

# Content Identification Service Understand and Classify your Data.

## PII Discovery and Identification.

Determining the amount and location of sensitive information across multiple unstructured data repositories is a significant undertaking and typically serves as a barrier to accomplishing this task. The Content Identification Service (CIS) allows organisations to take the first steps into discovering and understanding the amount of sensitive Personal Identifiable Information (PII) that exists within their unstructured data repositories. CIS provides an invaluable insight into the size and scale of your GDPR data compliance challenges, identifying sensitive information that is central and relevant to the success of GDPR compliance in a streamlined and automated fashion.

The CIS provides a detailed analysis and report, based on the data content of the ingested data from an identified data sample (Maximum 8TB). The study offers an invaluable insight into the level of data file duplication, data types, individual file size and ageing characteristics, while also classifying data items that contain PII/sensitive information.

## Service Deliverables.

CIS comprises of two principal phases;

**Phase 1: Understand and classify your data.**

Discover the location of documents that contain PII and company sensitive information.

**Phase 2: Take action.**

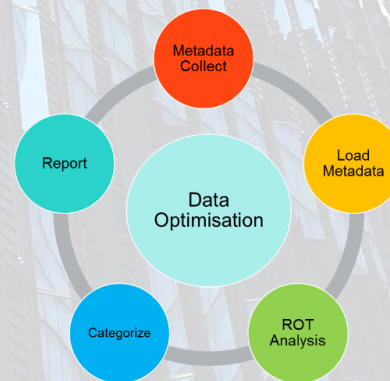
The following actions can highlight the business risks and challenges associated with data files that contain PII or sensitive information;

**Action 1.** Reduce the data management challenge (and associated storage costs and liabilities), by identifying Redundant, Obsolete, and Trivial (ROT) data items.

**Action 2.** Sort & Categorise the remaining data into information that is relevant to business, therefore, considered “in scope”.

**Action 3.** Classification of data items into simple risk domains, i.e. low, medium and high.

## Service Activities.



**Figure 1. Data Optimisation Methodology.**

Delivery of the CIS service requires the following tasks/actions and deliverables:

- Pre solution deployment meeting/ conference call. The call is generally an exchange/data gathering conference call to agree on the technical requirements.

- Statement of Works (SoW) describing the Project methodology, approach and software implementation will be issued.

- The SoW will include references to the;

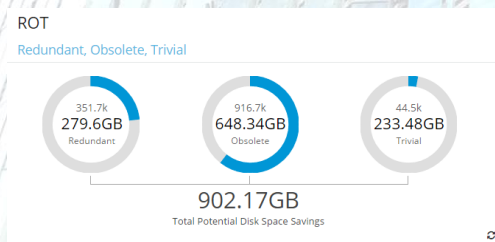
- .1. Unstructured data source.
- .2. Proposed “PII” education parameters.
- .3. Hardware infrastructure.

- The SoW will outline the technical requirements for the CIS implementation i.e. VMware clients, local storage, Operating system versions.

- Post the on-site software installation and configuration. The data ingestion process will be started.

- When the data ingestion and content indexing tasks have completed, the initial analysis will be performed;

- .1. High-level ROT data analysis.

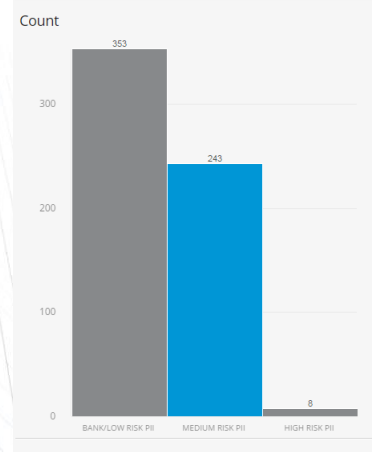


**Figure 2. Data ROT Analysis.**

- .2. PII and Sensitive information categorisation.

(Content Identification Service v2.3).

By Category Low Risk PII informa... -



**Figure 3 Low, Medium and High-Risk PII Classification.**

- .3. Determine the level of “structured” vs “unstructured” data.
- .4. Determine the percentage of file duplication.
- .5. Identify obsolete aged data files.

- Findings and recommendations to be presented and discussed after the customer has received the written report.

**Service Exclusions.**

It is important to note that this service excludes explicitly the following:

- Any remedial activity. However, if assistance is required to address some or all of the findings, it is suggested you request details of additional packages or consultancy services
- Any configuration changes.
- Capacity Planning or detailed monitoring of the system.

- A hardware/patch audit of the system.

- Availability of a Windows Service Account which can be associated with the associated installed software services.

### Delivery Pre-Requisites.

To ensure this service is delivered efficiently and effectively it is critical that the customer.

- Ensures that technical assistance is made available as and when requested by the specialist performing this study.
- The customer identifies the single (One) data repository (UNC path) that will be indexed.

### Software Pre-Requisites.

- 3x Vmware Virtual client machines each with Windows 2012/2016 installed and configured.

### Hardware Pre-Requisites

- Sufficient local storage to host the ingested data (typically, 5% of the volume of the Ingested data)

### Customer's environmental Pre-requisites.

As above, also including: