Data Protector Suite 2025

Data Protector Suite

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

database
TRENDS AND APPLICATIONS
TREND-SETTING PRODUCTS

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

IRI
Total Data Management

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

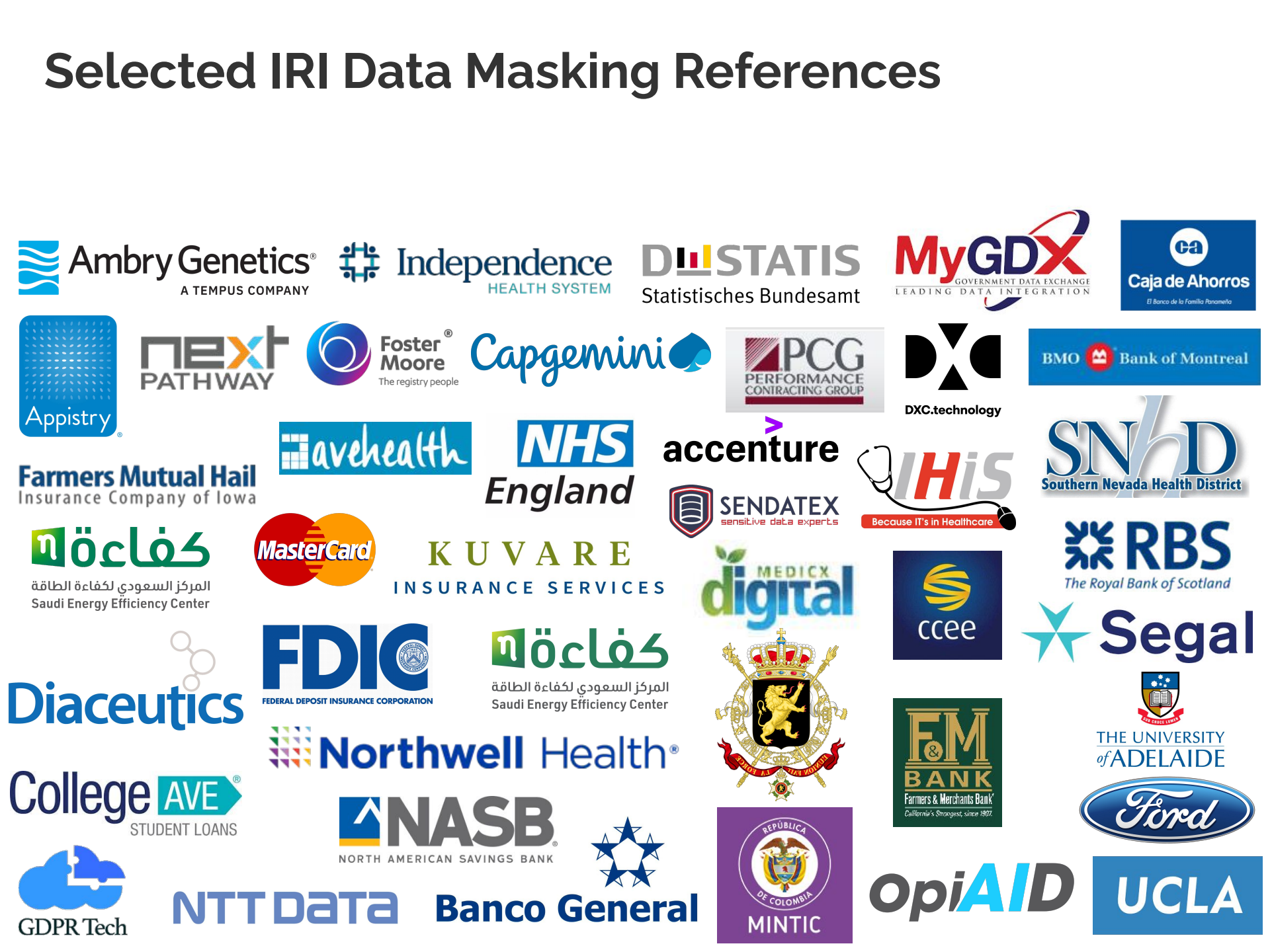
- Kafka & MQTT
- BI & Analytic Tools
- Cloud Stores
- Relational DBs
- NoSQL DBs
- Custom Reports
- DevOps
- Flat & EDI Files
- Logs, Excel, Images

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

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
Total Data Management

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

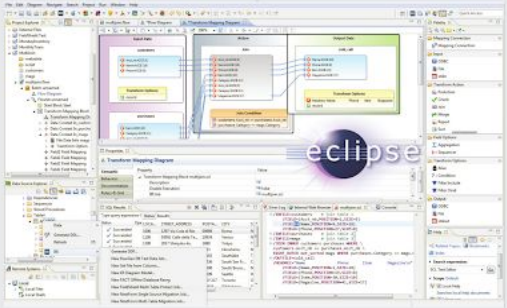
- Convert, federate, remap, and replicate legacy data
- Migrate data between databases and create new tables
- Change file formats, data types, and endian conditions
- Find, extract, and structure data in unstructured sources

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

© 2024 Innovative Routines International (IRI), Inc. All Rights Reserved.
CoSort, Voracity, NextForm, DarkShield, FieldShield, CellShield, Ripcurrent and RowGen
are trademarks or registered trademarks of IRI. FACT is a trademark of DataStreams Corp.
All other product, brand, or company names may be (registered) trademarks of their respective holders.

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

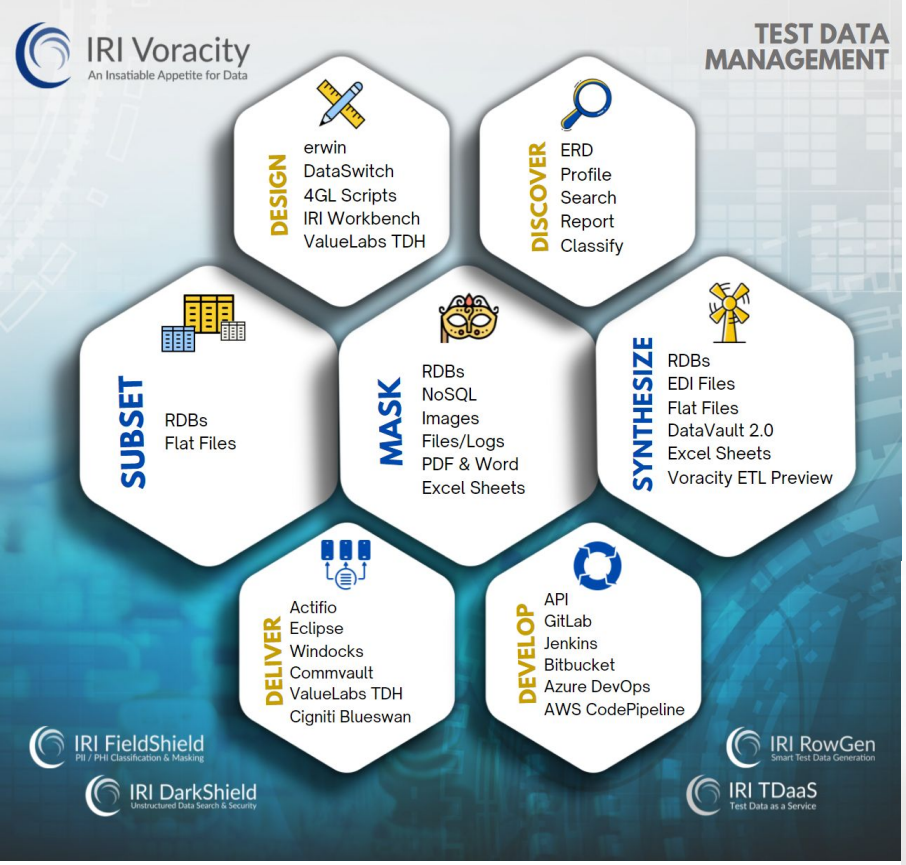
- Uses same data classes, masking rules, and GUI as FieldShield, but on semi- & unstructured data, too
- Offers multiple AI search matchers, plus multi-threading and load balancing to scale performance
- Supports special use cases like EDI loops, conditions, signatures, test documents, and textual ETL

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.



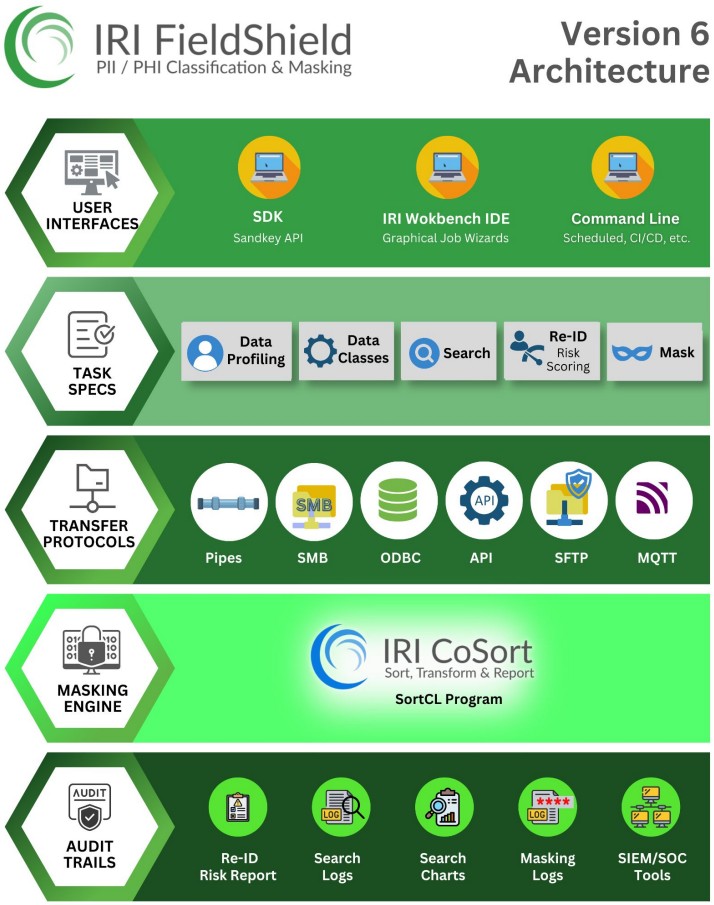
All these jobs can be designed, modified, shared, and run graphically in a rich design client, called IRI Workbench, built on Eclipse. This front-end also serves as a multi-DB administration hub and IDE for IRI and 3GL jobs, and hosts many free third party plugins like Git and Apache DS.

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

IRI
Total Data Management

© 2022 Innovative Routines International (IRI), Inc. All Rights Reserved.

IRI Data Masking Tool Architectures

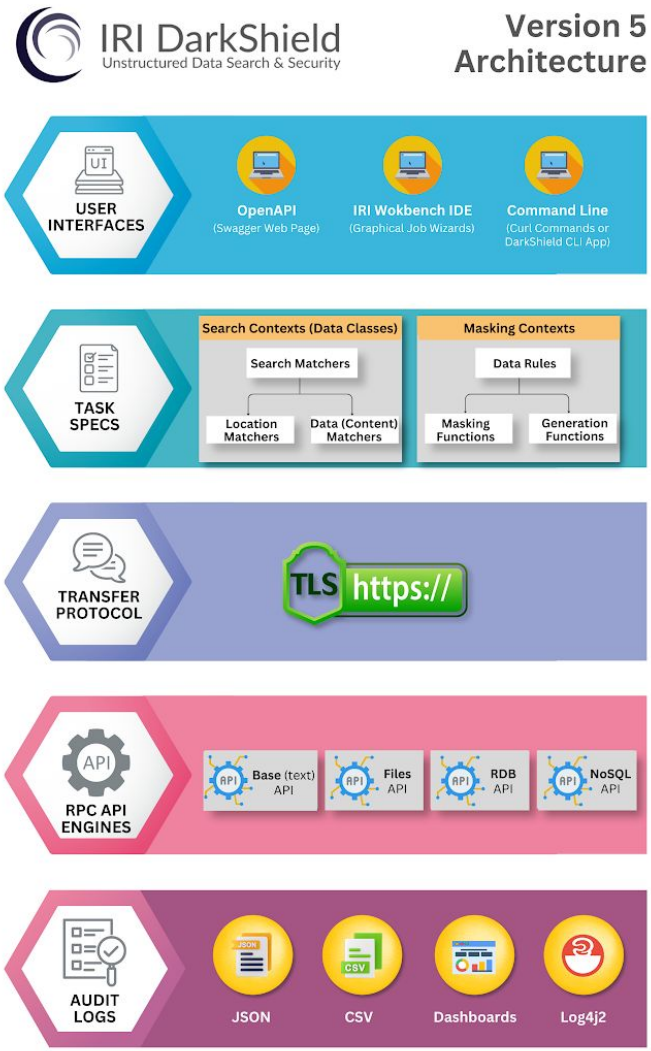


Both structured and unstructured data discovery functions -- including classification, search, and metadata creation -- are performed in IRI Workbench data discovery wizards.

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.

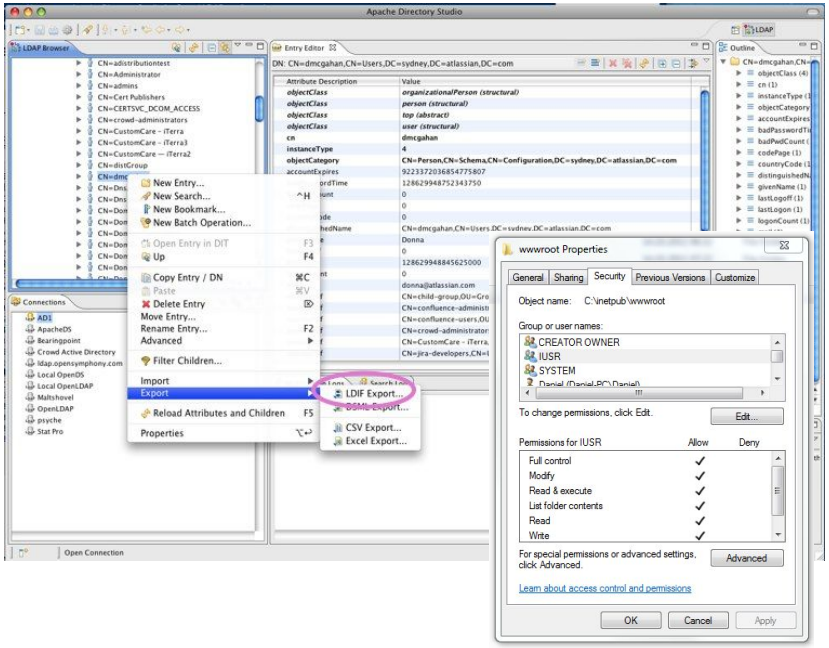
Finally, no server framework is required to orchestrate jobs. This fosters more granular allocation and tuning of resources, and the sharing of metadata artifacts.



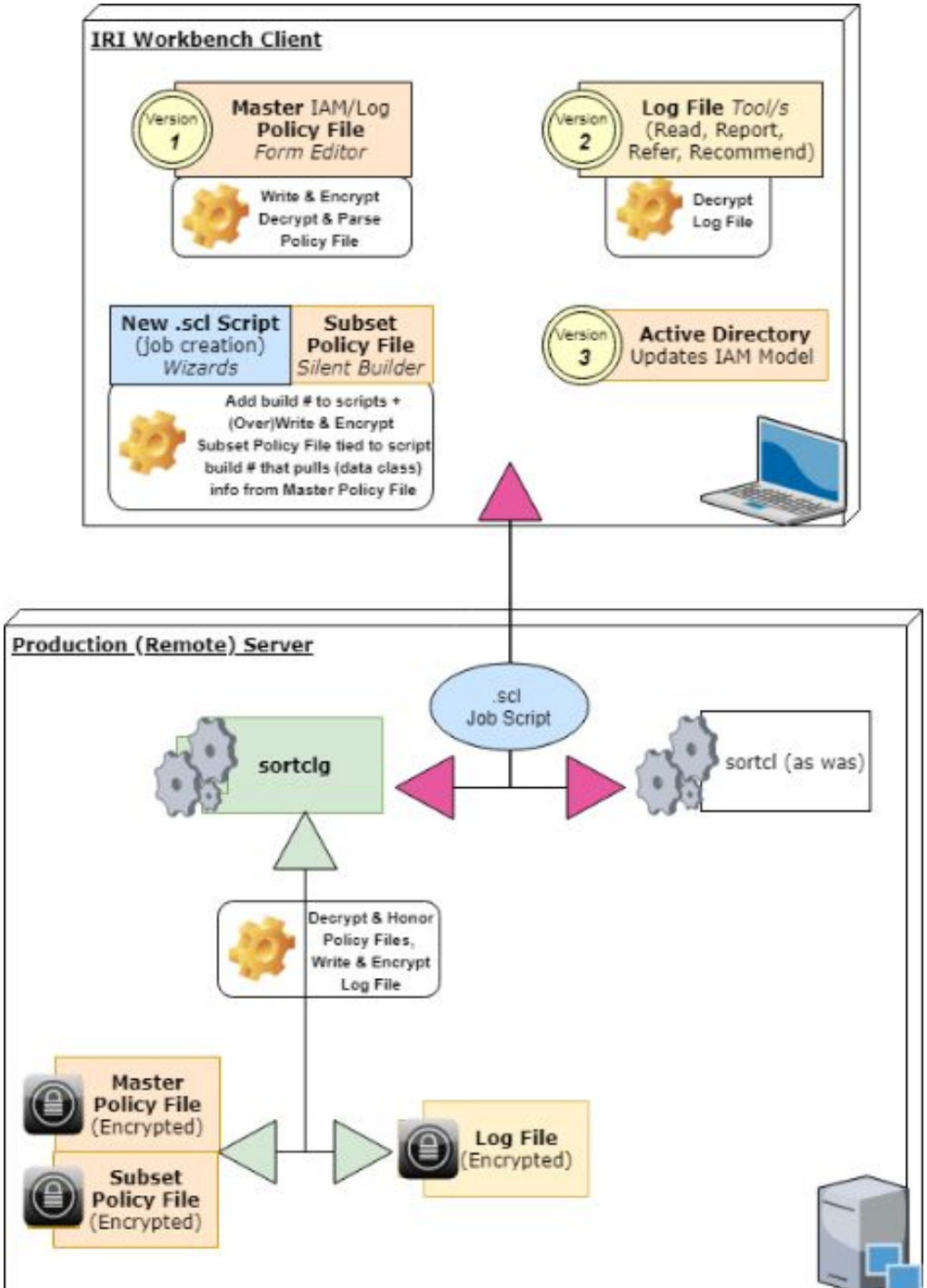
IRI
Total Data Management

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:

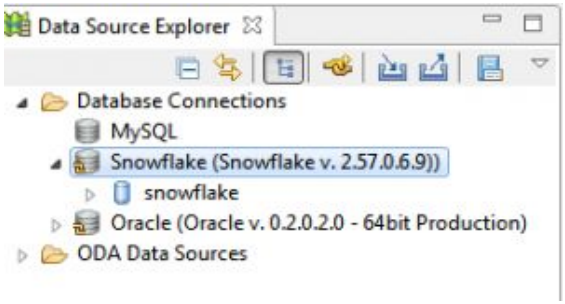


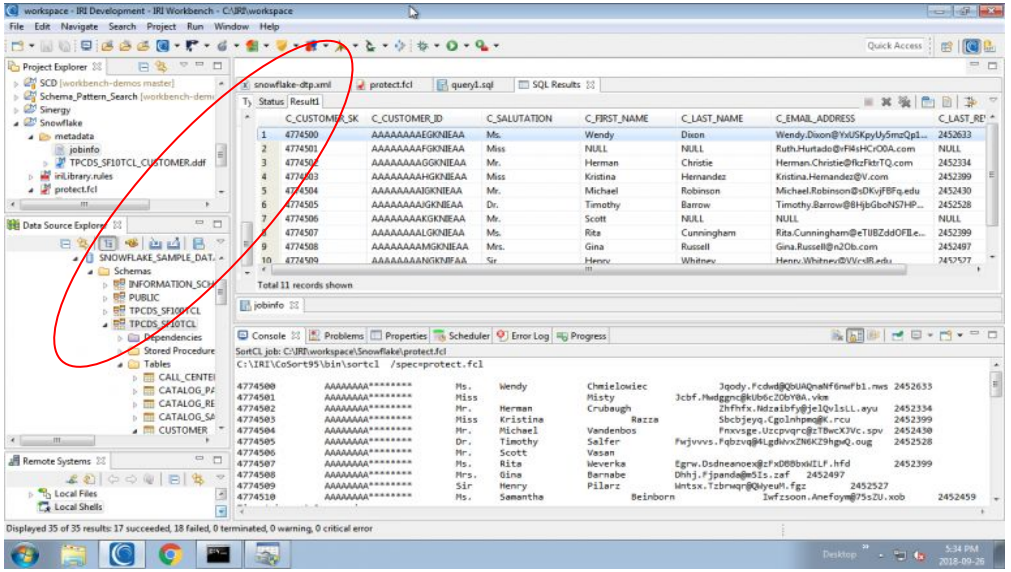
In late 2025, the IRI client/server governance system illustrated on the right will all you to assign and enforce RBACs to the same elements above, **and** to more granular elements like field names (mapped from data classes), functions, and perhaps specific data values (or ranges of values).

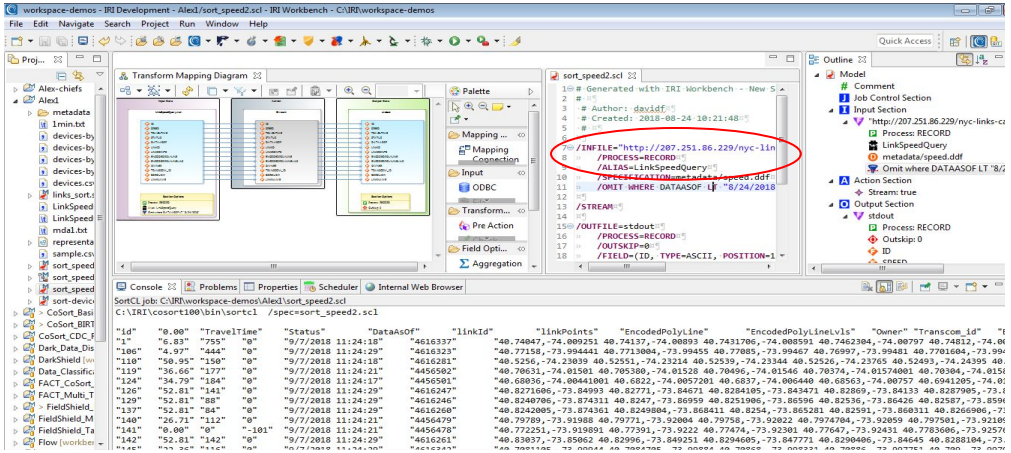


Cloud Data & Systems Support

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.







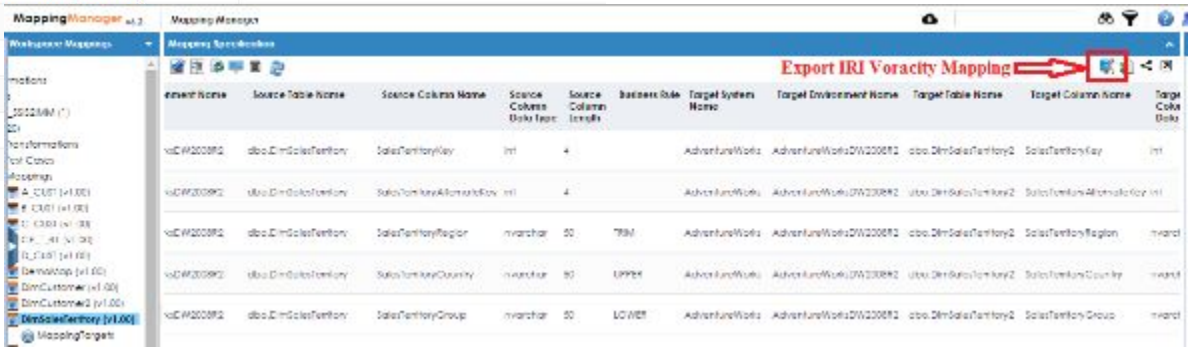
IRI
Total Data Management

Metadata Integrations

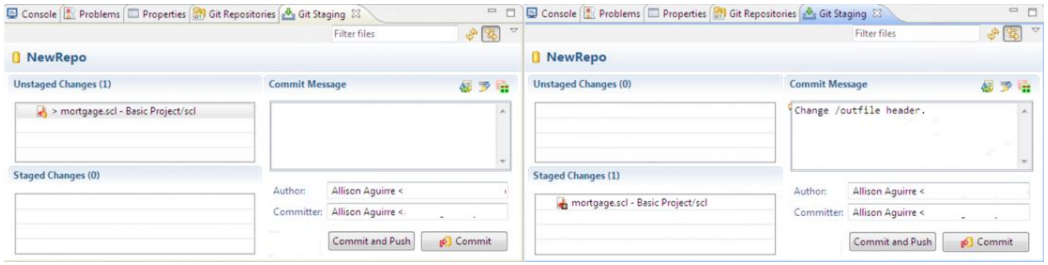
1. Voracity tooling *consumes* metadata from any structured source for data classification, profiling, search, de-ID, ETL etc.
2. FieldShield & RowGen job scripts also *produce* metadata for several **DB load utilities** in multi-DB masking & test data jobs.
3. Their data definition file metadata can 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, worfklows and metamodels, discovery configurations, search matchers and masking rules -- can also be team shared, secured and version controlled in **Git** et al



IRI
Total Data Management

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.*

IRI
Total Data Management

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*

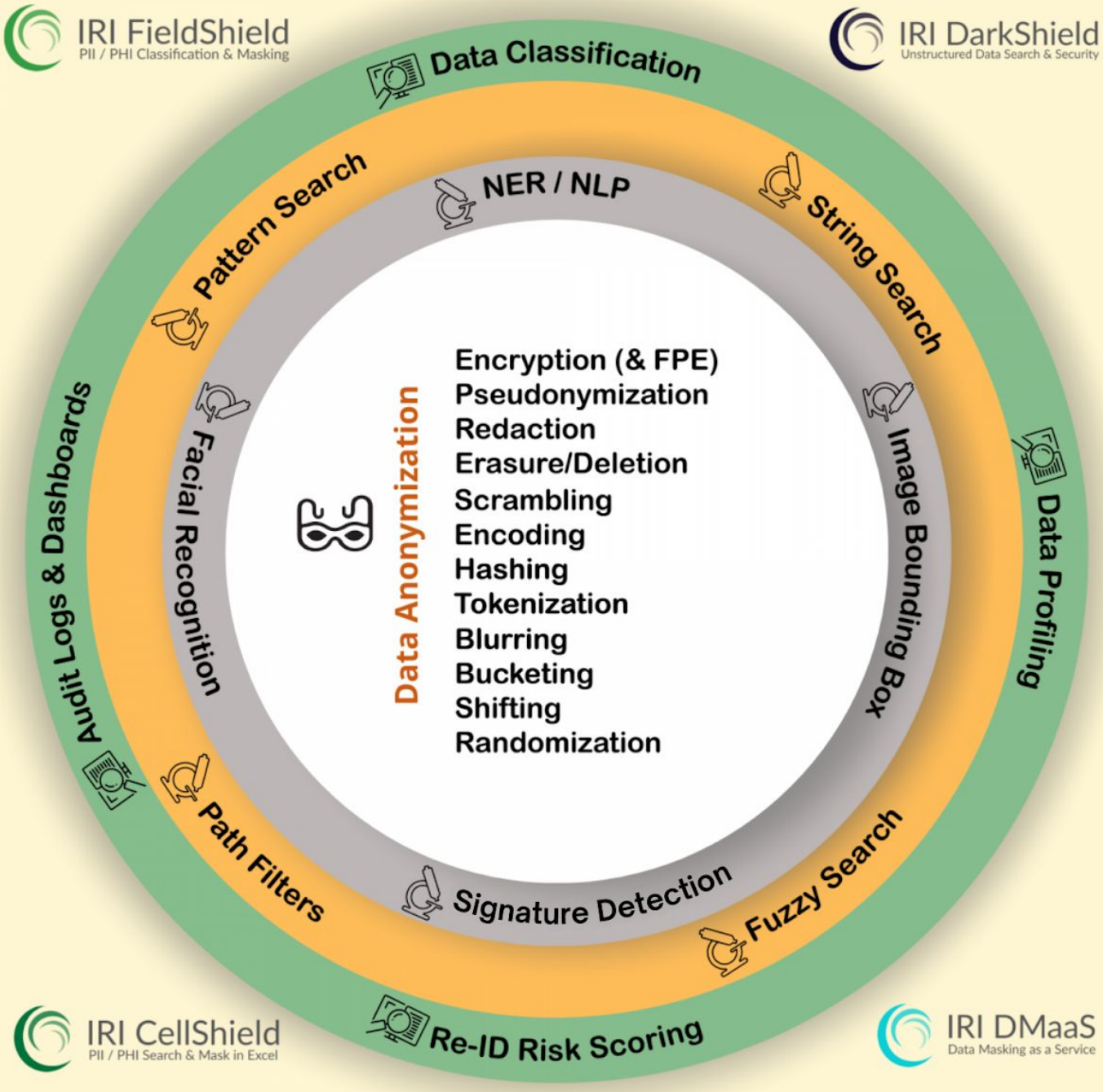
IRI
Total Data Management

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; CLOB columns in RDBs; PDF and Parquet, 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
Total Data Management


IRI FieldShield
PII / PHI Classification & Masking
IRI DarkShield
Unstructured Data Search & Security
Data Classification
Pattern Search
NER / NLP
String Search
Audit Logs & Dashboards
Facial Recognition
Image Bounding Box
Data Profiling
Data Anonymization
Encryption (& FPE)
Pseudonymization
Redaction
Erasure/Deletion
Scrambling
Encoding
Hashing
Tokenization
Blurring
Bucketing
Shifting
Randomization
Path Filters
Signature Detection
Fuzzy Search
Re-ID Risk Scoring
IRI CellShield
PII / PHI Search & Mask in Excel
IRI DMaaS
Data Masking as a Service
IRI
Total Data Management

IRI FieldShield

PII / PHI Classification & Masking

IRI Data Protector Suite

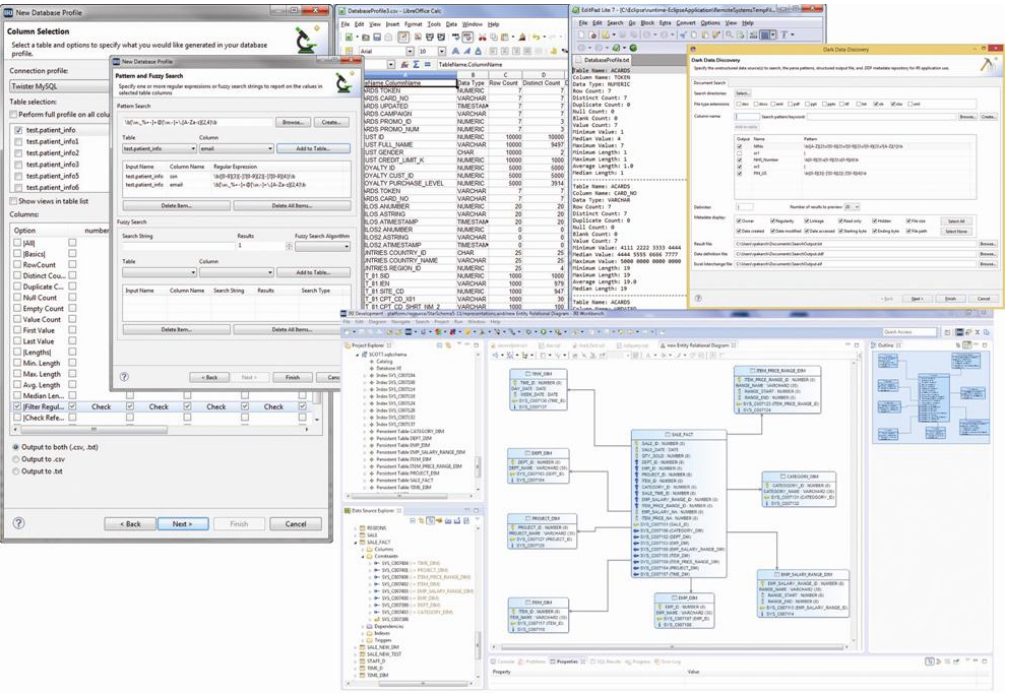
What FieldShield Does

- Finds and masks data in structured (flat) files, RDBs, Excel and ASN.1-encoded CDR files
- Supports complex business logic for custom masking needs, and special formats like ACH
- 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.

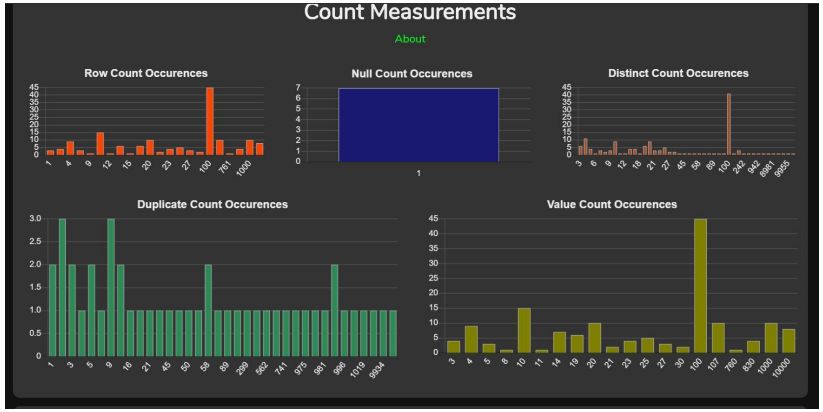
Fit-for-purpose GUI wizards deliver:

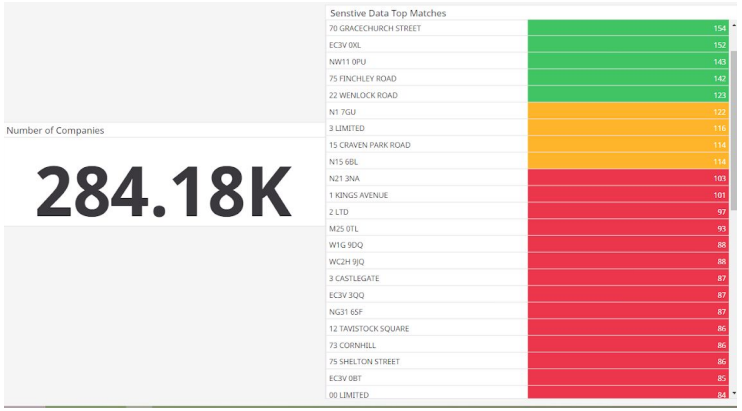


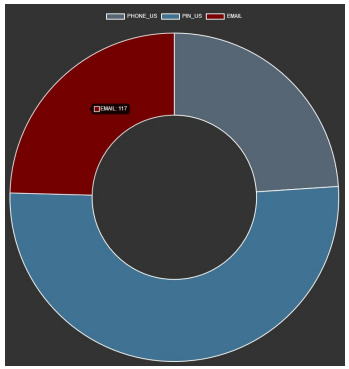
- DB and file data classification, with search and masking rule matchers
- DB profiling, ERDs, and table searches
- Flat-file profiling and value searches
- Data class searches through schema and directories for bulk discovery
- Metadata discovery and definition
- Dark data search and structuring, with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

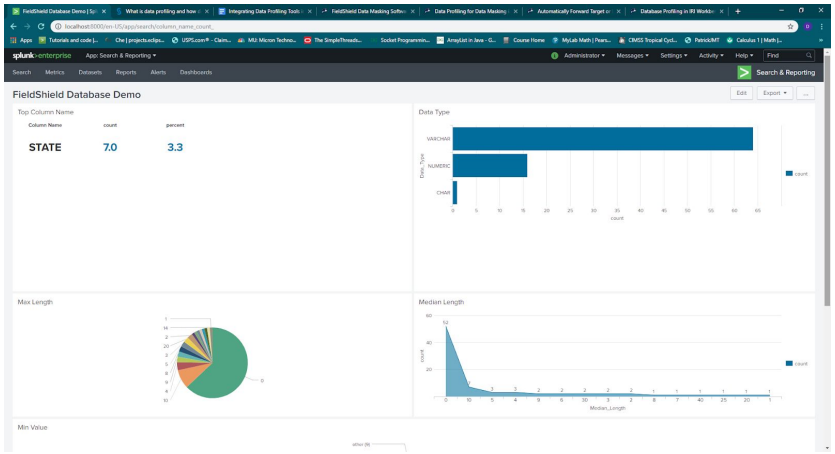
Search Result Reports, Dashboards & Exports

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:









IRI
Total Data Management

Multiple Masking Job Design Options

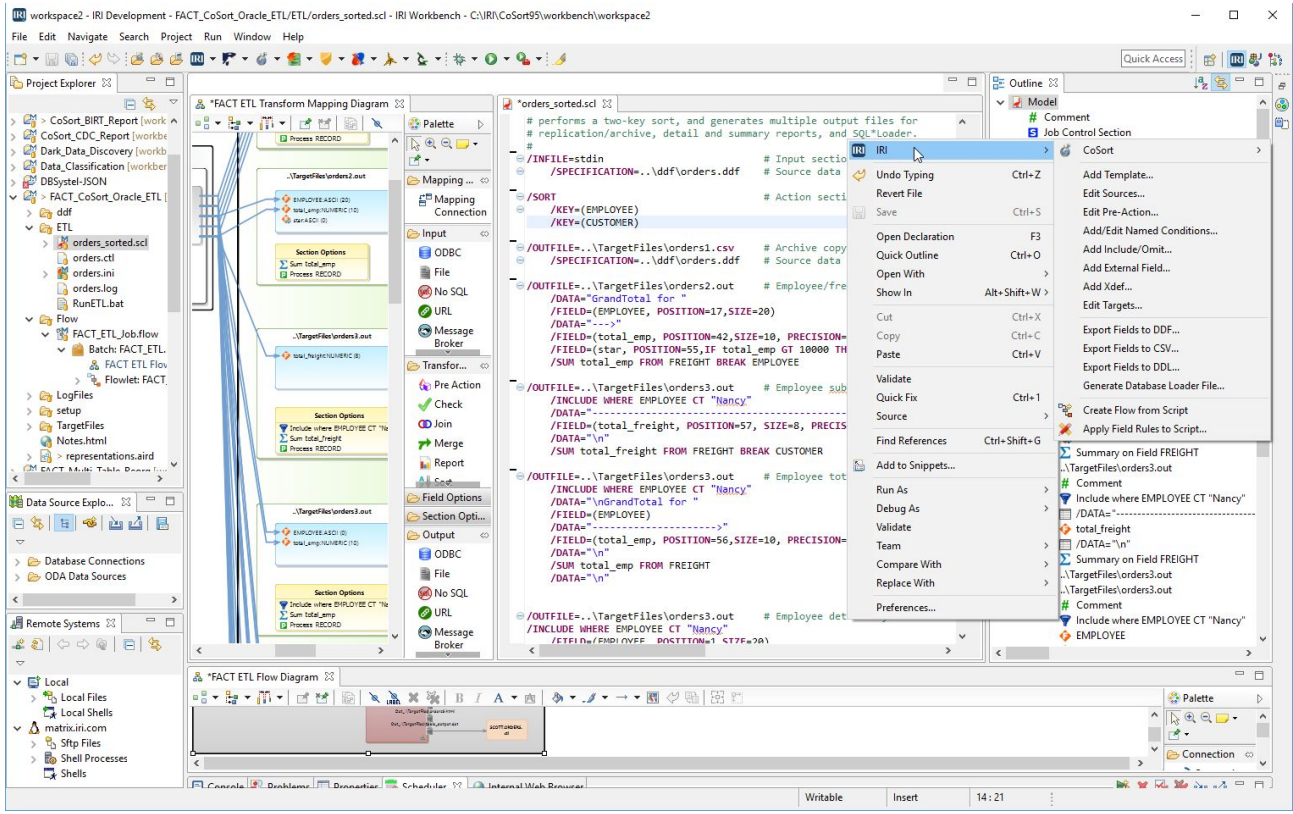
IRI FieldShield and other Voracity data masking, cleansing, transformation, migration, reporting, and wrangling jobs can be created and run *inside or outside* of IRI Workbench.

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**

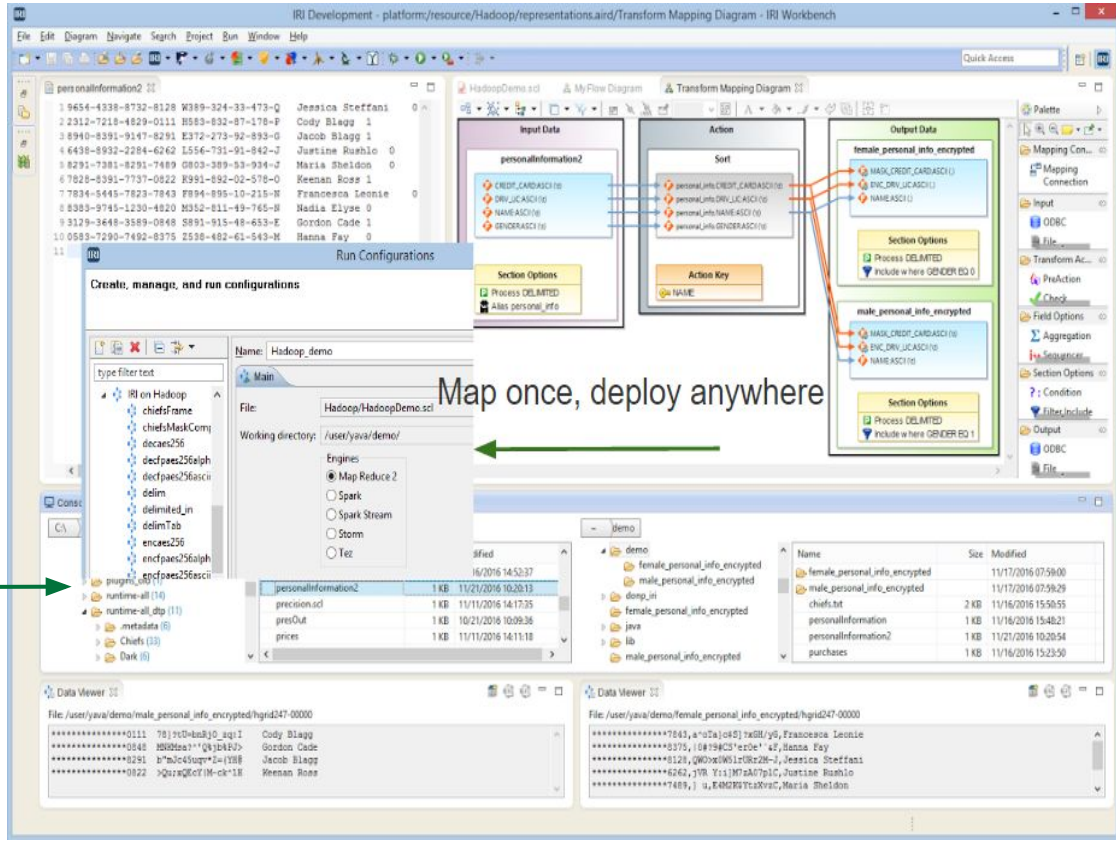
Job design methods supported *outside*:

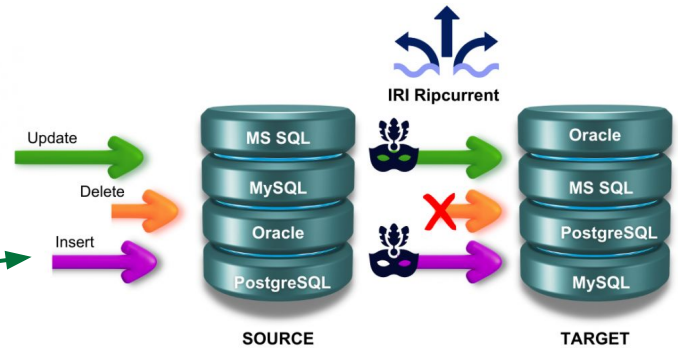
6) **DataSwitch no-code web app**
7) **erwin Mapping Manager**
8) **Value Labs Test Data Hub**
9) **Any external text editor**
10) **3GL app (system or API calls)**



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





IRI FieldShield Workflow

IRI FieldShield
PII / PHI Classification & Masking

Typical Workflow

IRI Workbench
Ergonomic Job Design Built on Eclipse™

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

JDBC produces IRI DDF Metadata, and ODBC moves data to/from mask engine

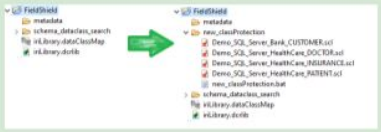
Create DDF Layouts in IRI Workbench Metadata Discovery Wizard

Step 2: **Define** Data Classes & Masking Rules (.dcrl 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

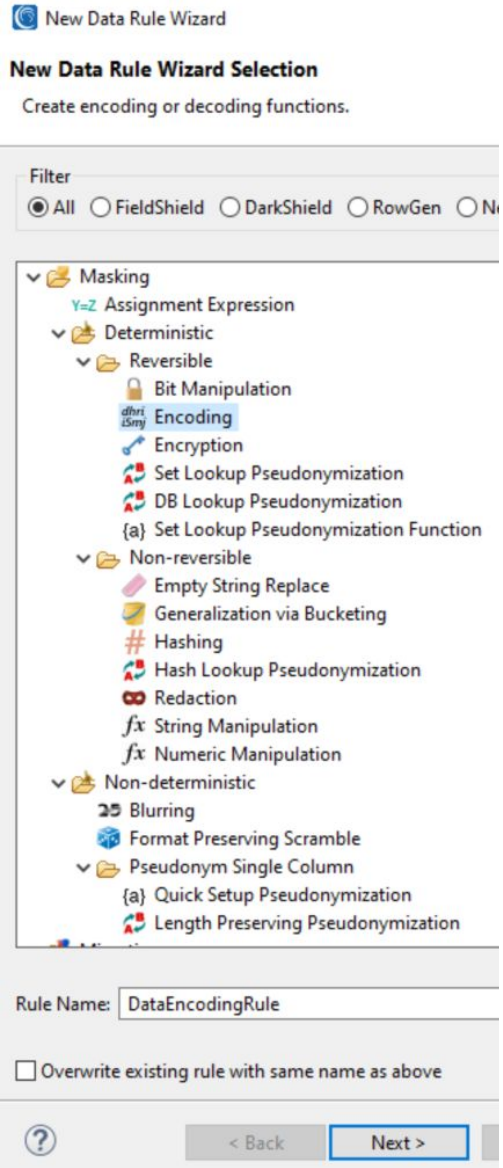
Step 5: **Mask** the RDB Schema or Files with Referential Integrity



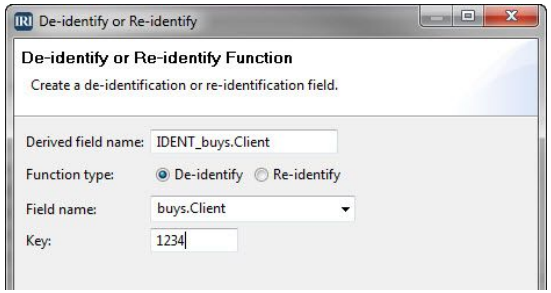
IRI
Total Data Management

es International (IRI), Inc. All Rights Reserved.

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

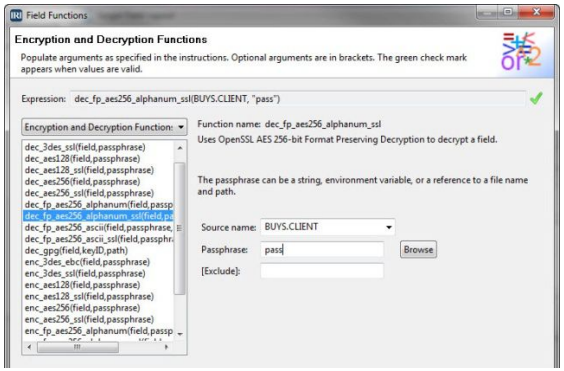


Bit Twiddling



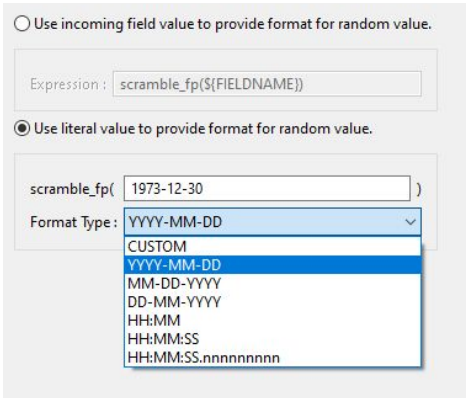
- For ASCII data
- Less secure
- Reversible

Encryption / Decryption

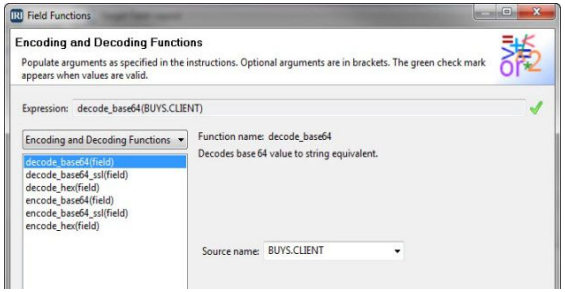


- Multiple algorithms
- AES-256 format-preserving
- Multiple key mgmt options

Format Preserving Scrambling



Encoding / Decoding

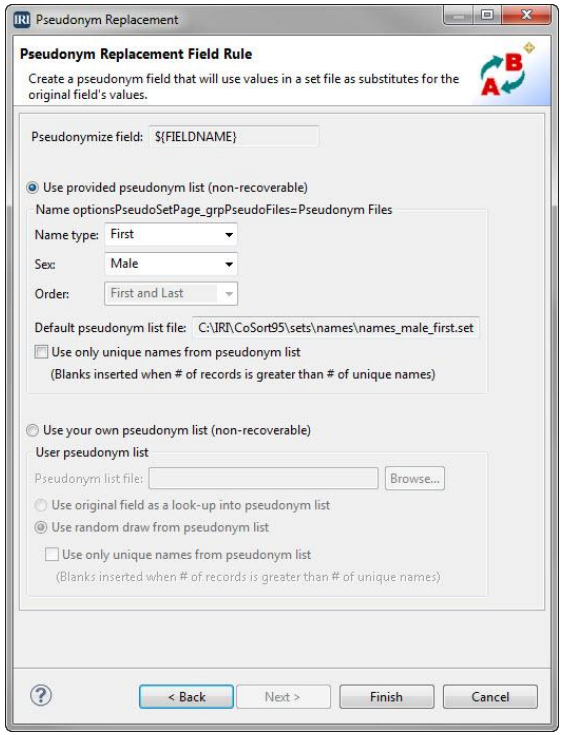


- Changes binary to ASCII
- Supports base64 & hex

IRI
Total Data Management

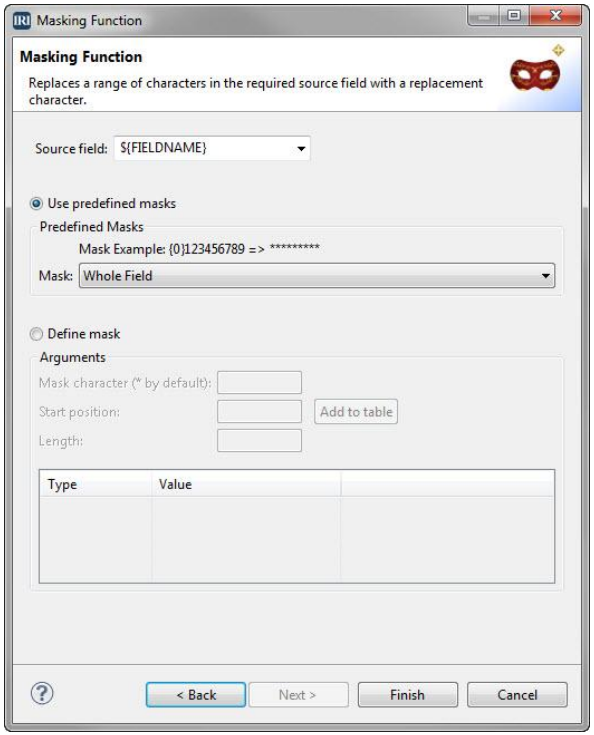
Static Data Masking Function Categories (5-7 of 15)

Pseudonymization



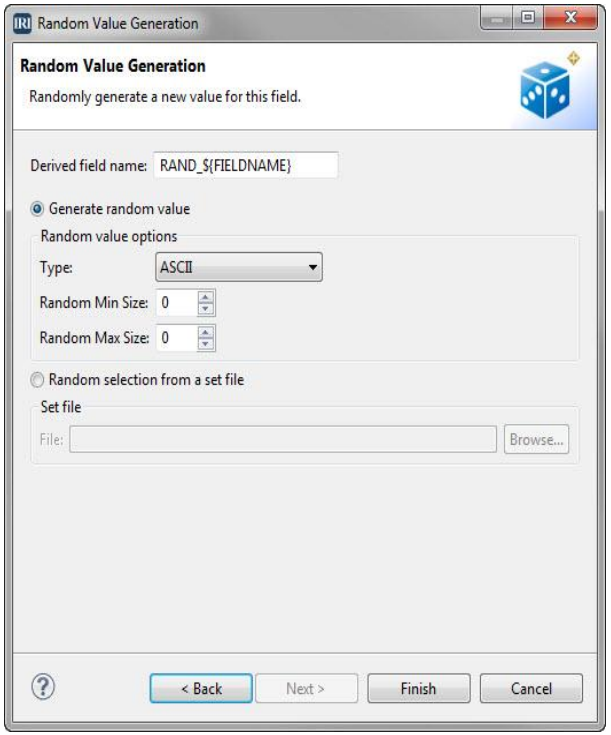
- Provides realistic names
- Reversible lookup values
- or Non-reversible selection

Redaction / Obfuscation



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

Randomization

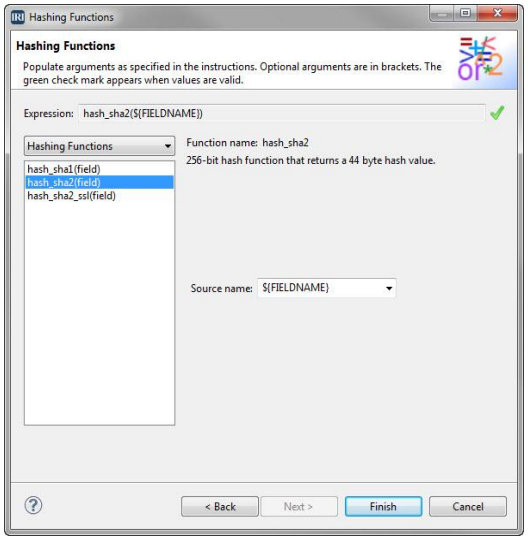


- Random data generation
- Random data selection
- Non-reversible

IRI
Total Data Management

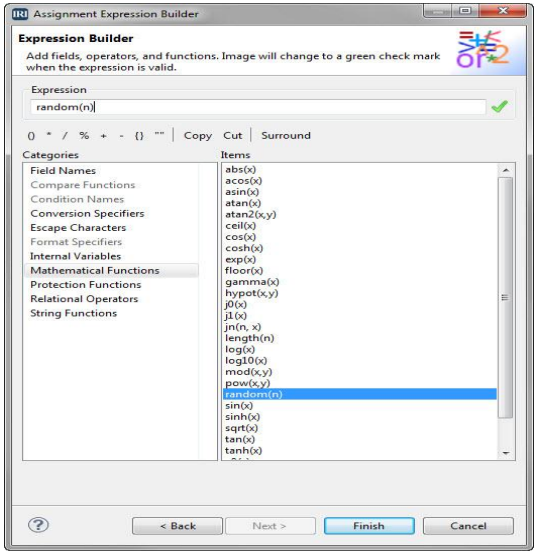
Static Data Masking Function Categories (8-15 of 15)

Hashing



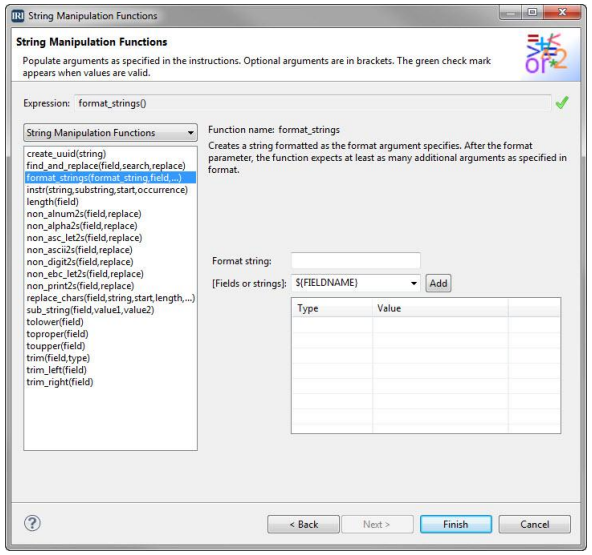
- SHA-1 & 2 cryptographic
- Returns hash of fieldstring
- Use for integrity checking

Expression Logic



- Mathematical operations
- PCRE logic
- Concatenations, etc.

String Manipulations



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

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

Bucketing / Binning: Generalize quasi-identifiers to to anonymize demographic traits

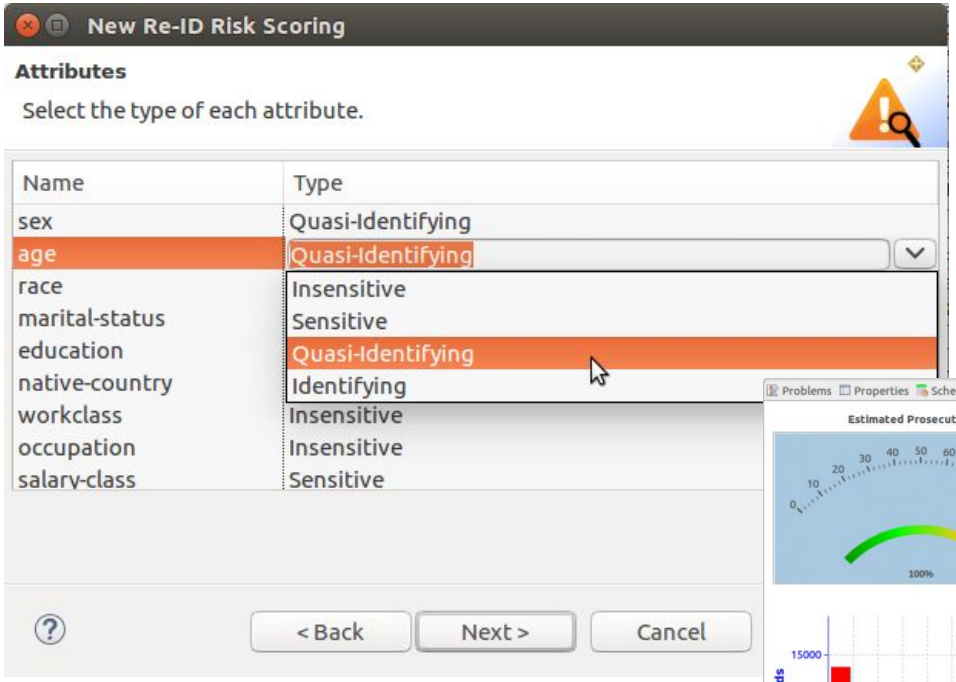
Tokenization: Substitute PANs with DB values for PCI-DSS

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

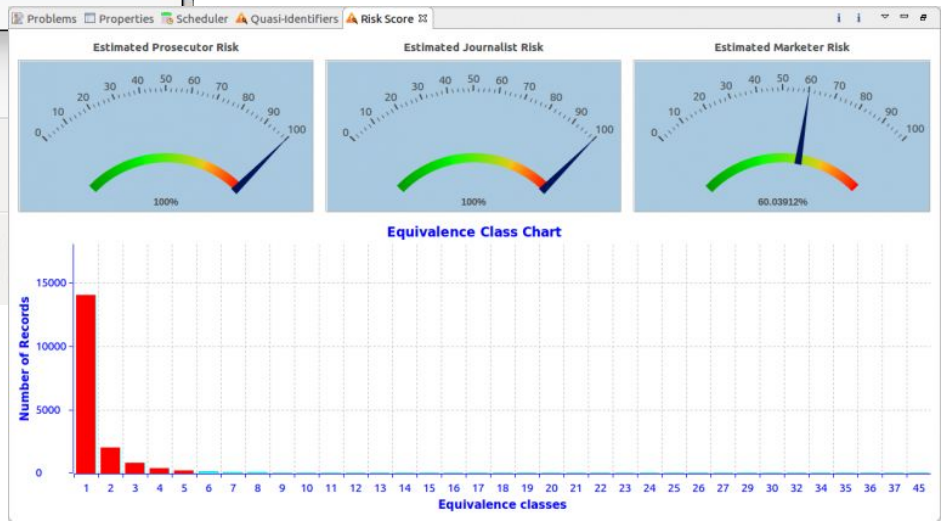
Custom Function: Nested, or user-written field, transforms.

IRI
Total Data Management

Re-ID Risk Measurement



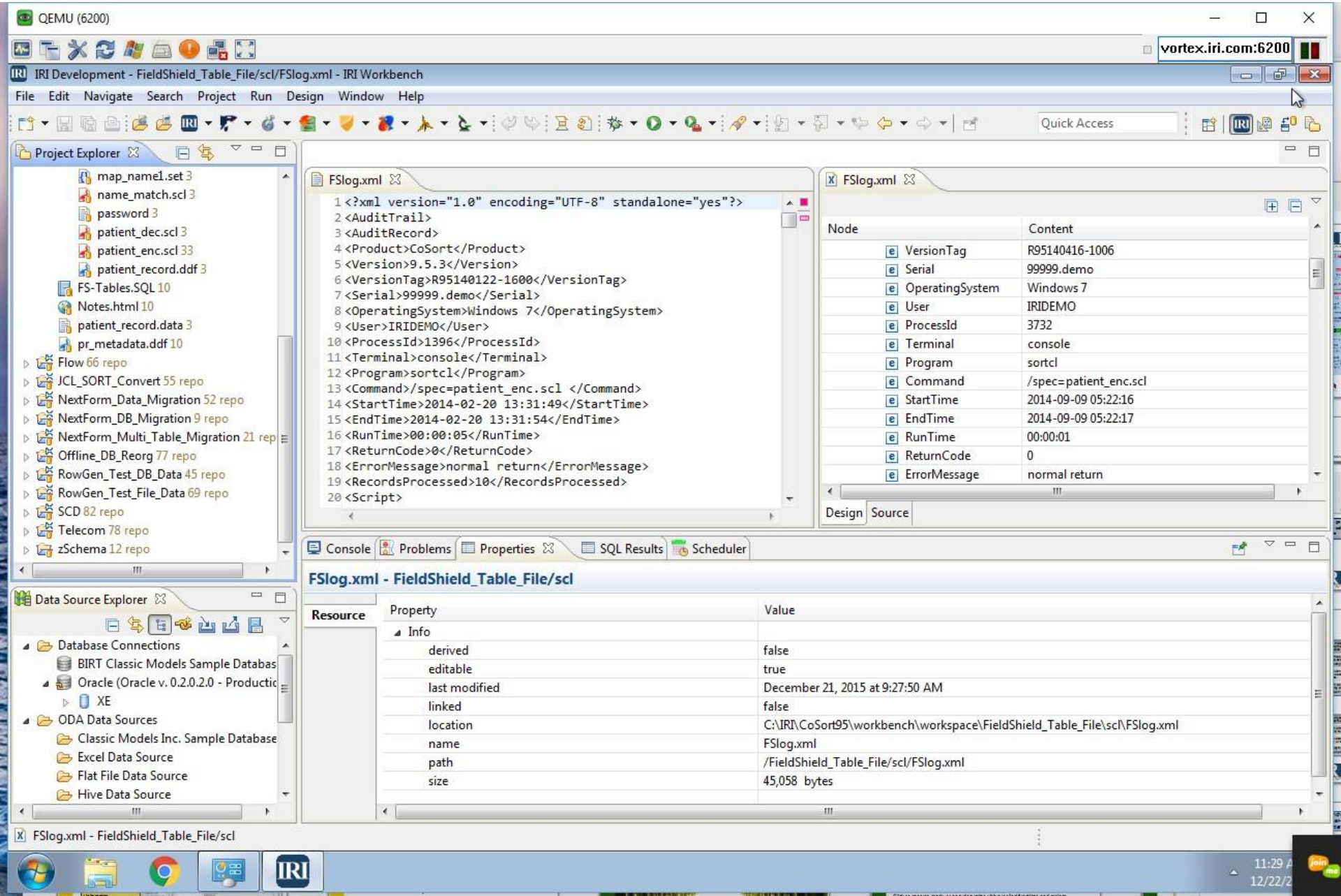
US HIPAA and FERPA regulations require that patient and student data sets used in research or marketing have a statistically certified “very small” chance of being re-identifiable.



- 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

IRI
Total Data Management

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



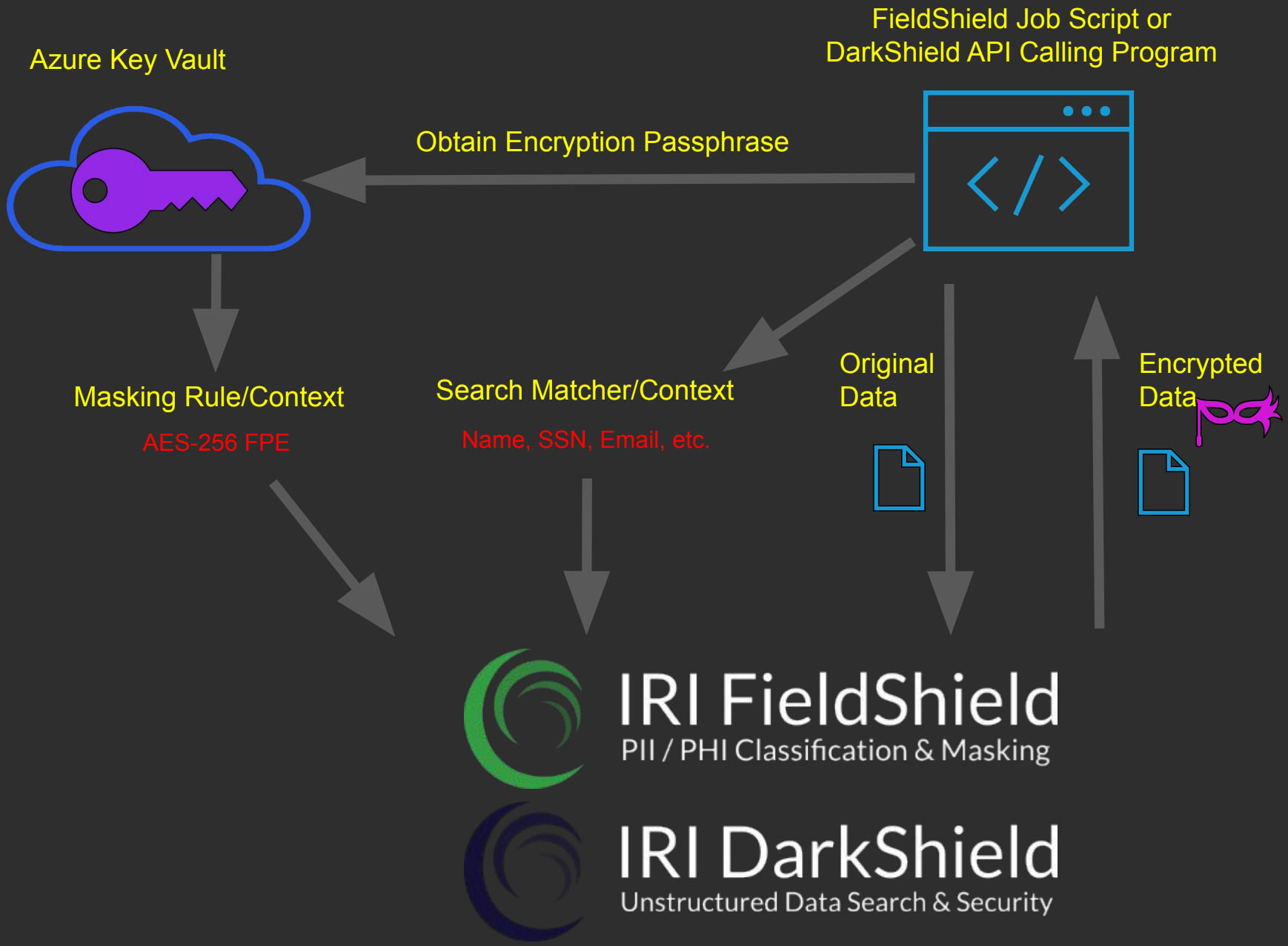
Dynamic & Real-Time Data Masking Options

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
  - Census / Tax
  - Defense
  - Hospitals
  - Pharmas
  - 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

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

Price Performance

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

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

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

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

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

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

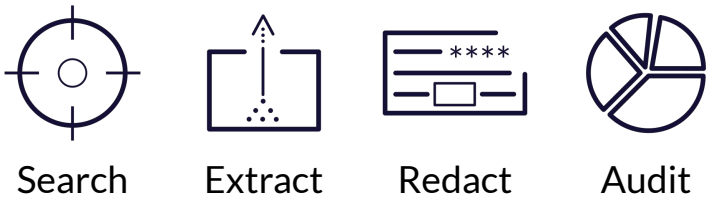
IRI
Total Data Management

IRI DarkShield

Unstructured Data Search & Security

IRI Data Protector Suite

What DarkShield Does



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

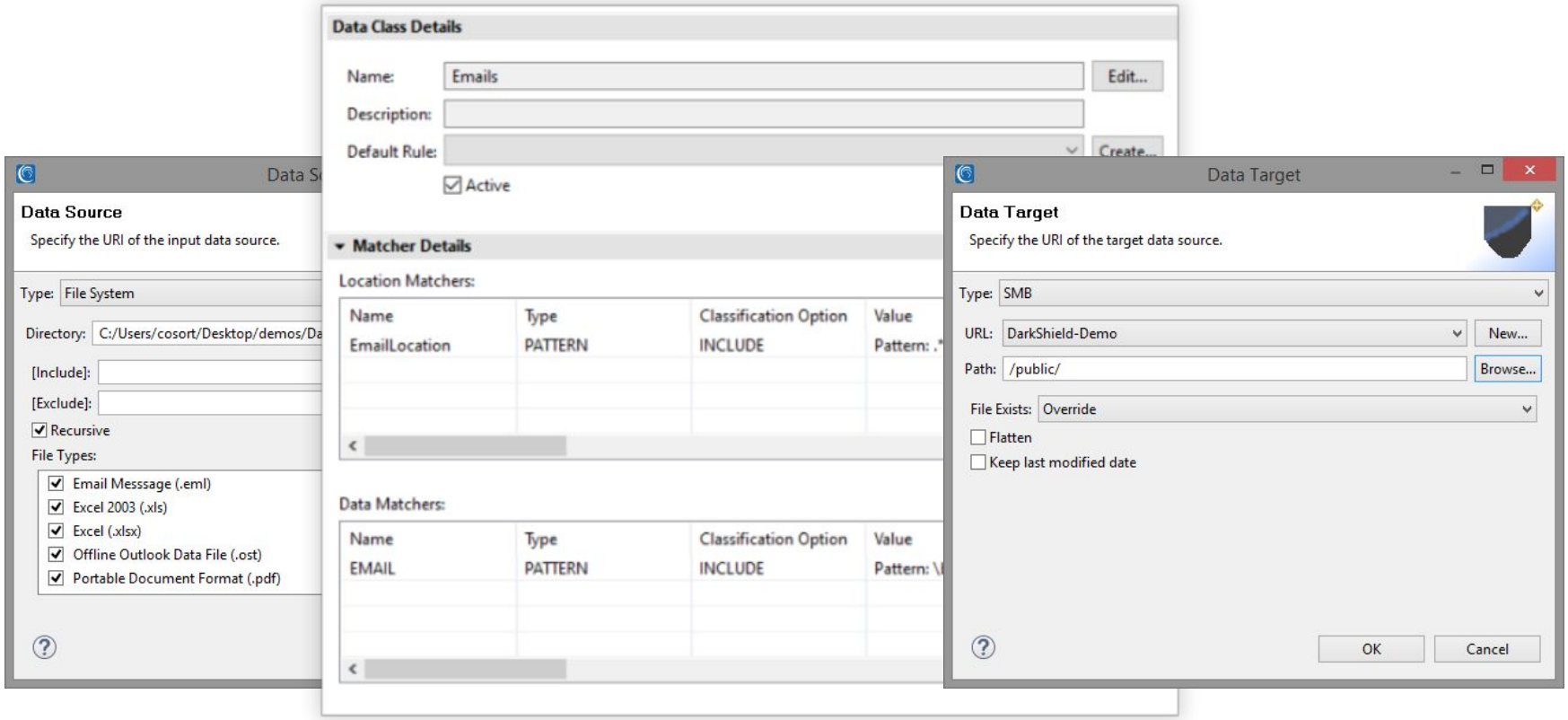
1. Combines PII discovery, delivery, deletion, and reporting in multiple structured, semi-structured, and/or unstructured source formats into one or few ergonomic operation(s)
2. Uses RegEx pattern matching, NER and other search methods for more accurate results
3. Supports multiple *right to be forgotten* 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

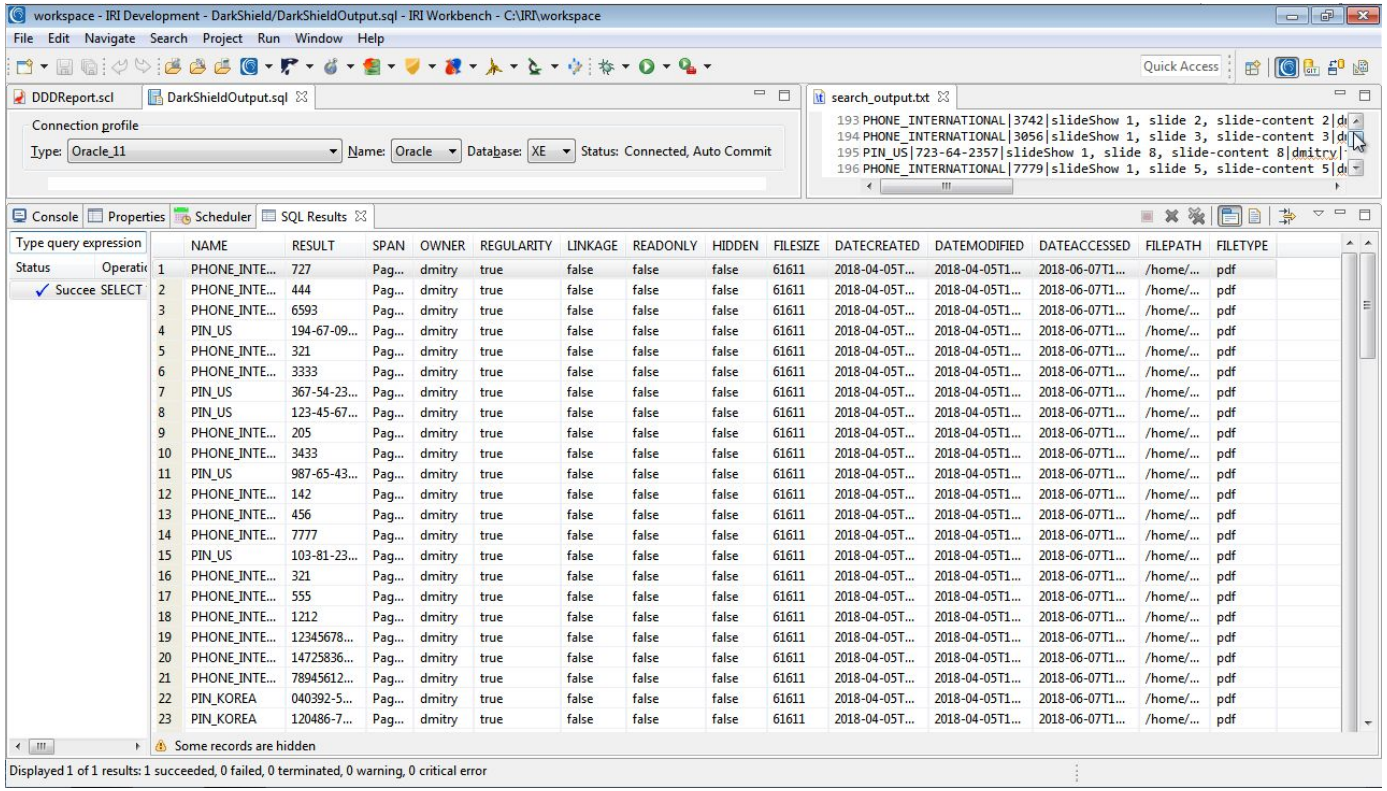
Search Extract Redact Audit



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.

IRI
Total Data Management

Extract



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

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

*example-redacted.json

```
[{
  "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":
  "gender
  "ip_add
}, {
  "id": 3
  "first_
  "last_n
  "email"
  "gender
  "ip_add
}]
```

```
<?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>
```

Commerce Bank

1000 Walnut
Kansas City MO 64106-3686

[redacted]
0701 Ogdsobii 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

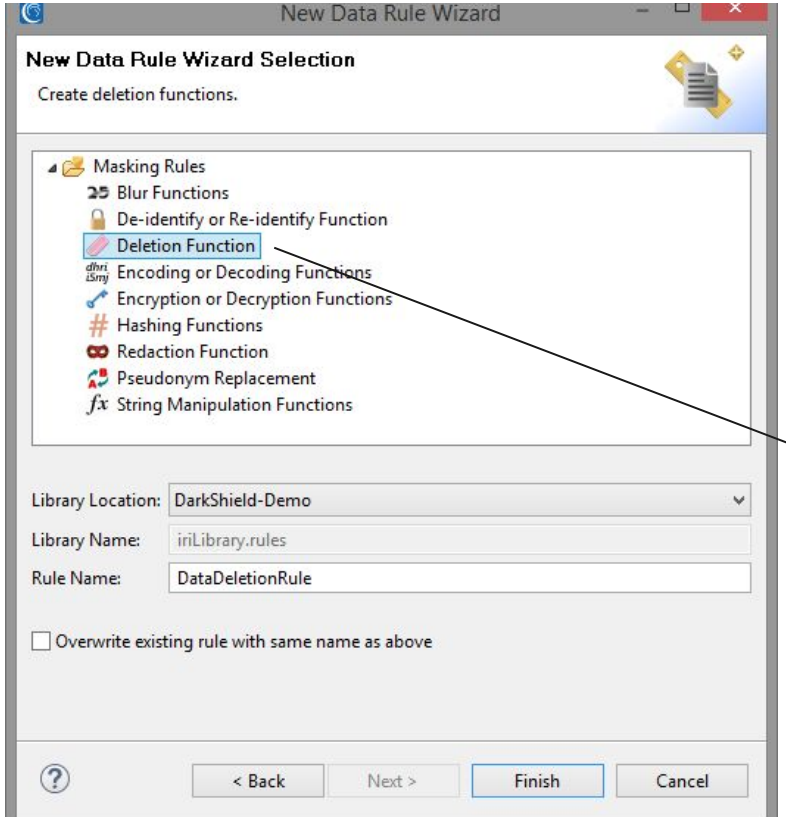
Beginning Balance on May 3, 2003	$7,126.11
Deposits & Other Credits	+3,615.08
ATM Withdrawals & Debits	-20.00
VISA Check Card Purchases & Debits	-0.00
Withdrawals & Other Debits	-0.00
Checks Paid	-200.00
Ending Balance on June 5, 2003	**$10,521.19**

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.

IRI
Total Data Management

Deletion Function

Search Extract Redact Audit



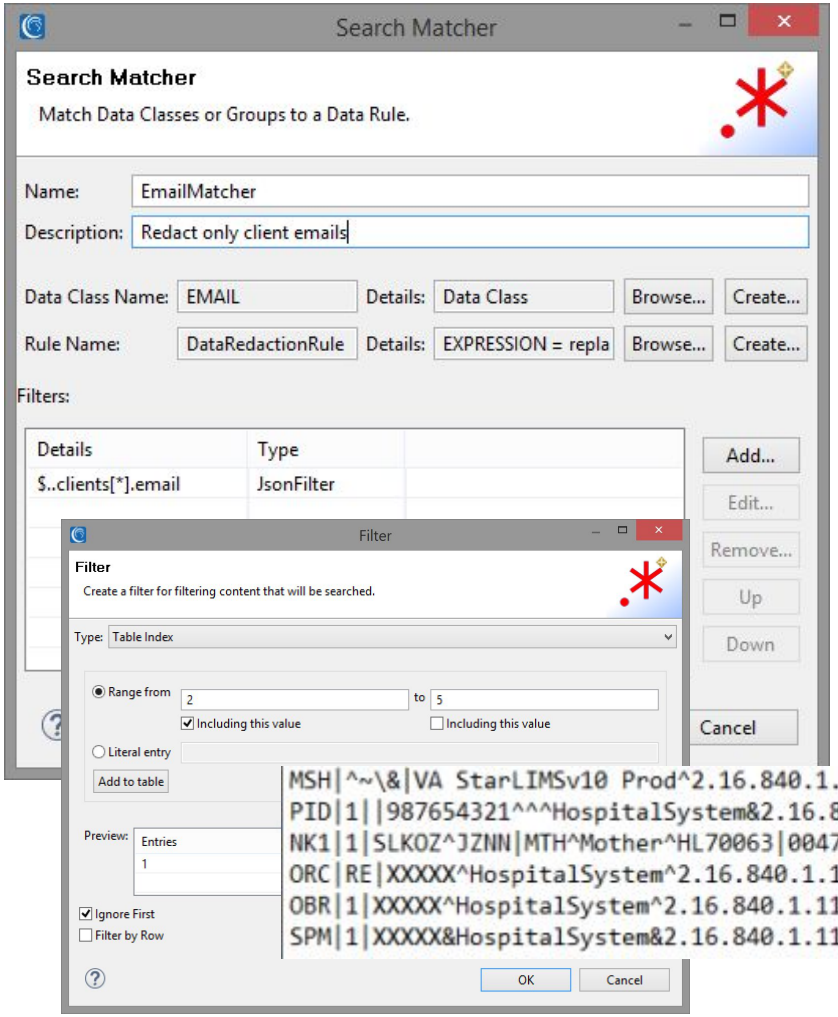
IRI FieldShield, DarkShield & CellShield and other features in Voracity combine to comply with GDPR (and thus CCPA, KVKK, etc.) provisions like:

- 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)

IRI Total Data Management

Search & Mask via Column & Path Filters

Search Extract Redact Audit



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

IRI
Total Data Management

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.

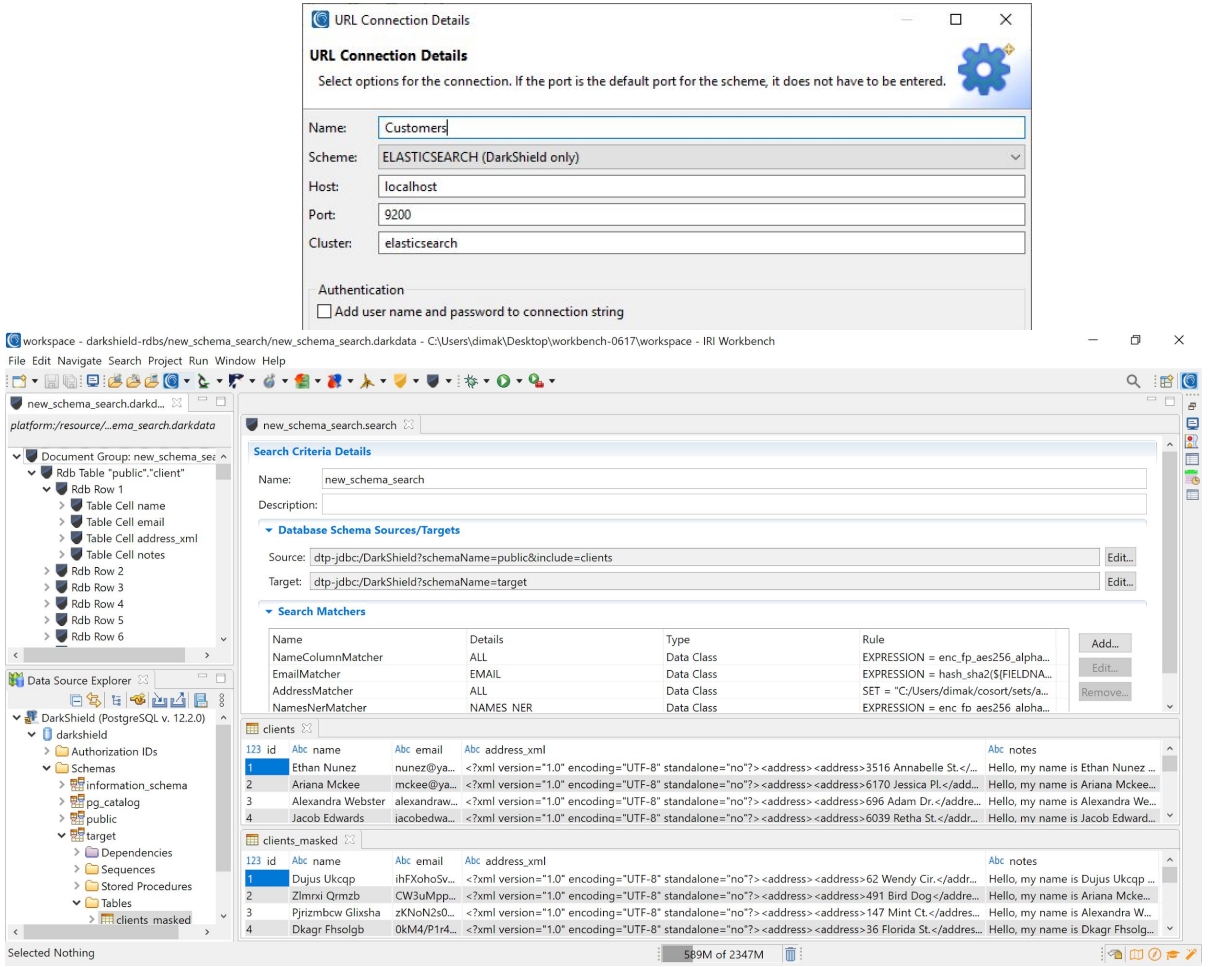
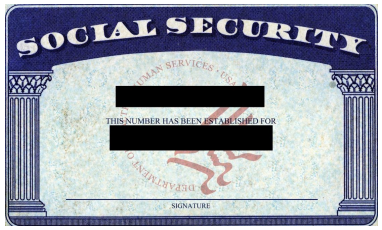
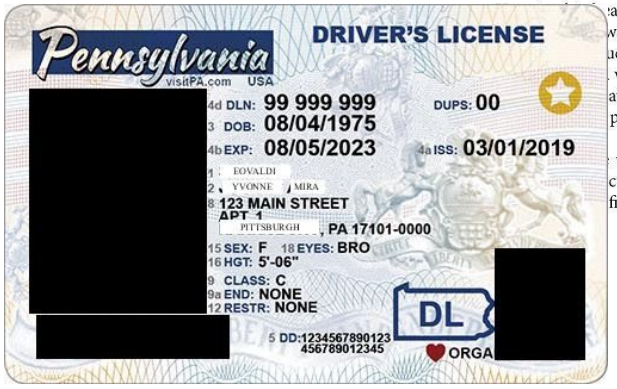


Image File Redaction or Value Replacement …

Search Extract Redact Audit





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.

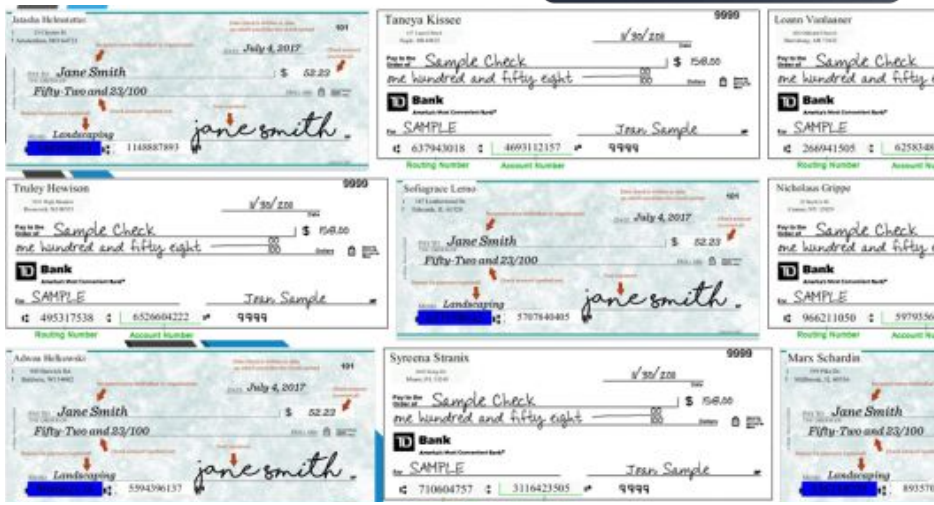
"Compose yourself," said I.

…adful smile to me,
…way the sheet. At si
…uch immense relie
… voice that was alr
…ated glass?" he ask
…place with someth

… with a smiling n
…cture and added
…first of a reddish h

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

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



IRI
Total Data Management

DICOM Medical Image De-ID / Anonymization

Search Extract Redact Audit

Before DarkShield:

	E	F	G
1	Subject ID	Study UID	Study Description
2	292821506	2.25.106461954783291641048254423668956446 19	XR CHEST AP PORTABLE for Douglas Davidson
3	292821506	2.25.106461954783291641048254423668956446 19	XR CHEST AP PORTABLE for Douglas Davidson
4	339833062	2.25.541629416691863029175036346640219638	NA

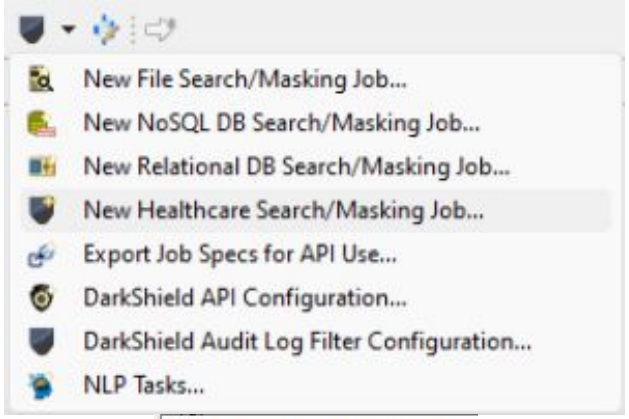
After DarkShield:

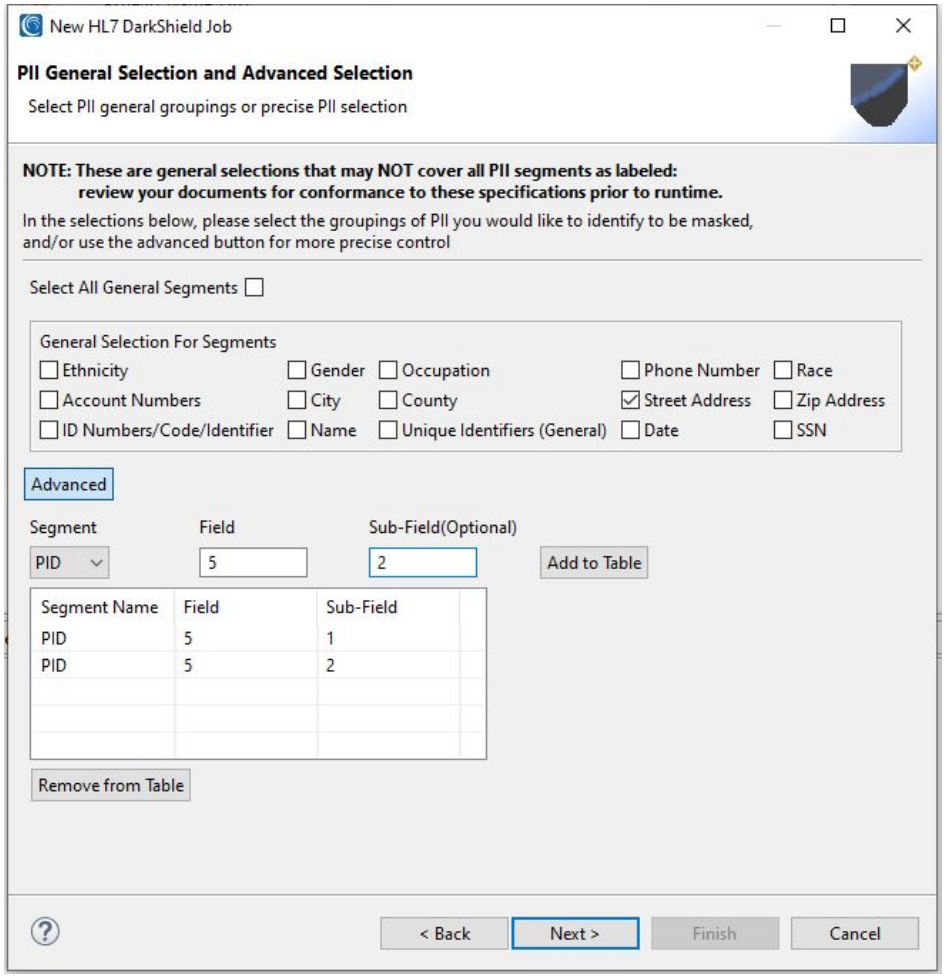
	C	D	E	F	G	H
1	3rd Party Analysis	Data Description URI	Subject ID	Study UID	Study Description	Study Date
2	NO	httpsdoi.org10.7937s17z-r072	703523712	2.25.106461990375634587137034423668956446198	XR CHEST AP PORTABLE for Axodquq Wmhpzjcx	7/13/201
3	NO	httpsdoi.org10.7937s17z-r072	703523712	2.25.106461990375634587137034423668956446198	XR CHEST AP PORTABLE for Axodquq Wmhpzjcx	7/13/201
4	NO	httpsdoi.org10.7937s17z-r072	204498801	2.25.541629416691863029175036346640219638	NA	7/5/200

IRI
Total Data Management

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

Search Extract Redact Audit





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

IRI
Total Data Management

Multilingual NER via Machine Learning

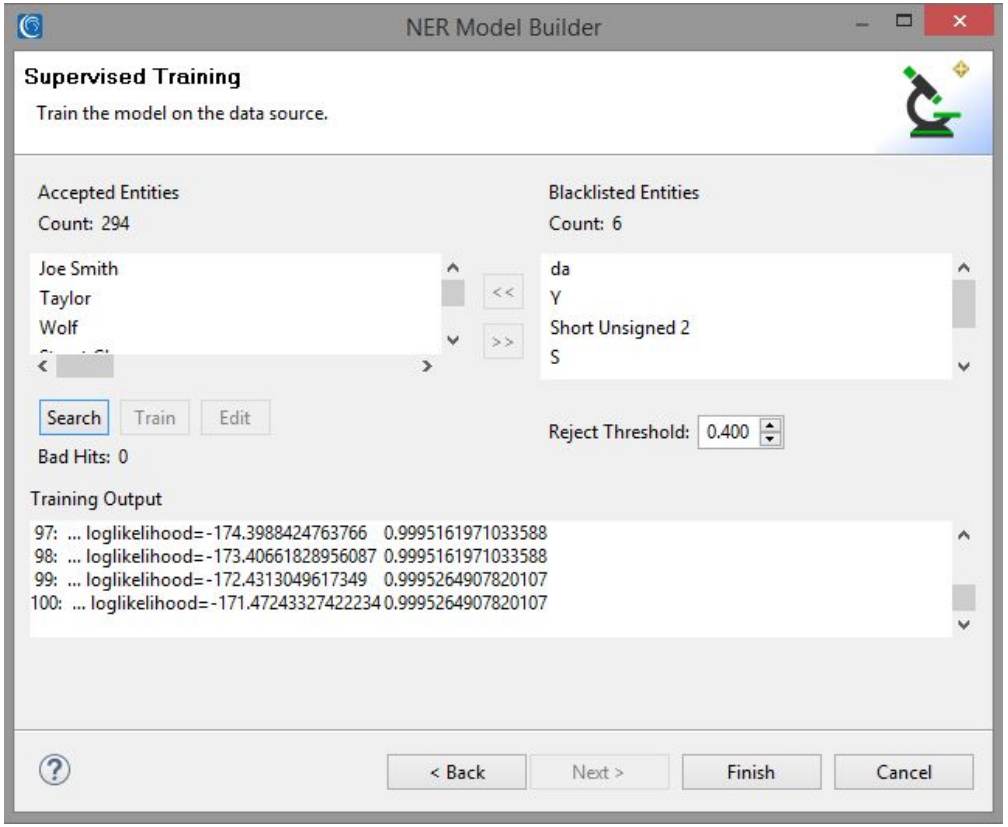
Search Extract Redact Audit

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.

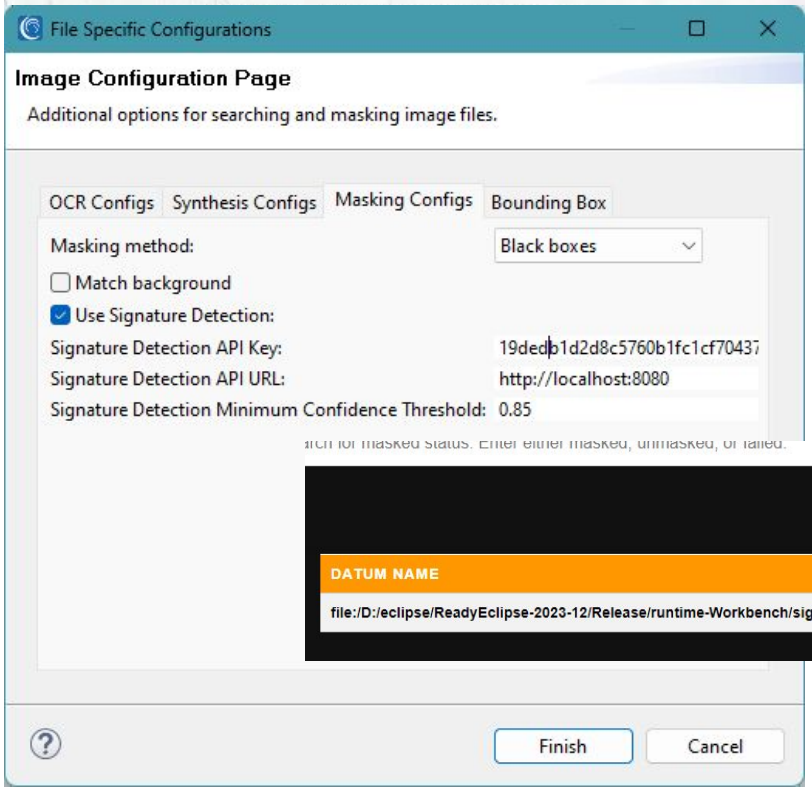
Compare this search method to other DarkShield search methods, like pattern and lookup matches, path filters, or bounding-box areas (for images).



IRI
Total Data Management

Signature Detection and Redaction

Search Extract Redact Audit



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

IRI
Total Data Management

Documented APIs

Swagger
Supported by SMARTBEAR

DarkShield API 1.3.0 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
- POST /api/darkshield/maskContext.destroy
- POST /api/darkshield/maskContext.mask
- POST /api/darkshield/searchContext.mask

Host a context which can be called whenever a search operation needs to be performed. The Context contains a list of search matchers that will be used to match on the text. Multiple contexts with the same name cannot be created.

Parameters Cancel

No parameters

Request body required application/json

The Search Context.

Examples: Create a Search Context with a mix of matchers.

```
{
  "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	**Response body**

```
{
  "failedResults": [],
  "maskedText": "Hello, my name is Wazz Edk and my email address is BqJA0RzCAWdtqXb3tJNBGB/RgNo3y+QKd0MsCjZsgMM=.",
  "results": [
    {
      "end": 26,
      "maskedResult": "Wazz Edk",
      "ruleName": "FpeNameRule",
      "start": 18
    },
    {
      "end": 95,
      "maskedResult": "BqJA0RzCAWdtqXb3tJNBGB/RgNo3y+QKd0MsCjZsgMM=",
      "ruleName": "HashEmailRule",
      "start": 51
    }
  ],
  "unmatchedAnnotations": []
}
```

Download

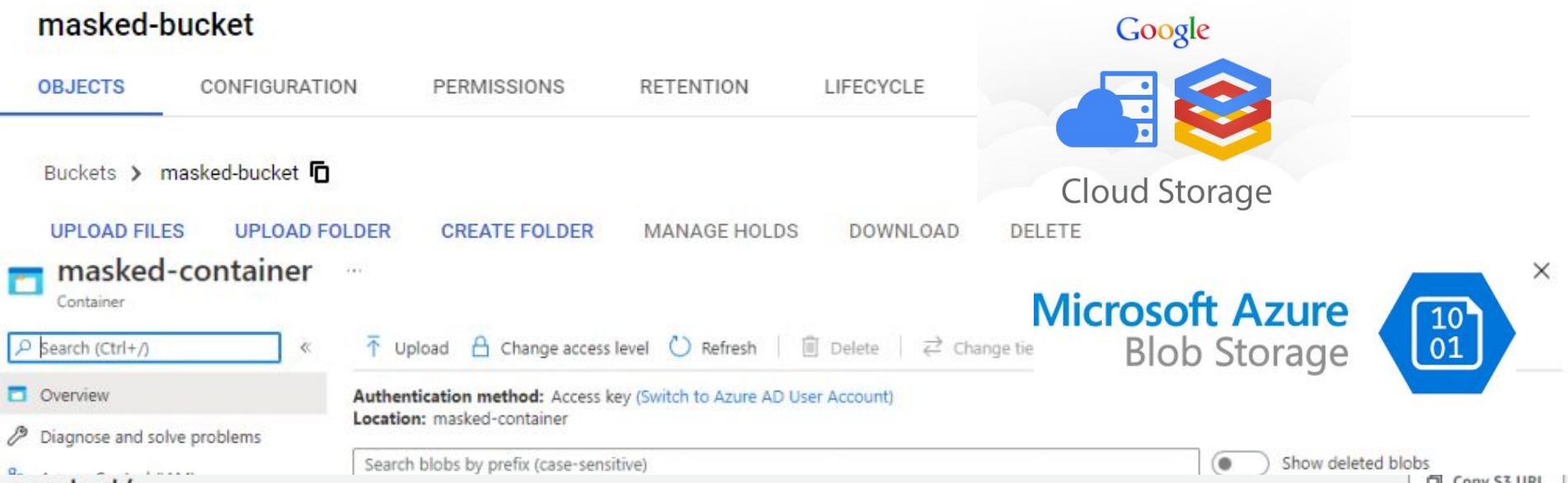
*Multi- and single (restful) endpoint search/mask context **APIs** for text, files, RDBs and NoSQL DBs.*

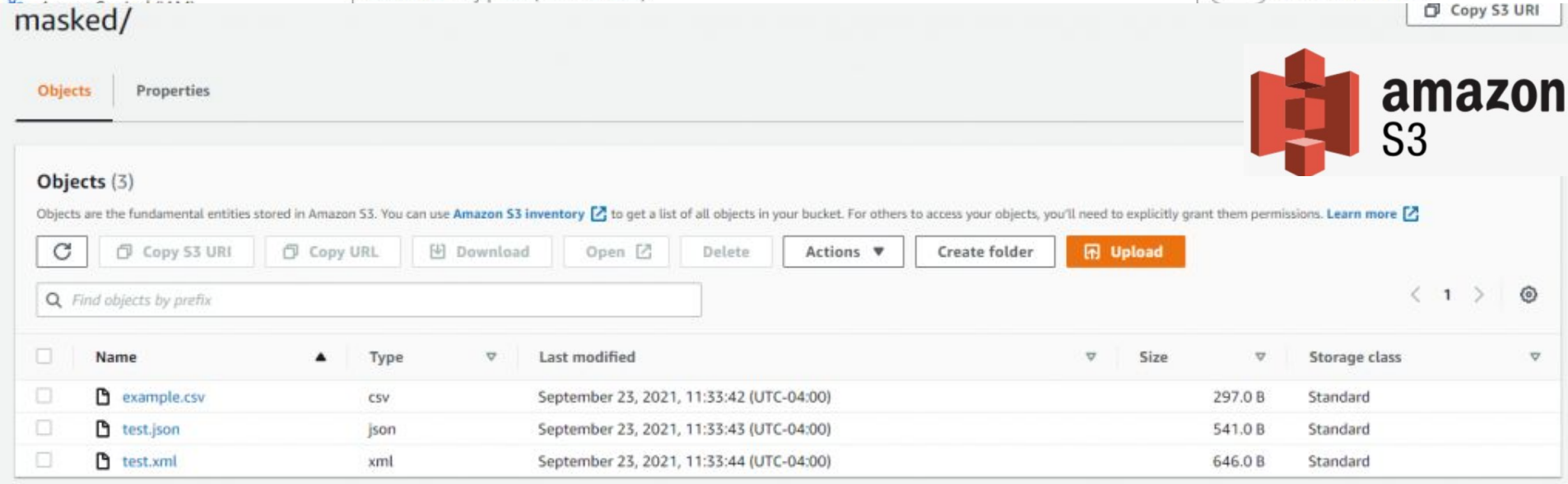
IRI
Total Data Management

Masking Cloud Files

GCP, Azure, S3, SharePoint, OneDrive

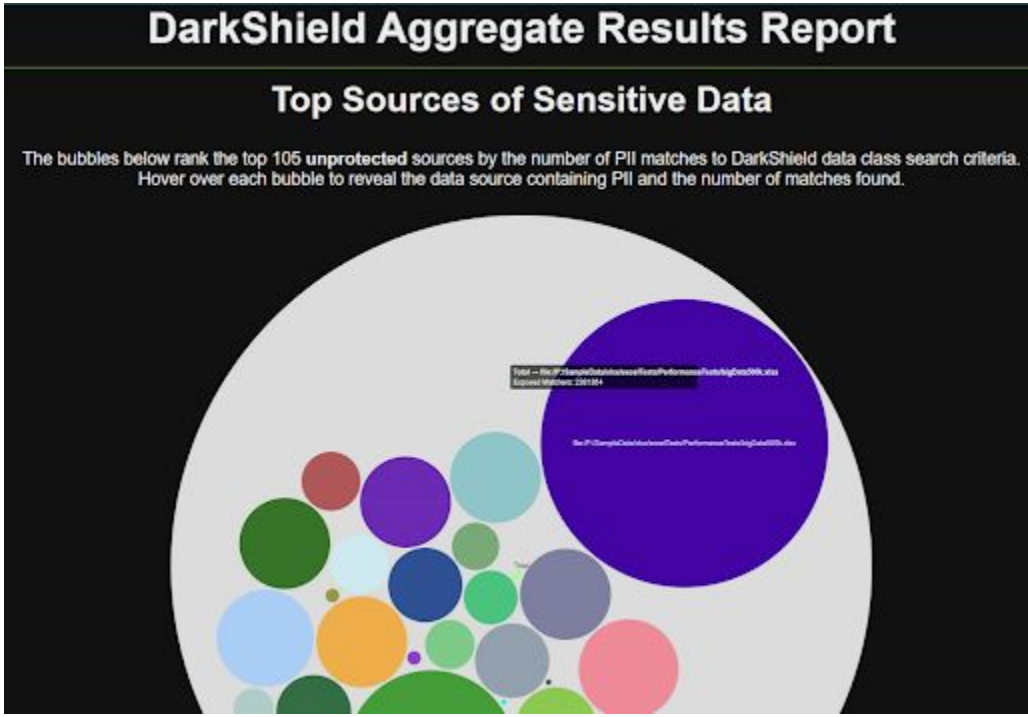
Search Extract Redact Audit





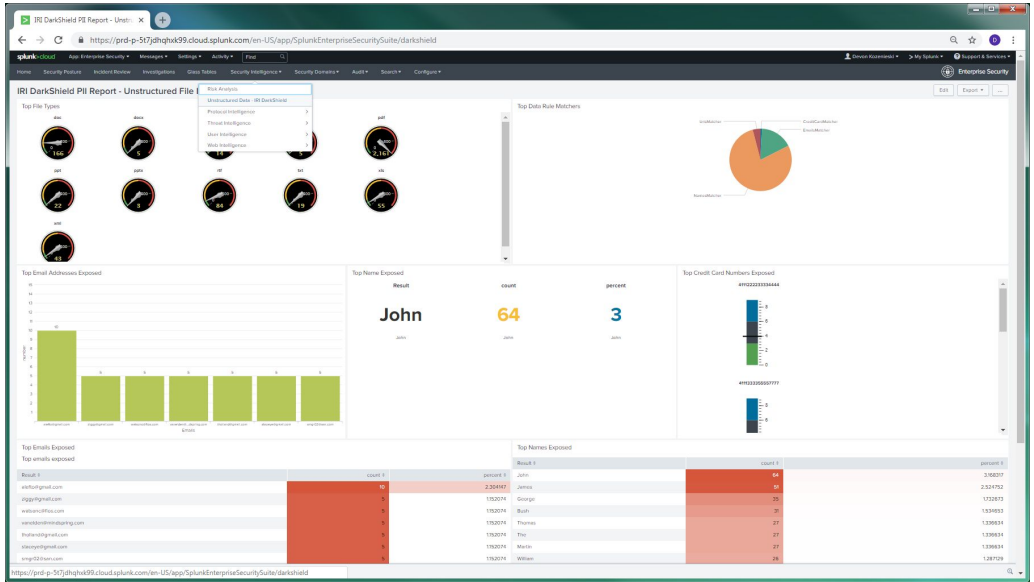
Audit

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
Total Data Management

IRI CellShield

PII / PHI Search & Mask in Excel

IRI Data Protector Suite

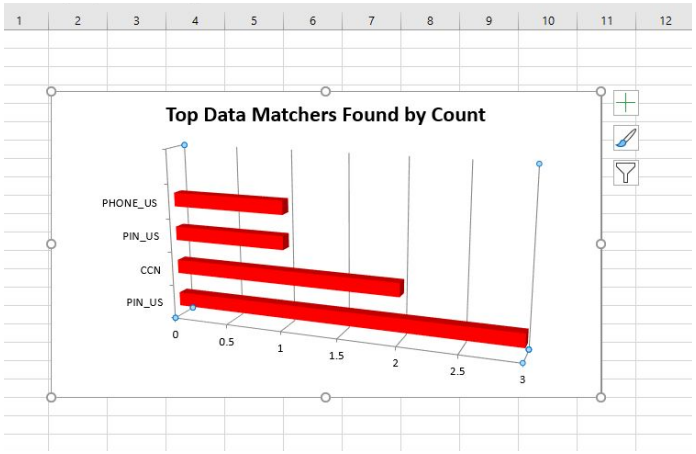
What CellShield EE Does

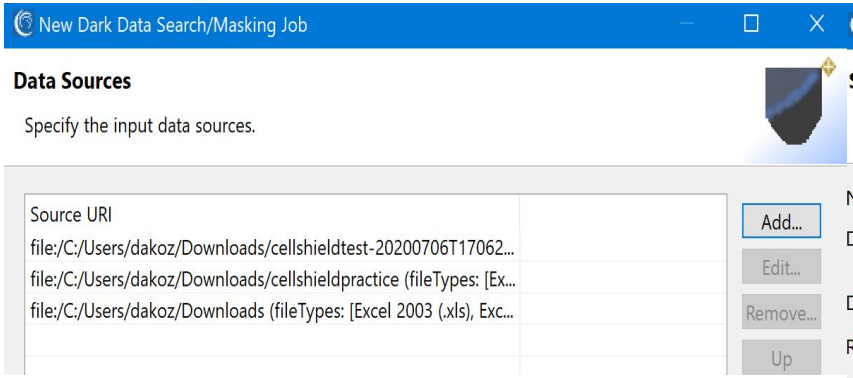
Note: FieldShield & DarkShield Support Excel, too.

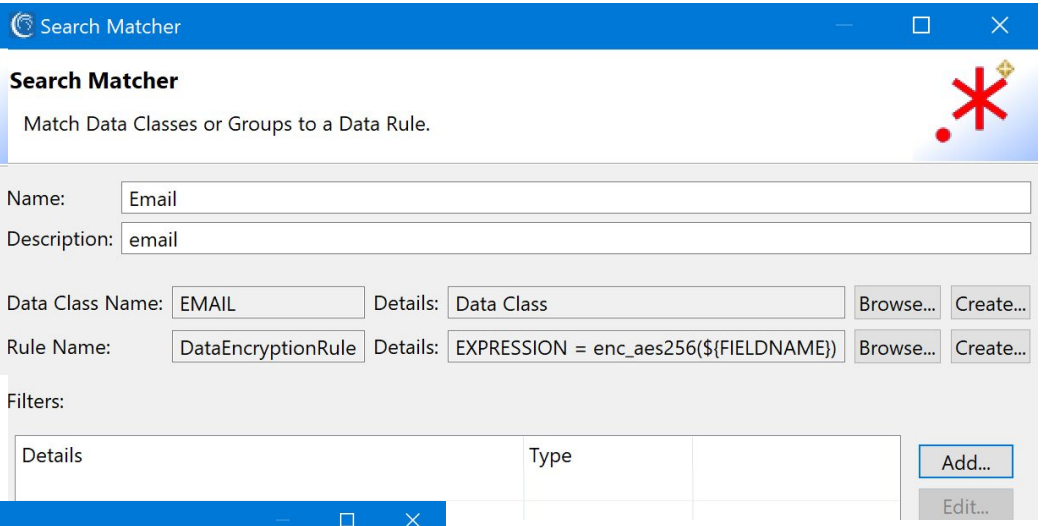
- Discovers, reports, and masks PII and perform audit actions in Excel 2010 & later
- Searches and secures PII in spreadsheets on one PC or throughout an SMB LAN
- Provides common and allow new search pattern definitions for PII formats
- Searches for strings in a dictionary, and find/fix PII *floating* 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.

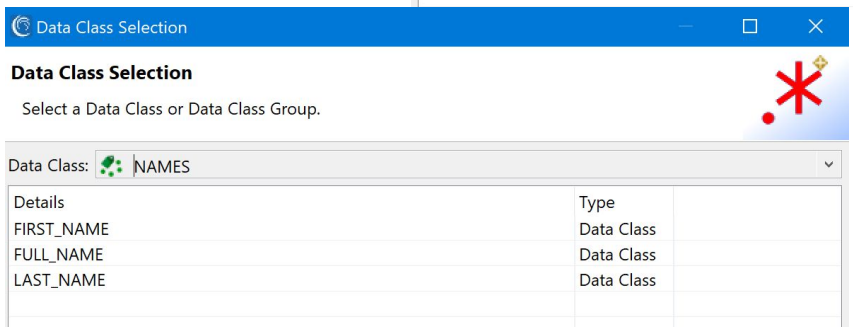
CellShield PII Discovery

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):





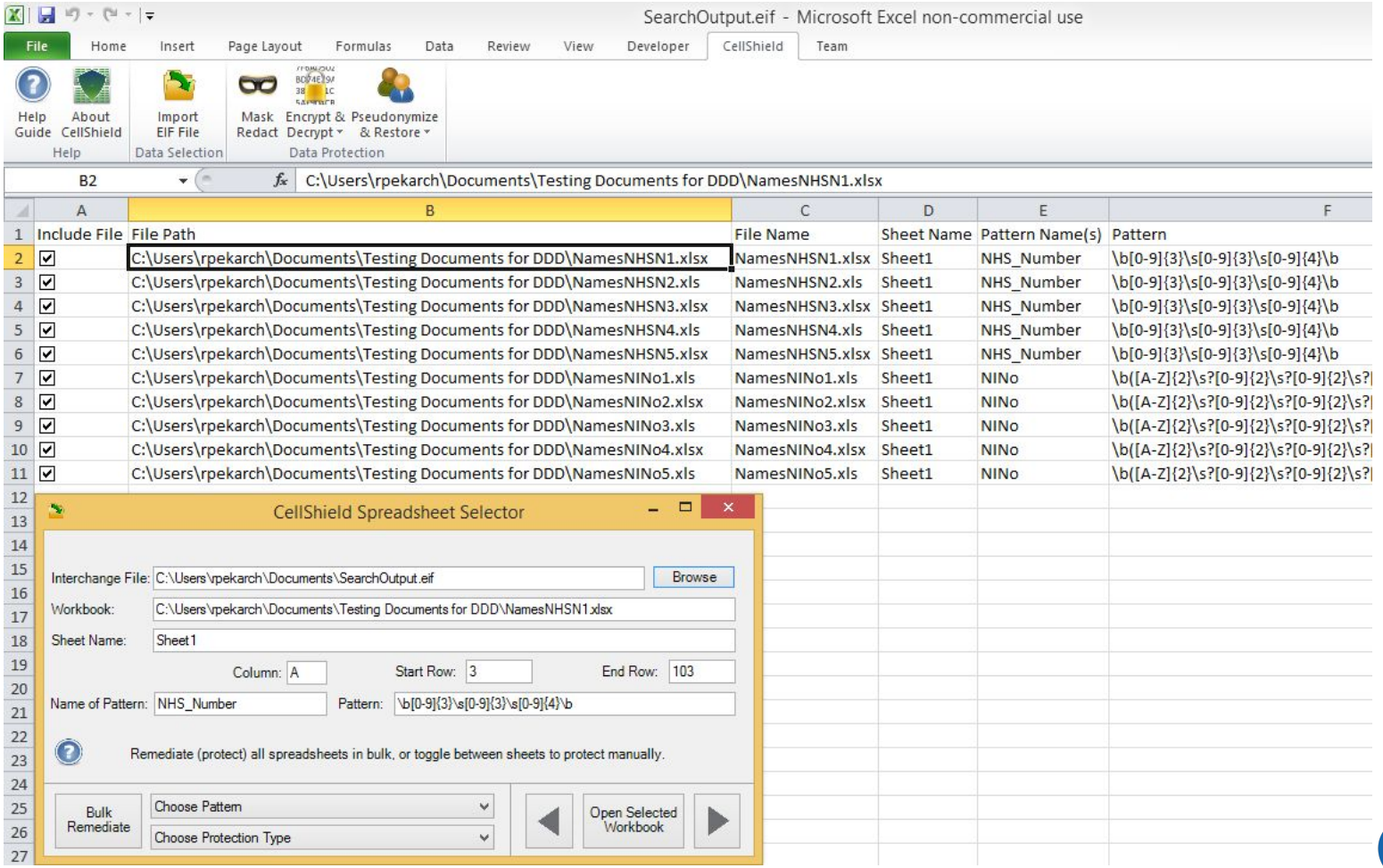




IRI
Total Data Management

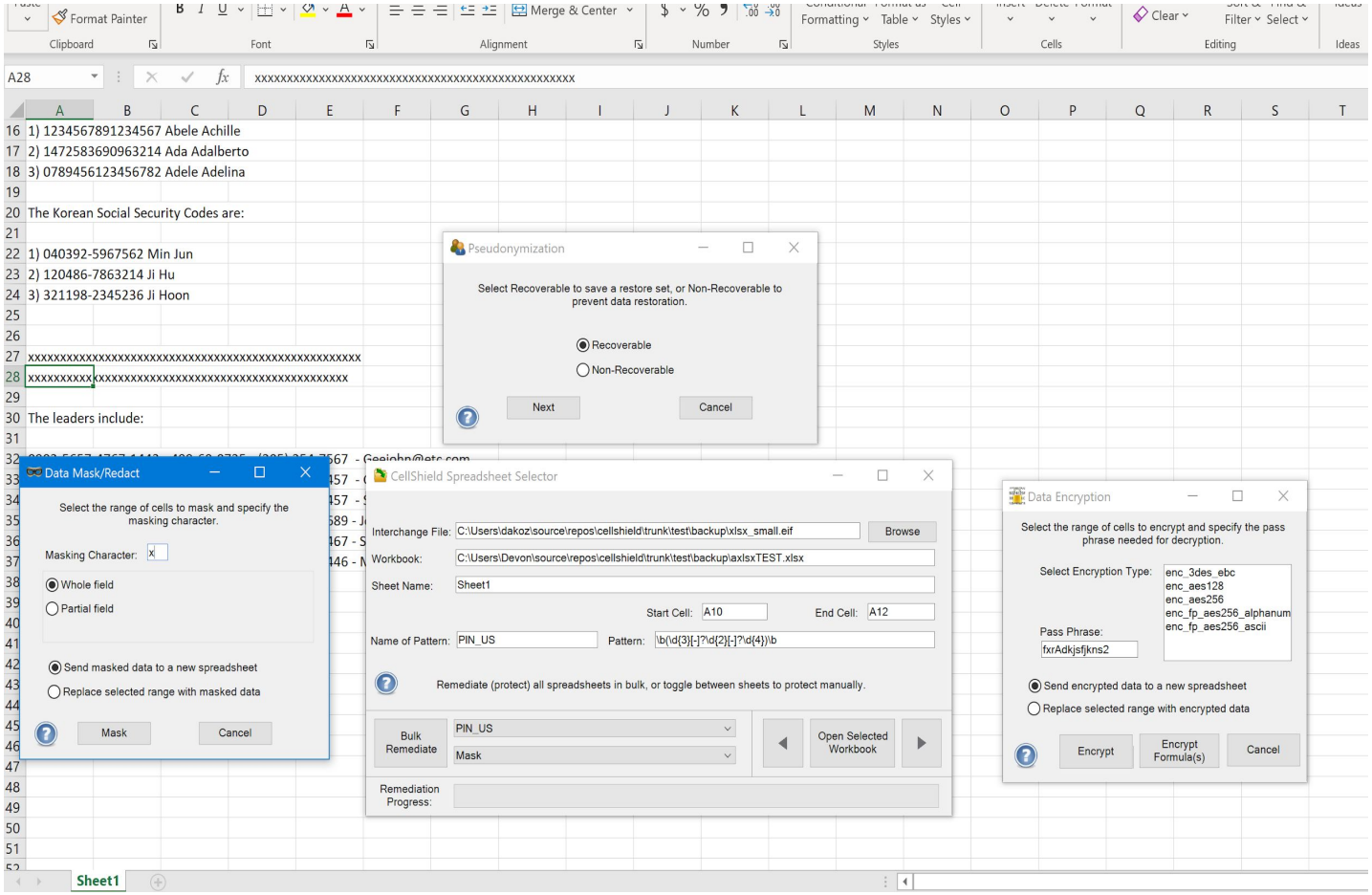
CellShield Search Report & Action Sheet

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



CellShield Masking Functions

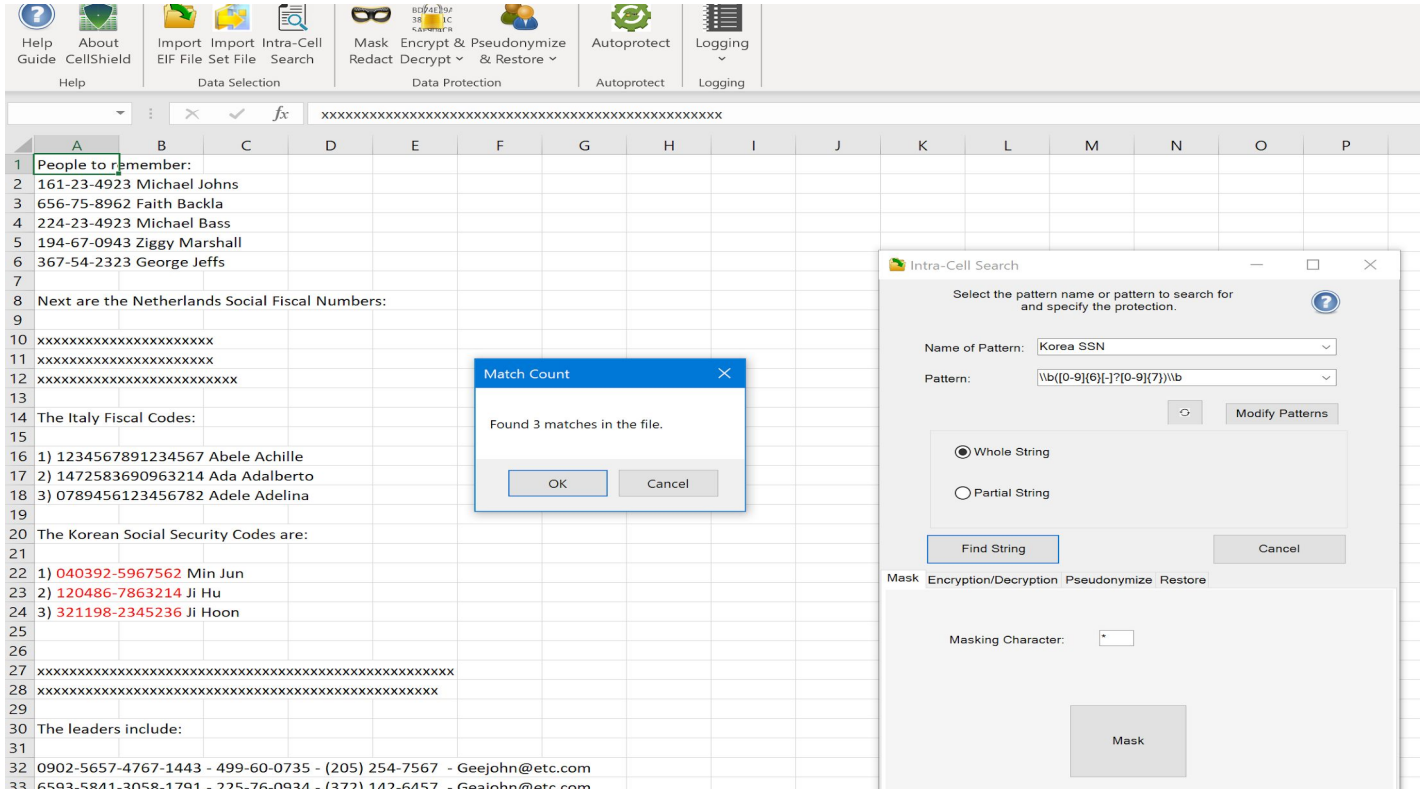
Perform point-and-click encryption and decryption, redaction (full or partial cell), or pseudonymization (reversible and non-reversible) of applicable ranges within the spreadsheets in the report. Formulas may also be encrypted and decrypted.



IRI
Total Data Management

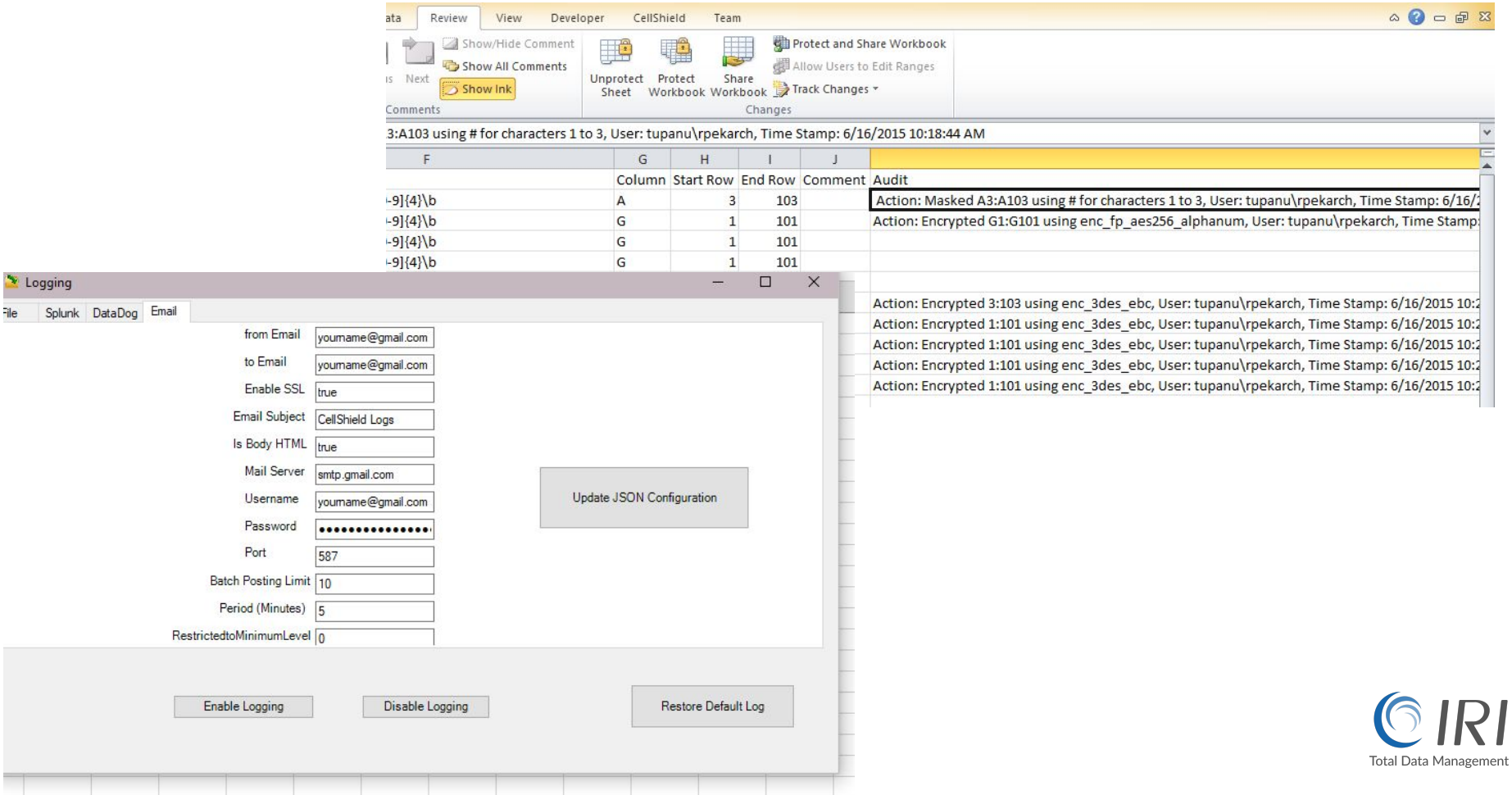
Intra-Cell Searching & Masking, Too

- 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



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.



IRI
Total Data Management

IRI DMaaS

Data Masking as a Service

IRI Data Protector Suite

What is DMaaS?

- 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 the IRI Voracity Platform

What RowGen Does

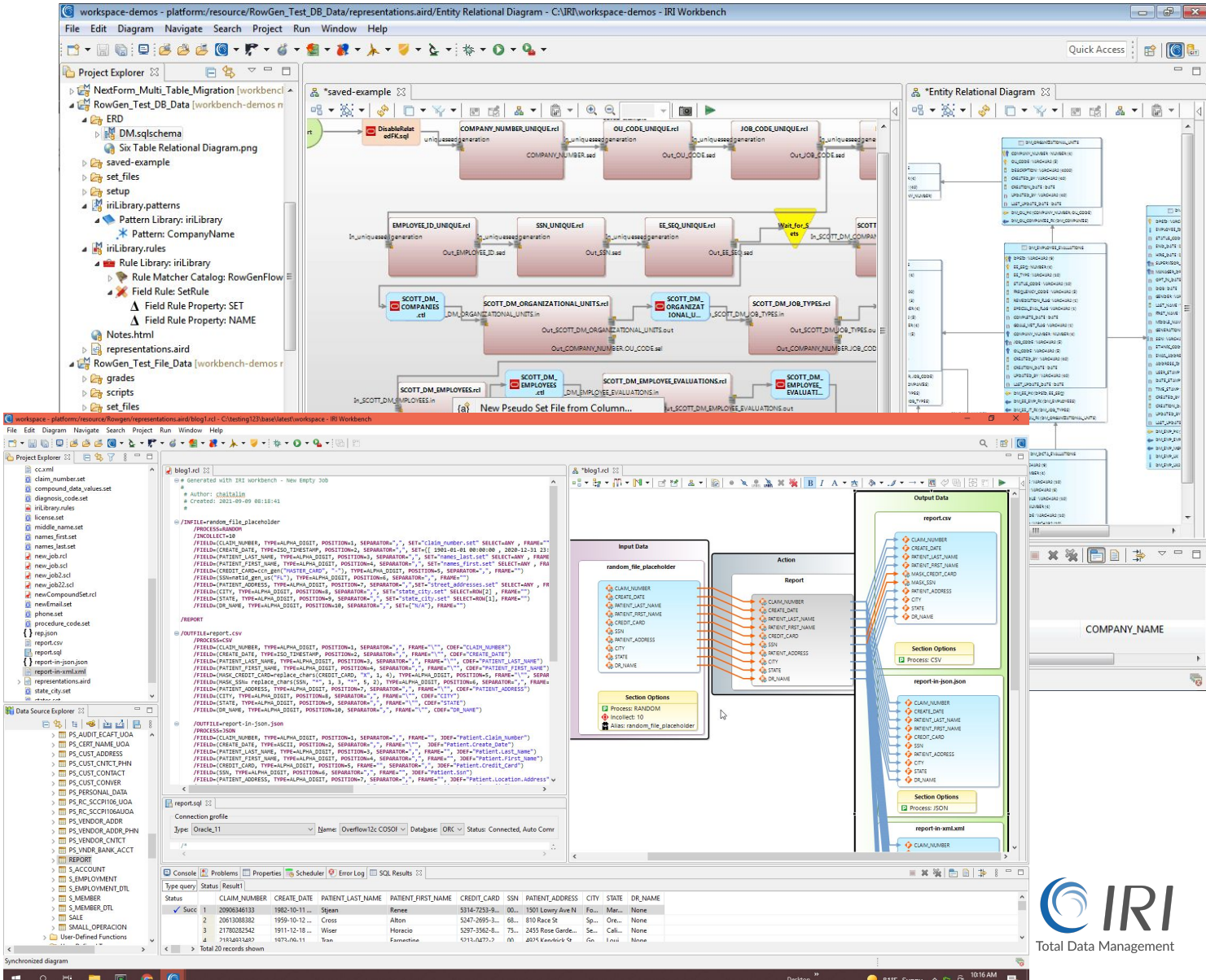
- Creates synthetic but realistic random and random-real test data simultaneously
- Improves DB prototypes, application quality, benchmarking, and outsourced operations
- Uses standard DB DDL, production file, and custom metadata to define layouts
- Preserves structural and referential integrity of real 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:

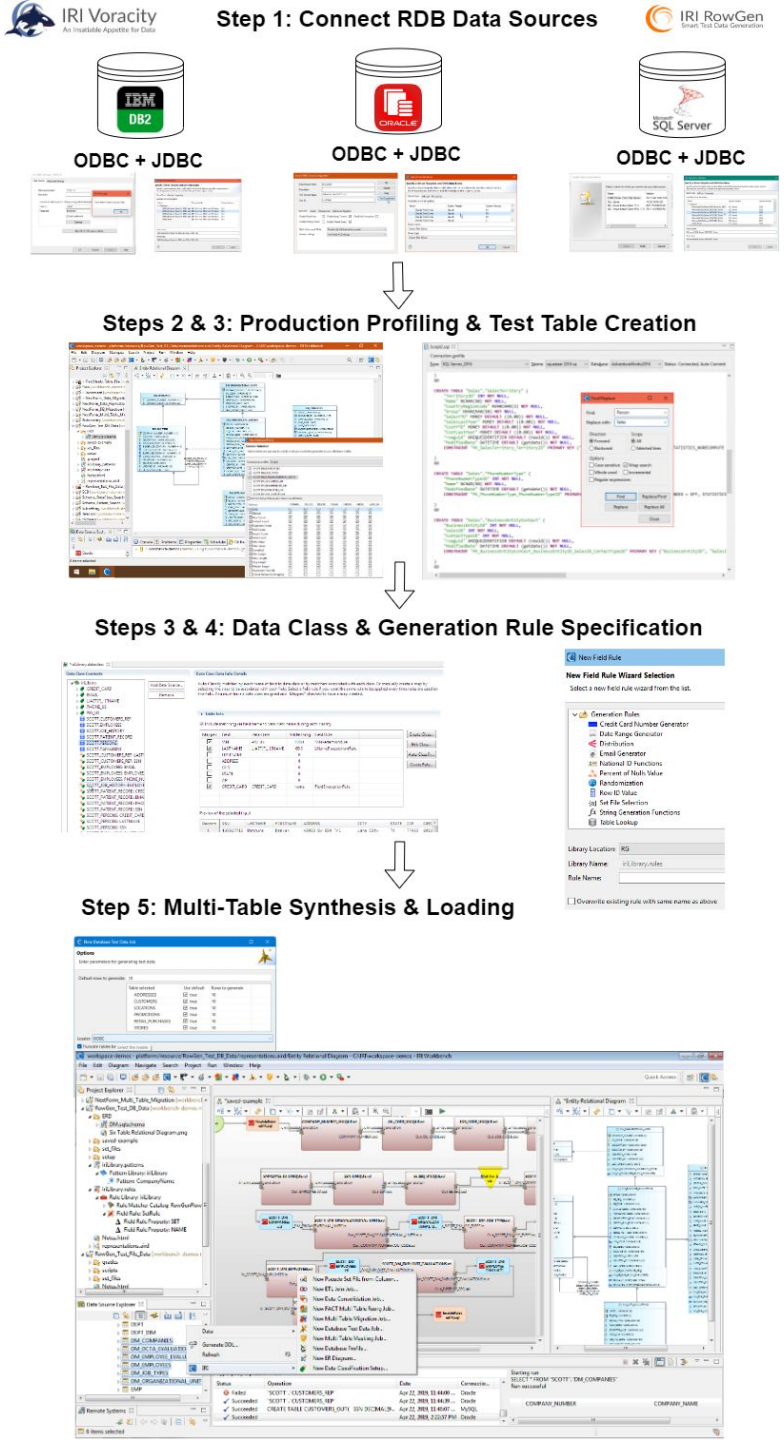
- RDBs
- ASN.1 CDRs
- CLF/ELF
- COBOL
- CSV / TSV
- Data Vault 2.0
- FHIR / HL7 / X12*
- Image*
- Excel
- Hadoop
- JSON
- LDIF
- NoSQL DBs
- PDF Forms *
- DOC/PPT(X) *
- XML

* via DarkShield



Total Data Management

IRI RowGen Workflow

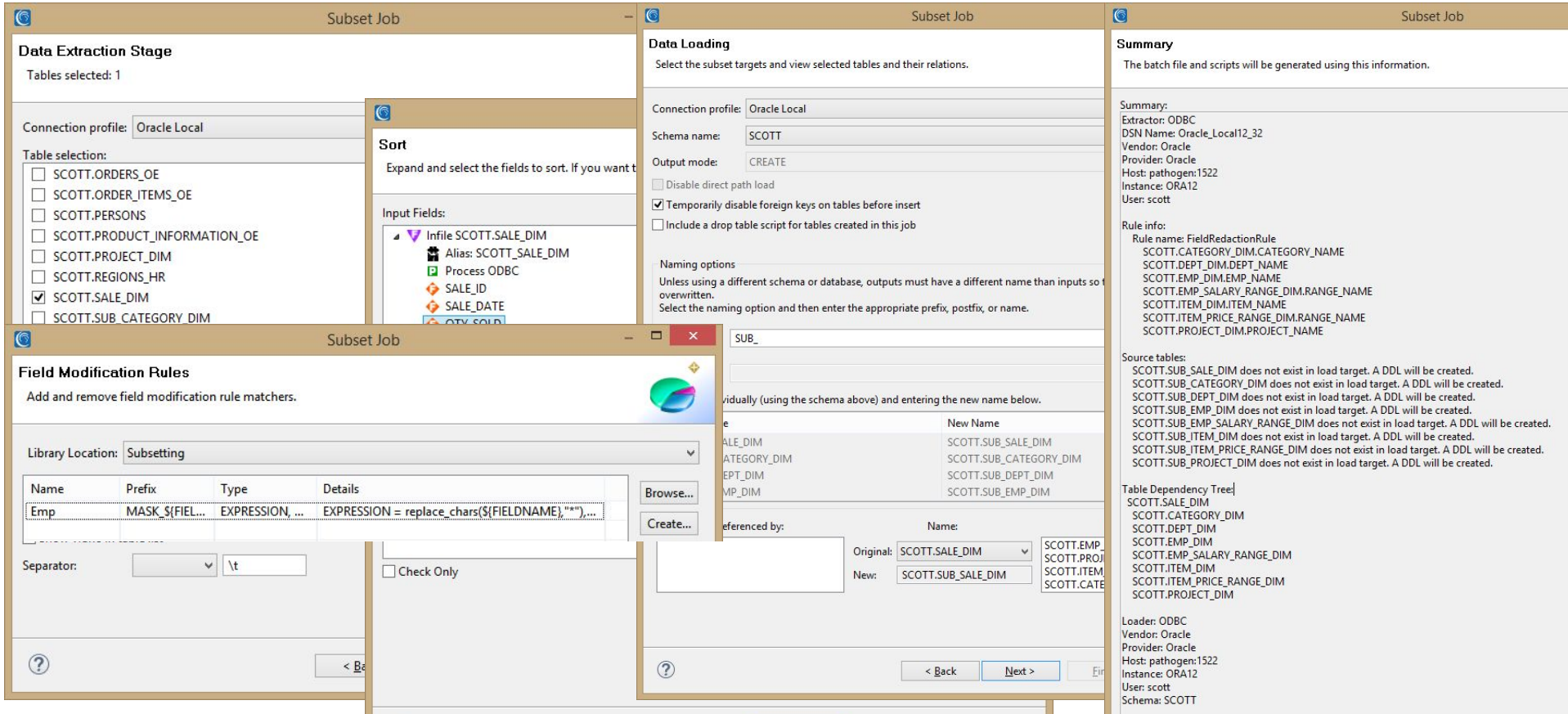


IRI
Total Data Management

© 2022 Innovative Routines International (IRI), Inc. All Rights Reserved.

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.



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

IRI
Total Data Management