

# Datafari

## Enterprise Search

*Your knowledge, now*

### Adding GenAI to Datafari Enterprise Search CE What and why

Cedric Ulmer - Président, BCS Search Solutions, Nov. 26th 2025

FRANCELABS  
**Datafari**  
Enterprise Search

# Agenda

- ▶ France Labs ?????!
- ▶ Enterprise Search: Definition
- ▶ Datafari Historic Capabilities
- ▶ Gen AI limitations in an Enterprise context
- ▶ RAG with Datafari and more
- ▶ Roadmap

# Who are France Labs

- ▶ Birth: 2011
- ▶ Released Datafari v1 in 2015 (and learning since then !)
- ▶ Based in Nice.
- ▶ **Recognised speakers** in Enterprise Search: Enterprise Search & Discovery Washington, Search Solutions London, WAICF Cannes, FOSDEM, OSXP Paris ...
- ▶ Part of the CNLL and APELL, resp. **French and European cluster for open source businesses.**
- ▶ Member of The Search Network
- ▶ Customers in Security, Banking, Public Sector, Energy...

# Our activities: SEARCH EXPERTS

## ► Software Maker

Datafari: our Enterprise Search Solution

## ► Technical Expertise

Datafari (and a bit of Solr)

- Audit
- Expertise
- Integration
- Workshops
- Training courses

# WHAT IS ENTERPRISE SEARCH ?

An **enterprise search** application enables employees to **find** all the **information** that the **company possesses** without the need to know where the **information** is stored **[securely]**.

Source: *Enterprise search*, Martin White





SUGGESTED QUERIES

Queries extending your current query terms

ALL CONTENT DOC (7)

Modification Date

From To GO

Authors

100 options

Source

☐ enron 9162

☐ aeap 2586

Show More >>

Extension

☐ Fichier Doc 7321

☐ Fichier PDF 2303

☐ Fichier Web 1414

Show More >>

Results 1 - 10 of 11776

Showing results for: anything

RELEVANCE



da7\_methodes.pdf

Mar s 20 22 Schéma Directeur d'Aménagement et de Gestion des Eaux (SDAGE) 2022-2027 du bassin Artois-Picardie Escaut, Somme & cours d'eau côtiers, Manche, Mer du Nord, Meuse (partie Sambr...

Attachment: [https://www.eau-artois-picardie.fr/sites/default/files/da7\\_methodes.pdf](https://www.eau-artois-picardie.fr/sites/default/files/da7_methodes.pdf)

[Open preview](#)

[Ask with AI](#)

Source: aeap



pdm\_2016-2021.pdf

PROGRAMME DE MESURES 2016-2021 Bassin Artois Picardie Districts hydrographiques Escaut, Somme et Côtiers Manche Mer du Nord Meuse (Partie Sambre) Programme de mesures 2016...

Attachment: [https://www.eau-artois-picardie.fr/sites/default/files/pdm\\_2016-2021.pdf](https://www.eau-artois-picardie.fr/sites/default/files/pdm_2016-2021.pdf)

[Open preview](#)

[Ask with AI](#)

Source: aeap

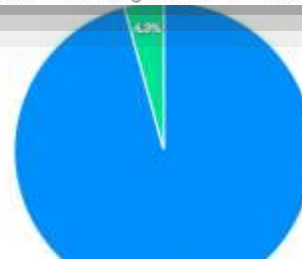


Cuiabá Integrated Energy Project

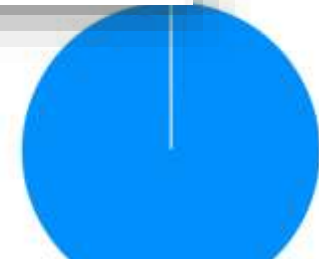
Cuiaba I Integrated Project Major Outstanding Items to close Financing Obtain Furnas Consent



Ask your content



Hit  
No Hit



No Click  
Clicked

Last 100 days Refresh every 10min Update

(count)

4001

295

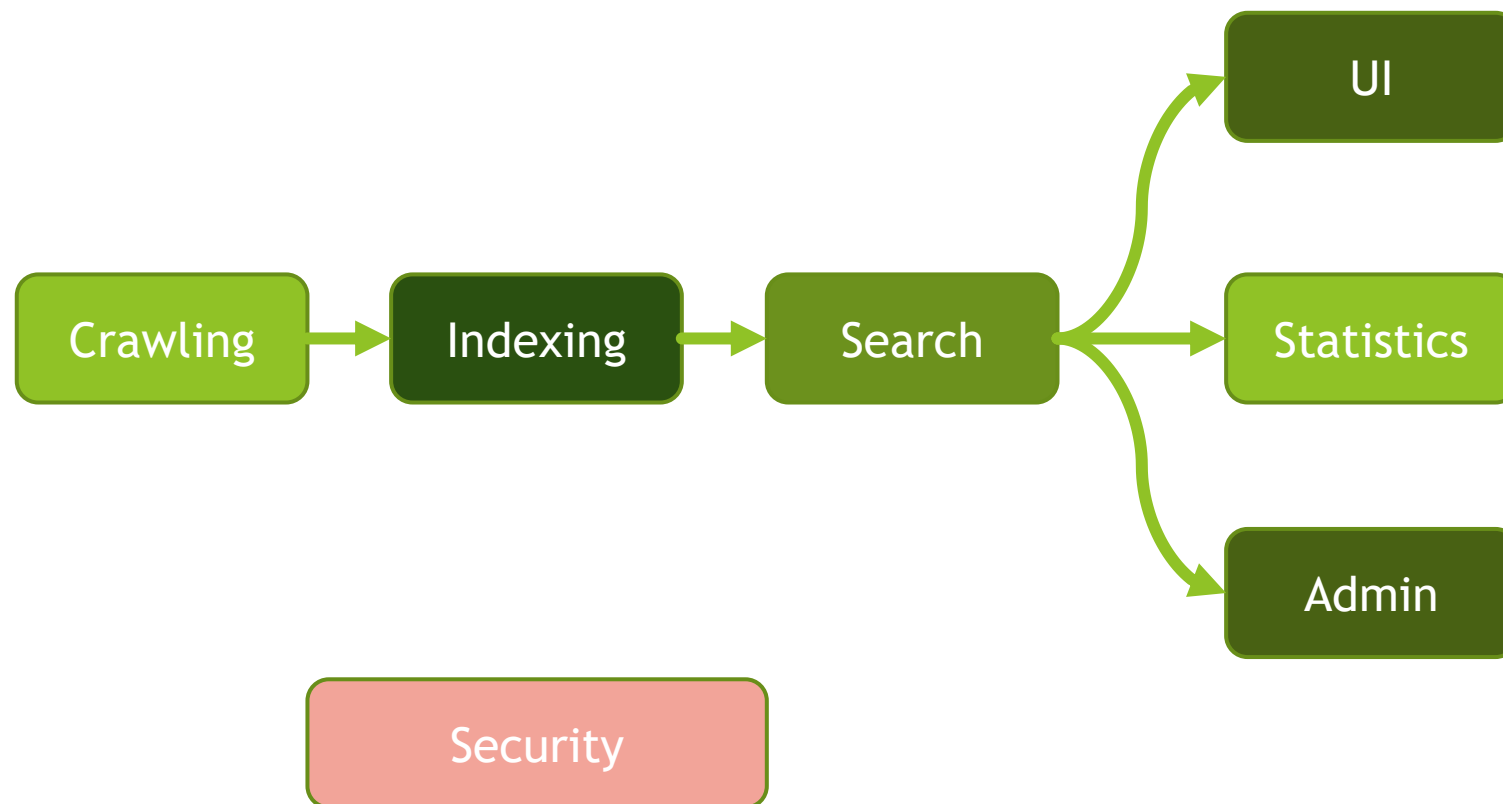
160

93

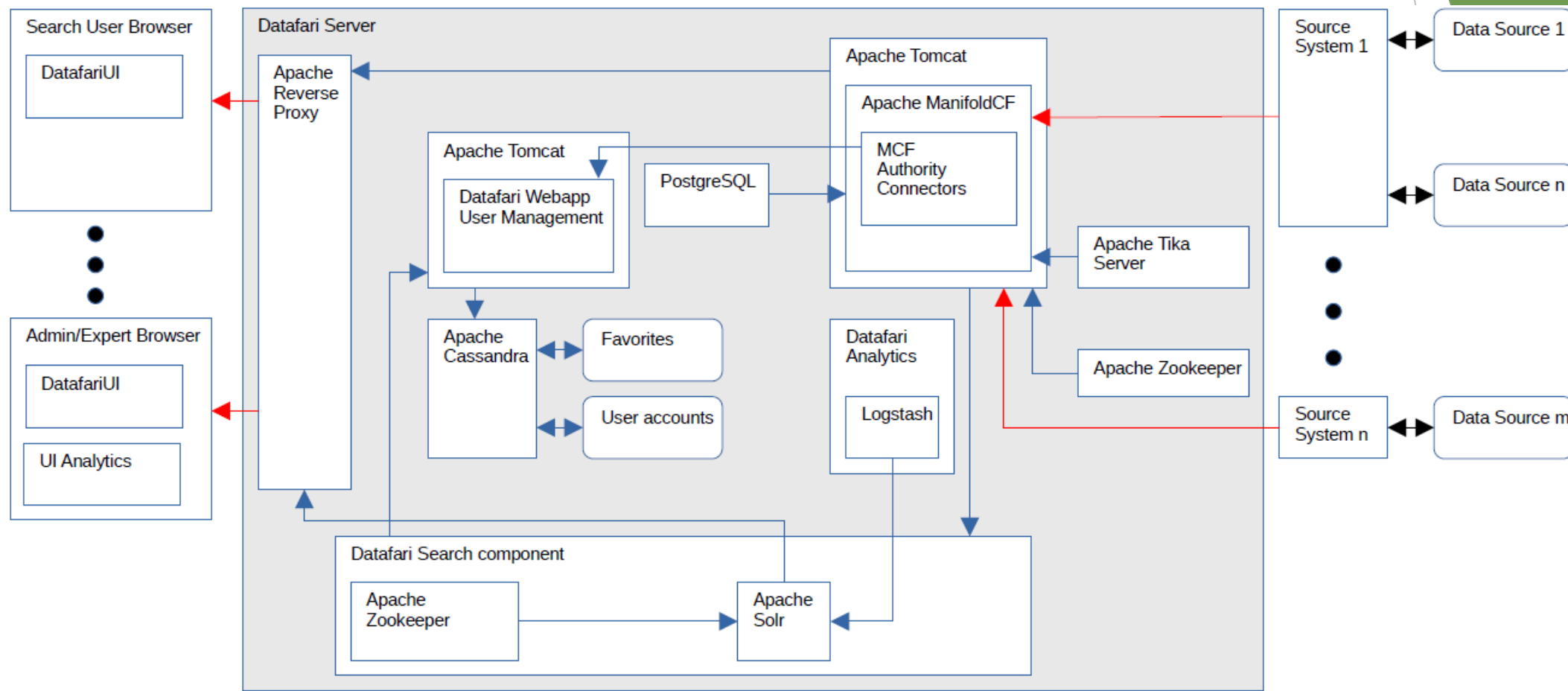
91

85

# Enterprise Search - How ? (historically)



# Datafari Architecture



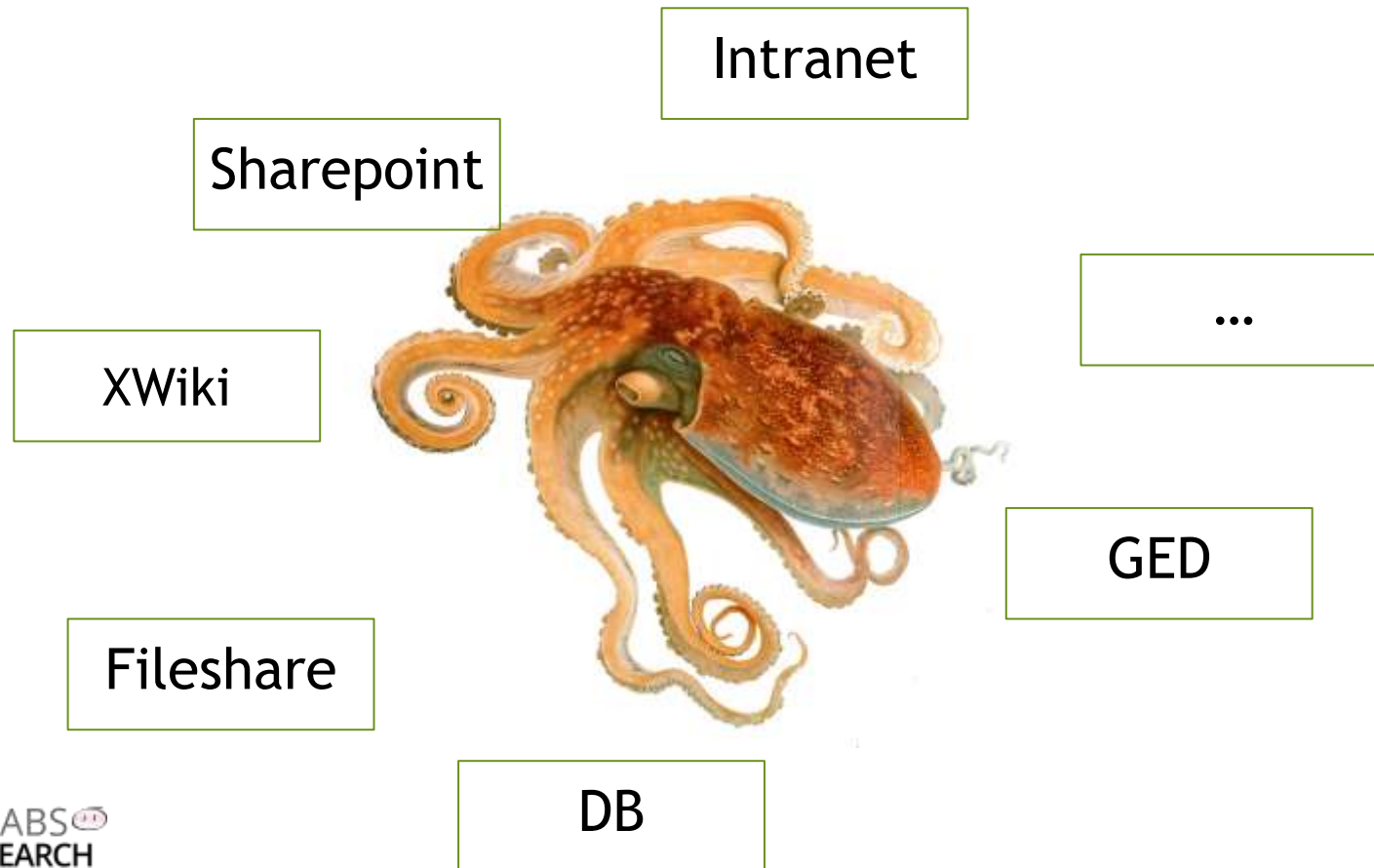


# Datafari - Historic functionalities

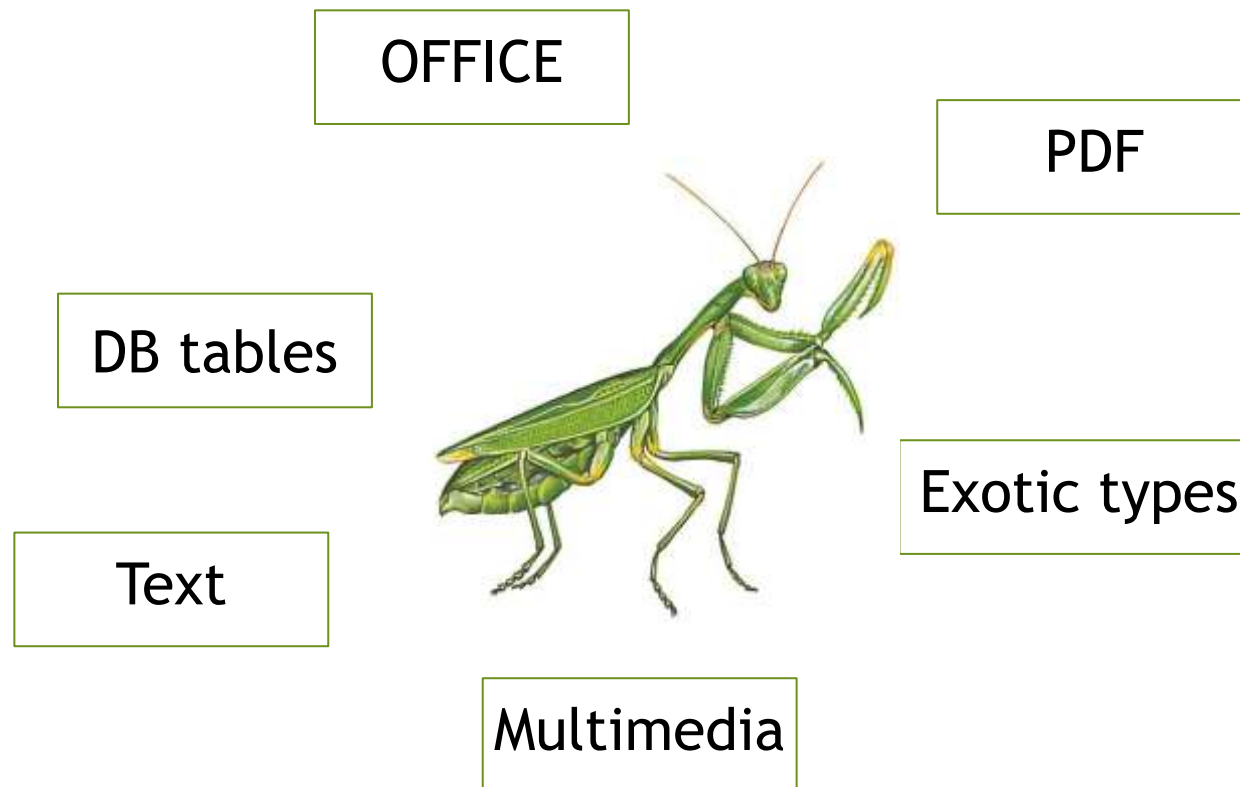
The screenshot displays the Datafari search interface with several key features highlighted by green arrows and labels:

- Search bar:** Located at the top right of the interface.
- Tabs:** Located on the left side, showing 'ALL CONTENT' and 'DOC (7)'. A dropdown menu for 'Modification Date' is also visible.
- Autocomplete:** A dropdown menu showing suggestions for the search term 'enr', including 'enron' and 'enrononline'.
- Autosuggest:** A dropdown menu showing suggestions for the search term 'enr', including 'The Nash equilibrium', 'Amendment', and 'Enron283\_9\_04\_01.doc'.
- Facets:** Located on the left side, showing filters for 'Source' (enron, aeap) and 'Extension' (Fichier Doc, Fichier PDF, Fichier Web).
- Result with + snippet + highlighting + previewing:** A search result for 'pdm\_2016-2021.pdf' is shown, including a snippet, highlighting, and a preview link.
- More options:** A list of additional features: '+Export', '+Saved search', '+Alerts', and '+Advanced search'.

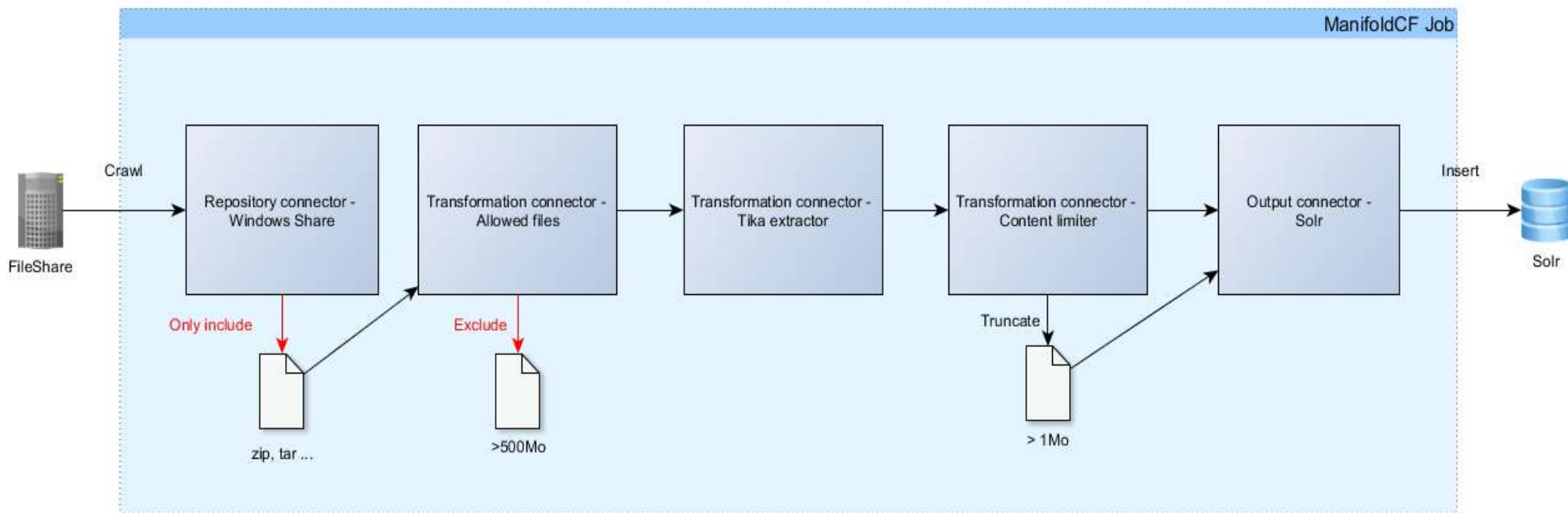
# Knowledge crawling - Multisources



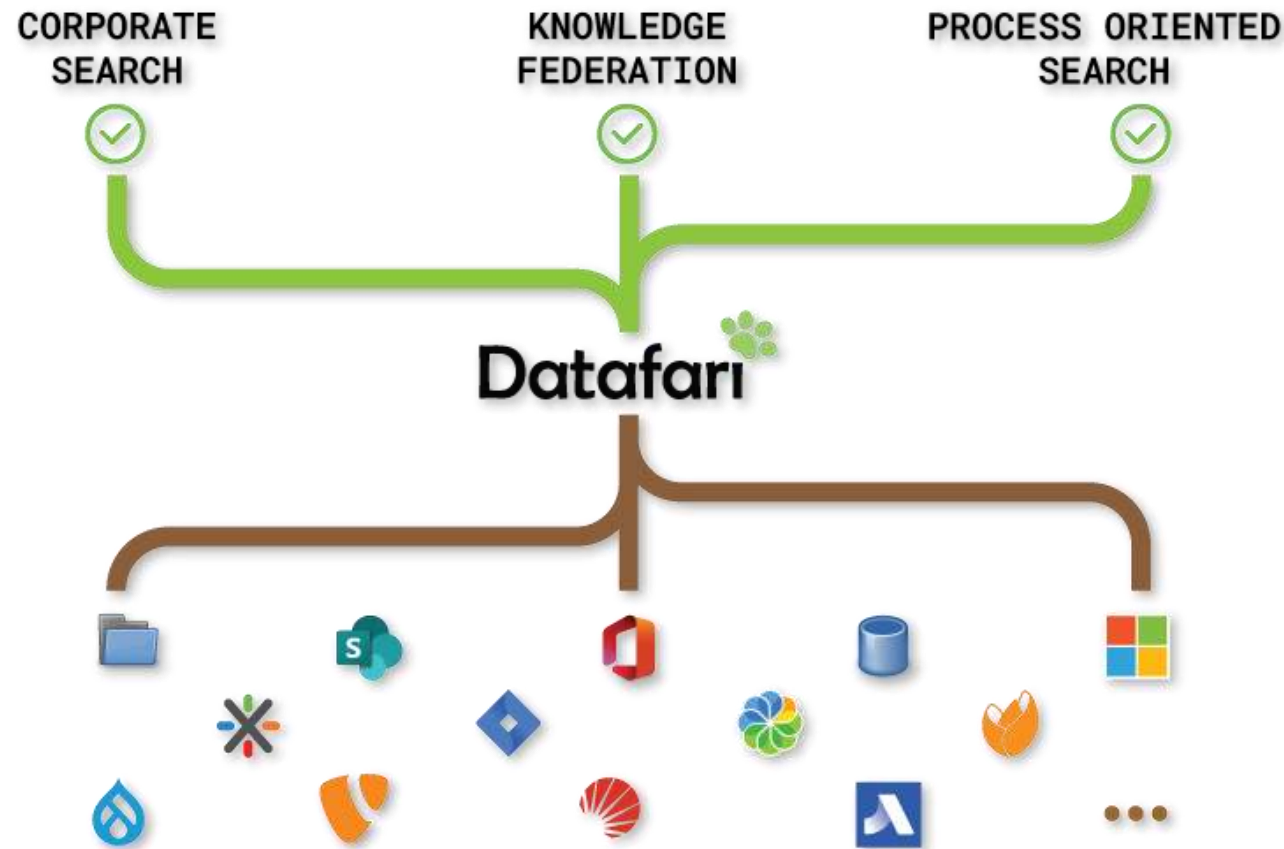
# Knowledge crawling - Multiformats



# Knowledge crawling - Pipeline



**Conclusion** - Overall vision on all the internal documents, respecting access controls



# GenAI revolution

- ▶ genAI can summarize documents
- ▶ genAI can generate answers
- ▶ genAI can create code
- ▶ genAI can do many things...
- ▶ genAI can pollute as hell
- ▶ BUT...





# GenAI limitations for security

- ▶ Trained LLMs generate answers with ALL the documents they learned from
- ▶ No respect of documents level permissions
- ▶ Impossible to use if documents level permissions required: typically the case in companies



# GenAI limitations for freshness

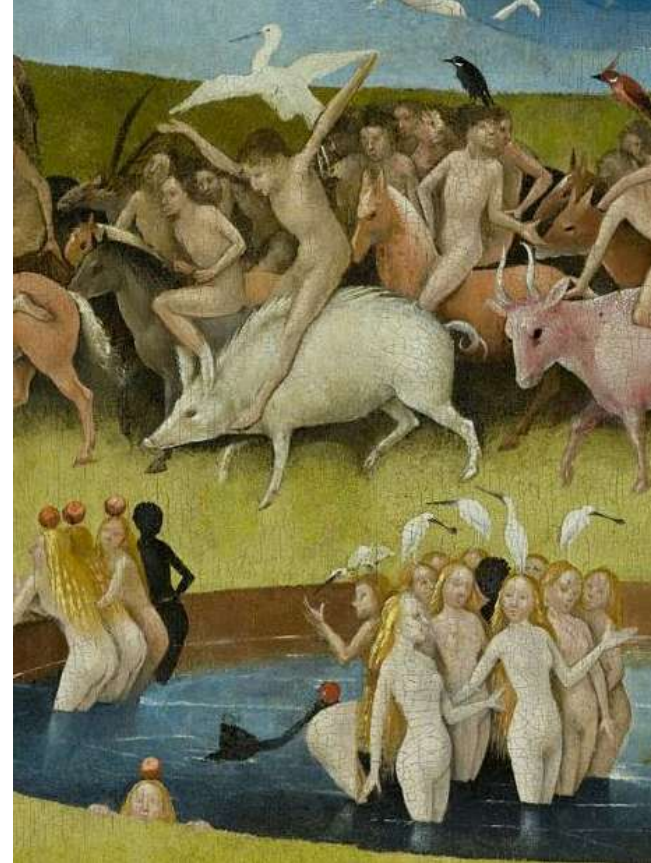
- ▶ Trained LLMs cannot easily manage changes in documents already ingested (or deletions)
- ▶ Training an LLM can be costly





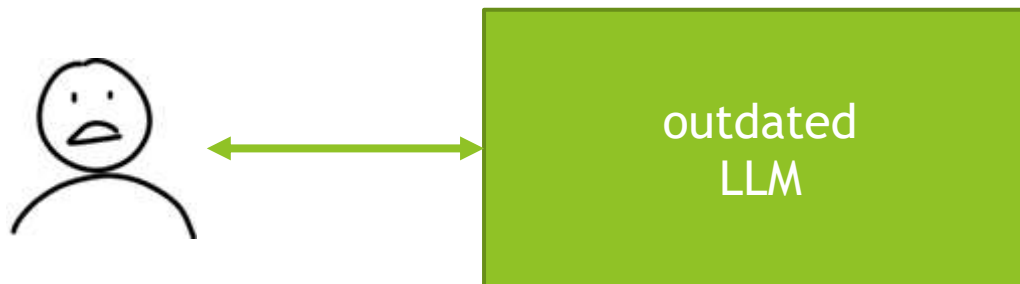
# SOLUTION: RAG

- ▶ Add a « retrieval » step before contacting the LLM
- ▶ Objective: provide more context to the question asked
- ▶ Retrieval step: usually a search (token based, vector or hybrid)

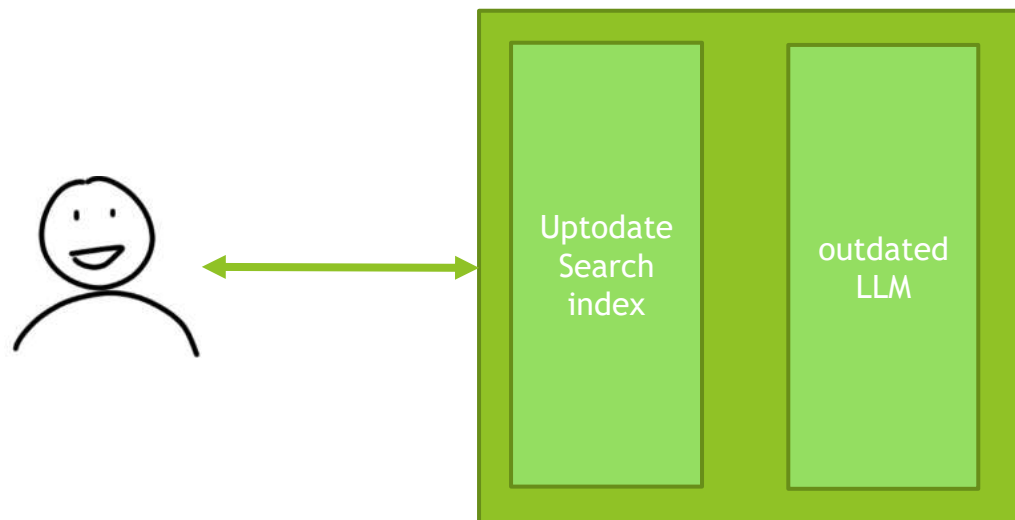


# SOLUTION: RAG

- The world before RAG

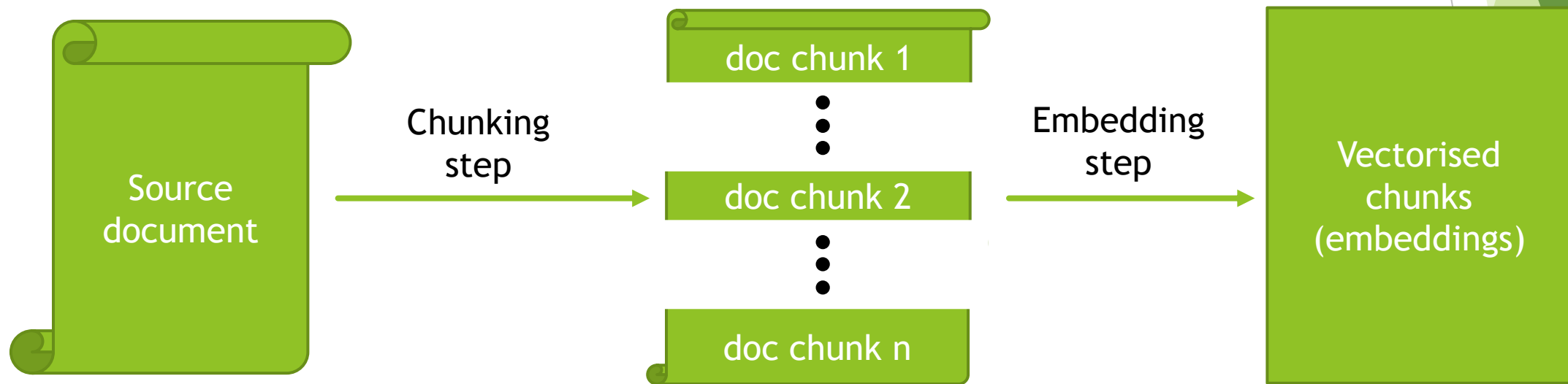


- The world with RAG



# SOLUTION: RAG

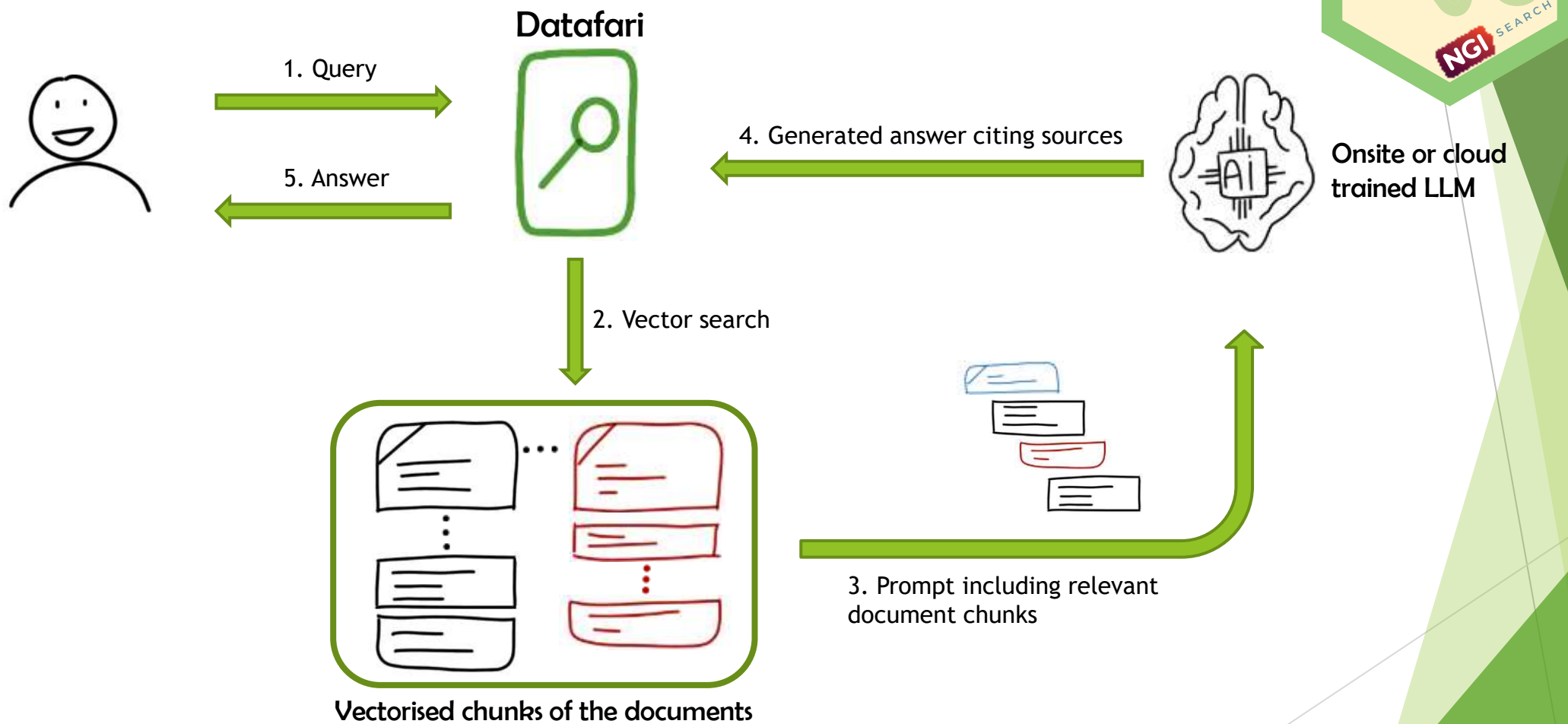
- Requires documents chunking and embedding for vector search



# WHY RAG

- ▶ « retrieval » takes permissions into account
- ▶ Search step (retrieval) easier to keep uptodate

# COMBINING DATAFARI AND RAG

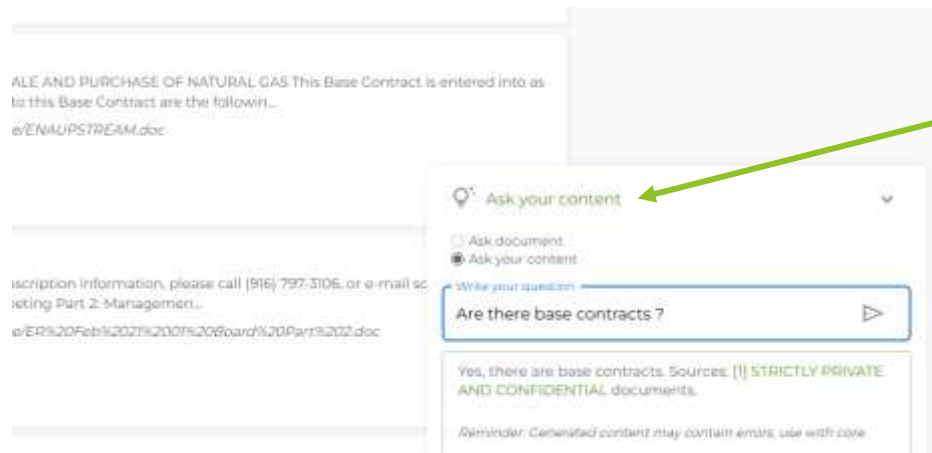


# DATAFARI AND AI

## Current capabilities

- ▶ Ask questions on the overall accessible corpus
- ▶ Ask questions on specific documents
- ▶ Mono hop
- ▶ Use « quick actions »
- ▶ Enrich documents at the indexing phase

# DATAFARI RAG - what does it look like ?



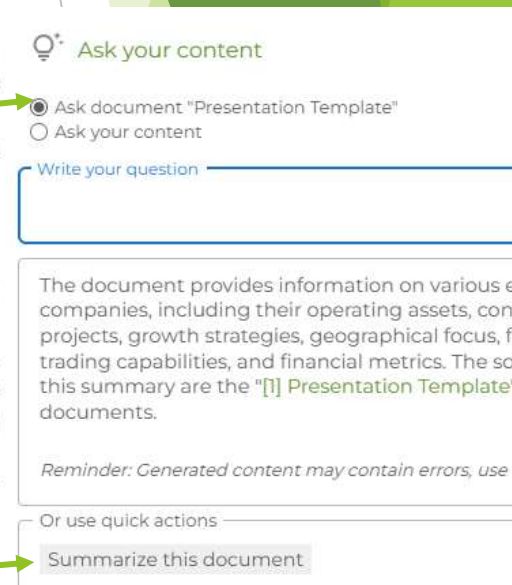
Ask questions  
over your corpus



Ask questions  
on a specific  
document


Presentation Template  
Newco CALpine Southern Energy\* Aes Orion NRG\*\*  
75 MW 3,355 MW 17,814 MW(3) 27,936 MW 5,227 MW  
Construction 5,328 MW 12,153 MW 7,775 MW...

NON-DISCLOSURE AGREEMENT  
NON-DISCLOSURE AGREEMENT This Non-Disclosure /  
Agreement") dated March 16, 2001 between the parties  
governs the disclosure of Confidential Information pur:  
...  
Attachment: file:///filesare.datafari.com/share/Enron%  
Open preview  
Ask with AI  
Source: share







Quick actions  
on individual  
docs

# DATAFARI AI - Ask With AI

 **Datafari**

Recherche

 **CONNEXION**

OUTILS DE RECHERCHE

Bienvenue à la démonstration de la recherche d'entreprise Datafari. L'objectif de Datafari est de simplifier votre vie quotidienne au travail, en vous permettant de rechercher des documents et des informations dans toute votre entreprise, sans avoir à parcourir les silos un par un.

Pour commencer votre expérience de recherche, saisissez simplement un ou plusieurs mots-clés dans la barre de recherche ci-dessus, et laissez Datafari faire le travail pour vous. Comme il s'agit d'une démo, veuillez consulter la troisième colonne pour quelques mots-clés qui ont des résultats. Datafari vous amènera à une page de liste de résultats, équipée de nombreuses fonctionnalités telles que des facettes pour naviguer dans vos résultats.

Datafari est sécurisé par conception, et en tant que tel, vous ne pourrez rechercher que les documents auxquels vous êtes autorisé à consulter, et vos communications sont cryptées via SSL.

Profitez de votre recherche !

Comment rechercher

Vous êtes probablement familier avec les paradigmes de recherche sur le Web, et Datafari s'en inspire pour que votre courbe d'apprentissage soit aussi courte que nécessaire. Tout comme vous le faites avec votre recherche Web quotidienne, saisissez simplement un ou plusieurs mots-clés dans la barre de recherche ci-dessus.

Par défaut, Datafari opte pour un opérateur AND caché, donc il recherchera les documents contenant tous vos mots-clés, à moins que vous n'ajoutiez explicitement un opérateur OR entre les mots-clés. Vous pouvez trouver plus d'informations sur les opérateurs de recherche disponibles dans la page d'aide de la page de résultats.

Datafari est équipé d'une fonctionnalité d'autocomplétion, qui proposera des mots présents dans les documents et commençant par les caractères que vous tapez actuellement. Cela vous permet de gagner du temps et de réduire les erreurs en ne voyant que les mots qui existent réellement dans l'index de recherche.


Quelles sources cet index Datafari indexe-t-il ?

Dans cette démo, nous avons indexé un partage de fichiers SMB et deux sites Web. Le partage de fichiers contient un corpus provenant d'une ancienne entreprise de services publics, basée aux États-Unis. Nous vous recommandons donc de rechercher des termes tels que "énergie", "services publics", "facture" ou similaires. Pour les sites Web, nous indexons les sites Web de Datafari et de France Labs, vous pouvez donc rechercher des termes tels que "Datafari" ou "innovation" pour commencer.


Amusez-vous en tant que Datafarien !



# DATAFARI AI - RAG

 Datafari

Recherche

    [CONNEXION](#)

OUTILS DE RECHERCHE

Bienvenue à la démonstration de la recherche d'entreprise Datafari. L'objectif de Datafari est de simplifier votre vie quotidienne au travail, en vous permettant de rechercher des documents et des informations dans toute votre entreprise, sans avoir à parcourir les silos un par un.

Pour commencer votre expérience de recherche, saisissez simplement un ou plusieurs mots-clés dans la barre de recherche ci-dessus, et laissez Datafari faire le travail pour vous. Comme il s'agit d'une démonstration, veuillez consulter la troisième colonne pour quelques mots-clés qui ont des résultats. Datafari vous amènera à une page de liste de résultats, équipée de nombreuses fonctionnalités telles que des facettes pour naviguer dans vos résultats.

Datafari est sécurisé par conception, et en tant que tel, vous ne pourrez rechercher que les documents auxquels vous êtes autorisé à consulter, et vos communications sont cryptées via SSL.

Profitez de votre recherche !

Comment rechercher

Vous êtes probablement familier avec les paradigmes de recherche sur le Web, et Datafari s'en inspire pour que votre courbe d'apprentissage soit aussi courte que nécessaire. Tout comme vous le faites avec votre recherche Web quotidienne, saisissez simplement un ou plusieurs mots-clés dans la barre de recherche ci-dessus.

Par défaut, Datafari opte pour un opérateur AND caché, donc il recherchera les documents contenant tous vos mots-clés, à moins que vous n'ajoutiez explicitement un opérateur OR entre les mots-clés. Vous pouvez trouver plus d'informations sur les opérateurs de recherche disponibles dans la page d'aide de la page de résultats.

Datafari est équipé d'une fonctionnalité d'autocomplétion, qui proposera des mots présents dans les documents et commençant par les caractères que vous tapez actuellement. Cela vous permet de gagner du temps et de réduire les erreurs en ne voyant que les mots qui existent réellement dans l'index de recherche.

Quelles sources cet index Datafari indexe-t-il ?

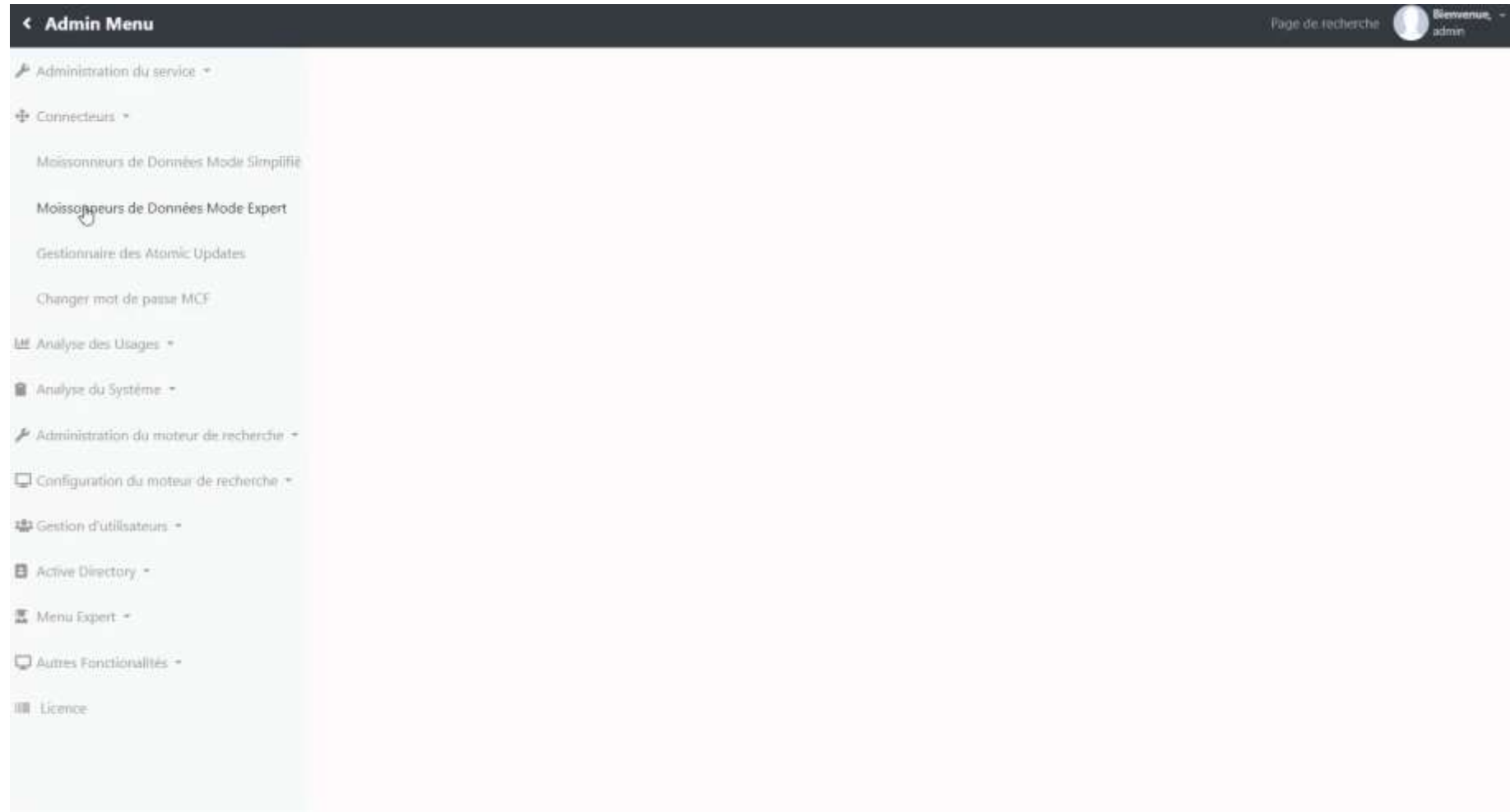
Dans cette démonstration, nous avons indexé un partage de fichiers SMB et deux sites Web. Le partage de fichiers contient un corpus provenant d'une ancienne entreprise de services publics, basée aux États-Unis. Nous vous recommandons donc de rechercher des termes tels que "énergie", "services publics", "facture" ou similaires. Pour les sites Web, nous indexons les sites Web de Datafari et de France Labs, vous pouvez donc rechercher des termes tels que "Datafari" ou "Innovation" pour commencer.

Amusez-vous en tant que Datafari !

# DATAFARI AI - Enriching documents

- ▶ Include genAI based analysis at indexing phase
- ▶ For instance: summaries, translation, entity extractions...

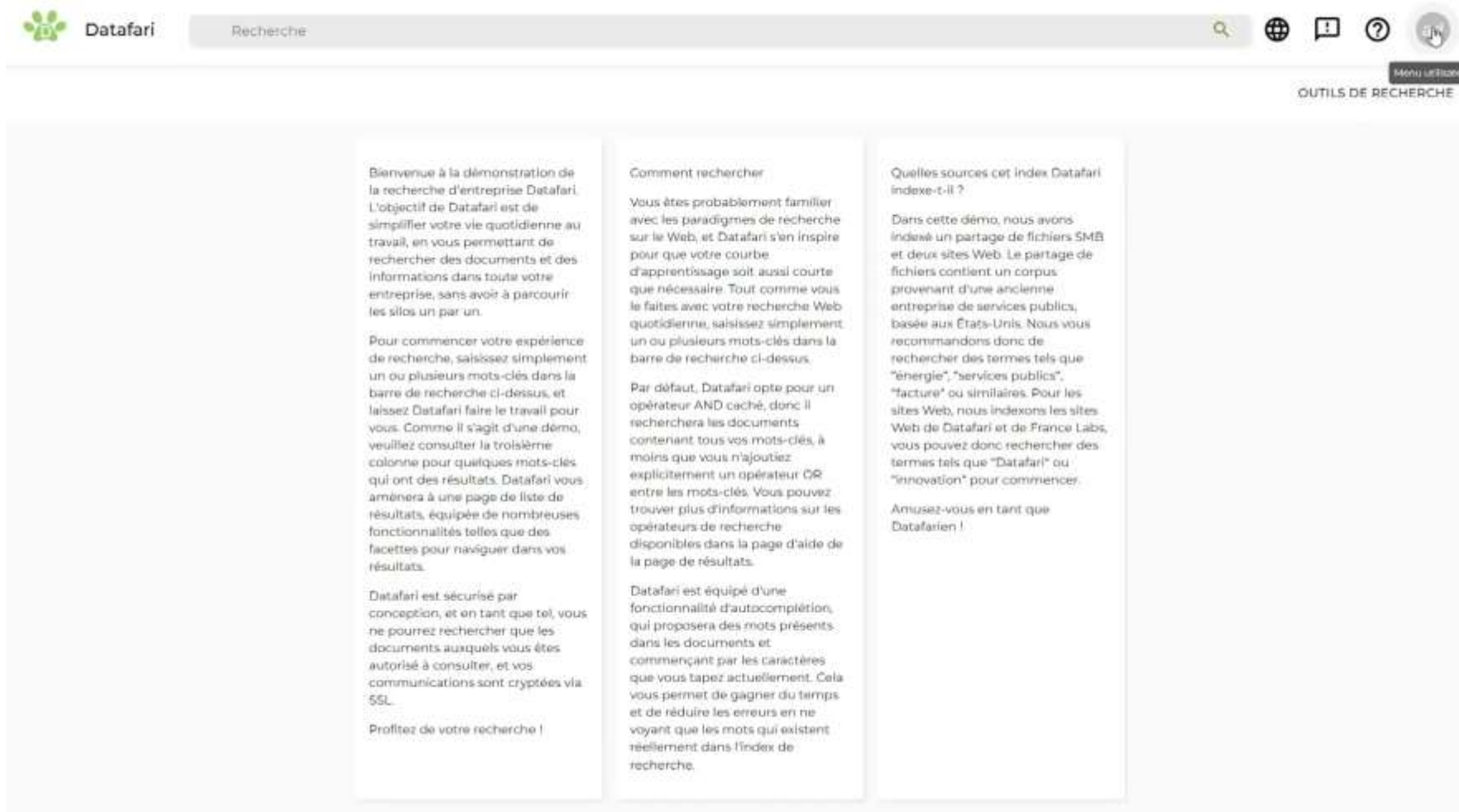
# DATAFARI AI - Enriching documents



# DATAFARI CE - Setting up your RAG

Step by step

# RAG Step 1 - Crawling



**Datafari** Recherche

Menu utilisateur

OUTILS DE RECHERCHE

Bienvenue à la démonstration de la recherche d'entreprise Datafari. L'objectif de Datafari est de simplifier votre vie quotidienne au travail, en vous permettant de rechercher des documents et des informations dans toute votre entreprise, sans avoir à parcourir les silos un par un.

Pour commencer votre expérience de recherche, saisissez simplement un ou plusieurs mots-clés dans la barre de recherche ci-dessus, et laissez Datafari faire le travail pour vous. Comme il s'agit d'une démo, veuillez consulter la troisième colonne pour quelques mots-clés qui ont des résultats. Datafari vous amènera à une page de liste de résultats, équipée de nombreuses fonctionnalités telles que des facettes pour naviguer dans vos résultats.

Datafari est sécurisé par conception, et en tant que tel, vous ne pourrez rechercher que les documents auxquels vous êtes autorisé à consulter, et vos communications sont cryptées via SSL.

Profitez de votre recherche !

Comment rechercher

Vous êtes probablement familier avec les paradigmes de recherche sur le Web, et Datafari s'en inspire pour que votre courbe d'apprentissage soit aussi courte que nécessaire. Tout comme vous le faites avec votre recherche Web quotidienne, saisissez simplement un ou plusieurs mots-clés dans la barre de recherche ci-dessus.

Par défaut, Datafari opte pour un opérateur AND caché, donc il recherchera les documents contenant tous vos mots-clés, à moins que vous n'ajoutiez explicitement un opérateur OR entre les mots-clés. Vous pouvez trouver plus d'informations sur les opérateurs de recherche disponibles dans la page d'aide de la page de résultats.

Datafari est équipé d'une fonctionnalité d'autocomplétion, qui proposera des mots présents dans les documents et commençant par les caractères que vous tapez actuellement. Cela vous permet de gagner du temps et de réduire les erreurs en ne voyant que les mots qui existent réellement dans l'index de recherche.

Quelles sources cet index Datafari indexe-t-il ?

Dans cette démo, nous avons indexé un partage de fichiers SMB et deux sites Web. Le partage de fichiers contient un corpus provenant d'une ancienne entreprise de services publics, basée aux États-Unis. Nous vous recommandons donc de rechercher des termes tels que "énergie", "services publics", "facture" ou similaires. Pour les sites Web, nous indexons les sites Web de Datafari et de France Labs, vous pouvez donc rechercher des termes tels que "Datafari" ou "innovation" pour commencer.

Amusez-vous en tant que Datafarien !

# RAG Step 2 - Configure the Embedding model [OPTIONAL]

- ▶ Edit the desired model from the configuration file (with graphical UI)

# RAG Step 3 - Configure the LLM model

## [OPTIONAL]

- Edit the desired model from the configuration file (no graphical UI yet)

# RAG Step 4 - It's ready !

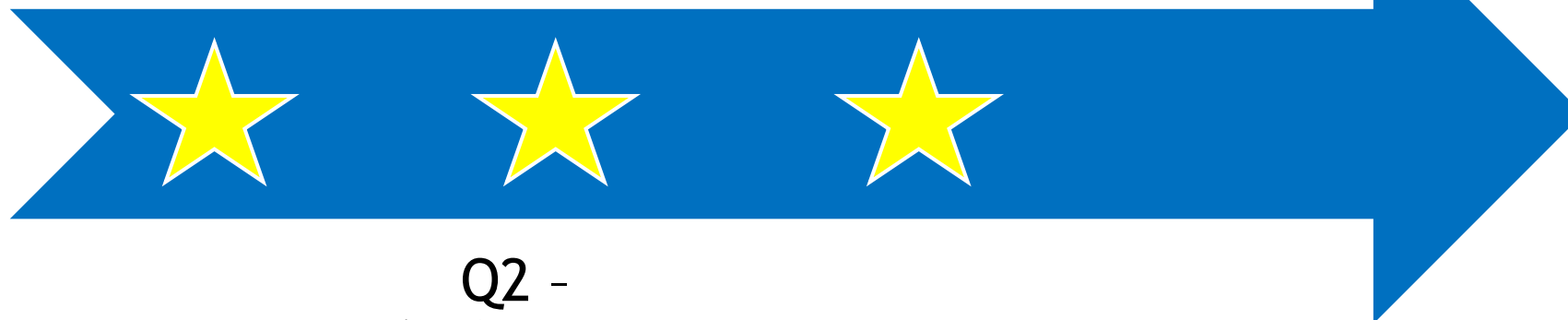
- ▶ Nothing more to activate 😊



# DATAFARI - Roadmap S1-2026

Q1 -  
Agentic  
AI

Q2 - Rework of the  
embedding phase

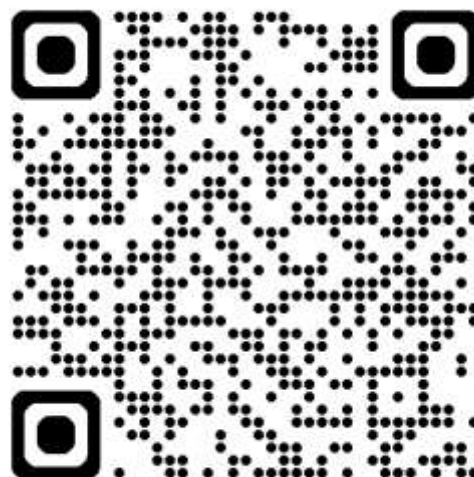


Q2 -  
Assistant-  
UI



# DATAFARI WHERE TO START ?

- ▶ <https://www.datafari.com/en/download.html>
- ▶ Using the container is the quickest way to have it running
- ▶ Lots of documentation on [wiki.datafari.com](http://wiki.datafari.com)
- ▶ Code is on gitlab / github :
  - ▶ [gitlab.datafari.com](http://gitlab.datafari.com)
  - ▶ [github.com/datafari](https://github.com/datafari)
- ▶ France Labs Linkedin account:



# QUESTIONS ?

cedric.ulmer@francelabs.com

X account: @francelabs

X tag: #datafari

FRANCELABS   
BIG DATA SEARCH  
AND ANALYTICS

**Datafari** 