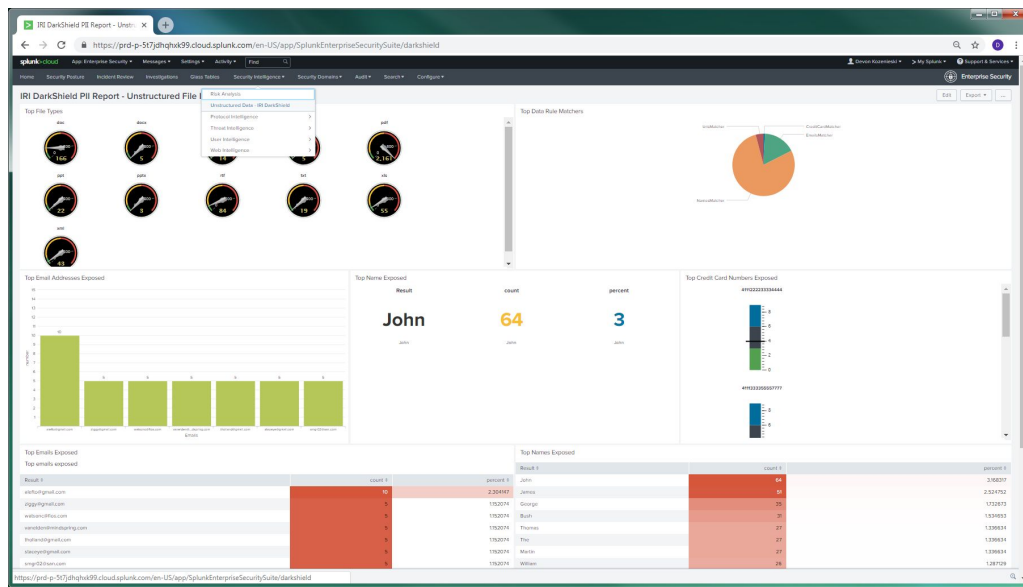
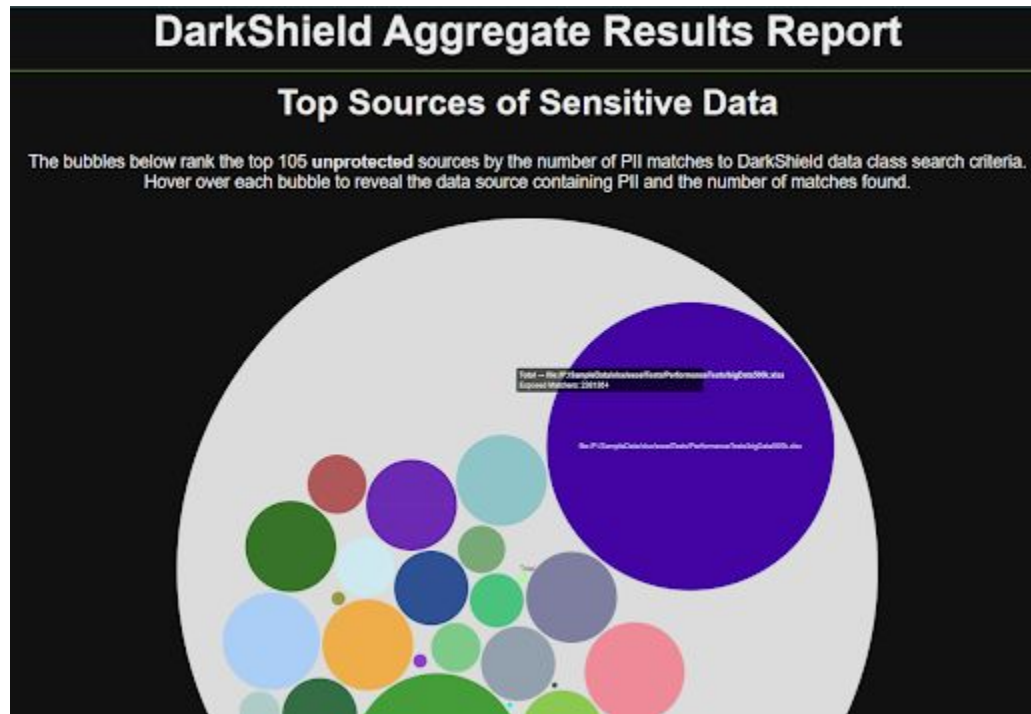
Name	Type	Last modified	Size	Storage class
example.csv	csv	September 23, 2021, 11:33:42 (UTC-04:00)	297.0 B	Standard
test.json	json	September 23, 2021, 11:33:43 (UTC-04:00)	541.0 B	Standard
test.xml	xml	September 23, 2021, 11:33:44 (UTC-04:00)	646.0 B	Standard



Easily query, analyze, and format the results of search and mask operation through built-in reports and this graphical display.

Or, export DarkShield log data for visualizations in BI tools like DataDog, or to SIEM environments like [Splunk ES](#), shown here.

It is also then possible to take actions through the Splunk [Adaptive Response](#) Framework or a Splunk [Phantom](#) [playbook](#).



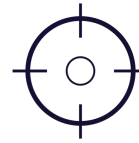


IRI CellShield
PII / PHI Search & Mask in Excel

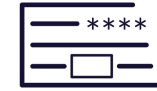
IRI Data Protector Suite

What CellShield EE Does

Note: [FieldShield & DarkShield Support Excel, too.](#)



Search



Mask

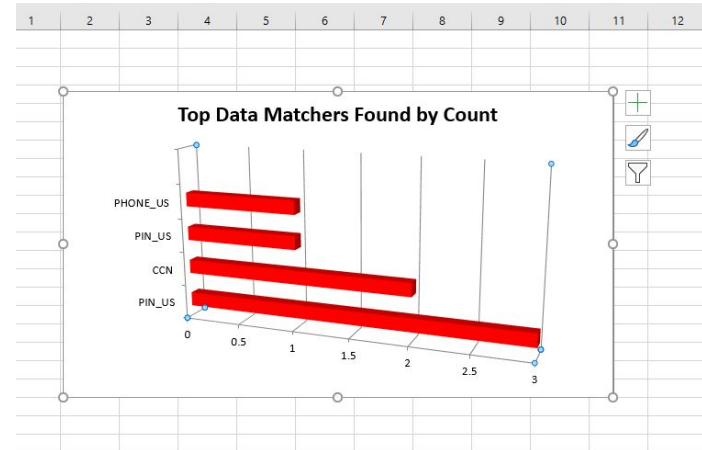


Extract
Report

- Discovers, reports, and masks PII and perform audit actions in Excel 2010 & later
- Searches and secures PII in spreadsheets on one PC or throughout an SMB LAN
- Provides common and allow new search pattern definitions for PII formats
- Searches for strings in a dictionary, and find/fix PII *floating* in cells
- Supports reuse and sharing of patterns in project or cloud repositories
- Generates a report of all patterns found and open it for action in a worksheet
- Opens applicable worksheets and highlights the located ranges for protection
- Encrypts, redacts, or pseudonymizes in one-pass with chosen functions and options
- Reveals data with the decryption key, or if reversible pseudonymization was used
- Overlays results directly into the affected cells, or in another worksheet
- Moves between, or bulk-remediates all, identified worksheets and ranges
- Auto-inserts protection details into an un-editable audit column in the report
- Logging capability is configurable through the user interface, and allows for audit reports, error messages, and selected ranges to be sent to any of the following logging sources: Excel audit column, email, Datadog, Splunk, file system.

CellShield PII Discovery

The dark data (DarkShield) search wizard in IRI Workbench searches network-wide for sensitive data in spreadsheets based on user-specified (plus popular and saved) Java regular expressions (patterns):



New Dark Data Search/Masking Job Search Matcher

Data Sources
Specify the input data sources.

Search Matcher
Match Data Classes or Groups to a Data Rule.

Source URI

- file:/C:/Users/dakoz/Downloads/cellshieldtest-20200706T17062...
- file:/C:/Users/dakoz/Downloads/cellshieldpractice (fileTypes: [Ex...
- file:/C:/Users/dakoz/Downloads (fileTypes: [Excel 2003 (.xls), Exc...

Name:

Description:

Data Class Name: Details: Browse... Create...

Rule Name: Details: Browse... Create...

Filters:

Details	Type

Data Class Selection
Select a Data Class or Data Class Group.

Data Class:

Details	Type
FIRST_NAME	Data Class
FULL_NAME	Data Class
LAST_NAME	Data Class

CellShield Search Report & Action Sheet

The report produced by the profiling wizard opens in a dynamic worksheet supported by an action dialog for protection and auditing activities:

The screenshot shows a Microsoft Excel spreadsheet titled "SearchOutput.eif - Microsoft Excel non-commercial use". The spreadsheet contains a table with the following data:

Include File	File Path	File Name	Sheet Name	Pattern Name(s)	Pattern
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNHSN1.xlsx	NamesNHSN1.xlsx	Sheet1	NHS_Number	\b[0-9]{3}\s[0-9]{3}\s[0-9]{4}\b
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNHSN2.xlsx	NamesNHSN2.xlsx	Sheet1	NHS_Number	\b[0-9]{3}\s[0-9]{3}\s[0-9]{4}\b
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNHSN3.xlsx	NamesNHSN3.xlsx	Sheet1	NHS_Number	\b[0-9]{3}\s[0-9]{3}\s[0-9]{4}\b
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNHSN4.xlsx	NamesNHSN4.xlsx	Sheet1	NHS_Number	\b[0-9]{3}\s[0-9]{3}\s[0-9]{4}\b
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNHSN5.xlsx	NamesNHSN5.xlsx	Sheet1	NHS_Number	\b[0-9]{3}\s[0-9]{3}\s[0-9]{4}\b
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNINo1.xls	NamesNINo1.xls	Sheet1	NINo	\b([A-Z]{2})\s?[0-9]{2}\s?[0-9]{2}\s?
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNINo2.xlsx	NamesNINo2.xlsx	Sheet1	NINo	\b([A-Z]{2})\s?[0-9]{2}\s?[0-9]{2}\s?
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNINo3.xls	NamesNINo3.xls	Sheet1	NINo	\b([A-Z]{2})\s?[0-9]{2}\s?[0-9]{2}\s?
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNINo4.xlsx	NamesNINo4.xlsx	Sheet1	NINo	\b([A-Z]{2})\s?[0-9]{2}\s?[0-9]{2}\s?
<input checked="" type="checkbox"/>	C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNINo5.xls	NamesNINo5.xls	Sheet1	NINo	\b([A-Z]{2})\s?[0-9]{2}\s?[0-9]{2}\s?

Overlaid on the spreadsheet is the "CellShield Spreadsheet Selector" dialog box. It contains the following fields and options:

- Interchange File: C:\Users\rpekarch\Documents\SearchOutput.eif
- Workbook: C:\Users\rpekarch\Documents\Testing Documents for DDD\NamesNHSN1.xlsx
- Sheet Name: Sheet1
- Column: A, Start Row: 3, End Row: 103
- Name of Pattern: NHS_Number, Pattern: \b[0-9]{3}\s[0-9]{3}\s[0-9]{4}\b
- Remediate (protect) all spreadsheets in bulk, or toggle between sheets to protect manually.
- Buttons: Bulk Remediate, Choose Pattern, Choose Protection Type, Open Selected Workbook

Intra-Cell Searching & Masking, Too

- Feature finds and protects floating PII, *ad hoc*, or *en masse*
- Available protections include encryption, masking, and pseudonymization
- Encryption and pseudonymization are reversible through the decryption and recover options, respectively

The screenshot displays the Intra-Cell Search application interface. The main window shows a spreadsheet with the following content:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	People to remember:															
2	161-23-4923	Michael Johns														
3	656-75-8962	Faith Backla														
4	224-23-4923	Michael Bass														
5	194-67-0943	Ziggy Marshall														
6	367-54-2323	George Jeffs														
7																
8	Next are the Netherlands Social Fiscal Numbers:															
9																
10	xxxxxxxxxxxxxxxxxxxxxxxx															
11	xxxxxxxxxxxxxxxxxxxxxxxx															
12	xxxxxxxxxxxxxxxxxxxxxxxx															
13	The Italy Fiscal Codes:															
14																
15																
16	1)	1234567891234567	Abele Achille													
17	2)	1472583690963214	Ada Adalberto													
18	3)	0789456123456782	Adele Adelina													
19																
20	The Korean Social Security Codes are:															
21																
22	1)	040392-5967562	Min Jun													
23	2)	120486-7863214	Ji Hu													
24	3)	321198-2345236	Ji Hoon													
25																
26																
27	xxxxxxxxxxxxxxxxxxxxxxxx															
28	xxxxxxxxxxxxxxxxxxxxxxxx															
29																
30	The leaders include:															
31																
32	0902-5657-4767-1443 - 499-60-0735 - (205) 254-7567 -	Geejohn@etc.com														
33	6593-5841-3058-1791 - 225-76-0934 - (372) 142-6457 -	Geejohn@etc.com														

The application interface includes a menu bar with options: Help Guide, About CellShield, Import EIF File, Import Set File, Intra-Cell Search, Mask Redact, Encrypt & Decrypt, Pseudonymize & Restore, Autoprotect, and Logging. A toolbar contains icons for these functions. A formula bar shows a long string of 'x' characters. A 'Match Count' dialog box is open, displaying 'Found 3 matches in the file.' with 'OK' and 'Cancel' buttons. The 'Intra-Cell Search' dialog box is also open, showing the 'Name of Pattern' as 'Korea SSN' and the 'Pattern' as '\b([0-9]{6}-[0-9]{7})\b'. The 'Whole String' radio button is selected. The 'Mask' section shows 'Encryption/Decryption', 'Pseudonymize', and 'Restore' options, with a 'Masking Character' field and a 'Mask' button.

CellShield Audit Log Options

An uneditable log entry for the function applied to each pattern identified in the report is automatically added onto each action. Based on logging settings, this information may also be sent to a file, to Splunk, to Datadog, or to an email address.

The screenshot shows the Microsoft Excel interface with the CellShield ribbon active. The ribbon includes options for 'Show/Hide Comment', 'Show All Comments', 'Show Ink', 'Unprotect Sheet', 'Protect Workbook', 'Share Workbook', 'Protect and Share Workbook', 'Allow Users to Edit Ranges', and 'Track Changes'. Below the ribbon, a status bar indicates '3:A103 using # for characters 1 to 3, User: tupanu\rpekarch, Time Stamp: 6/16/2015 10:18:44 AM'. A table is displayed with the following columns: F, G, H, I, J, and Audit. The table contains several rows of data, including actions like 'Action: Masked A3:A103 using # for characters 1 to 3, User: tupanu\rpekarch, Time Stamp: 6/16/2015 10:2...' and 'Action: Encrypted G1:G101 using enc_fp_aes256_alphanum, User: tupanu\rpekarch, Time Stamp: 6/16/2015 10:2...'. Below the Excel window, a 'Logging' dialog box is open, showing configuration options for logging to various destinations.

F	G	H	I	J	Audit
	Column	Start Row	End Row	Comment	
-9]{4}\b	A	3	103		Action: Masked A3:A103 using # for characters 1 to 3, User: tupanu\rpekarch, Time Stamp: 6/16/2015 10:2...
-9]{4}\b	G	1	101		Action: Encrypted G1:G101 using enc_fp_aes256_alphanum, User: tupanu\rpekarch, Time Stamp: 6/16/2015 10:2...
-9]{4}\b	G	1	101		Action: Encrypted 3:103 using enc_3des_etc, User: tupanu\rpekarch, Time Stamp: 6/16/2015 10:2...
-9]{4}\b	G	1	101		Action: Encrypted 1:101 using enc_3des_etc, User: tupanu\rpekarch, Time Stamp: 6/16/2015 10:2...
-9]{4}\b	G	1	101		Action: Encrypted 1:101 using enc_3des_etc, User: tupanu\rpekarch, Time Stamp: 6/16/2015 10:2...
-9]{4}\b	G	1	101		Action: Encrypted 1:101 using enc_3des_etc, User: tupanu\rpekarch, Time Stamp: 6/16/2015 10:2...

Logging

File Splunk DataDog Email

from Email

to Email

Enable SSL

Email Subject

Is Body HTML

Mail Server

Username

Password

Port

Batch Posting Limit

Period (Minutes)

RestrictedtoMinimumLevel

Update JSON Configuration

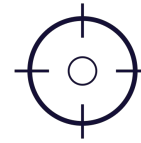
Enable Logging Disable Logging Restore Default Log



IRI DMaaS
Data Masking as a Service

IRI Data Protector Suite

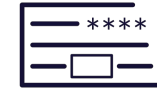
What is DMaaS?



Search



Extract



Mask



Report

- IRI Data Masking as a Service (DMaaS) is **not** SaaS, but professional masking services
- DMaaS makes use of the proven IRI 'shield' software products described above
- Certified IRI experts classify, discover, and de-identify PII of concern in supported silos
- Also available: HIPAA re-ID risk scoring and anonymization, and 'fake PII' for testing
- IRI services are performed under a SoW with NDA, BAA, or other data security terms
- All data access, classification, discovery (search) and masking operations are logged
- The IRI software and your data (in your on-premise or cloud infrastructure) is accessed remotely under your supervision and never transferred outside your firewall/control.
- Billing is hourly or daily, with project rates available; IRI software costs are subsumed in the service fee unless the software will stay behind for later use.

User Profiles

- DBAs and sysadmins responsible for PAN, PHI, PII or other sensitive information
- Sites needing standard data classification and consistent masking functions
- CISOs without sufficient internal IT resources to do this work internally
- Data governance and C-suite officers subject to compliance audits

Use Cases

RBS / Tesco (PCI DSS)

- Produced and implemented custom encryption for testing data in M&A

Confidential (HIPAA)

- Cataloged and de-identified protected health information

University of Adelaide (Privacy Act)

- Data classification, search, and de-identification of PII in massive PeopleSoft financial, HR, and campus test data schemas in Oracle



IRI RowGen
Smart Test Data Generation

*Also available with IRI Data Protector or
Manager Suites, and in the IRI Voracity Platform*

What RowGen Does

- Creates synthetic but realistic random and random-real test data simultaneously
- Improves DB prototypes, application quality, benchmarking, and outsourced operations
- Uses standard DB DDL, production file, and custom metadata to define layouts
- Preserves structural and referential integrity of real EDW DBs for testing
- Produces data in any type, structure, volume, value range, and if condition
- Synthesizes composite data values and custom (master) data formats
- Generates computationally valid and invalid NID (Codice Fiscale, etc.) SSNs, CCNs
- Sets and graphs test data value distributions (linear, normal, random, etc.)
- Applies common attribute rules (like lookups) rules for pattern-matched field names
- Filters, transforms, and pre-sorts test data while it's being generated
- Writes loader metadata and perform direct path loads for test DB populations
- Builds test flat-file and custom/structured detail and summary report targets
- Subsets and masks databases automatically for test purposes
- Provides SDK functions for generating test data in Java apps and Hadoop
- Works with RDB, IRI, and third-party metadata, plus many CI/CD, cloning and TDM tools

Synthesize Data Using DB/File Metadata or DarkShield

Build Test Data into:

- RDBs
- ASN.1 CDRs
- CLF/ELF
- COBOL
- CSV / TSV
- Data Vault 2.0
- FHIR / HL7 / X12*
- Image*
- Excel
- Hadoop
- JSON
- NoSQL DBs
- PDF Forms *
- DOC/PPT(X) *
- XML

* via DarkShield

The screenshot displays the IRI RowGen software interface, which is used for synthesizing test data. The interface is divided into several panes:

- Project Explorer:** Shows the project structure, including files like 'DM.sqlschema', 'Six Table Relational Diagram.png', and 'Entity Relational Diagram'.
- Entity Relational Diagram:** A central diagram showing relationships between tables. Tables include 'COMPANY_NUMBER_UNIQUEcd', 'OU_CODE_UNIQUEcd', 'JOB_CODE_UNIQUEcd', 'EMPLOYEE_ID_UNIQUEcd', 'SSN_UNIQUEcd', 'EESQ_UNIQUEcd', 'SCOTT_DM_COMPANIES', 'SCOTT_DM_ORGANIZATIONAL_UNITS', 'SCOTT_DM_ORGANIZATIONAL_UNITS', 'SCOTT_DM_JOB_TYPES', 'SCOTT_DM_EMPLOYEES', 'SCOTT_DM_EMPLOYEE_EVALUATIONS', and 'SCOTT_DM_EMPLOYEE_EVALUATIONS'. Relationships are indicated by lines and labels like 'In:unique:generation'.
- Log File:** A pane showing the execution log, including the command 'rowgen -f report -i json -o report-in-json.json' and the resulting data.
- Data Report:** A pane showing the output data in a table format, with columns for CLAIM_NUMBER, CREATE_DATE, PATIENT_LAST_NAME, PATIENT_FIRST_NAME, CREDIT_CARD, SSN, PATIENT_ADDRESS, CITY, STATE, DR_NAME.

The log file content is as follows:

```

rowgen -f report -i json -o report-in-json.json
# generated with IRI workbench - New Empty Job
# Author: chaitin
# Created: 2023-09-09 08:10:41
#
# /F:/file-random_file_placeholder
# /PROCESS:RANDOM
# /FIELD:(CLAIM_NUMBER, TYPE=ALPHA_DIGIT, POSITION=1, SEPARATOR=",", SET="claim_number.set" SELECT=ANY, FRAME="")
# /FIELD:(CREATE_DATE, TYPE=ZD_TIMESTAMP, POSITION=2, SEPARATOR=",", SET="(2002-01-01 00:00:00, 2009-12-31 23:59:59)" SELECT=ANY, FRAME="")
# /FIELD:(PATIENT_LAST_NAME, TYPE=ALPHA_DIGIT, POSITION=3, SEPARATOR=",", SET="name.last.set" SELECT=ANY, FRAME="")
# /FIELD:(PATIENT_FIRST_NAME, TYPE=ALPHA_DIGIT, POSITION=4, SEPARATOR=",", SET="name.first.set" SELECT=ANY, FRAME="")
# /FIELD:(CREDIT_CARD, TYPE=ALPHA_DIGIT, POSITION=5, SEPARATOR=",", SET="credit_card.set" SELECT=ANY, FRAME="")
# /FIELD:(SSN, TYPE=ALPHA_DIGIT, POSITION=6, SEPARATOR=",", SET="ssn.set" SELECT=ANY, FRAME="")
# /FIELD:(PATIENT_ADDRESS, TYPE=ALPHA_DIGIT, POSITION=7, SEPARATOR=",", SET="street_addresses.set" SELECT=ANY, FRAME="")
# /FIELD:(CITY, TYPE=ALPHA_DIGIT, POSITION=8, SEPARATOR=",", SET="city.set" SELECT=ANY, FRAME="")
# /FIELD:(STATE, TYPE=ALPHA_DIGIT, POSITION=9, SEPARATOR=",", SET="state.city.set" SELECT=ANY, FRAME="")
# /FIELD:(DR_NAME, TYPE=ALPHA_DIGIT, POSITION=10, SEPARATOR=",", SET="(N/A)", FRAME="")
# /REPORT
# /OUTFILE:report.csv
# /PROCESS:CSV
# /FIELD:(CLAIM_NUMBER, TYPE=ALPHA_DIGIT, POSITION=1, SEPARATOR=",", FRAME="", COEF="CLAIM_NUMBER")
# /FIELD:(CREATE_DATE, TYPE=ZD_TIMESTAMP, POSITION=2, SEPARATOR=",", FRAME="", COEF="CREATE_DATE")
# /FIELD:(PATIENT_LAST_NAME, TYPE=ALPHA_DIGIT, POSITION=3, SEPARATOR=",", FRAME="", COEF="PATIENT_LAST_NAME")
# /FIELD:(PATIENT_FIRST_NAME, TYPE=ALPHA_DIGIT, POSITION=4, SEPARATOR=",", FRAME="", COEF="PATIENT_FIRST_NAME")
# /FIELD:(CREDIT_CARD, TYPE=ALPHA_DIGIT, POSITION=5, SEPARATOR=",", FRAME="", COEF="PATIENT_CREDIT_CARD")
# /FIELD:(SSN, TYPE=ALPHA_DIGIT, POSITION=6, SEPARATOR=",", FRAME="", COEF="PATIENT_SSN")
# /FIELD:(PATIENT_ADDRESS, TYPE=ALPHA_DIGIT, POSITION=7, SEPARATOR=",", FRAME="", COEF="PATIENT_ADDRESS")
# /FIELD:(CITY, TYPE=ALPHA_DIGIT, POSITION=8, SEPARATOR=",", FRAME="", COEF="CITY")
# /FIELD:(STATE, TYPE=ALPHA_DIGIT, POSITION=9, SEPARATOR=",", FRAME="", COEF="STATE")
# /FIELD:(DR_NAME, TYPE=ALPHA_DIGIT, POSITION=10, SEPARATOR=",", FRAME="", COEF="DR_NAME")
# /OUTFILE:report-in-json.json
# /PROCESS:JSON
# /FIELD:(CLAIM_NUMBER, TYPE=ALPHA_DIGIT, POSITION=1, SEPARATOR=",", FRAME="", COEF="PATIENT.CLAIM_NUMBER")
# /FIELD:(CREATE_DATE, TYPE=ZD_TIMESTAMP, POSITION=2, SEPARATOR=",", FRAME="", COEF="PATIENT.CREATE_DATE")
# /FIELD:(PATIENT_LAST_NAME, TYPE=ALPHA_DIGIT, POSITION=3, SEPARATOR=",", FRAME="", COEF="PATIENT.LAST_NAME")
# /FIELD:(PATIENT_FIRST_NAME, TYPE=ALPHA_DIGIT, POSITION=4, SEPARATOR=",", FRAME="", COEF="PATIENT.FIRST_NAME")
# /FIELD:(CREDIT_CARD, TYPE=ALPHA_DIGIT, POSITION=5, SEPARATOR=",", FRAME="", COEF="PATIENT.CREDIT_CARD")
# /FIELD:(SSN, TYPE=ALPHA_DIGIT, POSITION=6, SEPARATOR=",", FRAME="", COEF="PATIENT.SSN")
# /FIELD:(PATIENT_ADDRESS, TYPE=ALPHA_DIGIT, POSITION=7, SEPARATOR=",", FRAME="", COEF="PATIENT.ADDRESS")
# /FIELD:(CITY, TYPE=ALPHA_DIGIT, POSITION=8, SEPARATOR=",", FRAME="", COEF="PATIENT.CITY")
# /FIELD:(STATE, TYPE=ALPHA_DIGIT, POSITION=9, SEPARATOR=",", FRAME="", COEF="PATIENT.STATE")
# /FIELD:(DR_NAME, TYPE=ALPHA_DIGIT, POSITION=10, SEPARATOR=",", FRAME="", COEF="PATIENT.DR_NAME")

```

The data report shows the following results:

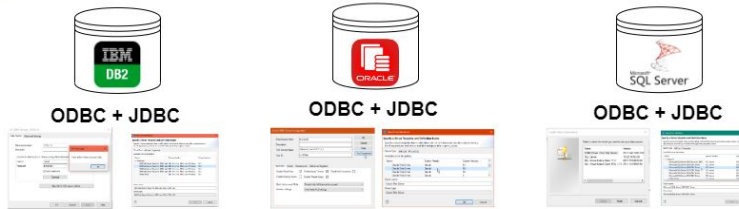
Type	Status	Result
1	Succ	1 20090346133 1982-10-11 Glenn Renee 5314-7253-5 00 1501 Lowry Ave N Fo Mar None
2	Succ	2 2061308382 1959-10-12 Cross Altan 2547-2695-3 68 810 Race St Sp Ore None
3	Succ	3 2170025242 1911-12-18 Wiser Horacio 6297-3562-8 75 2455 Race Garden Se Calif None
4	Succ	4 1842014240 1976-10-11 Tran Evanette 3163-8427-3 00 4015 Everlock Dr Ge None

The console shows the following output:

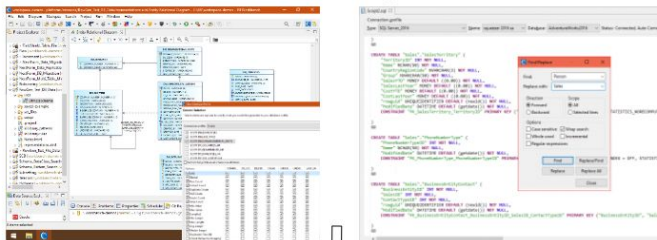
```

Total 20 records shown

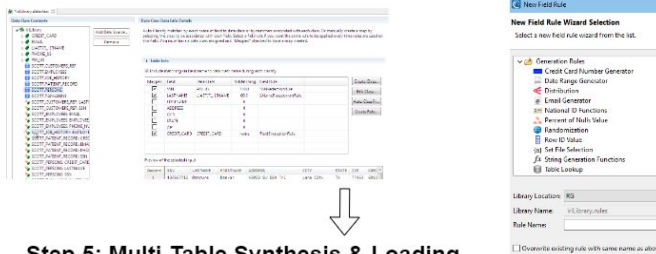
```



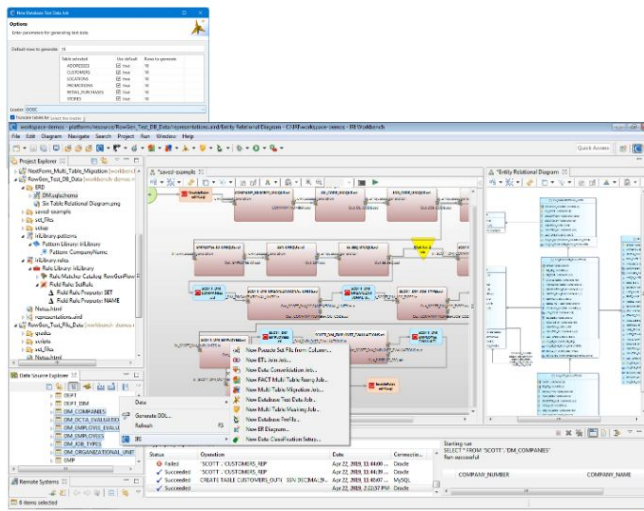
Steps 2 & 3: Production Profiling & Test Table Creation



Steps 3 & 4: Data Class & Generation Rule Specification



Step 5: Multi-Table Synthesis & Loading



IRI RowGen
Workflow

DB Subsetting, Masking Optional

Included table subsetting and test data generation wizards facilitate DB and EDW prototyping, as well as test data virtualization for DevOps. Masked and referentially-correct copies of production table extracts ensure that production data is safe and test data is realistic. Run these batch jobs from IRI Workbench or Value Labs TDH, the command line, or Windocks.

The image displays three screenshots of the IRI RowGen 'Subset Job' wizard interface, illustrating the process of creating a subset job.

Screenshot 1: Data Extraction Stage
 This window shows the 'Data Extraction Stage' where tables are selected for extraction. The 'Tables selected:' section lists several tables, with 'SCOTT.SALE_DIM' checked. The 'Connection profile:' is set to 'Oracle Local'. The 'Sort' section allows for expanding and selecting fields to sort, with 'Input Fields' including 'Infile SCOTT.SALE_DIM' and its alias 'SCOTT_SALE_DIM', along with 'Process ODBC', 'SALE_ID', 'SALE_DATE', and 'QTY_SOLD'.

Screenshot 2: Data Loading
 This window shows the 'Data Loading' stage where subset targets and selected tables are viewed. The 'Connection profile:' is 'Oracle Local', the 'Schema name:' is 'SCOTT', and the 'Output mode:' is 'CREATE'. Options include 'Disable direct path load' (unchecked), 'Temporarily disable foreign keys on tables before insert' (checked), and 'Include a drop table script for tables created in this job' (unchecked). The 'Naming options' section allows for selecting a naming option and entering a prefix, postfix, or name.

Screenshot 3: Field Modification Rules and Summary
 This window shows the 'Field Modification Rules' section, where a rule is defined for the 'Emp' field. The rule is: 'MASK_\${FIELDNAME}' with a type of 'EXPRESSION, ...' and a detail of 'EXPRESSION = replace_chars(\${FIELDNAME}, "*"')'. The 'Separator:' is set to '\t'. The 'Summary' section provides a summary of the batch file and scripts to be generated, including the extractor (ODBC), DSN Name (Oracle_Local12_32), vendor (Oracle), provider (Oracle), host (pathogen:1522), instance (ORA12), and user (scott). The 'Rule info:' section lists the rule name (FieldRedactionRule) and the tables to be created: SCOTT.CATEGORY_DIM, SCOTT.DEPT_DIM, SCOTT.EMP_DIM, SCOTT.EMP_SALARY_RANGE_DIM, SCOTT.ITEM_DIM, SCOTT.ITEM_PRICE_RANGE_DIM, and SCOTT.PROJECT_DIM. The 'Source tables:' section lists the tables that do not exist in the load target and will be created: SCOTT.SUB_SALE_DIM, SCOTT.SUB_CATEGORY_DIM, SCOTT.SUB_DEPT_DIM, SCOTT.SUB_EMP_DIM, SCOTT.SUB_ITEM_DIM, SCOTT.SUB_ITEM_PRICE_RANGE_DIM, and SCOTT.SUB_PROJECT_DIM. The 'Table Dependency Tree:' lists the tables and their dependencies: SCOTT.SALE_DIM, SCOTT.CATEGORY_DIM, SCOTT.DEPT_DIM, SCOTT.EMP_DIM, SCOTT.EMP_SALARY_RANGE_DIM, SCOTT.ITEM_DIM, SCOTT.ITEM_PRICE_RANGE_DIM, and SCOTT.PROJECT_DIM. The 'Loader:' is ODBC, the 'Vendor:' is Oracle, the 'Provider:' is Oracle, the 'Host:' is pathogen:1522, the 'Instance:' is ORA12, and the 'User:' is scott. The 'Schema:' is SCOTT.

User Profiles

Anyone doing DB testing, app development, stress-testing, or benchmarking, including:

- Developers (programmers)
- DBAs and DW (ETL) architects
- Analysts and consultants

Use Cases

Bank of Montreal

- Generates safe, realistic 20GB Oracles tables with RI for query testing

MasterCard Peru

- Synthesizes PAN and PII in files to support OLTP and app testing

Transitive UK

- Simultaneously creates and transforms data to test cross-OS virtualization

Key Differentiators

1. Big data generation and population performance for flat files, RDB and NoSQL DBs, Data Vault V2, HL7/X12 EDI files, ASN.1 CDR files, XLS/X spreadsheets, and even images and documents (in conjunction with DarkShield)
2. Embedded CoSort pre-sorting engine speeds VLDB loads
3. Synthetic data that's broader and safer than real data via multiple methods: <https://www.iri.com/blog/data-protection/making-realistic-test-data-production>
4. *Concurrent* test data manipulation (transformation) and custom report outputs
5. Simple, portable, and modifiable test data generation and auto-built DB loader scripts, all managed visually in Eclipse, and easily integrated into TDM pipelines and products (see next slide)
6. Metadata compatibility with IRI DDF, erwin SmartConnectors, and MIMB: to facilitate test data generation for 3rd-party BI, CRM, and ETL tools

TDaaS & TDM Options

1. Test Data as a Service (TDaaS), a remotely provided professional engagement leveraging RowGen or any of the data masking and subsetting features described above to provide highly customized test data without licensing or learning new technology.
2. Run IRI CLI jobs in CI/CD pipelines like [Jenkins](#), [GitLab](#), [Azure DevOps](#), [AWS](#), etc.
3. Run IRI jobs with these DB virtualization tools, which call our scripts at cloning time:
 - a. [Actifio](#)
 - b. [Commvault](#)
 - c. [Windocks](#)
4. On-demand TDM web apps are tightly integrated with IRI software too, including:
 - a. Cigniti BlueSwan
 - b. ValueLabs Test Data Hub (TDH)



Learn and Share

[IRI Data Masking Solutions](#)

[Data Masking How-to Articles](#)

[LinkedIn Data Masking Group](#)

[IRI Mask/Test Tech Talk Videos](#)

[IRI Test Data Solutions](#)

[Test Data How-to Articles](#)

[LinkedIn Test Data Group](#)

[Voracity Platform Resources](#)

