HELIACA
FINDER

INDICE

1. Introduzione 4

1.1. Ricerca 4

1.2. Classificazione 4

1.3. Report 4

1.3.1. Token 4

1.3.2. Dashboard 5

2. Modalità di utilizzo 7

2.1. Utilizzo rapido 7

2.2. Utilizzo integrale 7

3. Widget fondamentali per l'esecuzione della ricerca 8

3.1. Avvio della ricerca 8

3.2. Ripresa della ricerca 10

3.3. Pulsanti di Utilità 11

3.4. Treeviews 12

3.4.1. Treeview parametrico 12

3.4.2. Treeview Data Discovery 13

4. Frame 14

4.1. Home 14

4.2. Data classification 14

4.2.1. Tabella descrittiva 15

4.2.2. Tabella operativa 16

4.3. Data discovery 17

4.3.1. Impostazioni & Utilità 17

4.3.2. Keyword 18

4.3.3. Estensioni 18

4.3.4. Documenti 18

4.3.5. Dati comuni 18

4.3.6. Dati particolari 18

4.4. Locale 19

4.4.1. Barra di ricerca 19

4.4.2. Esecuzione ricerca locale 19

4.4.3. Remota 20

4.5. OneDrive 20
4.5.1. Login utente 20
4.5.2. Login enterprise application 20
4.5.2.1. Creazione dell'enterprise application 21
4.5.2.2. Assegnazione delle autorizzazioni 22
4.5.2.3. Inserimento delle credenziali per avvio ricerca 23
4.5.3. Esecuzione ricerca OneDrive 25
4.6. Sharepoint 26
4.6.1. Login enterprise application 26
4.6.2. Esecuzione ricerca Sharepoint 26
4.7. Outlook 26
4.7.1. Login mail 27
4.7.2. Login enterprise application 27
4.7.3. Esecuzione ricerca Outlook 27
4.8. Google Drive 28
4.8.1. Login utente 28
4.8.2. Login service account 28
4.8.2.1. Creazione del Service Account 28
4.8.2.2. Assegnazione delega a livello di dominio e assegnazione degli ambiti 30
4.8.2.3. Attivazione Admin SDK API 32
4.8.2.4. Utilizzo del Service Account 33
4.8.3. Esecuzione ricerca Google Drive 33
4.9. Gmail 34
4.9.1. Login mail 34
4.9.2. Login service account 34
4.9.3. Esecuzione ricerca Gmail 35

1. Introduzione

Heliaca Finder è un software per la ricerca, la classificazione e la generazione di report sintetici inerenti ai file e dati critici trovati. Le ricerche possono essere effettuate su diverse sorgenti: Locale, OneDrive, Sharepoint, Google Drive, Gmail e Outlook. Vengono classificate le categorie particolari dei dati personali che hanno una grande importanza ai fini del GDPR e alcune tipologie di documenti trovati, come fatture, certificati di malattia, ecc.

1.1. Ricerca

Il processo di ricerca dei file o elementi sensibili sulle sorgenti sopra elencate ha lo scopo di individuare dati sensibili, rientranti nella tipologia dei dati comuni secondo il GDPR (codice fiscale, carta d'identità, patente, ecc.), nel testo dei file, allegati o corpo delle e-mail analizzate.
È possibile scegliere nell'apposita sezione quali dati comuni ricercare e quali estensioni saranno contemplate dalla ricerca, nonché le parole chiave "keyword" che possono trovarsi nel testo degli elementi.
Il processo di ricerca negli archivi in cloud può essere effettuato secondo due modalità diverse:

- La modalità **singolo utente** o **cassetta postale** permette di analizzare il contenuto testuale dei file ed e-mail posseduti dall'utente o presenti nella cassetta postale.
- **L'Enterprise Application** (Microsoft) o **Service Account** (Google) permette invece di effettuare una ricerca su tutti gli utenti o cassette postali presenti nel dominio aziendale.

È possibile **riprendere** una ricerca sospesa o interrotta a seguito di un imprevisto selezionando la ricerca effettuata nell'apposita sezione.

1.2. Classificazione

Il processo di classificazione effettuato dal software ha l'obbiettivo di identificare e catalogare gli elementi trovati durante la ricerca secondo diversi criteri:

- Un criterio **personalizzato** che permette tramite l'utilizzo dell'apposito frame "Data Classification" di impostare una classificazione personalizzata in base alle esigenze del richiedente la ricerca.
- Un criterio basato sul tipo di **dato comune** e **dato particolare** identifica gli elementi in base alla rispettiva tipologia di dato trovato all'interno del testo, ad esempio (Codice fiscale, Partita Iva, ecc.) nel caso di dati comuni oppure (medico, biometrico, ecc.) nel caso di dati particolari.
- Un criterio basato sul tipo di **documento** permette di identificare i file o allegati trovati in base sempre al contenuto del testo, ad esempio (Fattura, Certificato di malattia, ecc.).

1.3. Report

La funzionalità di reportistica si occupa di trascrivere i risultati ottenuti dalla ricerca su tre diversi documenti:

- Un file tabellare in formato .xlsx chiamato File_sensibili.xlsx dove si trovano tutti i file sensibili con le relative informazioni incolonnate in modo rigoroso, utile per una revisione tecnica sui file trovati.
- Un file in formato .docx chiamato Data_discovery.docx dove si trova un riassunto della ricerca effettuata con un particolare focus sulle criticità trovate e sulle possibili soluzioni, utile nel caso in cui si voglia avere un resoconto rapido sui dati sensibili posseduti e i possibili rimedi. Da integrare poi in base alla situazione specifica del cliente.
- Un file in formato .pbix chiamato Dashboard [Sorgente Data Discovery].pbix che contiene al suo interno grafici interattivi, i quali danno la possibilità di estrapolare moltissime informazioni sul quantitativo e la postura dei dati sensibili in massimo 2 o 3 pagine.

Queste tre tipologie di report vengono creati automaticamente all'interno di una cartella avente il nome Data Discovery [Sorgente Data Discovery + data], cartella che si consiglia di comprimere in un file zip protetto da password, visto l'elevata criticità di questi documenti.

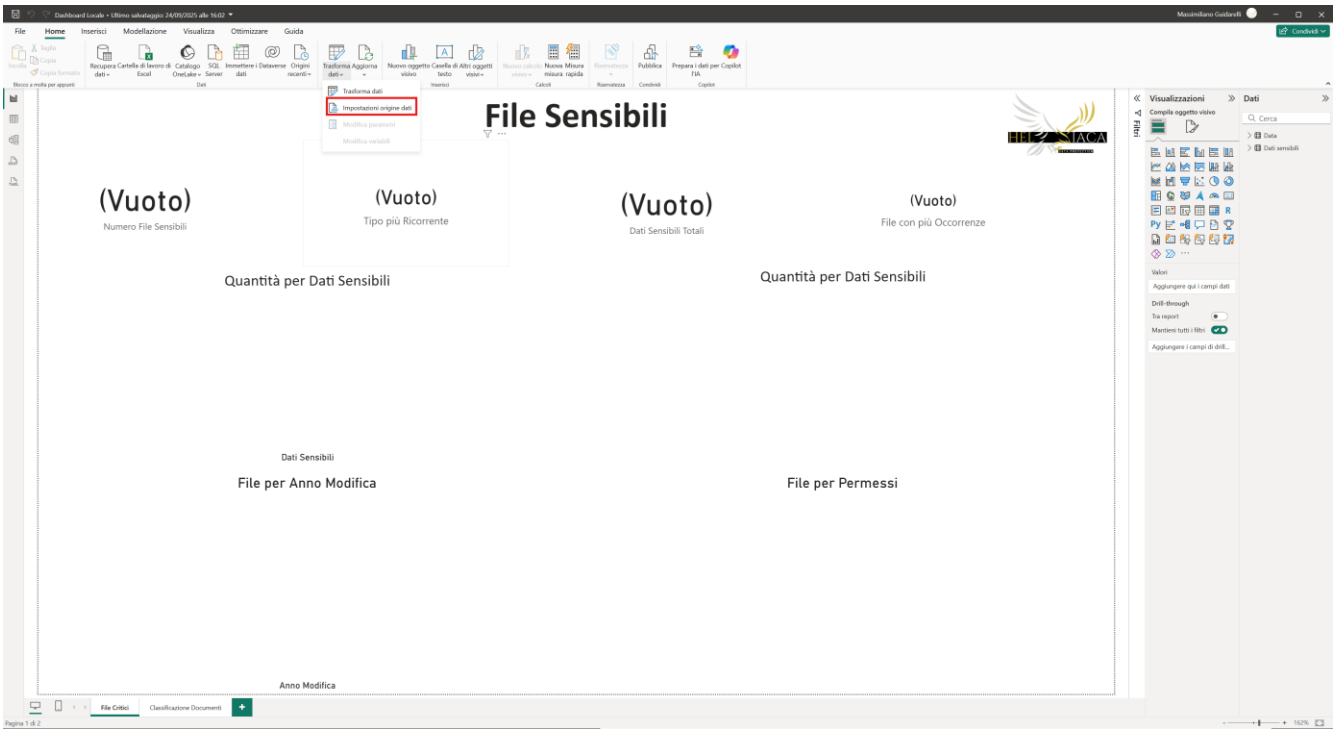
1.3.1. Token

È possibile estrapolare i report solo essendo in possesso di un TOKEN, il quale indica al software una data entro cui è possibile fare le estrazioni e se possono essere elaborati i report in formato .pbix.
Il token consiste in una stringa alfa-numerica contenuta nel file Report_token.json il quale si trova all'interno della cartella Save dentro a sua volta il percorso dove è installato il software Heliaca finder (di default "C:\Users\Heliaca\AppData\Local\Heliaca Finder".

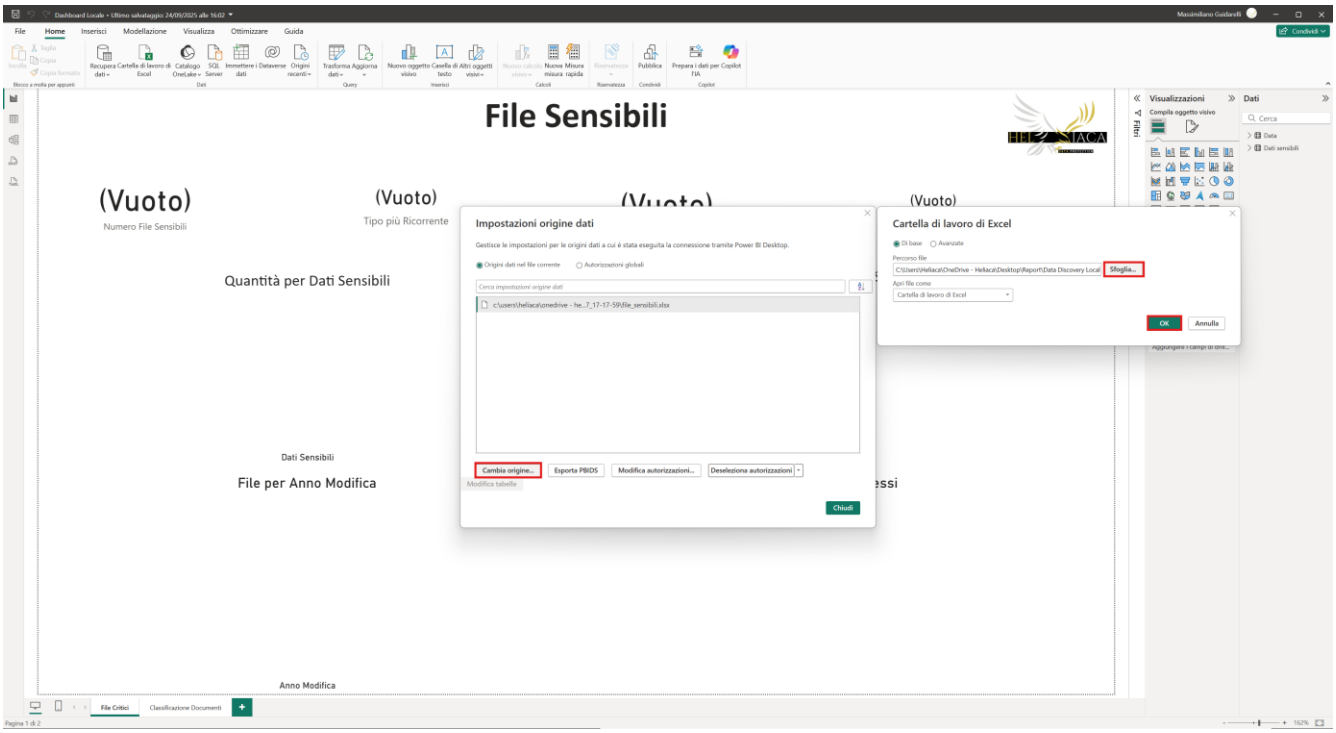
Il pulsante di report ROSSO indica che il token non è presente oppure è scaduto.
Il pulsante di report VERDE indica che il token è valido e si può, una volta finita la ricerca, procedere all'estrazione.

1.3.2. Dashboard

La dashboard per essere utilizzata ha bisogno di un passaggio manuale, ovvero deve essere selezionata la fonte dei dati che vengono utilizzati per costruire i grafici, i quali sono conservati nel report .xlsx.
Alla prima apertura la dashboard si presenta vuota. Cliccare su **Trasforma dati** eppoi su **Impostazioni origine dati**



Successivamente cliccare su **Cambia origine**, selezionare il report File_sensibili.xlsx che si trova accanto al file .pbix cliccando su **Sfoglia** e infine premere **Ok**.



L'ultimo passaggio consisterà nel cliccare su **Applica modifiche** (in alto a destra), dopo una breve attesa i grafici precedentemente vuoti saranno riempiti con i dati selezionati.

2. Modalità di utilizzo

Heliaca Finder è stato ideato per un utilizzo semplificato e rapido. È possibile però fare un tuning dei settings disponibili per rendere più adattabile e specifica la ricerca alle esigenze dell'utilizzatore/cliente.

2.1. Utilizzo rapido

In pochi click, in media 3, si può far partire una ricerca su diverse fonti sia locali che remote. Di default sono impostati per la ricerca tutti i:

- Dati comuni (eccetto Telefono e Cellulare i quali generano più falsi positivi data la loro forma semplice)
- Dati particolari
- Estensioni (eccetto quelle delle immagini che richiedono l'OCR)
- Documenti

Al **termine** della ricerca è possibile visionare i risultati trovati interagendo con il treeview e se in possesso di un Token estrarre i relativi report.

2.2. Utilizzo integrale

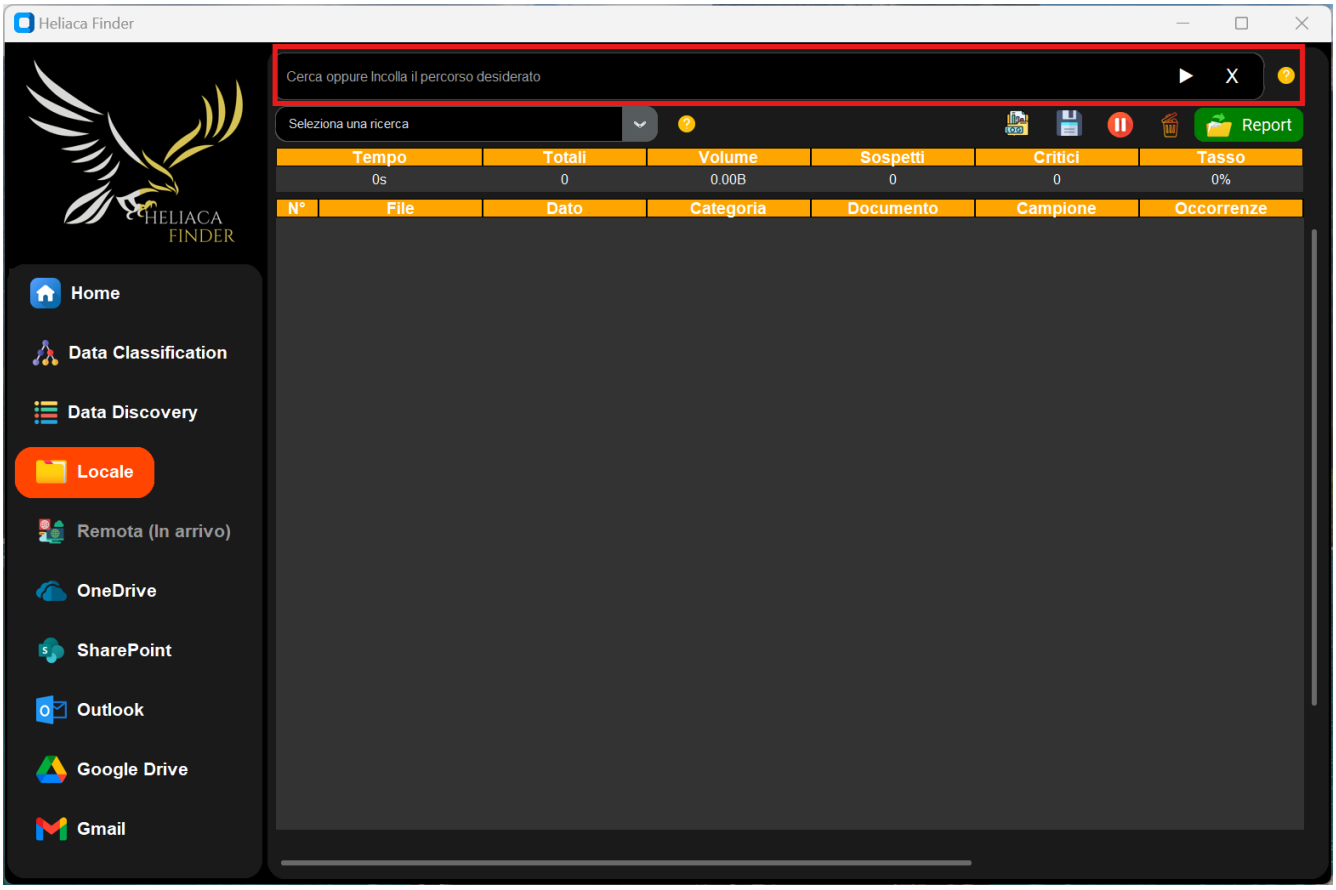
Se si vuole rendere la ricerca più adatta alle esigenze dell'utilizzatore/cliente e soprattutto più precisa è possibile effettuare diversi accorgimenti.
Nel frame dedicato alla **Data Classification** è possibile assegnare per ogni dato comune e ogni dato particolare un valore numerico da 1 a 4 interagendo con la relativa tabella. Questi valori indicano il grado di Valore, Riservatezza, Integrità e Disponibilità che un dato sensibile può avere nei confronti di un determinato cliente/azienda. Il tutto necessario per assegnare un diverso livello di classificazione ad ogni file critico trovato.
Nel frame **Data Discovery** è possibile effettuare un tuning preciso dei dati che si vogliono includere nella ricerca selezionando o deselezionando i relativi pulsanti, oppure introdurre la ricerca di keyword, attivare l'ocr, indicare una data di ricerca più stringente in cui i file sono stati modificati e infine una diversa precisione di individuazione e classificazione dei dati particolari.
Sempre al **termine** della ricerca è possibile visionare i risultati trovati interagendo con il treeview e se in possesso di un Token estrarre i relativi report.

3. Widget fondamentali per l'esecuzione della ricerca

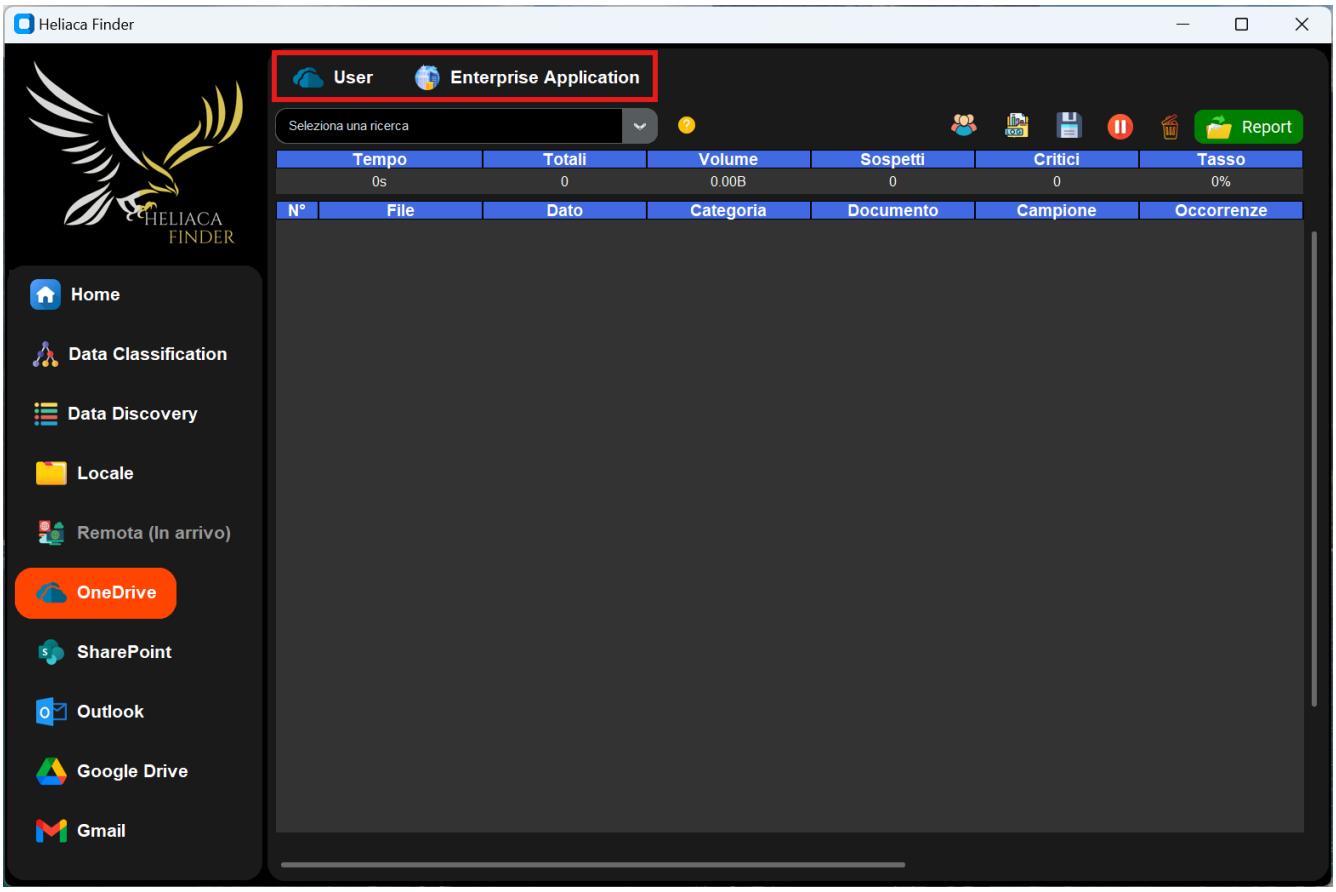
Ci sono in ogni frame dedicato alle varie tipologie di Data Discovery, delle "classi" di **widget** (oggetti dell'interfaccia grafica) importanti per il controllo del processo di ricerca.

3.1. Avvio della ricerca

Nella parte alta dei frame di Data Discovery si trovano i widget necessari per avviare il processo di ricerca.
Nel caso di una ricerca **Locale** troveremo la entry che permette di selezionare il percorso

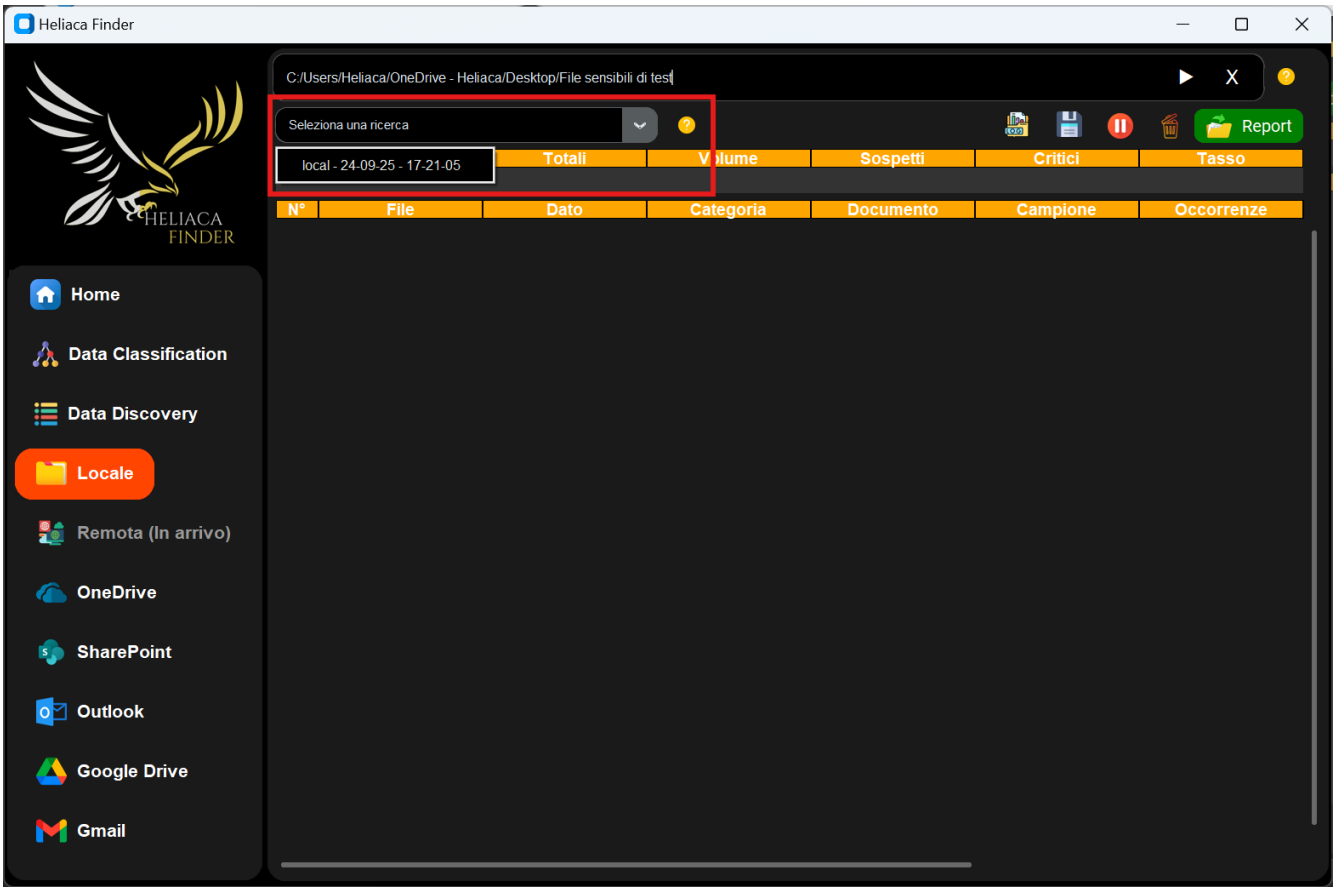


Nel caso di una ricerca sul cloud come **OneDrive** troveremo due pulsanti che servono per fare il login Utente o con Enterprise Application/Service Account.


Heliaca Finder
User
Enterprise Application
Seleziona una ricerca
Report
Tempo
Totali
Volume
Sospetti
Critici
Tasso
0s
0
0.00B
0
0
0%
N°
File
Dato
Categoria
Documento
Campione
Occorrenze
HELIACA FINDER
Home
Data Classification
Data Discovery
Locale
Remota (In arrivo)
OneDrive
SharePoint
Outlook
Google Drive
Gmail

3.2. Ripresa della ricerca

Heliaca Finder salva in modo continuo le ricerche effettuate su dei file testuali **criptati**, in modo da garantire un minor consumo di memoria RAM e soprattutto per dare la possibilità di riprendere la ricerca in caso venga interrotta in modo manuale oppure a causa di un imprevisto come un'interruzione di corrente.
Si può riprendere in modo immediato una ricerca precedentemente effettuata interagendo con il widget inerente, selezionandola dal menu a tendina.

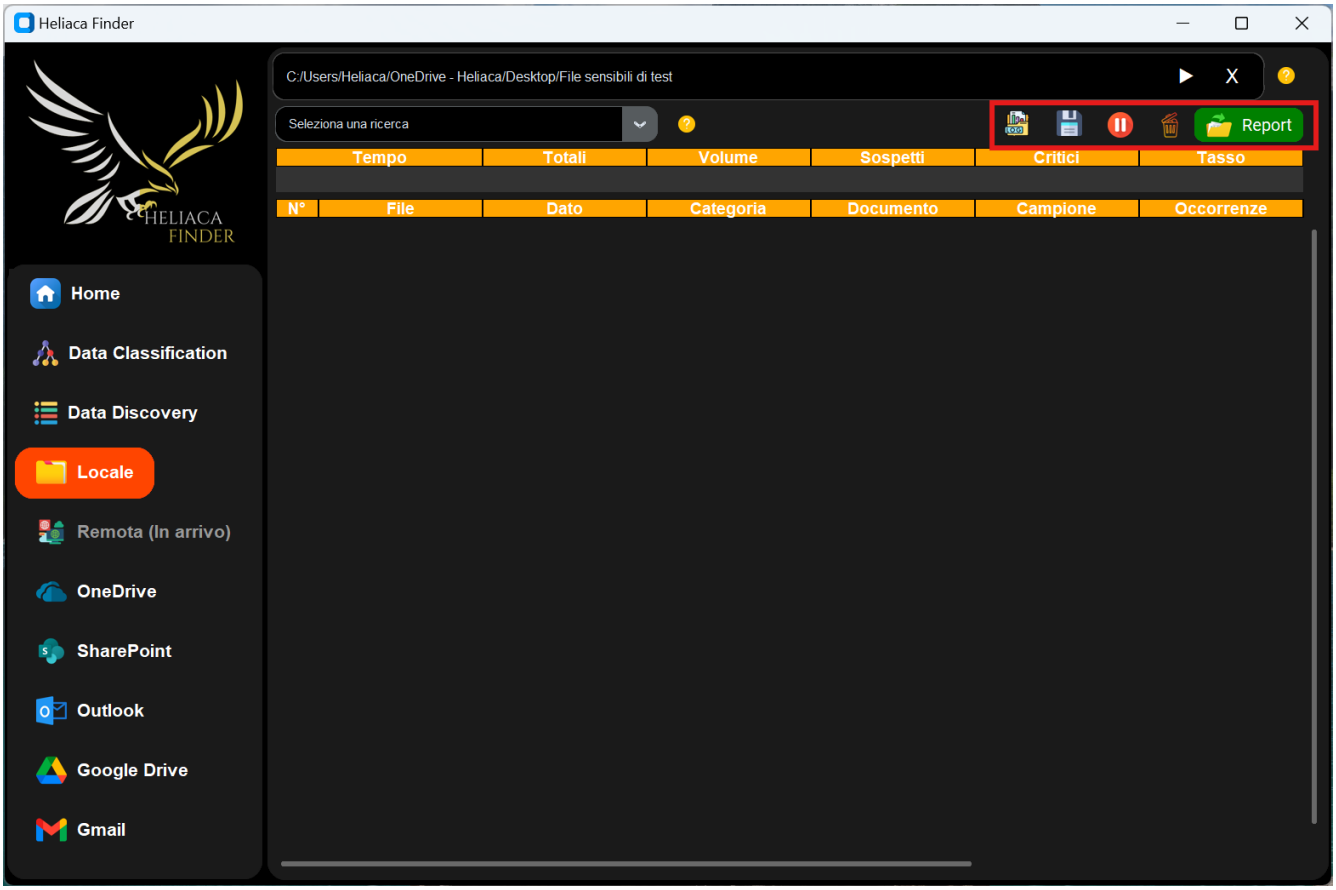


Il tempo di ripresa può **variare** a seconda della grandezza della ricerca, soprattutto nel caso di una ricerca sul cloud in quanto il processo di ripresa rielabora tutte le informazioni già analizzate (saltandole) fino ad arrivare al punto in cui si era fermata precedentemente. Come segnale di ripresa dell'analisi si può far riferimento all'incrementare del valore nel campo **Tempo**.

3.3. Pulsanti di Utilità

I pulsanti di "**Utilità**" agiscono sul controllo del processo di ricerca. Sono rispettivamente da sinistra a destra:

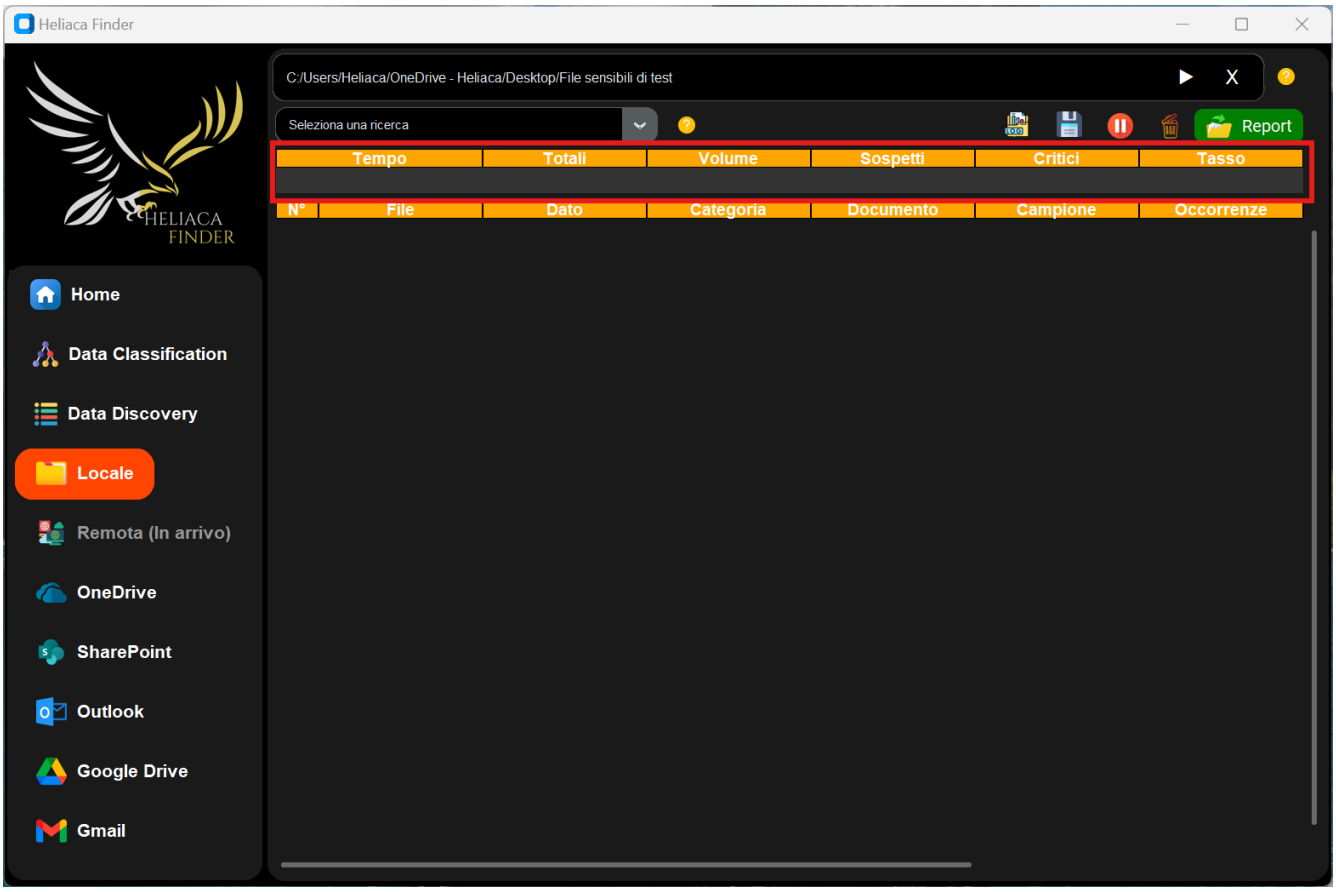
- **Log:** Apre la cartella Log dove sono tutti i log inerenti all'utilizzo di Heliaca Finder e all'esecuzione della ricerca.
- **Save:** Apre la cartella Save dove sono tutti i salvataggi delle ricerche effettuate e in caso il token per i report.
- **Pause:** Blocca l'esecuzione della ricerca.
- **Clean:** Pulisce i treeview dei paramentri, di data discovery e l'entry della ricerca dalle informazioni accumulate durante il funzionamento.
- **Report:** Estrapola i report una volta terminata la ricerca al 100%, passando dallo stato disattivo allo stato attivo (cliccabile). Il colore del pulsante cambia da rosso a verde in caso il token sia disponibile e in corso di validità. Una volta che viene premuto e quindi estrapolato un report vengono automaticamente cancellati i salvataggi inerenti alla ricerca.



3.4. Treeviews

I treeviews sono delle **tabelle** che contengono importanti informazioni riguardo la ricerca. Heliaca Finder utilizza due tipi di treeview.

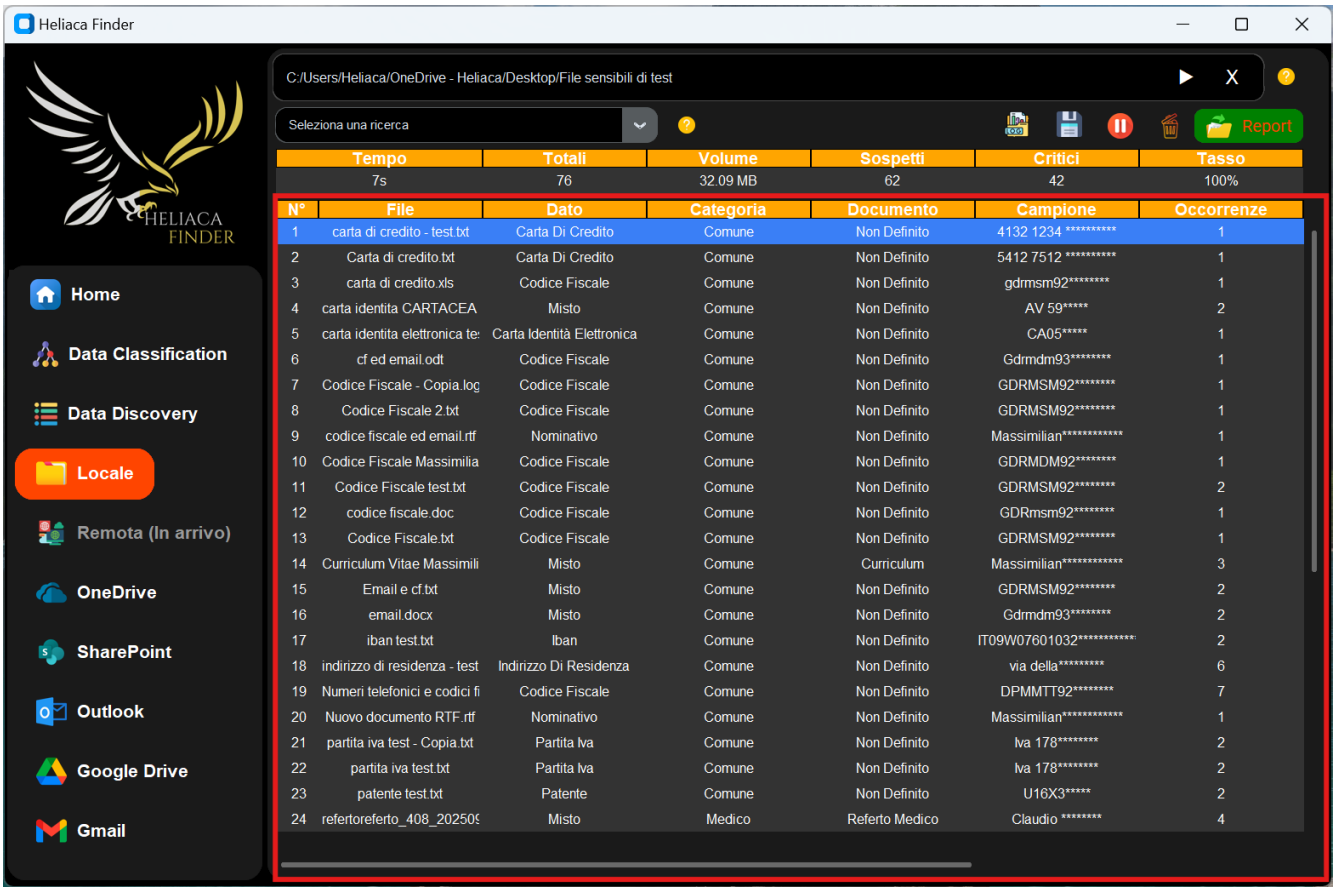
3.4.1. Treeview parametrico



Questa tabella contiene al suo interno tutti i parametri inerenti alla ricerca iniziata:

- **Tempo**: tempo trascorso dall'inizio della ricerca.
- **Totali**: quantitativo totale di file/elementi trovati all'interno della sorgente di ricerca.
- **Volume**: il volume di dati analizzato.
- **Sospetti**: il numero di file/elementi che rientrano nei criteri di ricerca, ovvero estensioni e dati comuni selezionati.
- **Critici**: il numero di file/elementi sensibili trovati all'interno della sorgente di ricerca.
- **Tasso**: la percentuale di completamento della ricerca.

3.4.2. Treeview Data Discovery



Tempo	Totali	Volume	Sospetti	Critici	Tasso
7s	76	32.09 MB	62	42	100%

N°	File	Dato	Categoria	Documento	Campione	Occorrenze
1	carta di credito - test.txt	Carta Di Credito	Comune	Non Definito	4132 1234 **********	1
2	Carta di credito.txt	Carta Di Credito	Comune	Non Definito	5412 7512 **********	1
3	carta di credito.xls	Codice Fiscale	Comune	Non Definito	gdrmsm92********	1
4	carta identita CARTACEA	Misto	Comune	Non Definito	AV 59*****	2
5	carta identita elettronica te:	Carta Identità Elettronica	Comune	Non Definito	CA05*****	1
6	cf ed email.odt	Codice Fiscale	Comune	Non Definito	Gdrmdm93********	1
7	Codice Fiscale - Copia.log	Codice Fiscale	Comune	Non Definito	GDRMSM92********	1
8	Codice Fiscale 2.txt	Codice Fiscale	Comune	Non Definito	GDRMSM92********	1
9	codice fiscale ed email.rtf	Nominativo	Comune	Non Definito	Massimilian**********	1
10	Codice Fiscale Massimilia	Codice Fiscale	Comune	Non Definito	GDRMDM92********	1
11	Codice Fiscale test.txt	Codice Fiscale	Comune	Non Definito	GDRMSM92********	2
12	codice fiscale.doc	Codice Fiscale	Comune	Non Definito	GDRmsm92********	1
13	Codice Fiscale.txt	Codice Fiscale	Comune	Non Definito	GDRMSM92********	1
14	Curriculum Vitae Massimili	Misto	Comune	Curriculum	Massimilian**********	3
15	Email e cf.txt	Misto	Comune	Non Definito	GDRMSM92********	2
16	email.docx	Misto	Comune	Non Definito	Gdrmdm93********	2
17	iban test.txt	Iban	Comune	Non Definito	IT09W07601032**********	2
18	indirizzo di residenza - test	Indirizzo Di Residenza	Comune	Non Definito	via della*********	6
19	Numeri telefonici e codici fi	Codice Fiscale	Comune	Non Definito	DPMMTT92********	7
20	Nuovo documento RTF.rtf	Nominativo	Comune	Non Definito	Massimilian**********	1
21	partita iva test - Copia.txt	Partita Iva	Comune	Non Definito	Iva 178********	2
22	partita iva test.txt	Partita Iva	Comune	Non Definito	Iva 178********	2
23	patente test.txt	Patente	Comune	Non Definito	U16X3*****	2
24	refertoreferto_408_202509	Misto	Medico	Referto Medico	Claudio ********	4

Questa tabella contiene tutte le informazioni inerenti ai file/elementi sensibili trovati distribuiti su più colonne, è possibile consultarla nella sua interezza scorrendo sia verticalmente che orizzontalmente utilizzando le apposite scrollbar.

Questo treeview è **interattivo**, ovvero cliccando su specifiche colonne è possibile ottenere funzionalità aggiuntive interessanti, come:

- Apertura del file (Click colonna **File**)
- Consultazione delle quantità relative alle tipologie di dato comune (Click colonna **Dato**)
- Consultazione delle quantità relative alle categorie di dato particolare (Click colonna **Categoria**)
- Consultazione delle occorrenze di dati sensibili contenute per file/elemento sensibile (Click colonna **Occorrenze**)

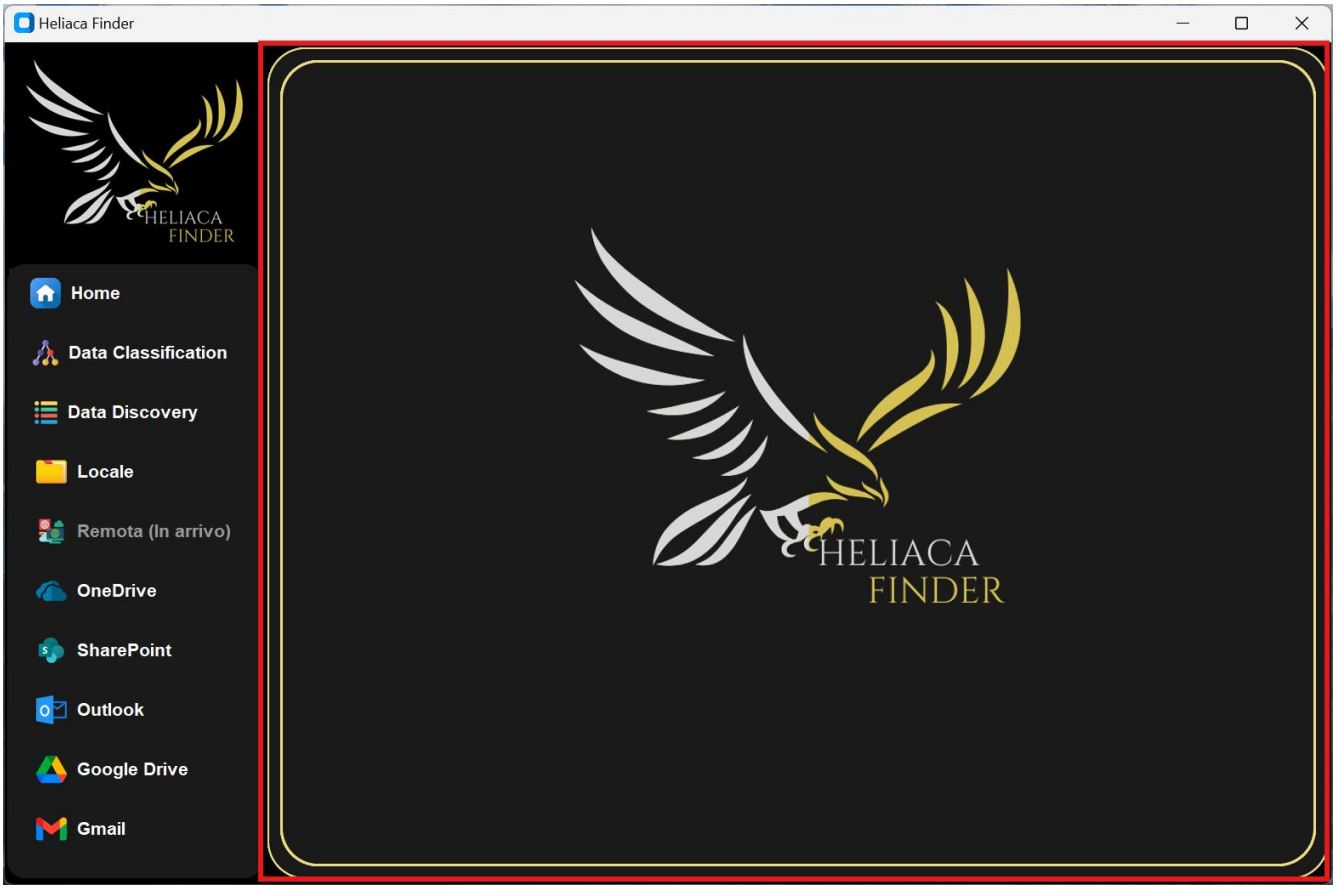
Con il click destro del mouse è possibile **eliminare** il file/elemento sensibile desiderato in caso fosse un falso positivo.

4. Frame

I frame sono delle “finestre” dove vengono creati e distribuiti i widget rientranti in una categoria.
Heliaca Finder utilizza un frame per ogni tipologia di ricerca e tre per contenere rispettivamente la Home, i settings della Data Discovery e la Data Classification.

4.1. Home

La **Home** è la prima finestra che viene mostrata all’apertura di Heliaca Finder. Contiene il logo del software e cliccando su di esso è possibile raggiungere la pagina dedicata.



4.2. Data classification

Il frame inerente alla Data Classification contiene al suo interno due tabelle necessarie per ultimare la fase di configurazione per la classificazione. La prima tabella ha la funzione di descrivere i parametri fondamentali per la classificazione dei file, mentre la seconda permette di incrementare/decrementare il valore attribuito.

4.2.1. Tabella descrittiva

La tabella riportata in apertura è finalizzata a descrivere il valore e l'impatto derivanti da un'eventuale compromissione delle diverse tipologie di dati sensibili in relazione all'utilizzatore o al cliente.
I parametri di riferimento adottati sono **Valore, Riservatezza, Integrità e Disponibilità**.
Per ciascun parametro sono definiti livelli graduati di impatto **Nullo, Basso, Medio e Alto** ai quali è attribuito un corrispondente punteggio numerico compreso tra **1 e 4**, utile a supportare una valutazione oggettiva e comparabile.

	Nullo - 1	Basso - 2	Medio - 3	Alto - 4
Valore	Nessun soggetto interno o esterno avrebbe interesse ad ottenere queste informazioni.	Alcuni soggetti interni o esterni potrebbero essere interessati a queste informazioni, ottenendo un vantaggio limitato dalla loro acquisizione.	Alcuni soggetti interni o esterni sono interessati a queste informazioni, la cui acquisizione potrebbe offrire un vantaggio significativo.	Alcuni soggetti interni o esterni sono interessati a queste informazioni, la cui acquisizione potrebbe offrire un vantaggio molto significativo.

	Valore	Riservatezza	Integrità	Disponibilità
Nominativo	1	1	1	1
Carta di credito	1	1	1	1
Iban	1	1	1	1
Partita iva	1	1	1	1
Codice fiscale	1	1	1	1
Tessera sanitaria	1	1	1	1
Email	1	1	1	1
Carta identità elettronica	1	1	1	1
Carta identità cartacea	1	1	1	1
Passaporto	1	1	1	1
Patente	1	1	1	1
Targa	1	1	1	1
Indirizzo di residenza	1	1	1	1
Cellulare	1	1	1	1
Fisso	1	1	1	1
Biometrico	1	1	1	1

4.2.2. Tabella operativa

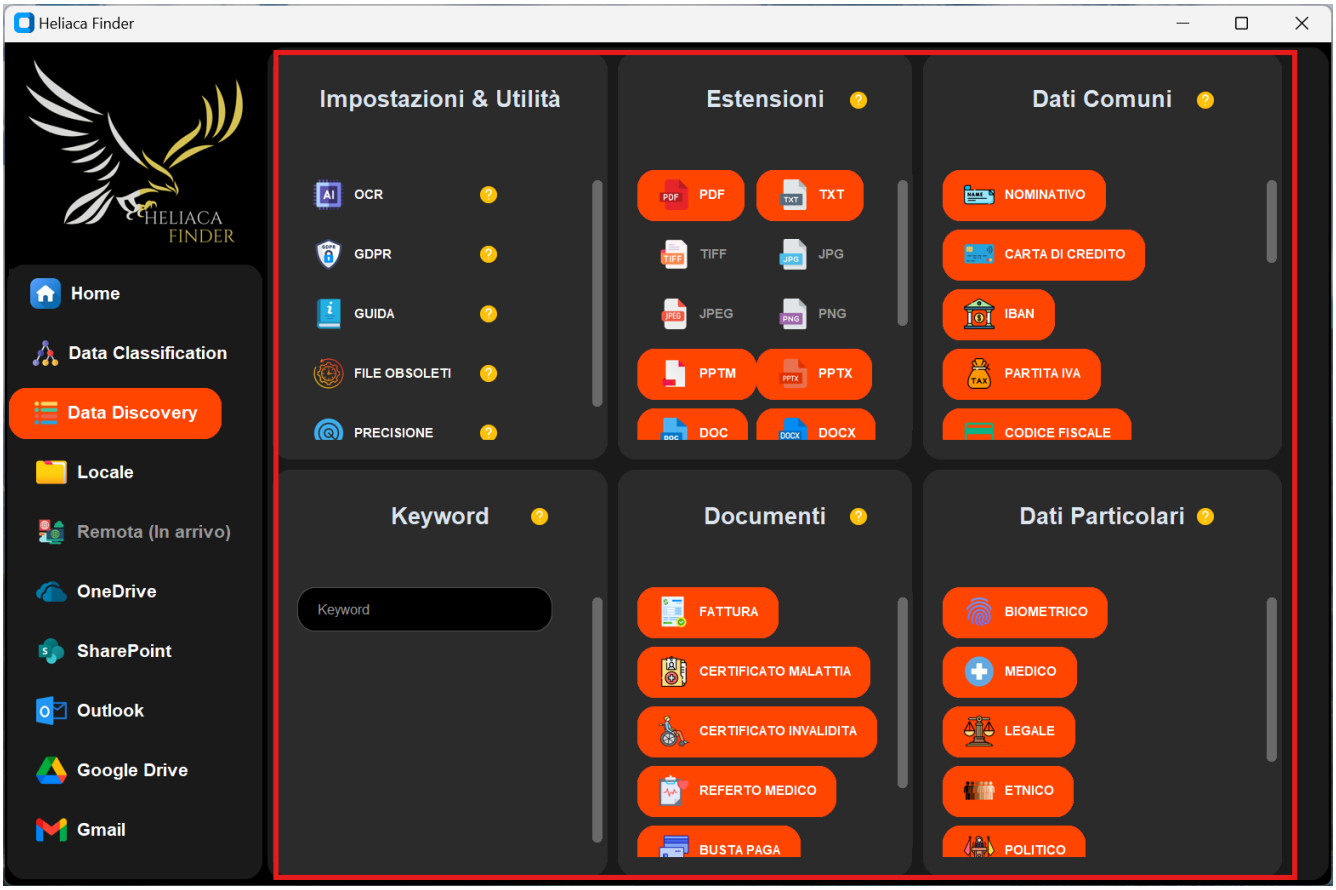
La seconda tabella è interattiva e fornisce la possibilità all'utilizzatore di modificare i parametri descritti precedentemente; cliccando sulla rispettiva cella il valore numerico cambia da 1 a 4 a modificare appunto il valore o l'impatto che quel dato sensibile può avere. Se invece si interagisce con l'intestazione delle colonne della tabella si può modificare il parametro descritto nella tabella descrittiva.

Esempio:
Si vuole modificare il valore e l'impatto che la compromissione dei codici fiscali dei clienti può avere sulla azienda X. Il primo passo è selezionare il parametro **Valore** e farsi guidare all'attribuzione dell'opportuno valore consultando la tabella descrittiva. Una volta individuato il giusto livello di valore si modifica il numero presente nella cella corrispondente alla colonna **Valore** e dato sensibile **Codice Fiscale**.

Il processo di classificazione nonostante possa risultare macchinoso, contribuisce in modo importante a migliorare la consapevolezza del parco di dati sensibili posseduti e a prendere provvedimenti tecnici nei loro confronti in modo più rigoroso e specifico.

4.3. Data discovery

Il frame Data Discovery contiene tutti i pulsanti/widget utili alla configurazione della ricerca, distribuiti in diverse finestre più piccole scorribili verticalmente, chiamate **Scrollable Frame**.



4.3.1. Impostazioni & Utilità

Il primo Scrollable Frame che incontriamo in alto a sinistra è quello inerente alle **Impostazioni & Utilità**, contiene al suo interno cinque pulsanti:

- **OCR**: Attiva la funzionalità di riconoscimento ottico dei caratteri; in questo modo è possibile estrarre il testo presente in file di tipo immagine e nei pdf in cui non si è riuscito ad estrarre il testo con la libreria standard. In aggiunta cliccando su di esso vengono attivati e selezionati automaticamente le estensioni dei file immagine, un altro click li disattiva e deseleziona.
- **GDPR**: Raggiunge il sito ufficiale dove è descritto il regolamento europeo sulla privacy (GDPR) nella sua versione più aggiornata.
- **GUIDA**: Scarica e apre la guida ufficiale di Heliaca Finder (questa guida).
- **FILE OBSOLETI**: Permette di impostare una data per filtrare i file durante la ricerca **locale**. I file che risultano modificati prima della data specificata nell'entry vengono inclusi nella ricerca.
- **PRECISIONE**: Permette di impostare un valore numerico nell'apposita entry, il quale va a modificare la precisione con cui i dati particolari vengono classificati. Un dato particolare utilizza per essere classificato un dizionario di termini, se il quantitativo di termini trovati in un file o elemento è maggiore o uguale del valore indicato il dato particolare verrà classificato correttamente. Minore è il valore di precisione indicato maggiore saranno i falsi positivi, maggiore il valore e minore saranno i falsi positivi. Il valore di default è 3, mentre l'intervallo di valori consigliati è tra 1 e 10.

4.3.2. Keyword

Interagendo con l'apposita entry contenuta nello Scrollable Frame **Keyword**, è possibile inserire parole chiavi che verranno cercate nel testo dei file. È una funzionalità fondamentale se si vuole trovare parole chiavi utilizzate solitamente per individuare file o documenti critici. Ad esempio, consideriamo il caso in cui una azienda abbia classificato i propri documenti scrivendovi all'interno la parola "Riservato", inserendo questa parola come keyword verranno trovati tutti i documenti contenenti questa parola.
Per aggiungere una keyword è sufficiente scriverla nella entry e successivamente premere Invio, per cancellarla utilizzare la X presente alla destra della keyword.
La keyword verrà inserita anche nella tabella operativa del frame Data Classification, in modo da dare la possibilità di attribuirgli un valore e impatto personalizzato.

4.3.3. Estensioni

Nello Scrollable Frame relativo alle **Estensioni** è possibile selezionare o deselezionare tutte le estensioni dei file che si vogliono includere nel processo di ricerca. Le estensioni dei file immagini sono disabilitate di default, ricorrere al pulsante OCR per attivarle.

4.3.4. Documenti

Nello Scrollable Frame relativo ai **Documenti** è possibile selezionare o deselezionare tutte le tipologie di documenti che vogliamo classificare durante il processo di ricerca. Se non venisse selezionato nessuna tipologia di documento i file verrebbero trovati comunque ma classificati come documenti **Non Definiti.**

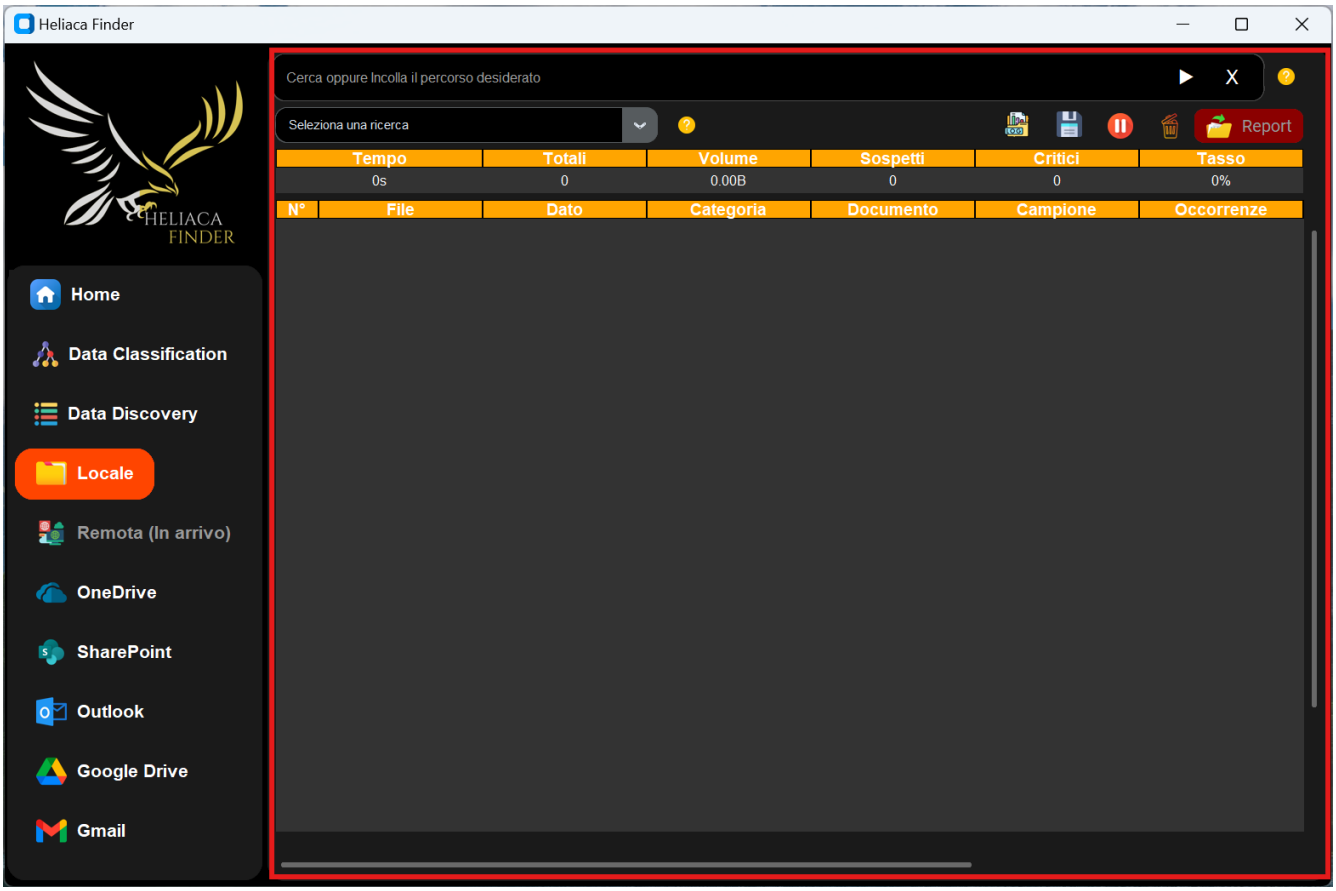
4.3.5. Dati comuni

Nello Scrollable Frame relativo ai **Dati Comuni** è possibile selezionare o deselezionare tutte le tipologie di dati comuni che vogliamo includere durante il processo di ricerca.
I dati comuni fanno riferimento alla tipologia base dei **dati personali**, i quali secondo il GDPR, possono identificare in modo diretto o indiretto una persona fisica.

4.3.6. Dati particolari

Nello Scrollable Frame relativo ai **Dati Particolari** è possibile selezionare o deselezionare tutte le tipologie di dati particolari che vogliamo classificare durante il processo di ricerca. Se non venisse selezionato nessuna tipologia di dato i file verrebbero trovati comunque ma classificati come dati particolari **Comuni.**
Le **categorie particolari di dati personali**, secondo il GDPR, sono una tipologia particolare di dati sensibili che vanno attenzionati maggiormente in merito al loro valore legale. Essi non possono essere trattati senza lo specifico consenso della persona a cui sono riferiti e in caso di violazione si può andare in contro a delle sanzioni amministrative la cui gravità e direttamente collegata all'entità della compromissione di questi dati.

4.4. Locale



Il frame inerente alla Data Discovery **Locale** contiene tutti i widget necessari per il controllo e la visualizzazione di una ricerca effettuata su un percorso locale.

4.4.1. Barra di ricerca

La barra di ricerca è una entry (un widget in cui è possibile inserire del testo), facendo **doppio click** su di essa o premendo il tasto **Play** è possibile richiamare il wizard per la selezione di un percorso locale. È altresì possibile incollare direttamente un percorso e successivamente premere Play; possono essere utilizzati anche i percorsi di rete a condizione che non siano protetti da password.
Per cancellare un percorso basta premere sulla **X**.

4.4.2. Esecuzione ricerca locale

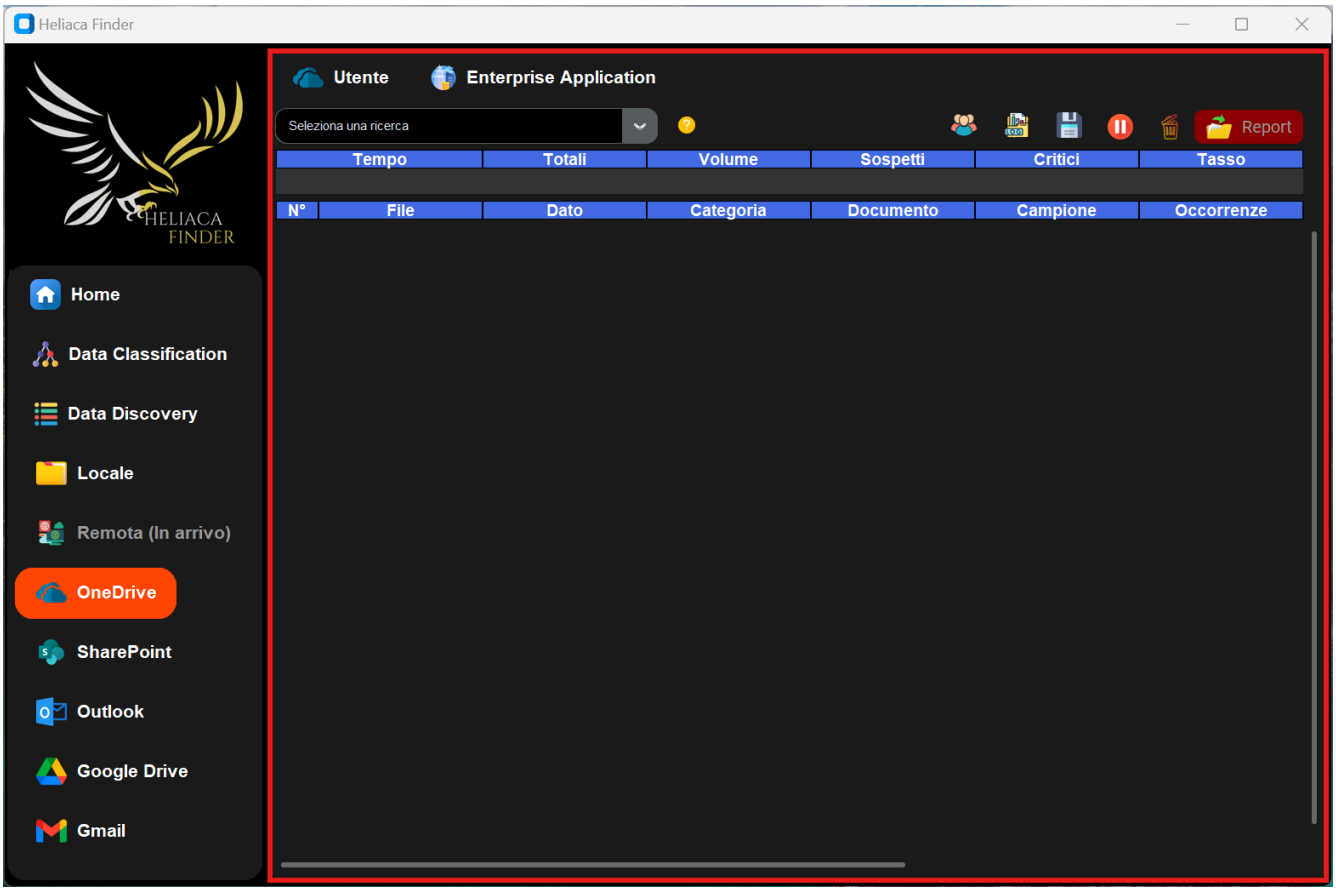
Una volta inserito il percorso (se valido), il processo di ricerca partirà immediatamente.
Inizialmente verrà fatto un conteggio totale dei file presenti nel percorso (step necessario per calcolare la percentuale di completamento) il quale impiegherà un ammontare di tempo in proporzione al quantitativo totale di file presenti. Terminato il conteggio inizia il vero e proprio processo di ricerca, dove tutti i file presenti nel percorso vengono dapprima filtrati. Se l'estensione del file e la sua data di modifica rientrano nei requisiti impostati nei settings allora il file verrà analizzato procedendo con l'estrazione del testo e la verifica della presenza di informazioni sensibili.
Tutto il processo è ottimizzato per non occupare memoria RAM in quanto tutte le informazioni vengono salvate e criptate su file json, questo anche per garantire la successiva ripresa dell'analisi.

Alcuni percorsi sono esclusi di default, come il percorso C://Windows oppure C://Programs per non rallentare significativamente la ricerca.

4.4.3. Remota

(In sviluppo)

4.5. OneDrive



Il frame inerente alla Data Discovery **OneDrive** contiene tutti i widget necessari per il controllo e la visualizzazione di una ricerca effettuata sul percorso remoto di OneDrive.

4.5.1. Login utente

Per intraprendere una ricerca è sufficiente premere sul pulsante **Utente**; in questo modo si attiverà il processo di login che una volta inserito utente e password per il cui si vuole scansionare i file, ed eventualmente sbloccato con l'authenticator, partirà la ricerca dei dati sensibili sui file disponibili.

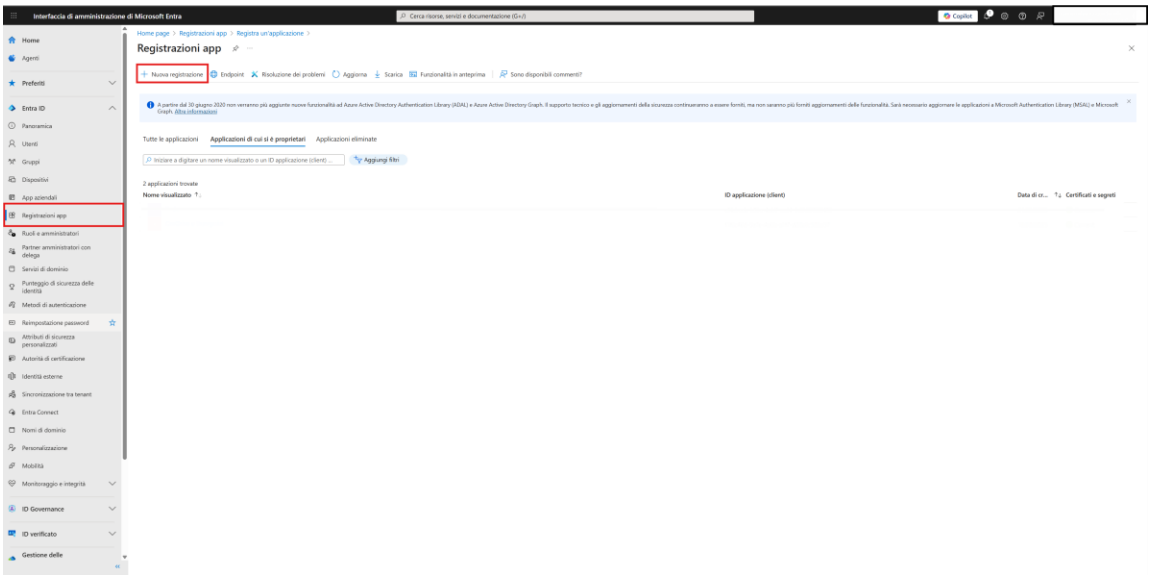
4.5.2. Login enterprise application

In caso si fosse interessati ad effettuare una ricerca su tutti gli utenti di un dominio aziendale in un colpo solo, è possibile registrare un enterprise application sul proprio dominio aziendale e successivamente utilizzare le relative credenziali per iniziare il processo di ricerca.

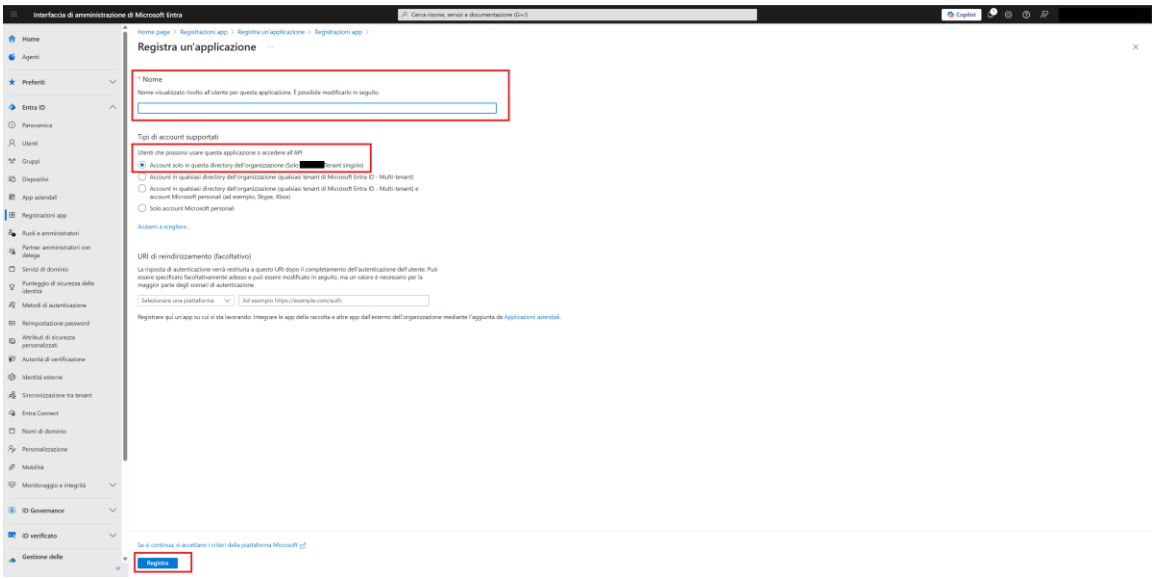
4.5.2.1. CREAZIONE DELL'ENTERPRISE APPLICATION

Per poter creare un enterprise application è necessario essere un amministratore del proprio dominio aziendale, di seguito gli step necessari per la creazione:

1. Collegarsi al sito https://entra.microsoft.com/#home
2. Effettuare l'accesso con l'account amministratore.
3. Click su **Registrazioni App** e **Nuova registrazione**.



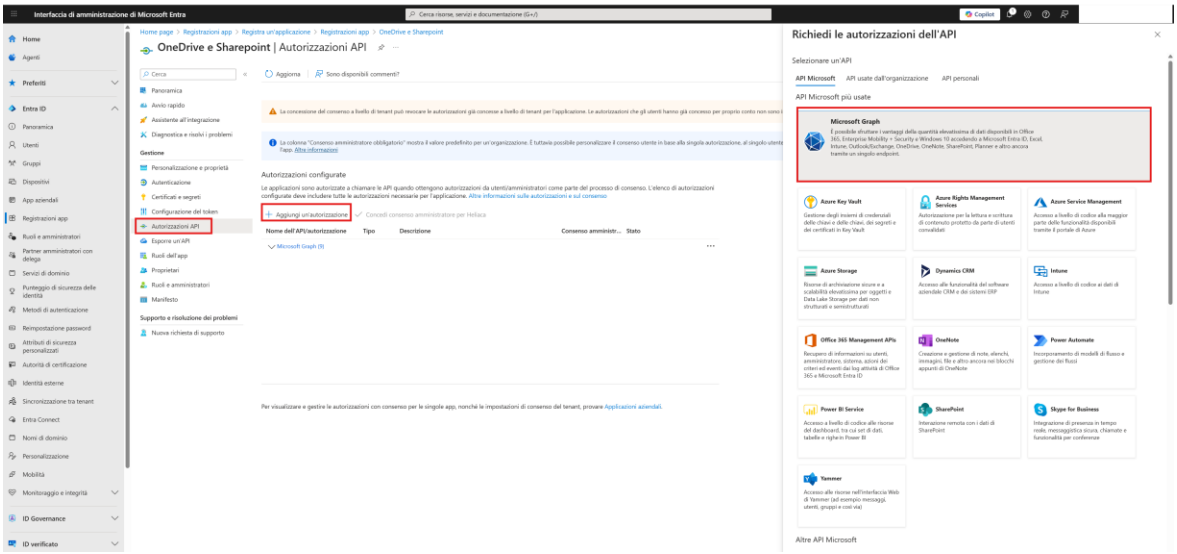
4. Registra un'applicazione inserendo il **Nome**, lasciare di default **"Account solo in questa directory dell'organizzazione (Solo xxx Tenant singolo)"** e clicca su **Registra**



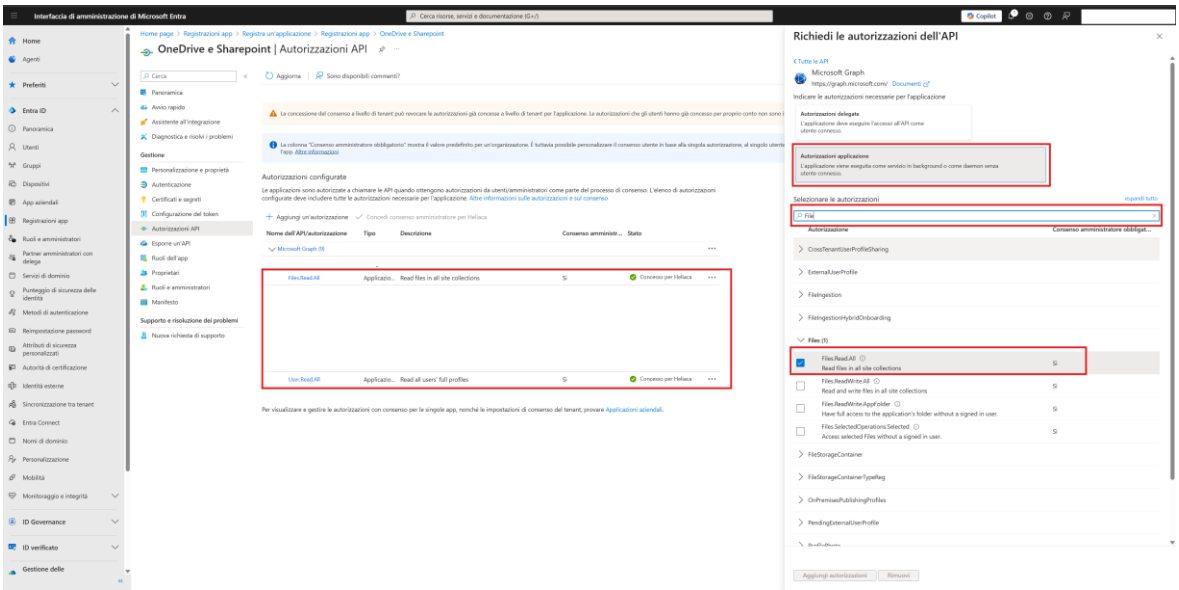
4.5.2.2. ASSEGNAZIONE DELLE AUTORIZZAZIONI

Una volta creata l'applicazione fare click su di essa per iniziare ad assegnare le **autorizzazioni** necessarie.

1. Clicca su **Autorizzazione API, Aggiungi un'autorizzazione** e **Microsoft Graph**



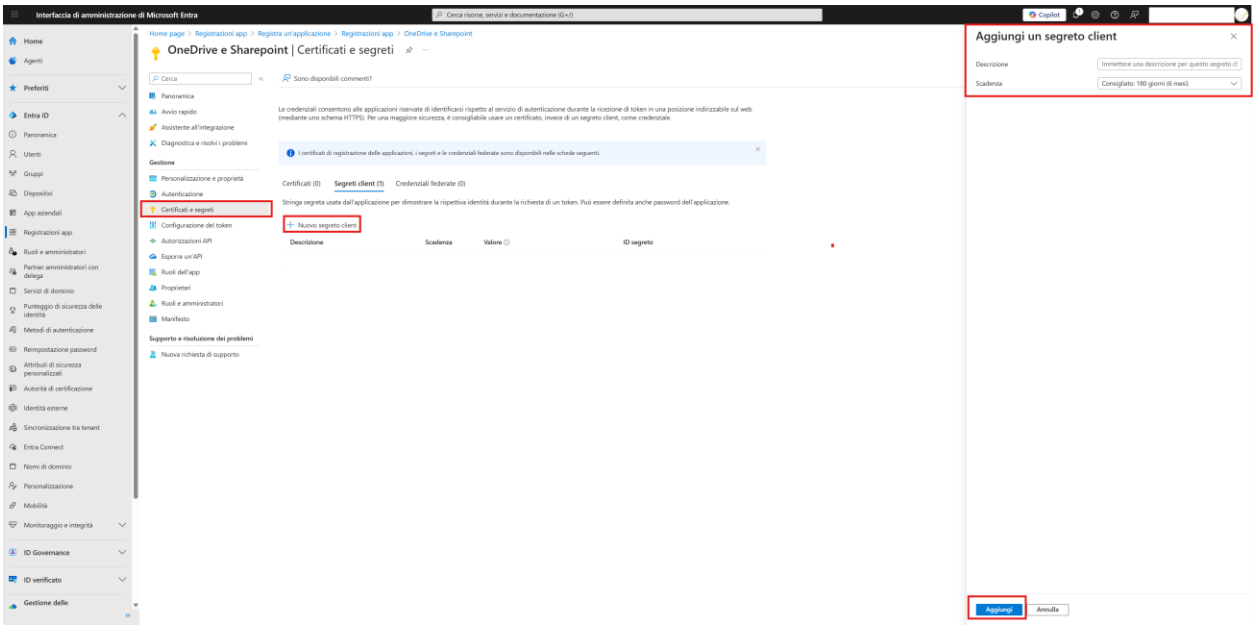
2. Cliccare su **Autorizzazioni applicazione**
3. Selezionare le autorizzazioni cercandole nella barra di ricerca e spuntarle utilizzando l'apposito checkbox, le autorizzazioni da concedere per poter accedere ai repository OneDrive di tutti gli utenti sono **Files.Read.All** e **User.Read.All**
4. Una volta concesse le autorizzazioni appariranno al centro dello schermo, dove andrà concesso il **consenso amministratore** per concludere la procedura.



4.5.2.3. INSERIMENTO DELLE CREDENZIALI PER AVVIO RICERCA

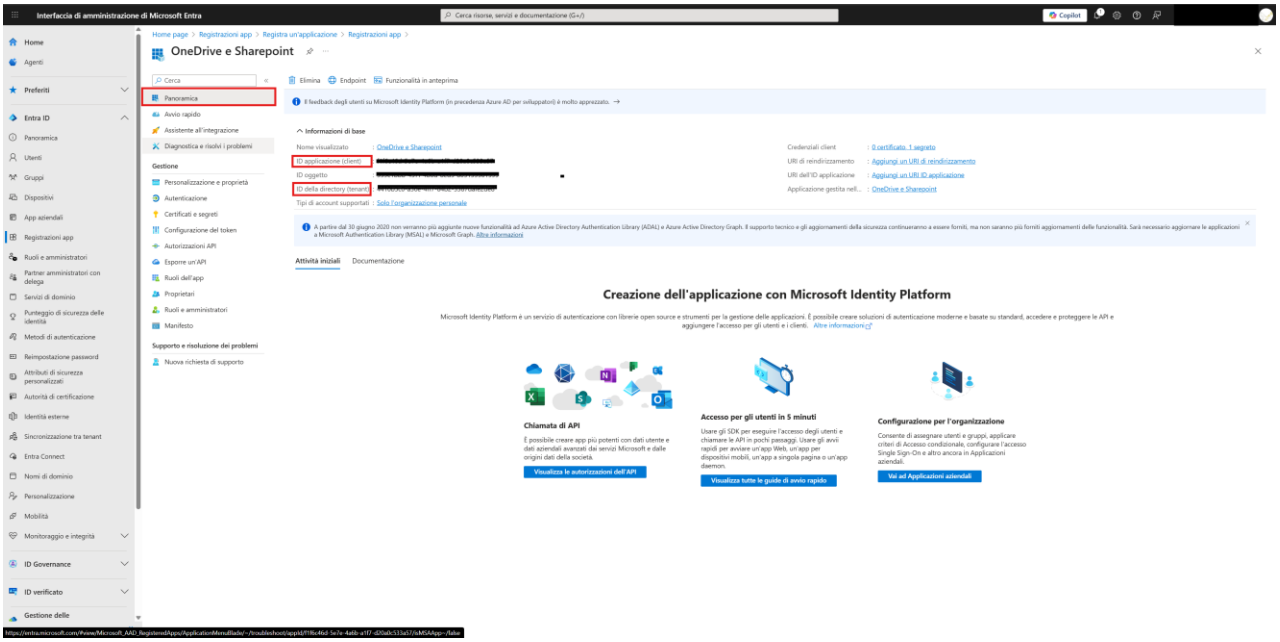
Una volta assegnate le autorizzazioni occorre creare la chiave segreta.

Cliccare su **Certificati e segreti**, **Nuovo segreto client** e assegnare una **Descrizione** e una **Scadenza** alla chiave, infine clicca su **Aggiungi**

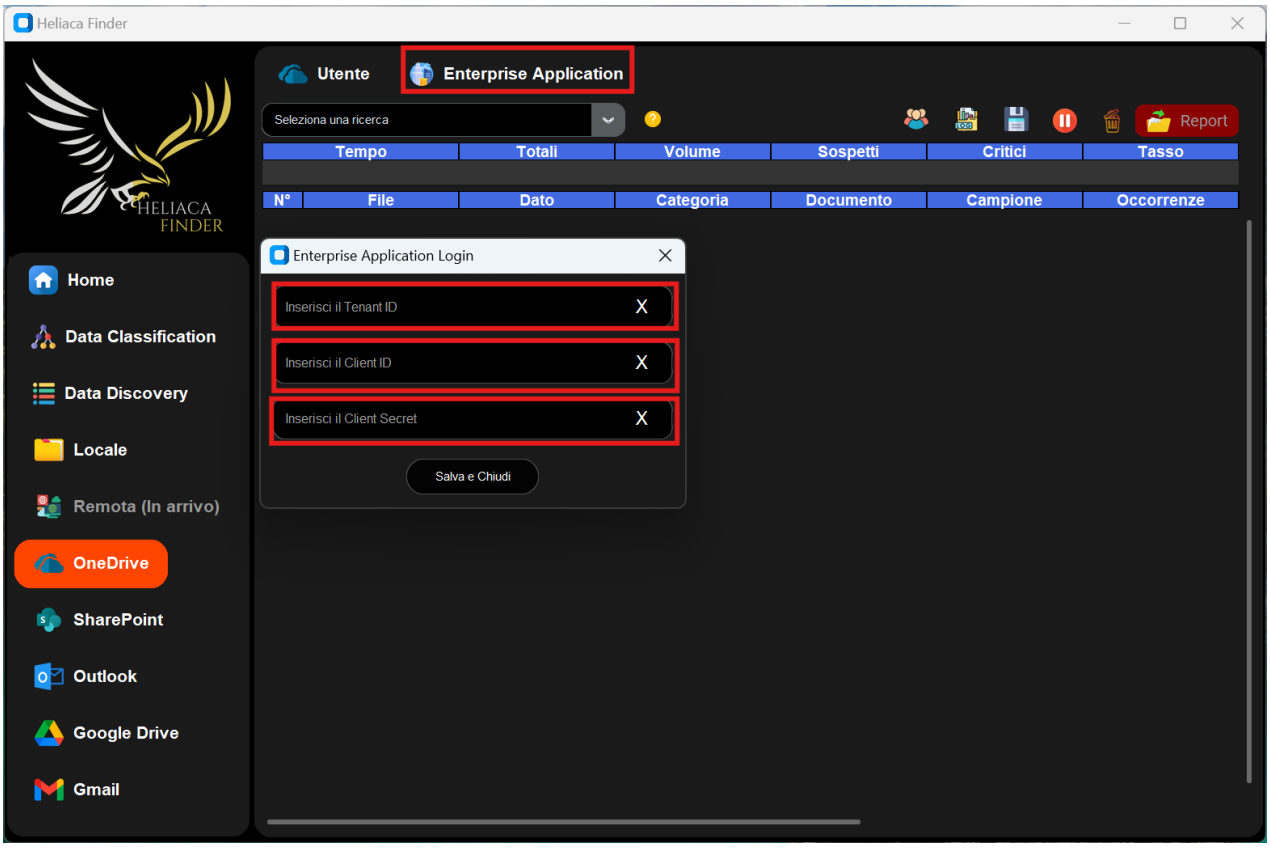


Ora appunta la chiave segreta che consiste in una stringa alfa-numerica presente sotto la colonna **Valore**, potrai copiarla solo una volta. In caso non l'avessi salvata occorre procedere ad una nuova creazione.

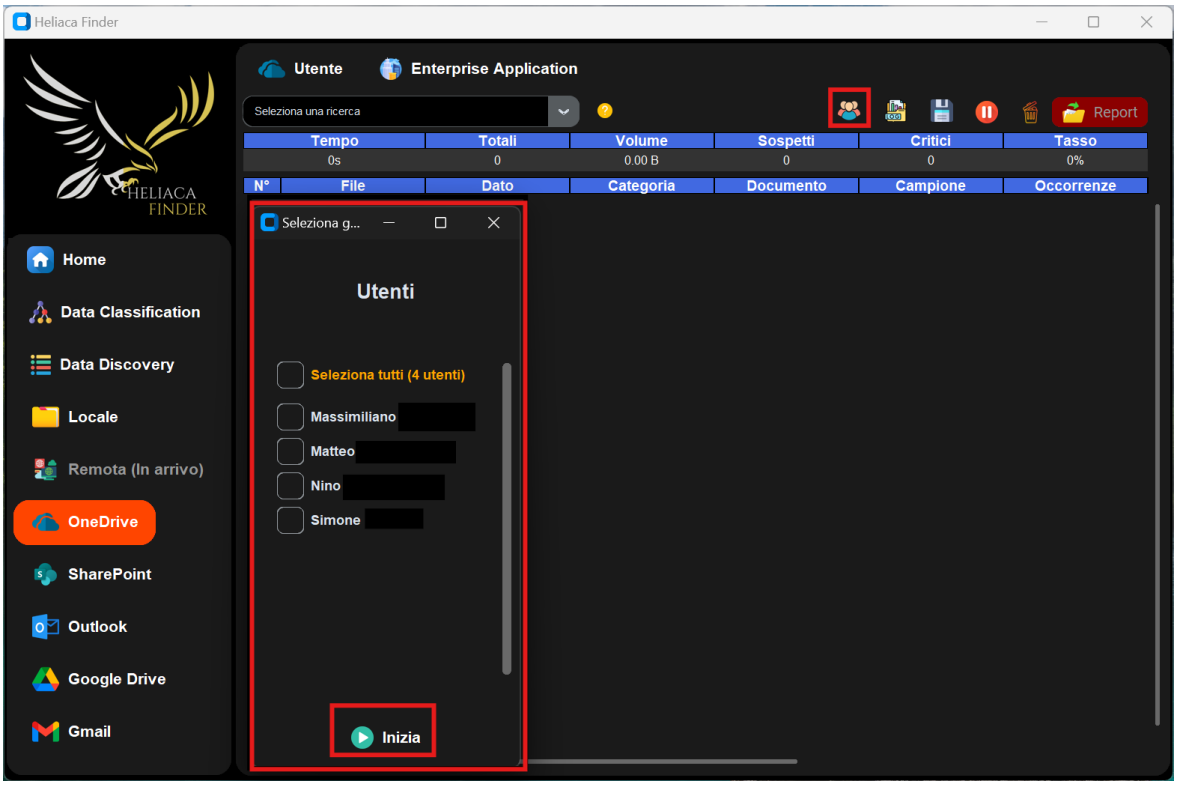
Annota anche i valori di **Tenant ID** e **Client ID** accedendo alla sezione **Panoramica**.



Ora possiamo tornare su Heliaca Finder per inserire le credenziali appena ottenute. Cliccando sul pulsante **Enterprise Application** apparità un top level, ovvero una piccola finestra, dove poter inserire le credenziali. Inserire rispettivamente il **Tenant ID**, il **Client ID** e infine il **Client Secret** che corrisponde alla chiave segreta creata poco fa.



Una volta terminato l'inserimento (è possibile cancellare una chiave cliccando sulla **x** in caso di errore) cliccare su **Salva e Chiudi.**
Ora dopo aver atteso il processo di login, il top level si chiuderà automaticamente e il pulsante relativo alla selezione degli **Utenti** diventerà VERDE ad indicare che la procedura di login è andata a buon fine.
Cliccando sul pulsante degli Utenti si aprirà un altro top level che permetterà di scegliere gli utenti da analizzare. Una volta terminata la scelta, cliccando su **Inizia** è possibile far partire la ricerca. È possibile **selezionare/deselezionare** con un solo click tutti gli utenti disponibili usando l'apposito check box.

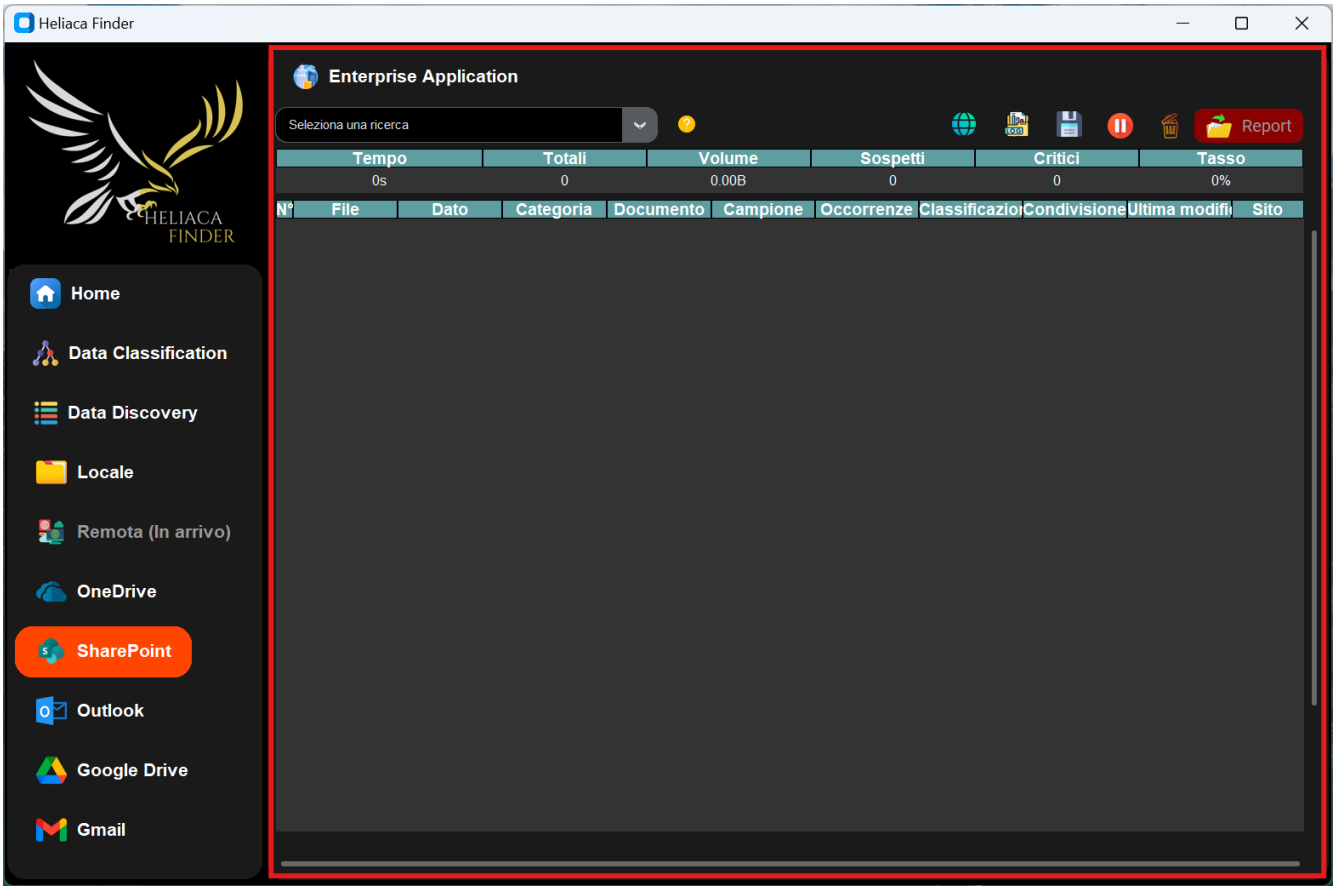


4.5.3. Esecuzione ricerca OneDrive

Una volta effettuato il login utente oppure utilizzando l'enterprise application, il processo di ricerca partirà immediatamente.
Inizialmente verrà calcolato il volume totale dei file posseduti dall'utente/utenti (step necessario per calcolare la percentuale di completamento).
Terminato il calcolo inizia il vero e proprio processo di ricerca, dove tutti i file presenti vengono dapprima filtrati. Se l'estensione del file rientra nei requisiti impostati nei settings allora il file verrà analizzato procedendo con l'estrazione del testo e la verifica della presenza di informazioni sensibili.
I file non vengono scaricati sul computer, ma viene immagazzinato momentaneamente sulla RAM il loro contenuto sotto forma di stringa BytesIO.

4.6. SHAREPOINT

Il frame inerente alla Data Discovery **Sharepoint** contiene tutti i widget necessari per il controllo e la visualizzazione di una ricerca effettuata sul percorso remoto di Sharepoint.



4.6.1. LOGIN ENTERPRISE APPLICATION

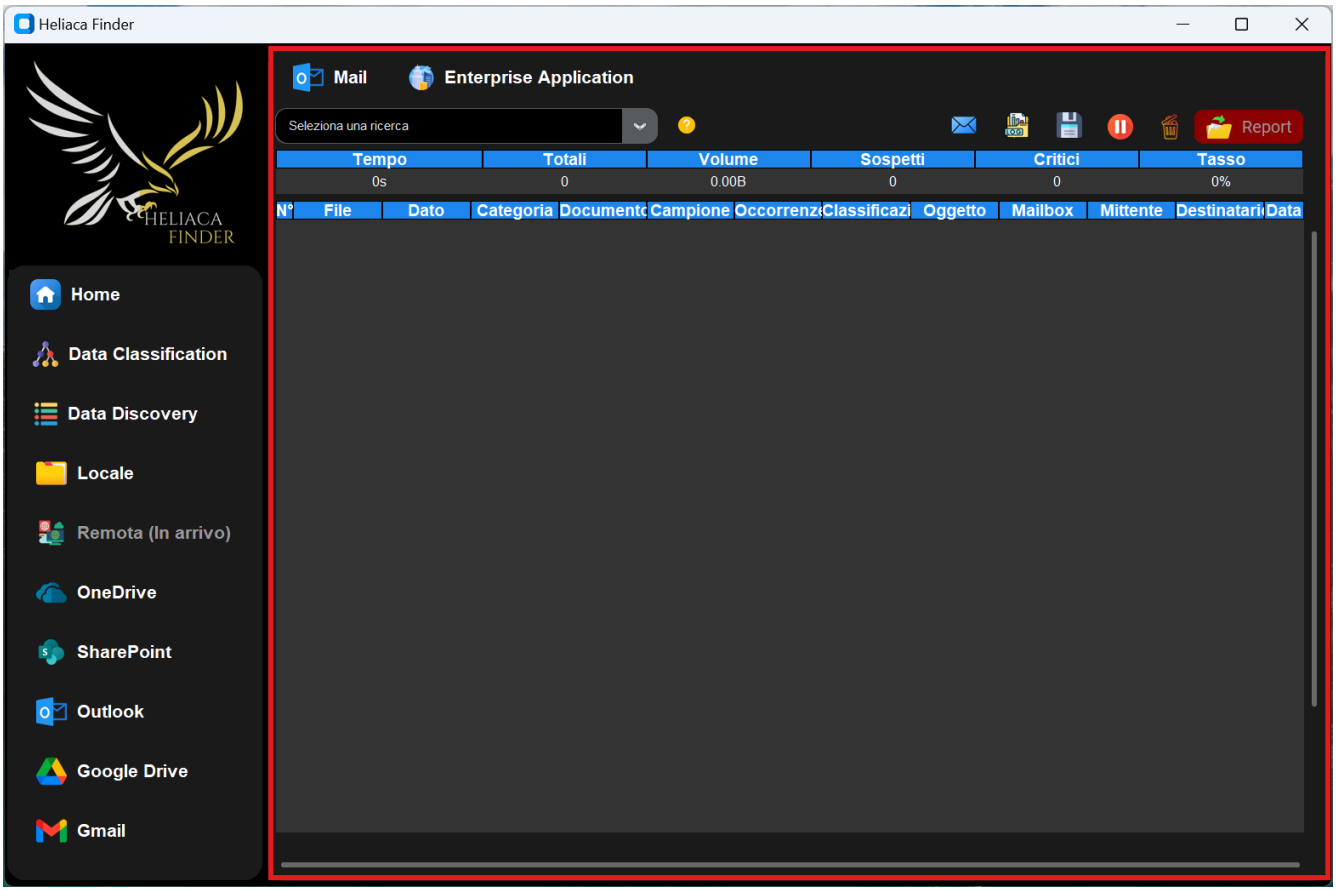
In caso si fosse interessati ad effettuare una ricerca su tutti i siti di un dominio aziendale in un colpo solo, è possibile registrare un enterprise application sul proprio dominio aziendale e successivamente utilizzare le relative credenziali per iniziare il processo di ricerca.
Per la configurazione di un enterprise application fare riferimento al paragrafo 4.5.2 dove è descritta l'intera procedura per effettuare il login per OneDrive.
L'unica differenza sarà costituita dalle **autorizzazioni** da assegnare all'enterprise application, nel caso di Sharepoint le autorizzazioni applicazioni da assegnare sono **Sites.Read.All, Sites.Selected** e **Files.Read.All.**

4.6.2. ESECUZIONE RICERCA SHAREPOINT

Una volta effettuato il login utilizzando l'enterprise application, il processo di ricerca partirà immediatamente.
Inizialmente verrà calcolato il volume totale dei file immagazzinati nei siti (step necessario per calcolare la percentuale di completamento).
Terminato il calcolo inizia il vero e proprio processo di ricerca, dove tutti i file presenti vengono dapprima filtrati. Se l'estensione del file rientra nei requisiti impostati nei settings allora il file verrà analizzato procedendo con l'estrazione del testo e la verifica della presenza di informazioni sensibili.
I file non vengono scaricati sul computer, ma viene immagazzinato momentaneamente sulla RAM il loro contenuto sotto forma di stringa BytesIO.

4.7. OUTLOOK

Il frame inerente alla Data Discovery **Outlook** contiene tutti i widget necessari per il controllo e la visualizzazione di una ricerca effettuata sul percorso remoto di Outlook.



4.7.1. Login mail

Per intraprendere una ricerca è sufficiente premere sul pulsante **Mail**; in questo modo si attiverà il processo di login che una volta inserito utente e password per il quale si vuole scansionare le e-mail e gli allegati, ed eventualmente sbloccato con l'authenticator, partirà la ricerca dei dati sensibili sugli elementi disponibili.

4.7.2. Login enterprise application

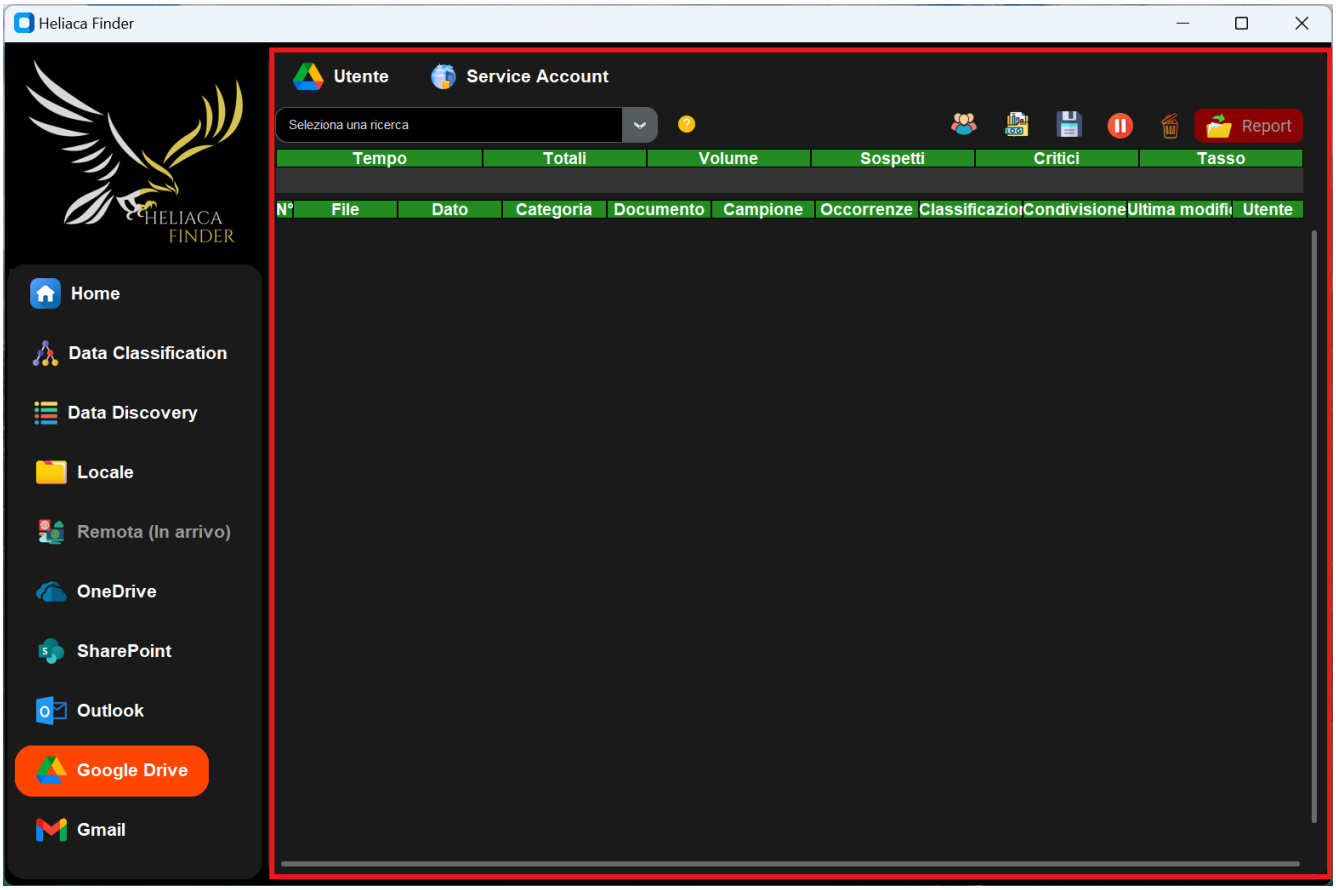
In caso si fosse interessati ad effettuare una ricerca su tutte le caselle postali di un dominio aziendale in un colpo solo, è possibile registrare un enterprise application sul proprio dominio aziendale e successivamente utilizzare le relative credenziali per iniziare il processo di ricerca.
Per la configurazione di un enterprise application fare riferimento al paragrafo 4.5.2 dove è descritta l'intera procedura per effettuare il login per OneDrive.
L'unica differenza sarà costituita dalle **autorizzazioni** da assegnare all'enterprise application, nel caso di Outlook le autorizzazioni applicazioni da assegnare sono **Mail.Read, Mail.ReadBasic.All** e **Files.Read.All.**

4.7.3. Esecuzione ricerca Outlook

Una volta effettuato il login, il processo di ricerca partirà immediatamente.
Inizialmente verrà calcolato il numero totale delle e-mail (step necessario per calcolare la percentuale di completamento). Terminato il calcolo inizia il vero e proprio processo di ricerca, dove viene estratto il contenuto testuale delle e-mail (titolo + oggetto + corpo). In caso l'e-mail contenga informazioni sensibili sarà considerata come elemento sensibile e mostrata nel treeview di data discovery. Si passa poi a verificare la presenza degli allegati nell'e-mail e in caso ne fosse provvista, l'allegato prima di essere scaricato viene filtrato in base alla sua estensione. Infine viene estratto il testo ed analizzato.
Né le e-mail né tantomeno gli allegati vengono scaricati sul computer, tutto viene immagazzinato momentaneamente sulla RAM sotto forma di stringa BytesIO.

4.8. Google Drive

Il frame inerente alla Data Discovery **Google Drive** contiene tutti i widget necessari per il controllo e la visualizzazione di una ricerca effettuata sul percorso remoto di Google Drive.



4.8.1. Login utente

Per intraprendere una ricerca è sufficiente premere sul pulsante **Utente**; in questo modo si attiverà il processo di login che una volta inserito utente e password per il cui si vuole scansionare i file, ed eventualmente sbloccato con l'authenticator, partirà la ricerca dei dati sensibili sui file disponibili.

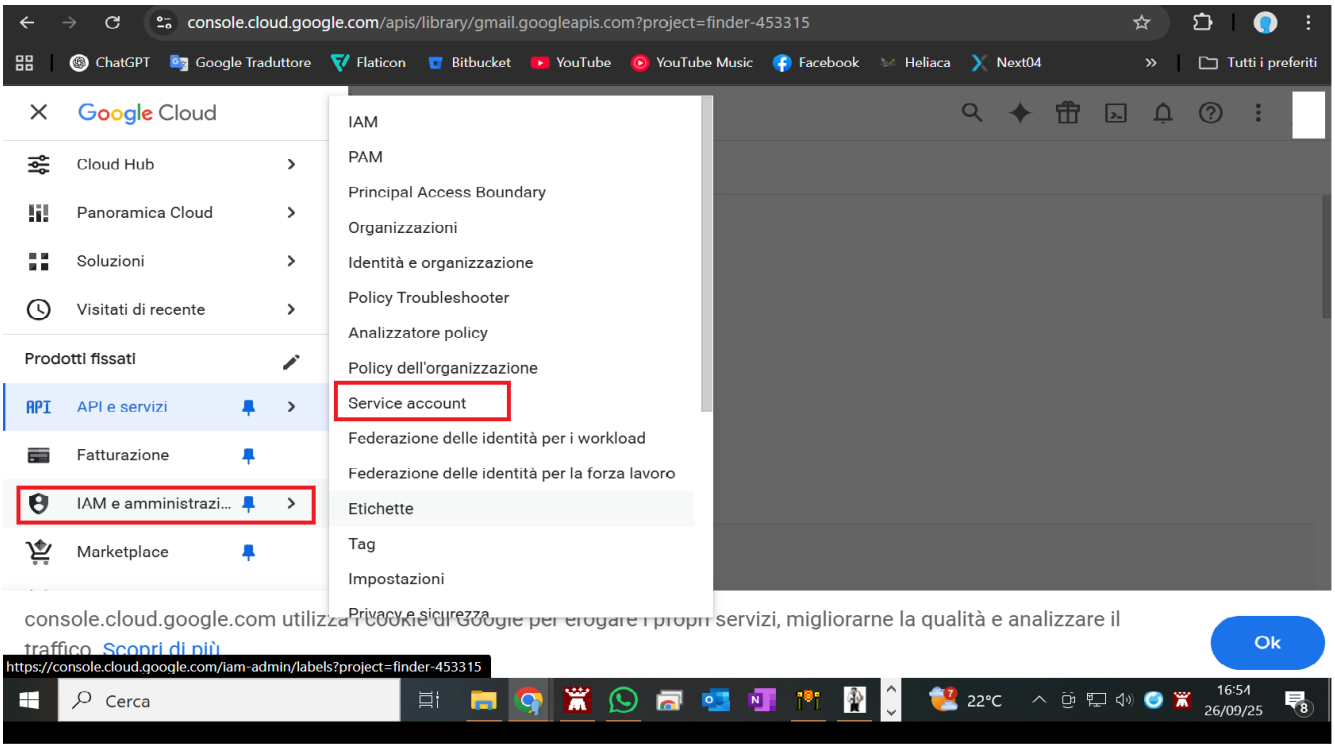
4.8.2. Login service account

In caso si fosse interessati ad effettuare una ricerca su tutti gli utenti di un dominio aziendale in un colpo solo, è possibile registrare un **Service Account** sul proprio dominio aziendale e successivamente utilizzare un profilo super amministratore e un file json con le credenziali per iniziare il processo di ricerca.

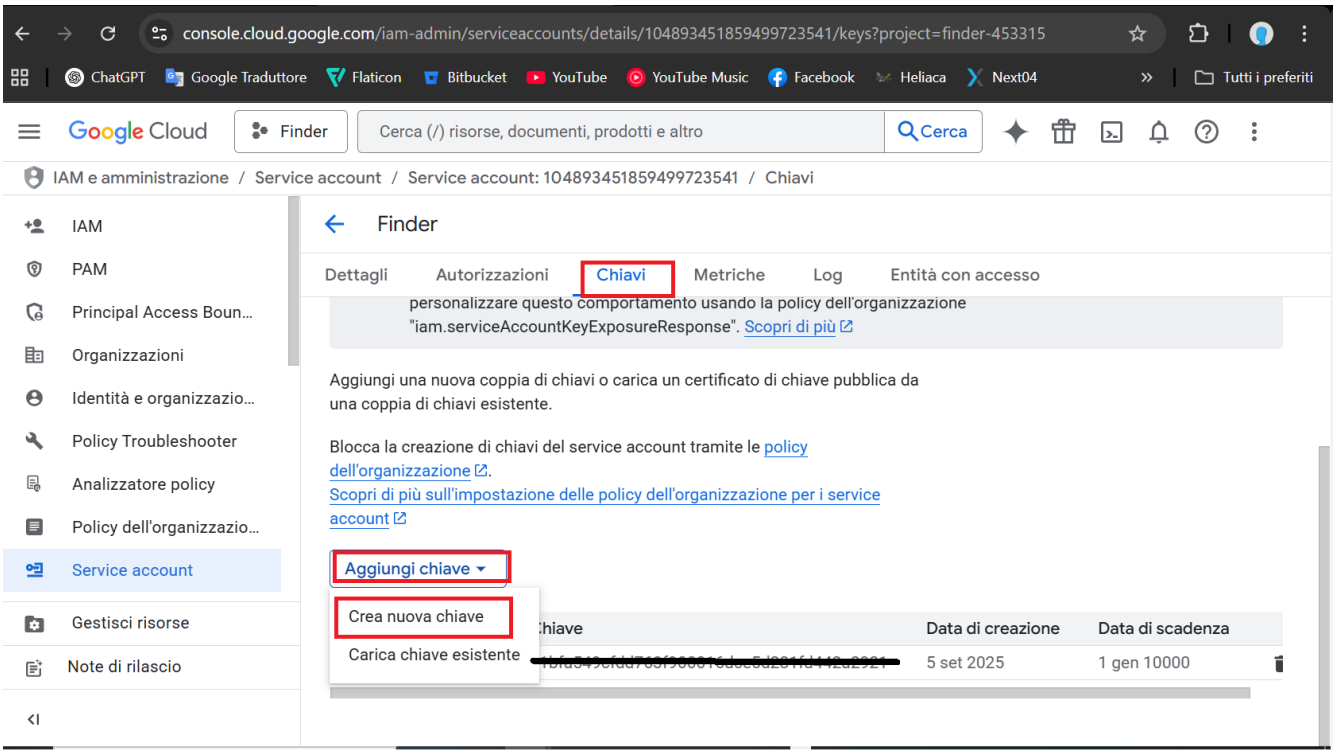
4.8.2.1. Creazione del Service Account

Per creare un Service Account sul proprio dominio aziendale Google, seguire i seguenti passaggi:

1. Collegarsi al sito: https://console.cloud.google.com
2. Clicca in alto a sinistra sulle tre linee.
3. Fai click su **IAM e amministrazione** e **Service Account**



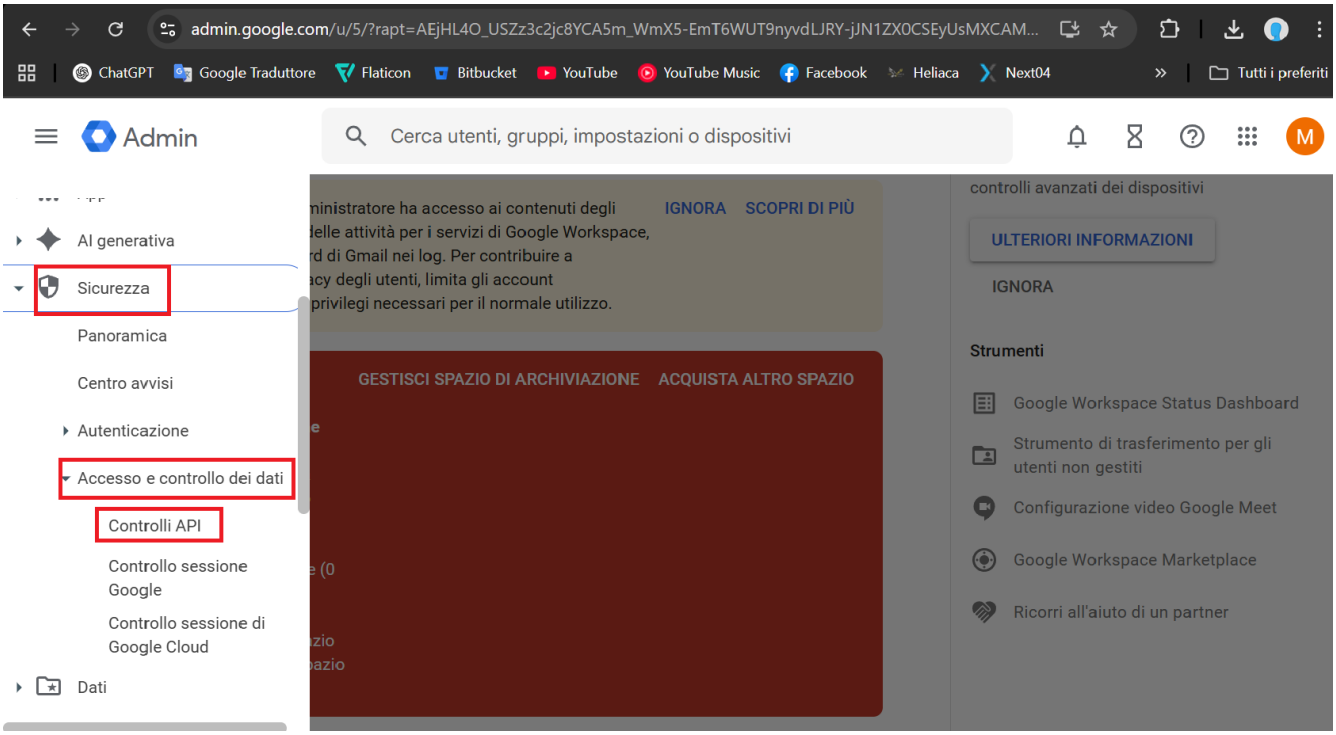
4. Fai click su **Crea service account**
5. Ora inserisci il **Nome, Id** e la **Descrizione** del service account e fai click su **Crea e continua**
6. Saltare la configurazione facoltativa di **Autorizzazioni** e **Entità con accesso**
7. Clicca su **Fine**
8. Appuntarsi **l'ID client** del Service Account
9. Ora clicca sul **Service Account** appena creato
10. Fai click su **Chiavi**, **Aggiungi chiave** e infine **Crea nuova chiave**
11. Ora seleziona **JSON** eppoi fai click su **Crea**
12. Conserva il file **JSON**, esso conterrà tutte le informazioni di accesso al dominio aziendale. Sarà poi utilizzato da Heliaca Finder per accedere tramite Service Account



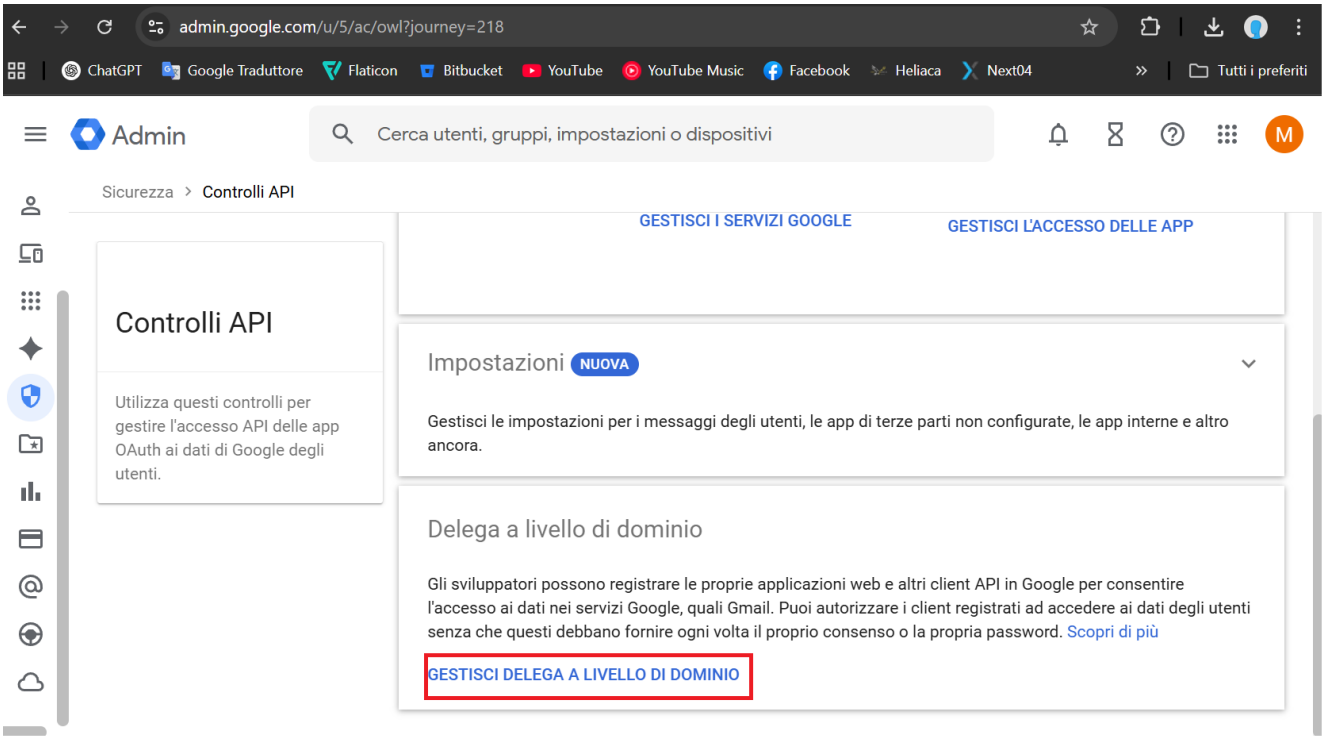
4.8.2.2. ASSEGNAZIONE DELEGA A LIVELLO DI DOMINIO E ASSEGNAZIONE DEGLI AMBITI

Un passaggio fondamentale consiste nel concedere la delega a livello di dominio e assegnare gli ambiti.

1. Collegarsi al sito come utente amministratore: https://admin.google.com
2. Fare click su **Sicurezza**, **Accesso e controllo dei dati** e **Controlli API**



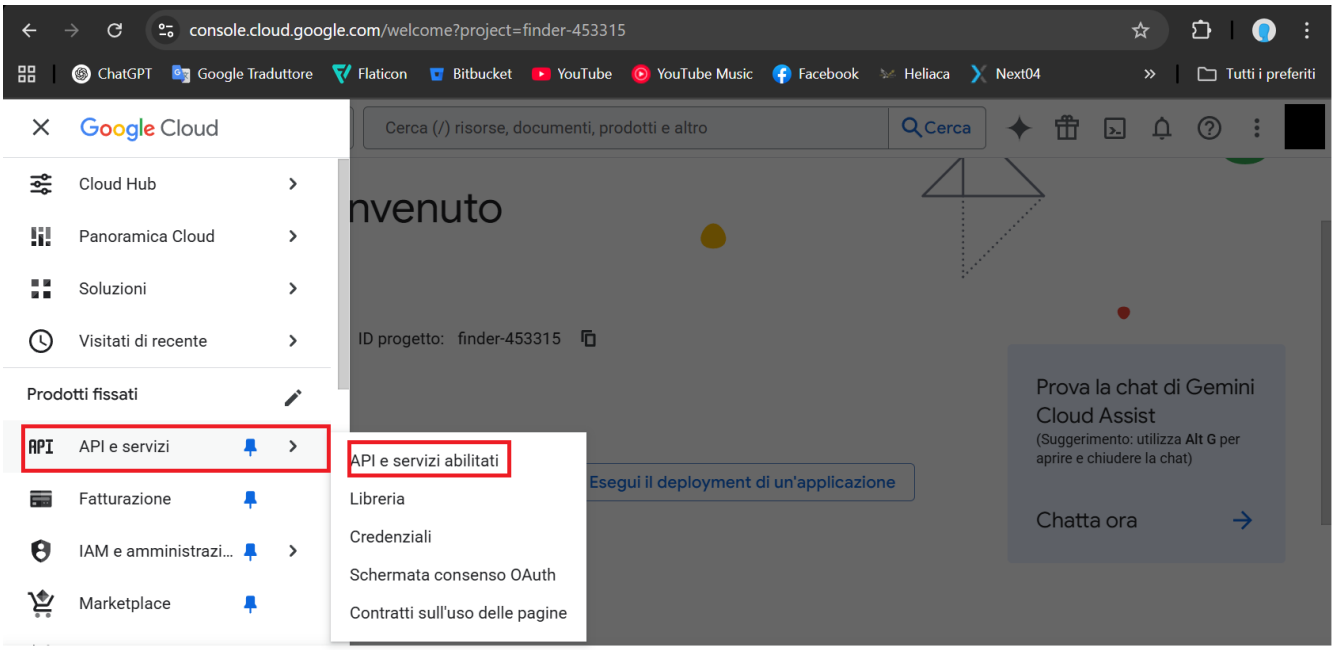
3. Scorrere in basso fino a trovare **GESTISCI DELEGA A LIVELLO DI DOMINIO** e cliccare.



4. Cliccare su **Aggiungi nuovo**, inserire l'ID client del Service Account
5. Nel campo degli ambiti inserire per Google Drive **https://www.googleapis.com/auth/drive.readonly** e **https://www.googleapis.com/auth/admin.directory.user.readonly**
6. Fai click su **AUTORIZZA**

4.8.2.3. Attivazione Admin SDK API

1. Collegarsi al sito: https://console.cloud.google.com
2. Cliccare in alto a sinistra sulle tre linee, poi **API e servizi** e **API e servizi abilitati**



console.cloud.google.com utilizza i cookie di Google per erogare i propri servizi, migliorarne la qualità e analizzare il traffico. Scopri di più.

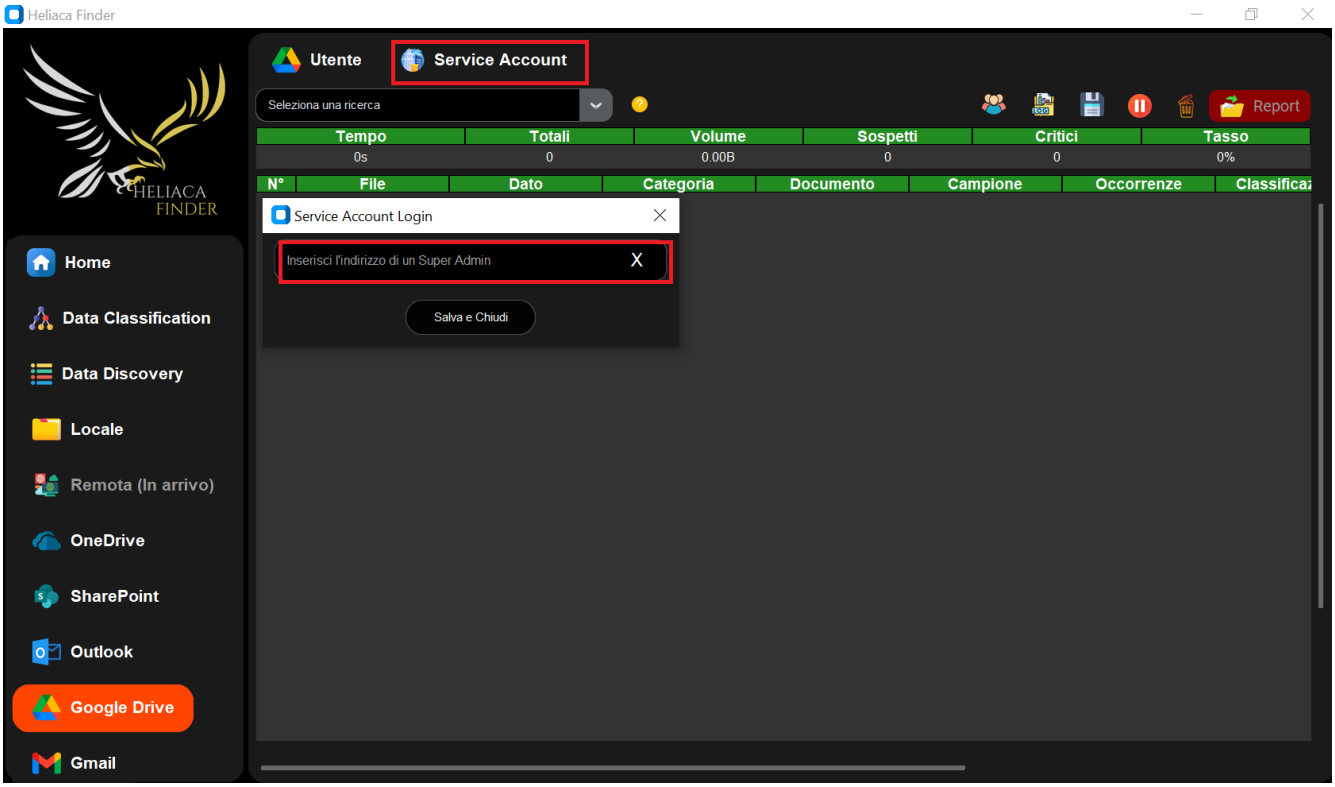
Ok

3. Ora fai click su **Abilita API e servizi**
4. Cerca nella barra di ricerca **Admin SDK API**
5. Clicca sul risultato e **Attiva**
6. Cerca ora nella barra di ricerca **Google Drive API**
7. Clicca sul risultato e **Attiva**

4.8.2.4. Utilizzo del Service Account

Una volta creato il Service Account e ricavato il file json con le credenziali è possibile far partire la ricerca utilizzando la funzionalità Service Account su Heliaca Finder.

1. Fare click su **Service Account**
2. Inserire l'indirizzo di un **Super Admin**
3. Fai click su **Salva e Chiudi**



4. Selezionare il file **.json** contenente le credenziali, precedentemente ottenuto dalla console google.
5. Attendere il processo di login, una volta effettuato la finestra si chiuderà e diventerà verde il pulsante relativo agli **utenti**.
6. Cliccare sul pulsante **utenti**
7. Selezionare gli **utenti** da scansionare e fai click su **Inizia**

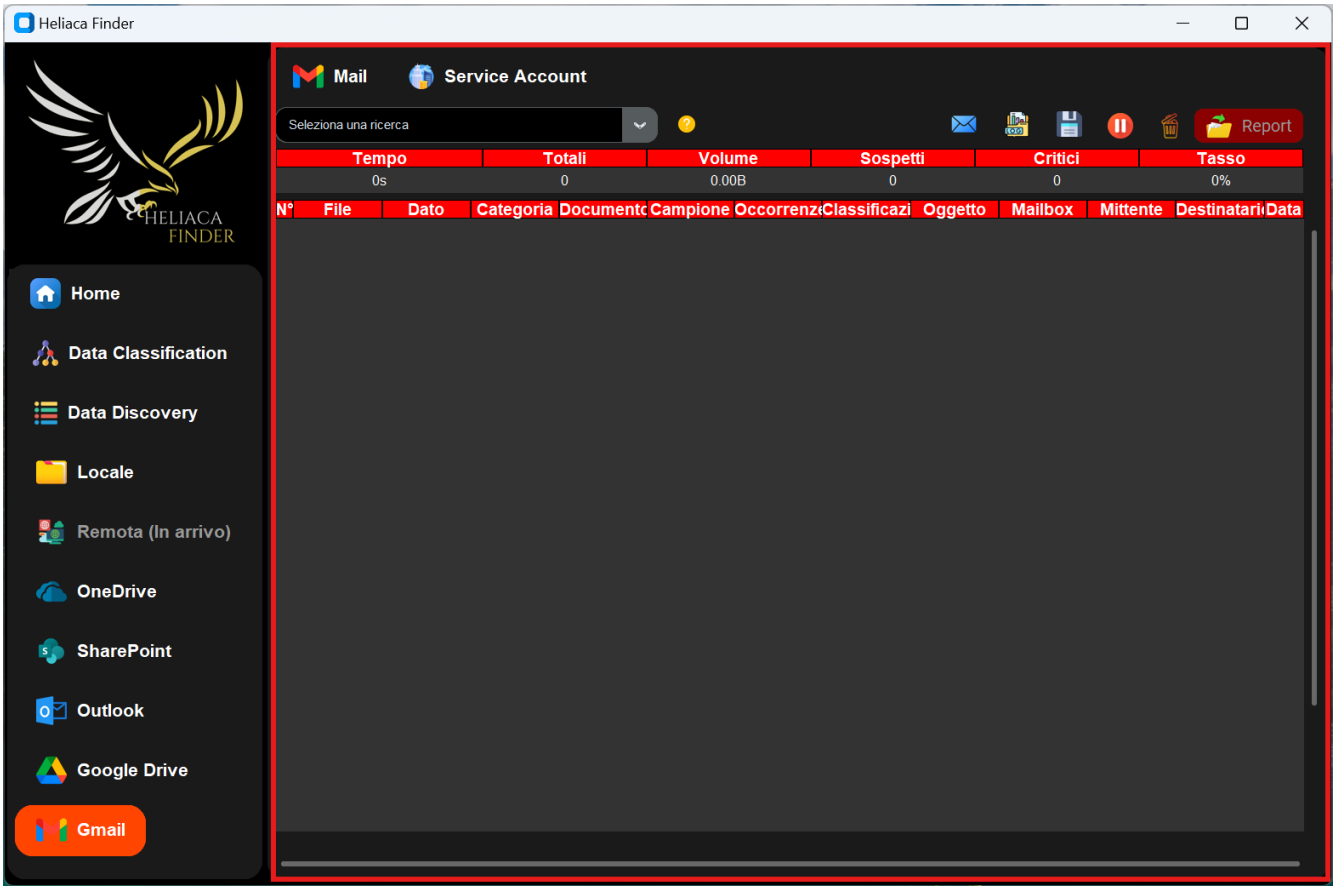
Ora la ricerca dei file sensibili sugli utenti selezionati è partita.

4.8.3. Esecuzione ricerca Google Drive

Una volta effettuato il login utente oppure utilizzando il Service Account, il processo di ricerca partirà immediatamente.
Inizialmente verrà calcolato il volume totale dei file posseduti dall'utente/utenti (step necessario per calcolare la percentuale di completamento).
Terminato il calcolo inizia il vero e proprio processo di ricerca, dove tutti i file presenti vengono dapprima filtrati. Se l'estensione del file rientra nei requisiti impostati nei settings allora il file verrà analizzato procedendo con l'estrazione del testo e la verifica della presenza di informazioni sensibili.
I file non vengono scaricati sul computer, ma viene immagazzinato momentaneamente sulla RAM il loro contenuto sotto forma di stringa BytesIO.

4.9. GMAIL

Il frame inerente alla Data Discovery **Gmail** contiene tutti i widget necessari per il controllo e la visualizzazione di una ricerca effettuata sul percorso remoto di Gmail.



4.9.1. LOGIN MAIL

Per intraprendere una ricerca è sufficiente premere sul pulsante **Mail**; in questo modo si attiverà il processo di login che una volta inserito utente e password per il quale si vuole scansionare le e-mail e gli allegati, ed eventualmente sbloccato con l'authenticator, partirà la ricerca dei dati sensibili sui file disponibili.

4.9.2. LOGIN SERVICE ACCOUNT

In caso si fosse interessati ad effettuare una ricerca su tutte le caselle postali di un dominio aziendale in un colpo solo, è possibile registrare un Service Account sul proprio dominio aziendale e successivamente utilizzare un profilo amministrato ed il file .json con le credenziali per iniziare il processo di ricerca.
Per la configurazione di un Service Account fare riferimento al paragrafo 4.8.2 dove è descritta l'intera procedura fino ad effettuare il login per Google Drive.
L'unica differenza sarà costituita dall'abilitazione della **Gmail API** dal Google Cloud Console e dagli **ambiti** da assegnare al Service Account, nel caso di Gmail gli ambiti da assegnare sono:

- https://www.googleapis.com/auth/admin.directory.user.readonly
- https://www.googleapis.com/auth/gmail.readonly

4.9.3. Esecuzione ricerca Gmail

Una volta effettuato il login, il processo di ricerca partirà immediatamente.
Inizialmente verrà calcolato il numero totale delle e-mail (step necessario per calcolare la percentuale di completamento). Terminato il calcolo inizia il vero e proprio processo di ricerca, dove viene estratto il contenuto testuale delle e-mail (titolo + oggetto + corpo). In caso l'e-mail contenga informazioni sensibili verrà considerata come elemento sensibile e mostrata nel treeview di data discovery. Si passa poi a verificare la presenza degli allegati nell'email e in caso ne fosse provvista, l'allegato prima di essere scaricato viene filtrato in base alla sua estensione.
Alla fine viene estratto il testo ed analizzato.
Né le e-mail né tantomeno gli allegati vengono scaricati sul computer, tutto viene immagazzinato momentaneamente sulla RAM sotto forma di stringa BytesIO.