IRI CellShield

PII / PHI Search & Mask in Excel

Version 2
Product Overview

What is CellShield EE?

IRI CellShield® Enterprise Edition (CellShield EE) is data-centric security software for Excel® spreadsheets.[1] CellShield EE combines data discovery in Eclipse and remediation in Excel to automate the classification, search, masking, and change-auditing of personally identifiable information (PII) in spreadsheets located in multiple computers on a local area network (LAN).

CellShield EE is one of three enterprise-class standalone static data masking products from IRI. These products are also members of the IRI Data Protector suite, and included at no additional cost in the IRI Voracity data management platform. They are designed and updated to comply with current US and international data privacy laws, and reduce the impact of data breaches.

While it is also possible to support Excel data in the course of IRI FieldShield (RDB and flat-file) and IRI DarkShield (semi- and unstructured) data discovery and de-identification, CellShield EE was purpose-built for Excel users concerned only with PII in spreadsheets, and who prefer to see and mask data directly in the Excel environment. CellShield EE provides point-and-click functionality to protect columns with reversible and non-reversible masking functions within a single sheet, or simultaneously across thousands of sheets on network drives.

CellShield EE shares the same data classes, search methods, and core masking functions as FieldShield and DarkShield. Thus, data masked in one platform can be unmasked in another.

CellShield EE exceeds the security and scope of a single password by combining and automating the capabilities described in this booklet. There is also a low-cost, no-evaluation Personal Edition (CellShield PE) available from IRI for masking data manually in a sheet, or one file at a time with the Autoprotect feature.

Booklet Contents & Help

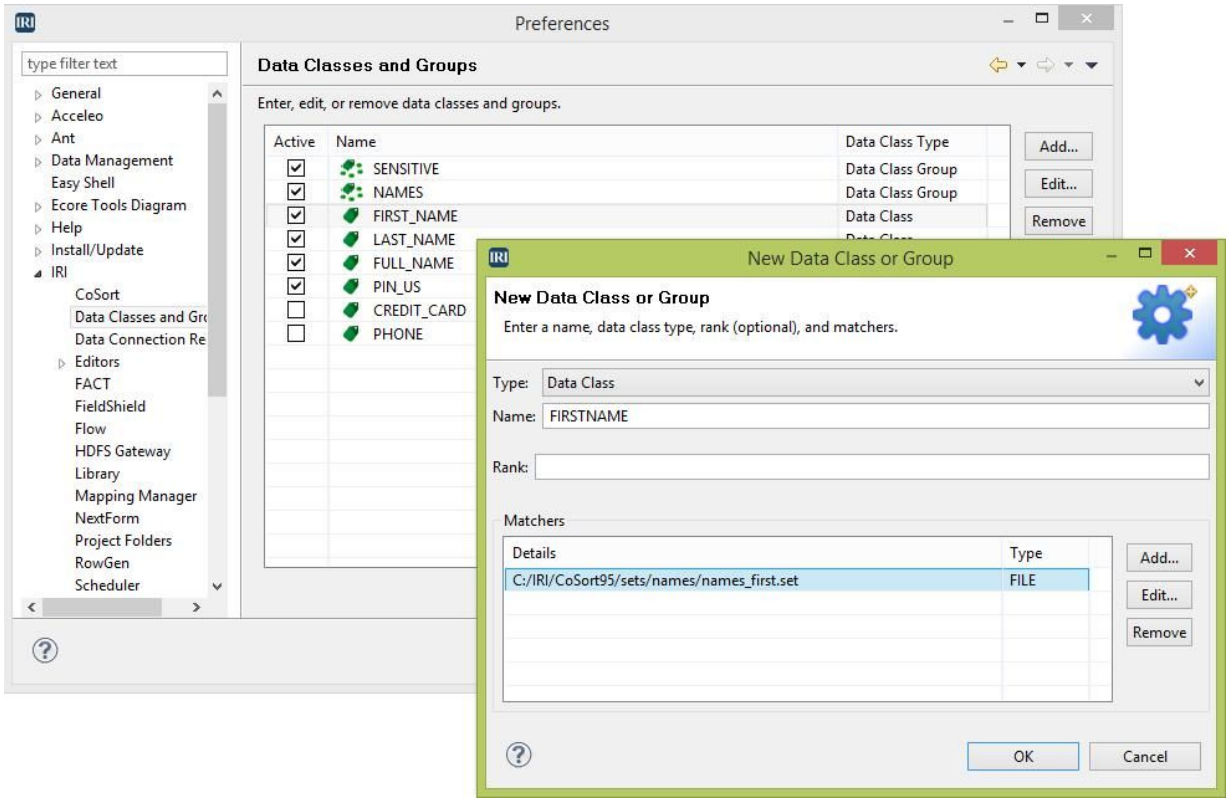
This product primer introduces details, and links to the specific how-to articles in the IRI blog site, for the following major CellShield EE functions:

For all CellShield EE technical or commercial inquiries, please email cellshield@iri.com.

[1] CellShield is a registered trademark of Innovative Routines International (IRI), Inc. Excel is a registered trademark of Microsoft Corporation. IRI and Microsoft are independent, but IRI is a member of the MSDN.

PII / Data Classification

Cellshield EE uses the data classification facilities in IRI Workbench as FieldShield and DarkShield to define and catalog one or more items of PII. These items, such as credit card numbers, names, or addresses can belong to individually named and saved Data Classes or Data Class Groups.



In CellShield EE, these reusable data classes or groups are further characterized by the search methods used by the Dark Data Search/Masking process run from IRI Workbench. One or more of the following methods can be used to find data in your sources that belong to your data class/group:

1. PATTERN - Strings conforming to IRI-supplied or custom-defined Java Regular Expression (RegEx) patterns (e.g., NIDs, email addresses, phone numbers)

2. FILE - Exact matches to strings in a lookup file/table (e.g., countries, products)

3. PATH - Uses Excel column names to filter specific data in context and save time

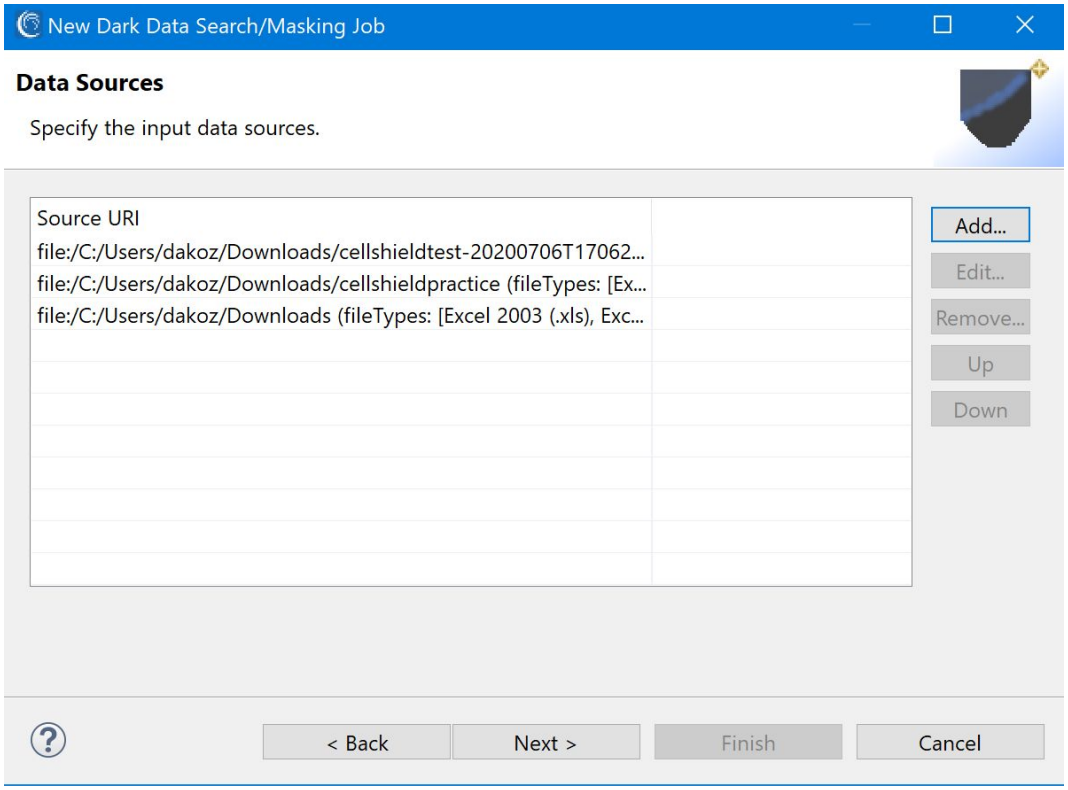
4. MODEL - Named-Entity Recognition (NER), based on machined-trained Natural Language Processing (NLP) Models (e.g., first and last names, street addresses).

Thus the search method(s) used should locate the items in each class that exist in every spreadsheet to be scanned in the next step.

Multi-Sheet Data & Metadata Discovery

The Dark Data Search/Masking wizard in avilable in the IRI Workbench DarkShield menu scans, extracts, and reports on sensitive data in .XLS and .XLSX spreadsheets and many other unstructured file formats. The search is performed on the basis of patterns, or defined data classes (see above).

The search can run for one or more folders, or across multiple disk drives and nodes on a LAN. Performance scales linearly, as the search is multi-threaded to leverage available cores on the platform running IRI Workbench.

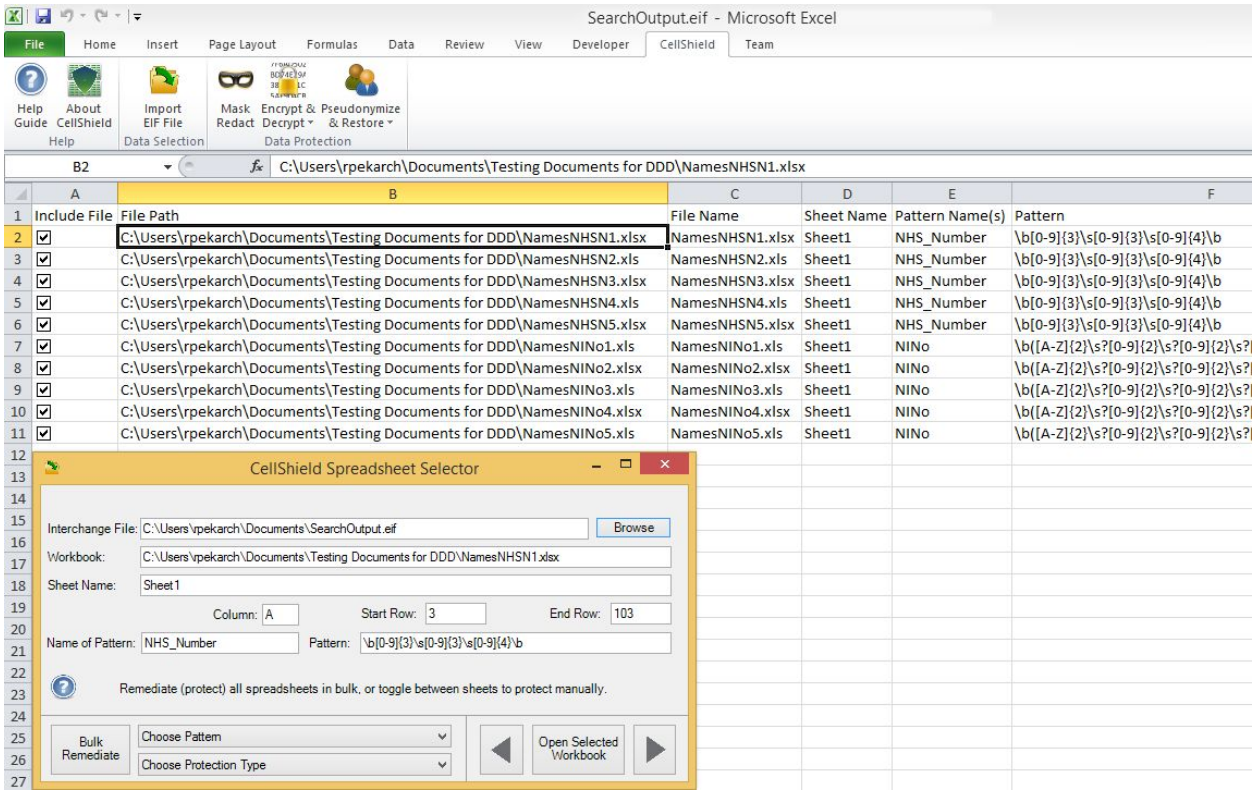


The search results obtained through the process defined in the wizard are written to delimited output files that can be used for database queries, reporting via the SortCL program in the IRI CoSort package or IRI Voracity platform, SIEM tool indexes, and off course in Excel for masking the found data by CellShield.

More specifically for CellShield EE purposes, an .EIF (Excel Interchange File) produced by the search wizard is used and described next, for reporting and bulk-masking results inside Excel.

Multi-Sheet Search Reporting

Once CellShield EE has been licensed and installed in Excel, the .EIF described above can be imported and opened. The search results open into a reporting and remediation worksheet profiling and locating all data discovered in the IRI Workbench Dark Data Search/Masking wizard.



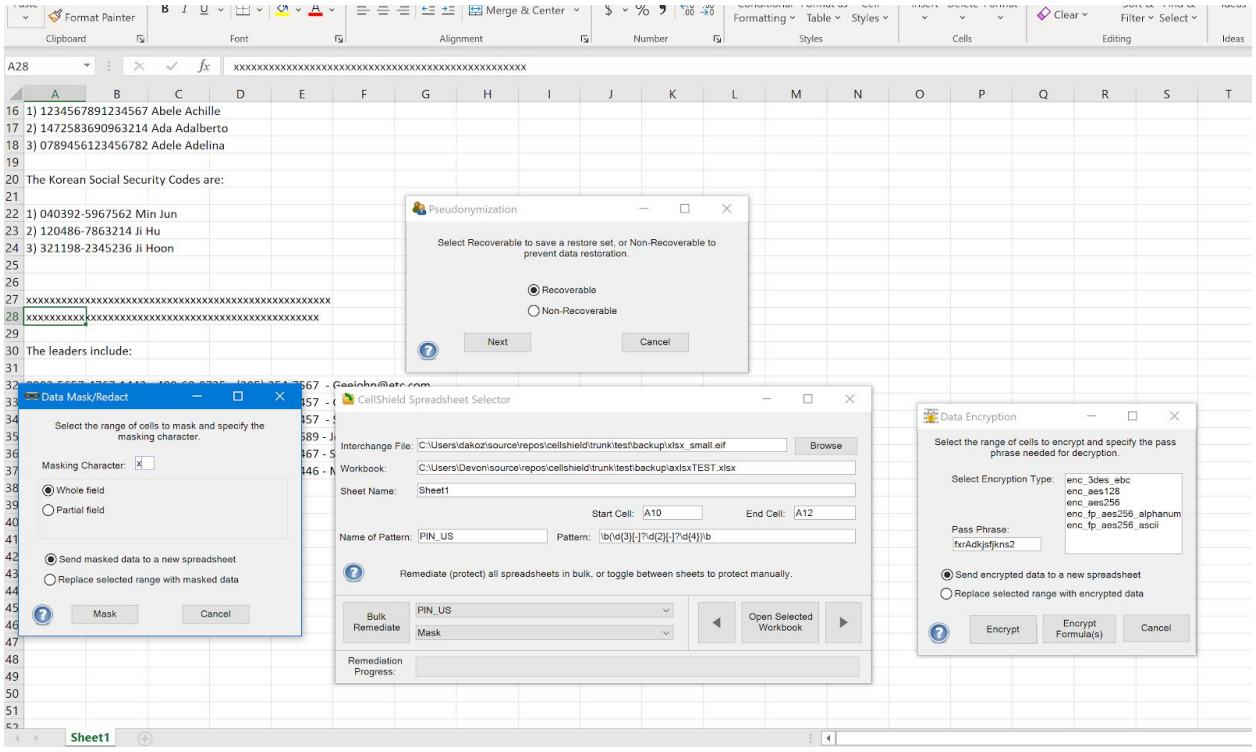
It is from this report sheet -- and the "Spreadsheet Selector" dialog that interacts with it -- where protection and logging actions against the pattern-matched columns are taken, one or more sheets at a time.

Ad Hoc or Bulk Data Masking

Built-in CellShield EE functional dialogs for encryption, redaction, or pseudonymization are used to provide point-and-click data masking on the data in ranges you select manually, or those that are pre-selected in a bulk remediation process controlled by the ‘Spreadsheet Selector’ dialog.

The masking function you choose for each pattern or range should be based on your business rules and the need each column has for security, reversibility, and appearance. The options for manual or bulk treatment are:

1. encryption (and decryption), 3DES or AES, including format-preserving encryption
2. redaction (full or partial cell, non-reversible)
3. pseudonymization (reversible or not) using original values or those in a lookup set



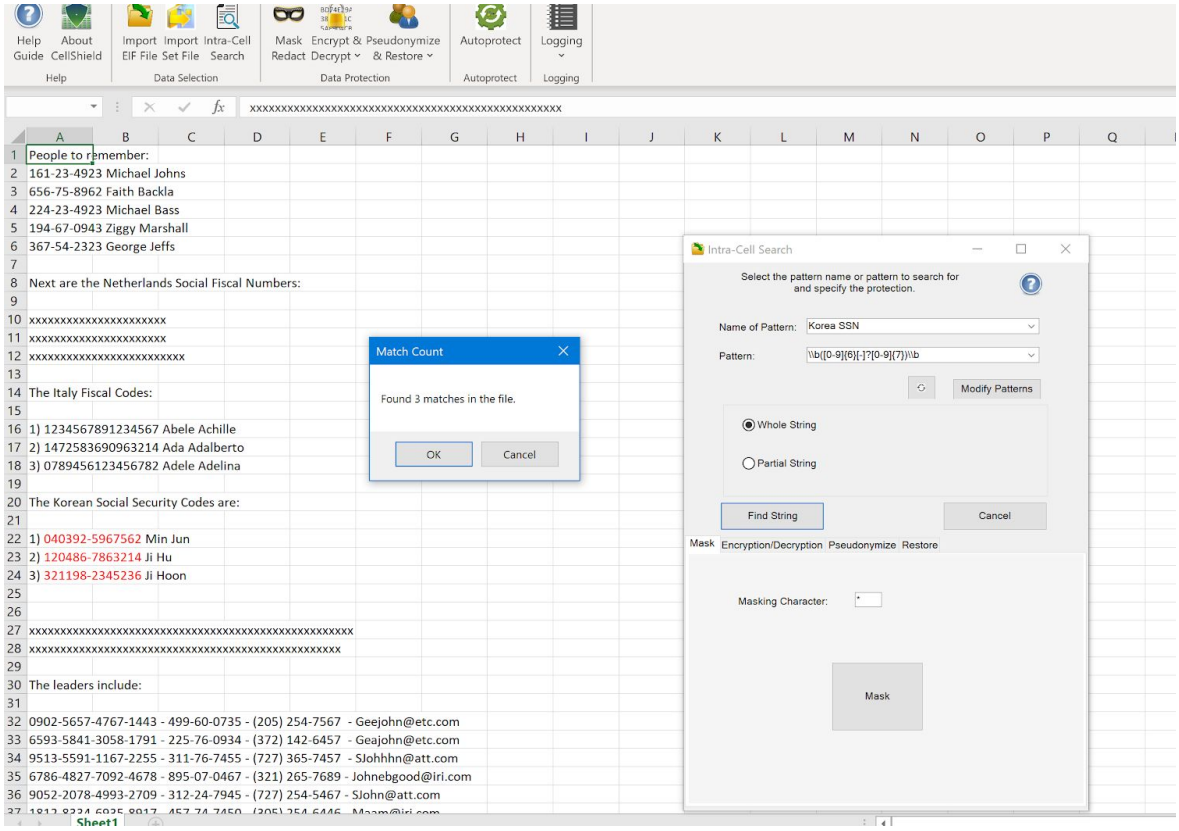
CellShield EE users can toggle between individual workbooks and ranges, or select the rows desired and mask them all in the same way with one click. Masking is either in-situ where the original values are overwritten, or the protected results are written into another sheet (recommended) in the same relative cell locations.

Intra-Cell Searching & Masking

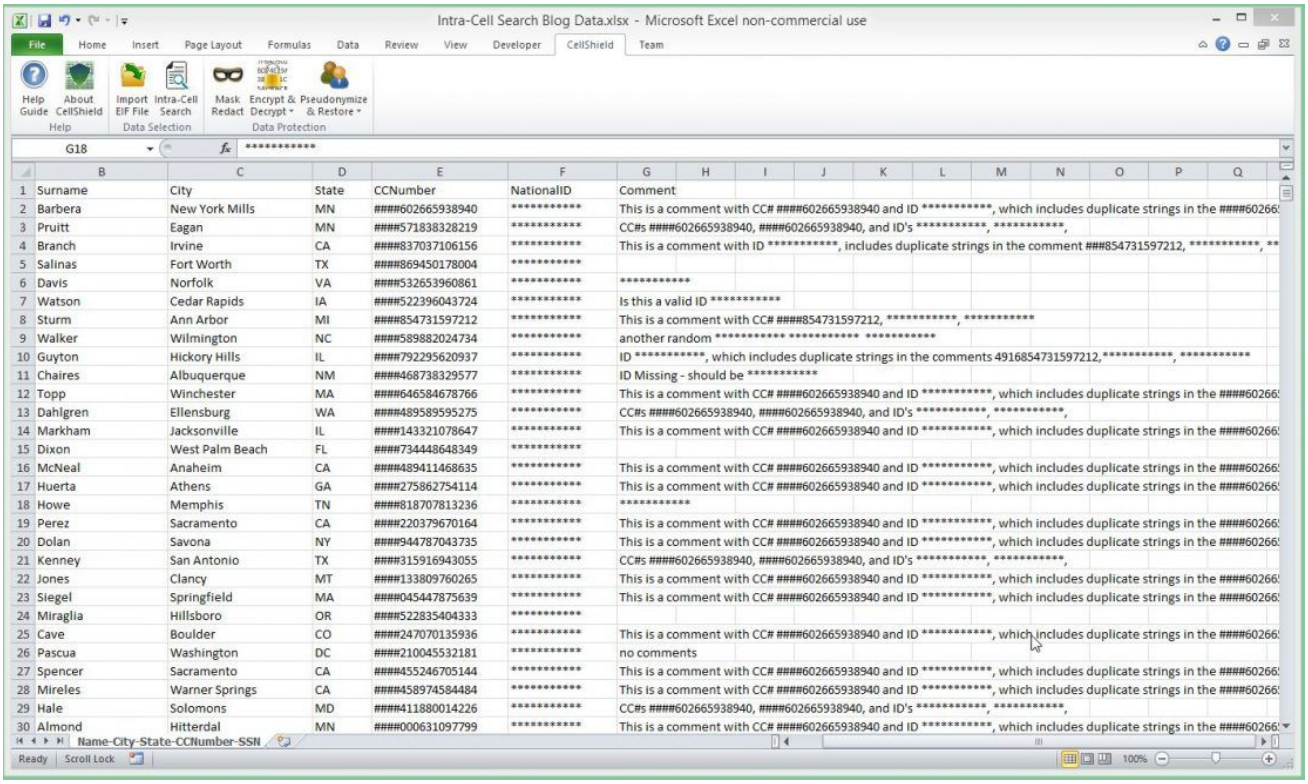
CellShield EE also includes an Excel-side “Intra-Cell” feature to automatically find and mask sensitive data floating within unstructured cell contents; e.g., a comments cell with free text.

A fit-for-purpose wizard in CellShield EE that runs in Excel supports the definition of -- or the selection of a supplied -- pattern to use for searching. These regular expression patterns can be used to find phone numbers, credit cards, email addresses and other sensitive information inside a cell with other values.

A custom pattern can be specified, and custom patterns can also be added to the list of patterns to save through the Modify Patterns menu.



After a pattern is specified, you can elect to mask, encrypt, or pseudonymize all or part of the values discovered.



Encryption and pseudonymization are reversible through the decrypt and recover options. Masking is non-reversible.

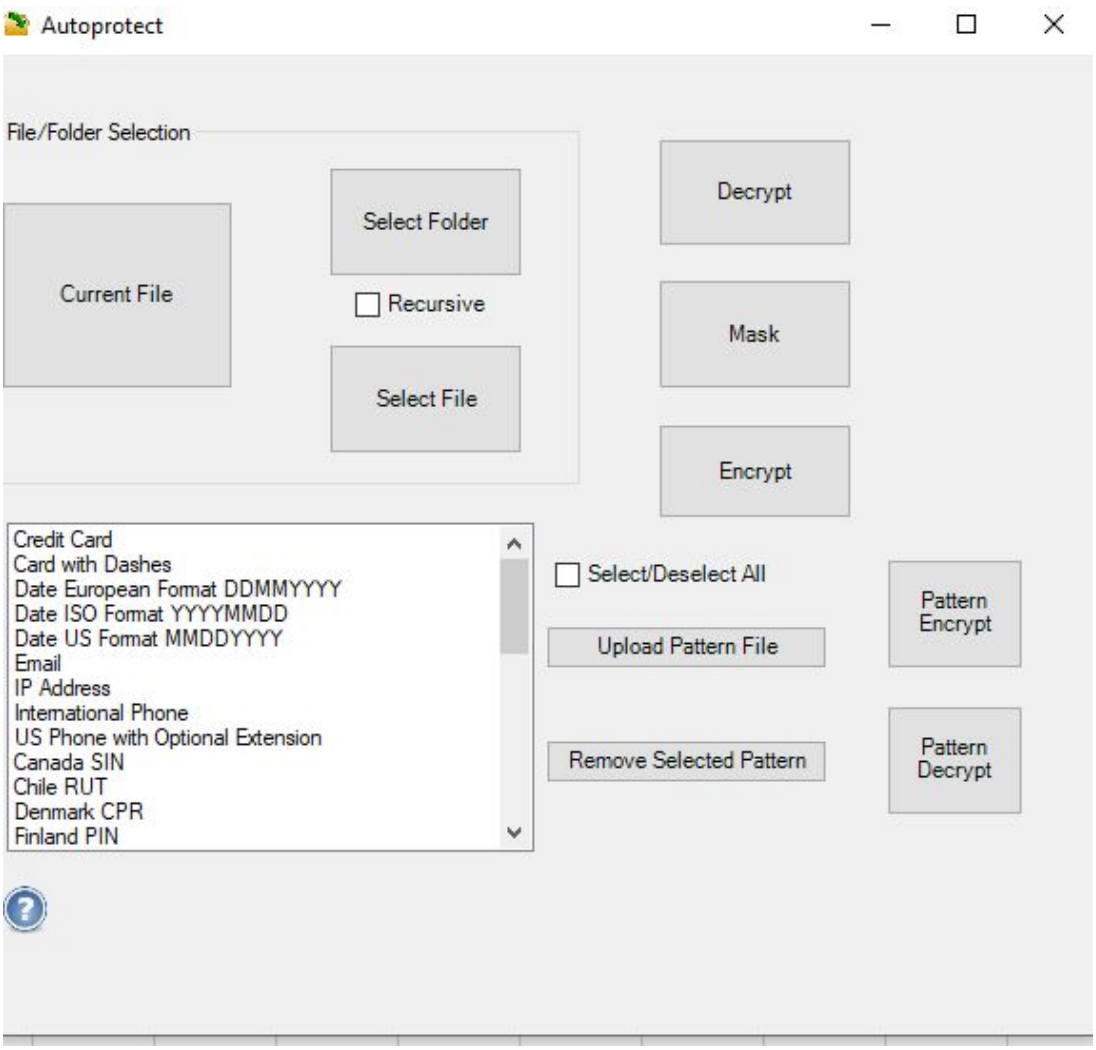
Autoprotect

The Autoprotect menu allows for simple protection of Excel files, either through protection of the entire file, or pattern encryption. To begin, you must first select a file, or folder containing Excel files.

A list of common patterns can be found in the bottom left corner of the menu, but this list can also be overwritten by uploading a custom pattern file. When pattern encrypting, only the patterns highlighted/selected in the list box in the bottom left corner of the menu will be used to search for matching text.

The Encrypt, Decrypt, and Mask buttons will perform these functions for all used cells in each sheet in every workbook that has been selected through the file/folder selectors. There is no need to specify encryption algorithm, password, masking character, or position with any of the Autoprotect menu functions.

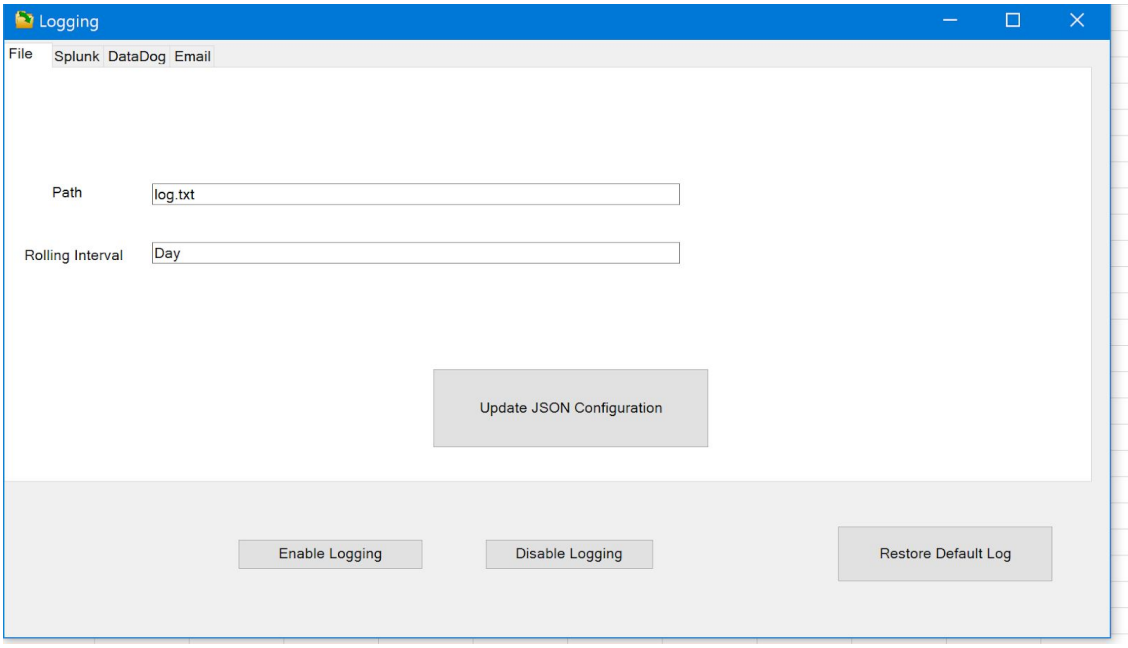
The screenshot below illustrates the Autoprotect options newly available in CellShield EE V2:



Logging

New in Version 2 of CellShield EE is the ability to specify logging sources via a JSON configuration file. The JSON file should be called appsettings.json and be located in the directory specified by the %CELLSHIELD_HOME% environment variable.

This file can be edited manually or interactively edited from the Logging Settings menu. Logging may also be turned on or off from this menu; it is set to off by default.

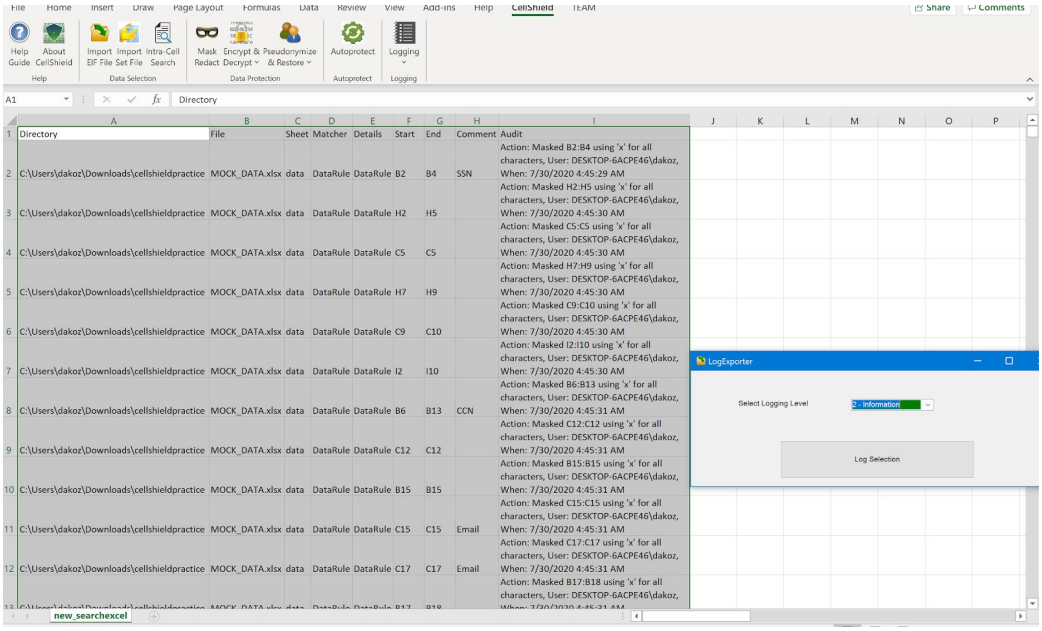


If the JSON configuration file is missing or malformed, a base skeleton can be regenerated at any time by clicking on the Restore Default Log button.

Specific information can be selected and logged at any time with the Export to Log menu. This means that even if logging was turned off during execution of a bulk remediation job, the audit details can still be logged later to any of the supported sources.

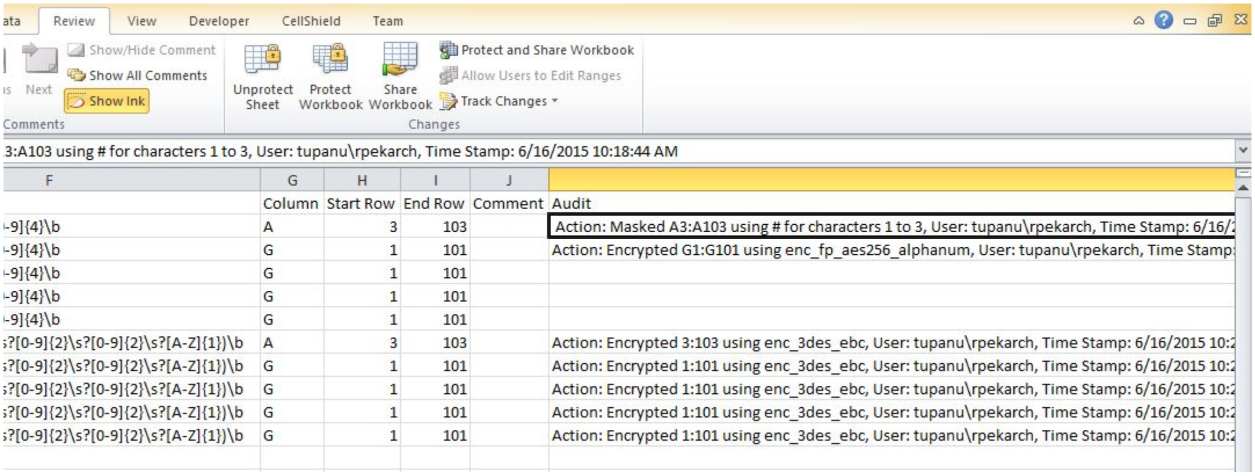
The content can be error messages, audit reports from bulk remediation, or selected content that has been manually exported. Alternatively, other logs not related to CellShield can be opened in Excel, selected and logged to one or all of the four sources supported in CellShield.

A logging level is specified to categorize the information being logged, especially in platforoms like Splunk and Datadog. Some logging sources have a default minimum logging level of 2 - Information. What this means is that if it is not otherwise specified in the JSON configuration file, no events will be logged if the level specified is below 2. Therefore, it is recommended that a logging level of at least 2 is used when utilizing the Log Exporter.



Audit Reporting

After each CellShield action is taken, an automatic log entry for each range updates the main report -- specifying what action was taken, when, where, and by whom -- to verify compliance with data privacy laws.



This sheet therefore becomes both a pre- and post-action report that can also be saved and secured offline. Its information may also be configured from the logging settings menu in CellShield to be additionally logged directly to: email, Datadog, Splunk, or local JSON files.

Product Evaluation & Licensing

IRI and its authorized representatives worldwide license CellShield EE for perpetual use on the basis of document volume, or free by subscription inside Voracity. CellShield EE is supported from evaluation and proof of concept to optimization and updates. Trial licenses are free for up to 30-days on physical or virtual Windows machines.

Maintenance (technical support and site-specific software updates) services are provided free of charge during the first year after installation. Subsequent annual maintenance is usually offered at 20% of the base license fee, and 24/7 technical support is available for a per-site premium.

U.S. educational and 501c(3) non-profit institutions qualify for a 10% license fee discount, and government agencies can buy CellShield EE from prime contractors on GSA schedule.

Confidential, site-specific cost estimates can be obtained through this form.

Professional Services

Beyond more complex CellShield EE operations, IRI provides help to classify and mask data at risk in other sources, and for many data management activities that IRI Voracity supports. An IRI professional services engagement allows you to leverage more than 100 collective years of data processing and IRI software experience behind a vast range of data-driven use cases.

Services include training, implementation, and support for activities Voracity performs, such as:

- *Big data wrangling* – packaging, protecting, and provisioning structured and unstructured data sets for analytic/BI, database (DB), and ETL operations
- *Data masking* – profiling, de-identification, encryption, tokenization and other services to aid data loss prevention, data governance, and data privacy law compliance efforts for data at risk in databases, spreadsheets and files -- structured, semi-structured, and unstructured -- on-premise or in the cloud
- *Risk scoring and certification* -- statistical analysis and legal advice pertaining to re-identification risk, HIPAA compliance, breach insurance and defense
- *Data replication and federation* – acquiring, re-mapping, and creating views
- *DB migration* – mapping data and relationships to new versions or platforms
- *Data conversion* – reformatting LDIF, XML, and COBOL index files (e.g., Vision, MF-ISAM), multi-byte characters, mainframe data types, and endians
- *Data quality and MDM* – outlier and fuzzy-matching value discovery and consolidation, format (re-)definition and validation, cleansing and enrichment
- *Program replacement* – translating cryptic and inefficient SQL, 3GL, ETL, legacy sort, and shell procedures into simple, portable, IRI 4GL text scripts
- *Test data management* – end-to-end services from DevOps needs definition through data generation and target persistence (without production data).

IRI

Total Data Management

INNOVATIVE ROUTINES INTERNATIONAL (IRI), INC.
2194 Highway A1A, Suite 303
Melbourne, FL 32937 USA
Phone +1-321-777-8889
www.iri.com