

# Mitigate Privacy Risks with Data Analytics

Insight AnalytiX - Privacy Risk Classifier

## Trends and Challenges

Enterprises today are faced with a diverse infrastructure that ranges from on-premise storage to the cloud. As organizations aim to leverage analytics and drive deep insights from their unstructured data, the growth of data and its complexity has left them exposed to risk and high costs.

Unstructured data is challenging to manage, as it does not easily lend itself to the more traditional models of data storage and analysis. However, if you can find a better way to manage this kind of data and understand its content, that unstructured data can be an extremely rich source of relevant information that can add tremendous value to your business.

Digital Subject Access Rights (DSAR) requests require enterprises to respond within **72 hrs.**

Enterprises are required to respond within **1 month** to GDPR requests.

## Why Unstructured Data?

In the traditional data space, structured data is what most have dealt with and are comfortable with. Not only is it easy to identify and understand what data elements you have based on your database schemas, but you also typically have a fairly good idea of where your data resides, because that data is managed. Additionally, there are tools today that help you to manage and catalog your data in the structured world.

Unlike this structured world, where information is stored in a predefined data model, unstructured data:

- Does not have a predefined data model and is not organized in a predefined manner
- Comes in many unrelated forms, such as documents, text files, spreadsheets, presentation files, video files, audio files, and pictures, which are the more common formats
- Is typically stored in individual files without an organizational application organizing and managing the relationships between those files

There is currently no easy way for us to understand what is really stored in the unstructured space without physically opening the files and looking. There are, of course, machine learning and artificial intelligence tools available to help you parse some of the unstructured world, but those type of projects require time, money, and the right data science team to train and productionize the model.

## Mitigate Data Risk with Insight AnalytiX - Privacy Risk Classifier

Insight AnalytiX is a platform that runs Privacy Risk Classifier, which recognizes files containing Personal Identifiable Information (PII). It processes more than 200 file types, including:

- Known file formats, such as XLSX, CSV, PPT, DOCX, and DOC
- Image formats like BMP, JFIF, or JPEG
- Scanned documents, as in PDF files

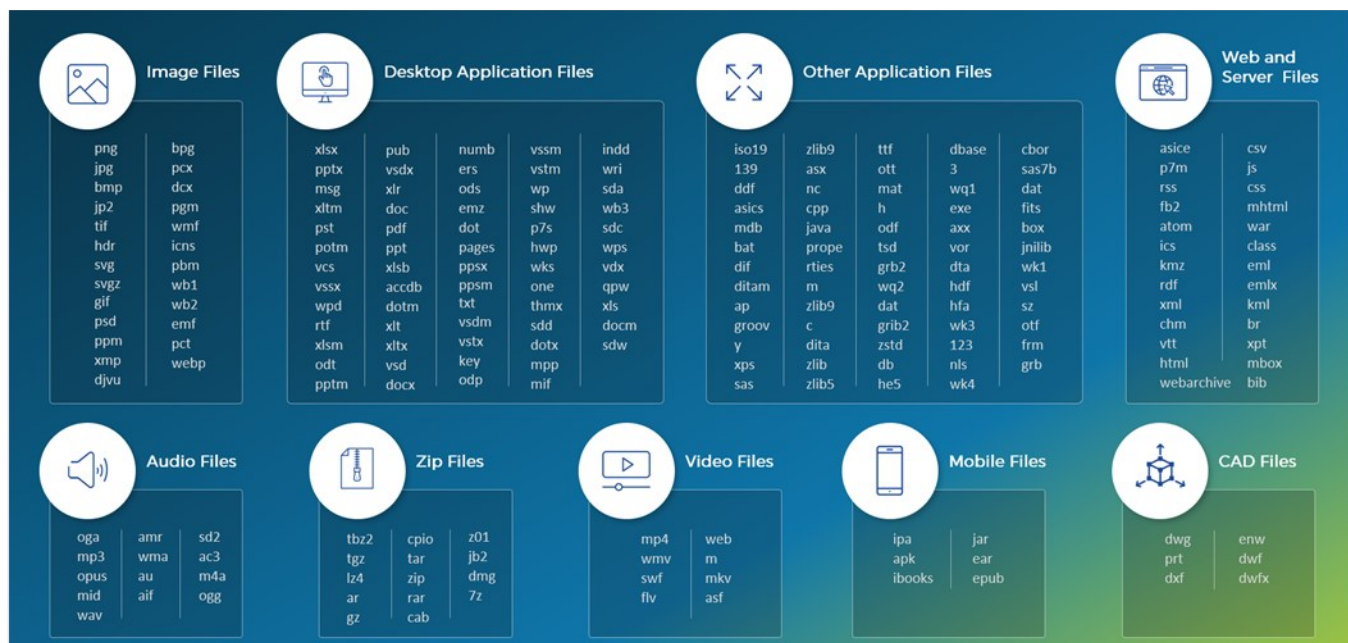


Figure 1: Insight AnalytiX Supports 200+ Unstructured Data Formats

Privacy Risk Classifier recognizes 49 different types of personally identifiable information, which include names, email addresses, mobile numbers, social security numbers (SSNs), and addresses.

This application is built on top of pattern recognition technology, keyword recognition technology, and artificial intelligence (AI). It comes with AI models, a pattern dictionary of 500+ advanced patterns, and a 200,000+ keyword dictionary. Role-based access (RBAC) is also part of the functionality, as Privacy Risk Classifier has a built-in authentication mechanism that can integrate with an Active Directory server.

## How Does it Work?

As a prerequisite, Insight AnalytiX requires StorageX to be installed in the same environment. A user creates an analysis set in StorageX before they can start using Insight AnalytiX's Privacy Risk Classifier. Privacy Risk Classifier works in coordination with StorageX and fetches lists of analysis sets from the StorageX server.

Once a user initiates data discovery, Privacy Risk Classifier connects to the network-attached storage (NAS) devices included in the selected analysis set and streams the files of that analysis set. It uses NFS (v3) and SMB (v2 & v3) protocols to stream files from NAS devices.

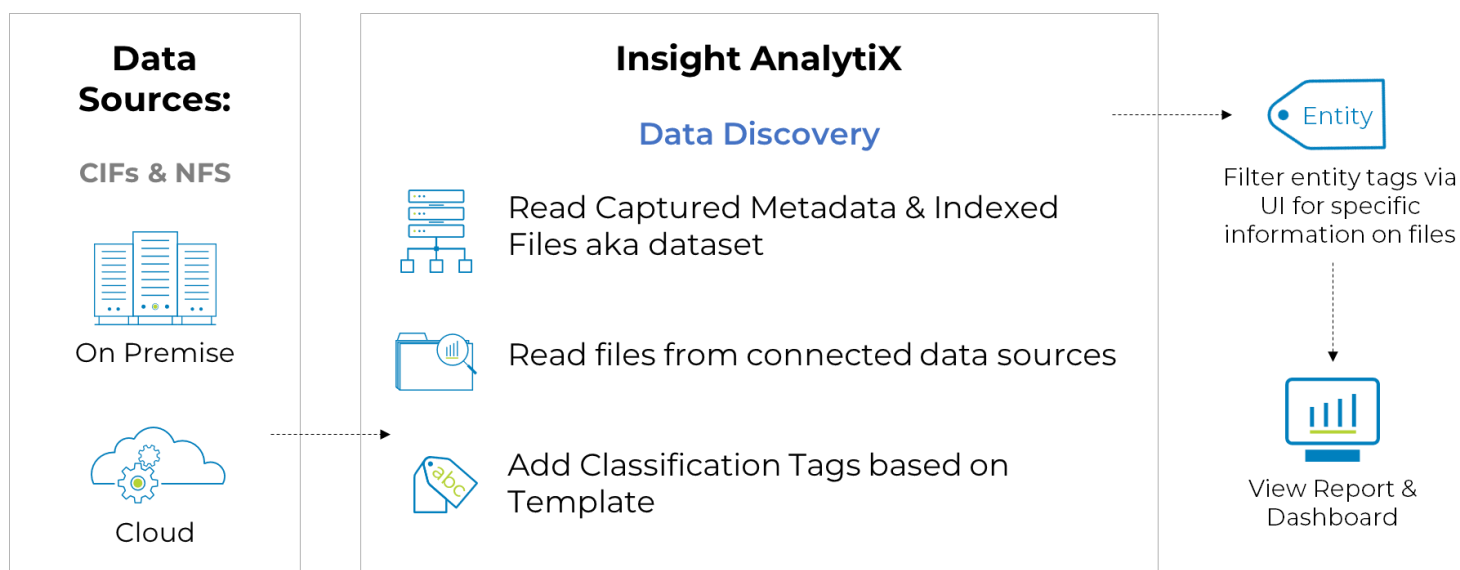


Figure 1: Insight AnalytiX Privacy Risk Classifier

## 1. Analyze & Classify your Unstructured Data

After an analysis set is processed, the user can select a template within Privacy Risk Classifier and view the data analysis. Each template includes a set of specific PII that are grouped for different types of use. Privacy Risk Classifier analyzes file data using the below metrics:

<b>PII Content</b>	Personal Files vs Non-Personal Files vs Files Not Analyzed
<b>Risk Analysis</b>	High-Risk Files vs Low-Risk Files
<b>Top 10 PII values found</b>	Top 10 PII values found for an analysis set
<b>Top 10 file types found</b>	Top 10 File Types values found for an analysis set
<b>File size discovered</b>	File Size ranges found for an analysis set
<b>Last Modified Date</b>	Files grouped in three groups based on these criteria: modified in last six month, modified in six to 12 months and modified before 12 months.

The user can then download reports in PDF, XML, XSL, and CSV file formats, as desired.

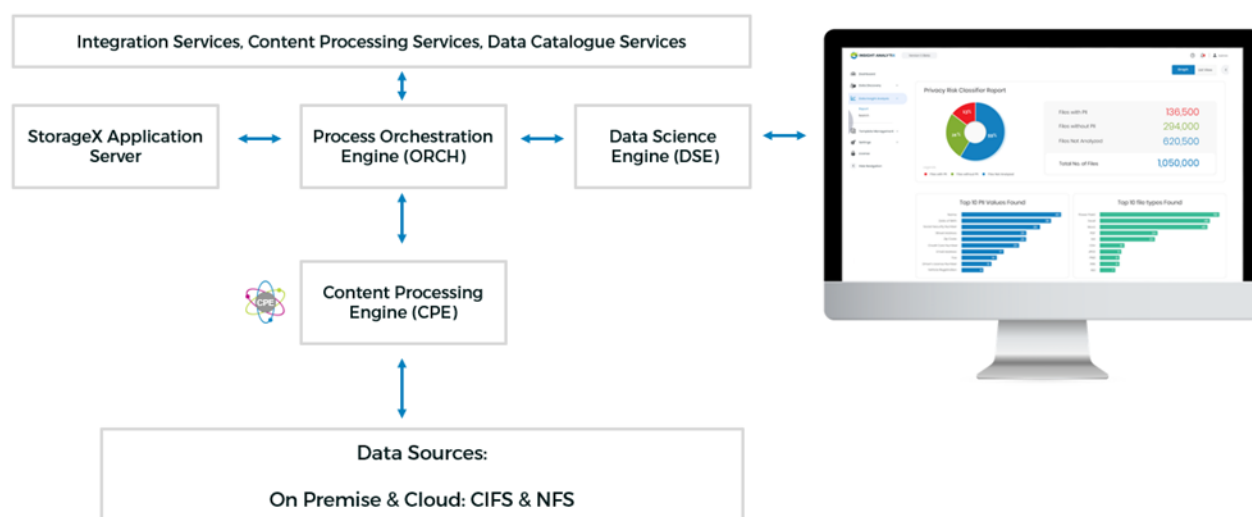


Figure 2: Main Components of Insight AnalytiX

## 2. Search

Insight AnalytiX has full text-search functionality. It allows a user to search a processed analysis set or all processed analysis sets in their environment. A user can also search with the help of “AND” and “OR” operators, giving the user the flexibility to search more intelligently. For example, using the AND operator, a user can make all words in the search text mandatory in the search results; in the case of the OR operator, a user can make at least one of the search text words mandatory in the search results.

The latest version of Insight AnalytiX allows users to generate a Data Insight report on a dataset by building advanced, multi-level logical expressions using a combination of logical operators. This ensures increased accuracy in PII/sensitive data discovery, backed by deep analytics (both descriptive and diagnostic).

Search returns the search results using file type-based filters, allowing users to see results for a specific file type. The user can export the results in CSV file format.

### 3. Role-Based Access

Insight AnalytiX includes three pre-defined roles and allows users to create roles based on their specific needs:

<b>System Administrator</b>	Access all Insight AnalytiX – Privacy Risk Classifier functionality
<b>User Administrator</b>	Configure users, roles, and application configuration
<b>User</b>	Create templates, run scans, download results, and search analysis sets

Insight AnalytiX comes with its own user authentication mechanism or can be configured to authenticate users against an existing Active Directory server.

### Conclusion

The key to managing your unstructured data – and therefore optimizing your storage footprint – is knowledge. Knowing what data you have in your environment is essential to figuring out what you can and should do with that data.

Since StorageX is tightly integrated with Insight AnalytiX, users have an unprecedented ability to dig deeply into their unstructured data and find data that could be potentially sensitive or need to be organized or stored differently.

Using Privacy Risk Classifier, you can quickly and easily find and categorize your data, then use the analysis results to determine what data may need to be migrated, what sensitive data may need to be secured, and what data should be removed entirely. This information powers data optimization on existing storage and helps you make better plans for future storage usage, as well as avoid potential liabilities.

Find answers to your questions by contacting [solutions@datdyn.com](mailto:solutions@datdyn.com)

See Insight AnalytiX in Action and [Book a Demo](#)

[Data Dynamics](#) is a leader in intelligent file management solutions that empower enterprises to analyze seamlessly, move, manage and modernize critical data across hybrid, cloud, and object-based storage infrastructures for true business transformation. For more information, please visit [www.datadynamicsinc.com](http://www.datadynamicsinc.com).



© 2022 Data Dynamics. The trademarks Data Dynamics, StorageX, Insight AnalytiX, ControlX, Universal Data Engine, UDE, Insight Analytics, Manage Your Data, Champion of Data, and Dynamic File Management are the registered or unregistered trademarks of Data Dynamics, Inc. All other trademarks are the property of their respective owners.