



Grafhydro : la connexion sémantique au service de la gestion des connaissances en ingénierie hydraulique

Yann REBOURS & Sébastien FREAL-SAISON

EDF Hydro / CIH

Nicolas CHAUVAT

Logilab

SemWebPro – 27/11/2025

1

Plantons le décor

Offres de service du Centre d'Ingénierie Hydraulique (CIH)

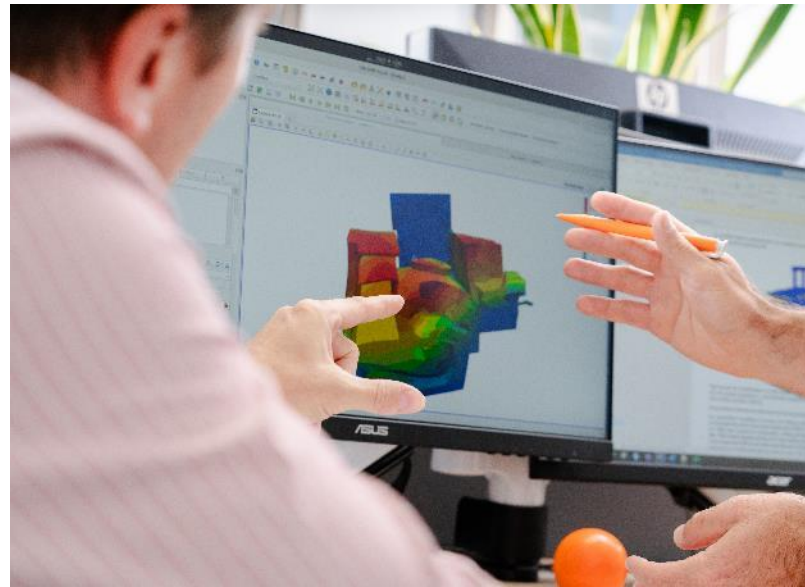
Conseiller



AMOA

(ASSISTANCE À MAÎTRISE D'OUVRAGE)

Étudier



Expertise,
études,
conception

Réaliser



MOE

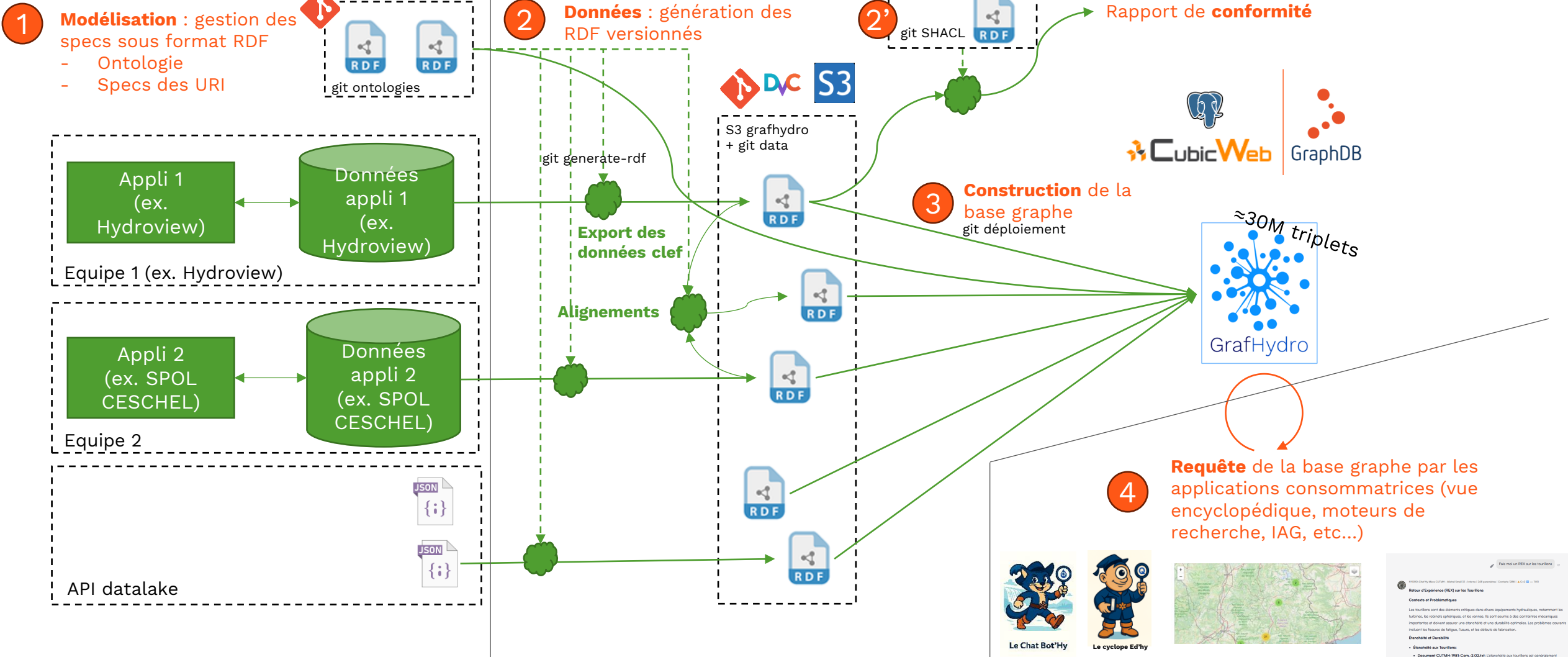
(MAÎTRISE D'ŒUVRE)

MOE de réalisation, travaux,
Maintien en Condition
Opérationnelle...

Quand ton collègue vient d'entendre pour la première fois « web sémantique », « ontologie » et « base graphe. »



Droit au but = infra & pipeline mises en place



#1 Identifiez les données sources

#2 Assemblez et publiez vos graphes

#3 Exploitez les graphes obtenus



Cf. la méthodologie <https://semgraph.logilab.fr>

2

Zoom sur les étapes

Modélisation

Définition de la modélisation = ontologie ; grammaire, structure

≈170 classes
≈650 propriétés

1a

Exemple de fichier TTL en utilisant l'ontologie OWL. Inclut la définition des URI des nœuds

```
52  rdfs:range ceschel:Armoire .
53
54  ceschel:a_pour_titulaire
55  a owl:ObjectProperty ;
56  rdfs:label "a pour titulaire" ;
57  rdfs:comment "Le schéma électrique a été produit par un titulaire" ;
58  rdfs:isDefinedBy ceschel: ;
59  rdfs:domain ceschel:Document ;
60  rdfs:range ceschel:Titulaire .
61
62  ceschel:is_about
63  a owl:ObjectProperty ;
64  rdfs:subPropertyOf grafhydro:is_about ;
65  rdfs:isDefinedBy ceschel: ;
66  rdfs:domain
67    ceschel:EntreeSortie,
68    organisation:Site ;
69  rdfs:range ceschel:Document .
70
71
72  ceschel:Document
73  a owl:Class ;
74  rdfs:label "Document CESCHEL" ;
75  rdfs:comment "CESCHEL signifie Cellule Schéma Electrique. Comme le nom est l'identifiant unique, la production de l'URI se fait sous le forme" ;
76  rdfs:isDefinedBy ceschel: .
77
78
79  ceschel:Armoire
80  a owl:Class ;
81  rdfs:label "Armoire électrique" ;
82  rdfs:comment "Les plans font référence à des armoires électriques. Cette référence est liée au plan. Il n'est pas certain que cette étiquette" ;
83  rdfs:isDefinedBy ceschel: .
84
85
86  ceschel:Titulaire
87  a owl:Class ;
88  rdfs:label "Titulaire" ;
89  rdfs:comment "Le titulaire est la société qui a réalisé le schéma électrique. L'URI se forme ainsi : https://hydro.data.edf.fr/document/sche" ;
90  rdfs:isDefinedBy ceschel: .
91
```

L'URI se forme ainsi :

[https://hydro.data.edf.fr/document/schema-electrique/entree-sortie/\[SITE-REFERENCE\]](https://hydro.data.edf.fr/document/schema-electrique/entree-sortie/[SITE-REFERENCE])

Par exemple :

<https://hydro.data.edf.fr/document/schema-electrique/entree-sortie/G.MAIH-V171>

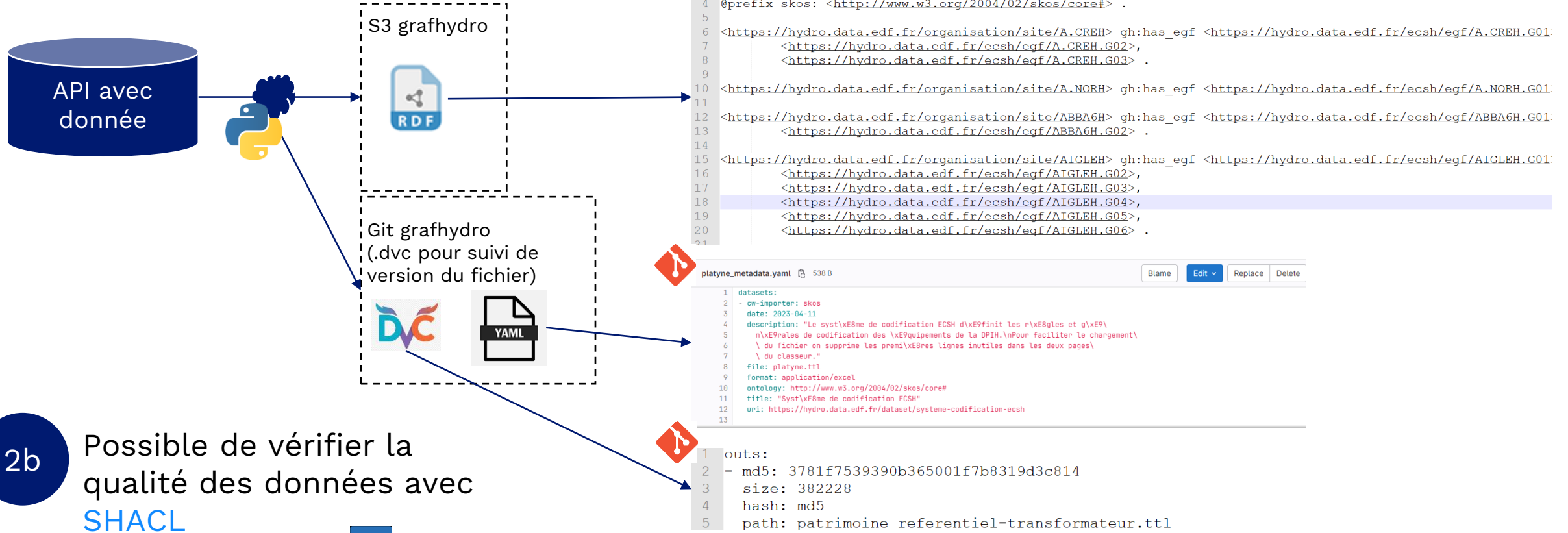
1b

Documentation HTML du modèle générée automatiquement à partir des fichiers OWL exprimés en TTL

Production des RDF

2a

Requête sur la source de données



2b

Possible de v rifier la
qualit  des donn es avec
[SHACL](#)

```
S3
@prefix cw: <https://cubicweb.org/shacl/> .
@prefix gh: <https://hydro.data.edf.fr/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix sh: <http://www.w3.org/ns/shacl#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

8 cw:Concept a sh:NodeShape ;
9   sh:targetClass gh:Concept .

10
11 cw:Equipement a sh:NodeShape ;
12   sh:property [ sh:maxCount 1 ;
13     sh:path rdfs:label ],
```

Gestion de version des données : illustration

```
(venv) λ git log --all --decorate --oneline --graph
* 6268cf5 (origin/astrolabe, astrolabe) Ajout URI personnes Astrolabe
* 57afa08 update: astrolabe
* 58bef28 (origin/hebdo_mro, hebdo_mro) MAJ TTL hebdo MRO

* 268d687 (origin/edd-alx, edd-alx) MAJ TTL avec 500 docs ALX EDD
* 82f744f Merge branch 'edd-alx' of
* 76db001 feat: add EDD_ALX metadafiles
* b149f9f update: edd-alx

* e2477f7 DVC
* a5d4339 update: edd-alx
* 0bb1d6d (origin/tagpi, tagpi) MAJ JSON tagpi après requête sur Dal'hy par Yann

* a15f5e6 (HEAD -> main, origin/main) feat: update person alignements
* 10c51ed feat: add tagpi_concepts files
* 1b1ade3 Nouveaux concepts métier PI
* ab4592d feat(hydroview) update hydroview rdf to obtain a unique is_related_to link
* 3188a7c chore: rename pdivision to pi-conceptsmetier
* 10b562c feat: rewrite pdivision datasets descriptions and mapping
* 39ca380 update: pdivision
* 2289529 JSON arbre métier PI
* e77936b feat(gdm): add git remote to push/pull from
* d55938c feat: fix hebdo_mro uri

* 61c8ce9 (origin/hors-prod) docs: ajout de doc pour la génération des données
* fde65d8 feat(hydroview) update rdf file with gu, geh and up link
* 600e505 feat(hydroview) add UniteProduction in hydroview mapping
* 85c1ab9 feat:hydroview update hydroview rdf with view type
* 7dad42b feat: add hydroview type mapping
* d40560e fix: barrage name is not imported
* 4dea9f9 (tag: v240905.0) feat: update rdf file for hydroview with barrage link
* 607e4cd feat: update for hydroview to add link with barrage
* 79fd8d3 feat: nettoyage des données inutilisées
```

Branches des évolutions dans les données (DEV)

Branche des données en PREPROD

Tag = données à publier en PROD

3

SPARQL Query & Update ⓘ

Groupes de production (EGF)

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2
3 SELECT DISTINCT ?turbine ?turbineLabel WHERE {
4   ?turbine a <https://hydro.data.edf.fr/ontologie/ecsh/EquipementTurbine>.
5   OPTIONAL {
6     ?turbine rdfs:label ?turbineLabel.
7   }
8 }
```

	turbine
1	https://hydro.data.edf.fr/ecsh/eqt/GLANDH.G01.TU11.TB11
2	https://hydro.data.edf.fr/ecsh/eqt/GLANDH.G01.TU21.TB21
3	https://hydro.data.edf.fr/ecsh/eqt/MALGOH.G01.TU11.TB11
4	https://hydro.data.edf.fr/ecsh/eqt/MALGOH.G01.TU21.TB21
5	https://hydro.data.edf.fr/ecsh/eqt/MALGOH.G02.TU12.TB12
6	https://hydro.data.edf.fr/ecsh/eqt/MALGOH.G02.TU22.TB22
7	https://hydro.data.edf.fr/ecsh/eqt/MALGOH.G03.TU13.TB13
8	https://hydro.data.edf.fr/ecsh/eqt/MALGOH.G03.TU23.TB23
9	https://hydro.data.edf.fr/ecsh/eqt/MALGOH.G04.TU14.TB14



2025 = SPARQL +
politique groupe mais
besoin d'une IHM dédiée



4

	Graftydro
Unité de Production : ALPES	JURA MAURIGNEN > REVERMONT > ALLEMONT (ALLEMH)
actions - site modifier	Any X,AA,AB,AC ORDERBY AA WHERE X is_instance_of Site, X code_ech=AA, X modification_date AB, X name AC, X id 72805 ALLEMONT (ALLEMH)
signets	code ECHN ALLEMH
[] Liste des préparés queries pour les signets >	siret 5520813178038
	segment GH
	coord_x_lambert_93 887084.1000000000
	coord_y_lambert_93 6559762.0000000000

Filtrer les requêtes Grafhydro

Personnes

Pour rechercher une personne sur VEOL ou référente

Personnes : référents

Recherche

Du

grande artere du canal du Forez

La Nive des Aldudes

Ruisseau du Tourmalet

La NESTE du Louron

Ruisseau du Dorinet

Dropdown_riviere

La Durance

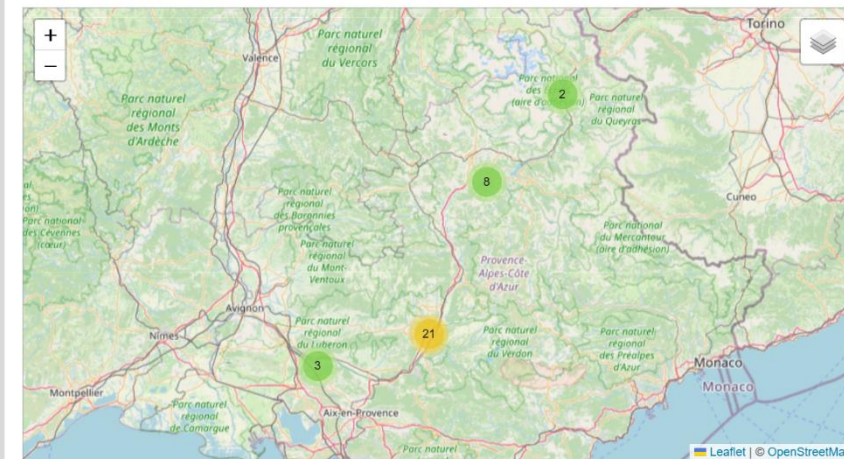
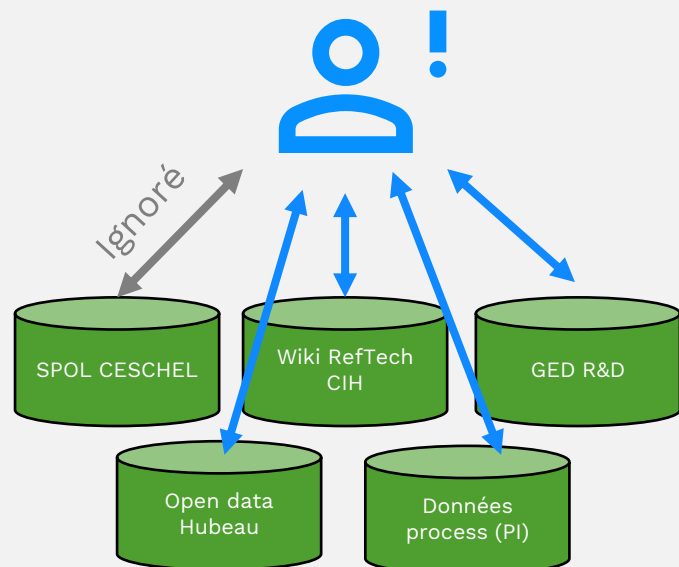


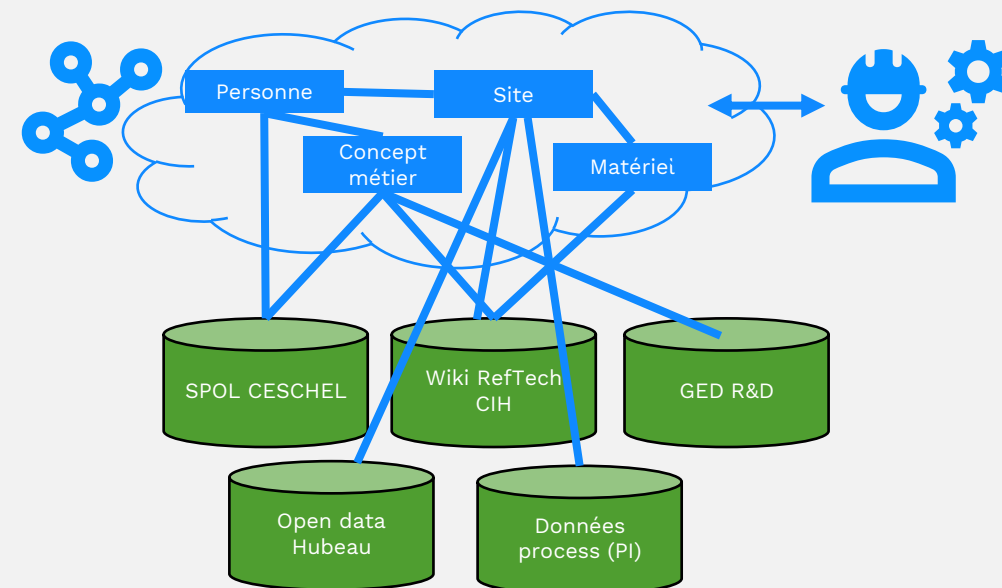
Tableau ☒ Carte

IHM de recherche avec des requêtes préparées

Résultat : connaissances visibles, identifiées, connectées



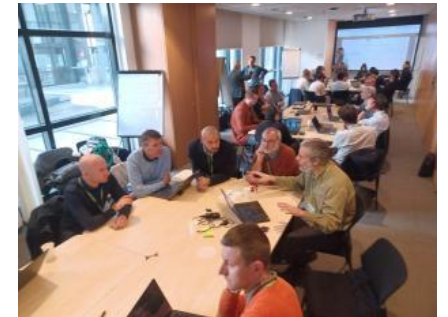
Notre problème : des données
1) ignorées
2) noyées
3) disjointes
et des applications peu liées entre elles



Notre besoin : des données
1) visibles
2) identifiées & univoques
3) connectées
et une navigation plus fluide entre applications

Impact de ce graphe de connaissances

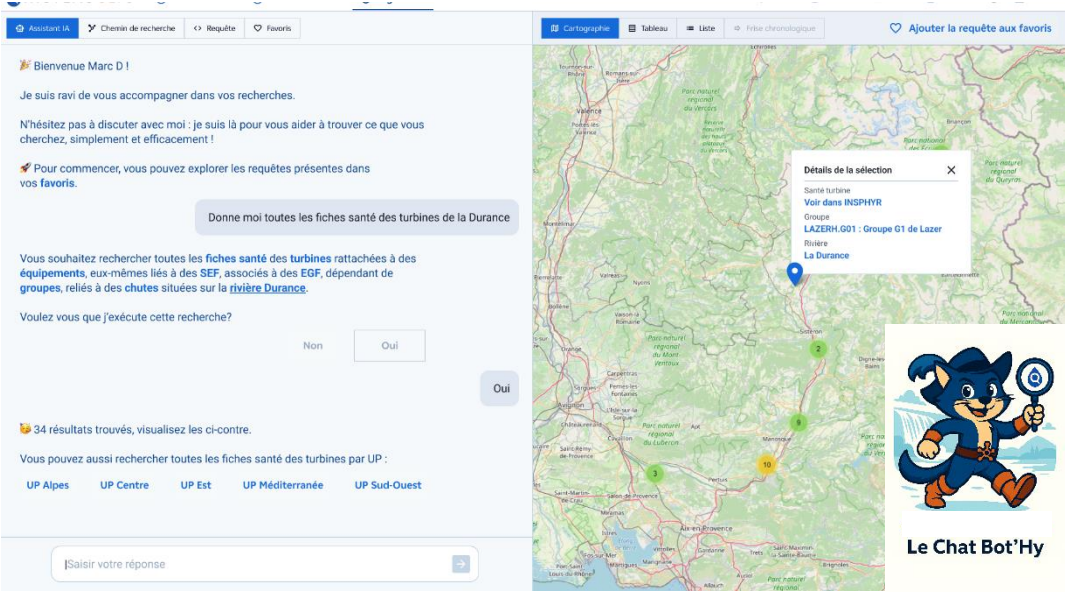
- **Maturité** = au seuil de l'industrialisation
- **Nombre personnes impactées** :
 - 140+ utilisateurs sur un chatbot qui pointe vers des pages encyclopédiques
 - 15+ utilisateurs directs sur la recherche avancée
 - 40+ personnes impliquées dans un événement base graphe interne EDF
- **Bénéfices métier** à court terme :
 - **Données organisées + alignements utiles immédiatement** pour applis pour les exploitants
 - **Requêtes avancées** pour des responsables référentiel ou maintenance :
 - « Tu viens de m'apprendre des choses sur le contenu de notre référentiel »
 - « Truc de dingue »
- **Bénéfices filière SI / data / métier** :
 - **Compréhension** : objet d'acculturation puissant (Mr Krabs !)
 - Montrer que **c'est possible** : convainquant pour opérationnels & dirigeants
 - Créer des **relations humaines**
- **Gains en synthèse** :
 - **Coût faible** : quelques dizaines de k€/an en RUN sur une techno éprouvée, peu gourmande en ressources
 - **Pérenne** : les données sont utiles sur la durée (bien rangées, avec la doc associées et accessibles par un standard ouvert)
 - **Gains XXL** : retrouver rapidement la bonne info est juste essentiel



3

Perspectives

Plus souple : nouvelle ergonomie de recherche



Bienvenue Marc D !

Je suis ravi de vous accompagner dans vos recherches.

N'hésitez pas à discuter avec moi : je suis là pour vous aider à trouver ce que vous cherchez, simplement et efficacement !

Pour commencer, vous pouvez explorer les requêtes présentes dans vos favoris.

Donne moi toutes les fiches santé des turbines de la Durance

Vous souhaitez rechercher toutes les fiches santé des turbines rattachées à des équipements, eux-mêmes liés à des SEF, associés à des EGF, dépendant de groupes, reliés à des chutes situées sur la rivière Durance.

Voulez vous que j'exécute cette recherche?

Non

Oui

34 résultats trouvés, visualisez les ci-contre.

Vous pouvez aussi rechercher toutes les fiches santé des turbines par UP :

UP Alpes UP Centre UP Est UP Méditerranée UP Sud-Ouest

Saisir votre réponse

Détails de la sélection

Santé turbine

Voir dans INSPHYR

Groupe LAZERH.G01 : Groupe G1 de Lazer

Rivière La Durance

Le Chat Bot'Hy

Interface mixte conversationnelle + navigationnelle

Services génériques sans données métier



Mistral Small 3.1

LLM exposé par EDF

$T_T \rightarrow \{ \}$

$T_T \rightarrow \{ \}$

Agent générique text2SPARQL



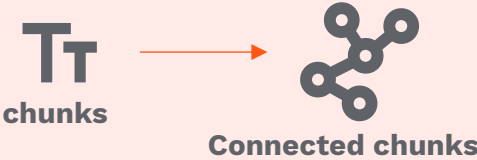
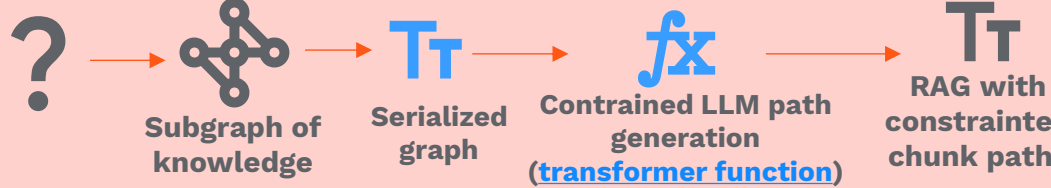
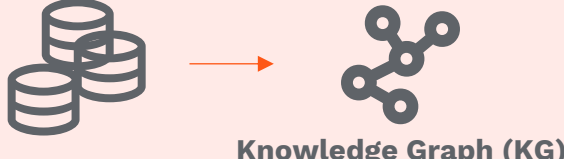
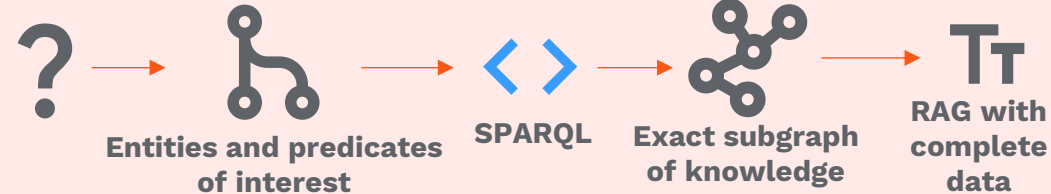
$\{ \} \leftrightarrow \{ \}$



Grafhydro ou autre base graphe

Données spécifiques au métier

Aperçu de quelques stratégies de retrieving testées à EDF

Techno	Intégration des données (embedding)	Récupération des données (retrieving)
RAG classique	 Tt chunks → 123 → Vectorized chunks	 ? → 123 → similar chunks → Tt RAG with raw chunks
GraphRAG Microsoft	 Tt chunks → Connected chunks	 ? → Subgraph of chunks → Subgraph reduction with community summaries → Tt RAG with distilled chunks
GCR (graph constrained reasoning)	 Knowledge Graph (KG)	 ? → Subgraph of knowledge → Serialized graph (Tt) → Constrained LLM path generation (transformer function) (fx) → Tt RAG with constrained chunk paths
Text-to- SPARQL	 Knowledge Graph (KG)	 ? → Entities and predicates of interest → SPARQL → Exact subgraph of knowledge → Tt RAG with complete data

Jointure de graphes

- **Travaux indépendants** de deux directions : avec des ID de type <https://optimisation.data.edf.fr/> et <https://hydro.data.edf.fr/>
- **Alignement** (bon sens métier, regex, un peu de LLM) :
 - <https://optimisation.data.edf.fr/RefOptim/referentiels/Site/Entity/HYDRAU+AVRIEUX> skos:exactMatch <https://hydro.data.edf.fr/organisation/site/AVRIEH>
 - <https://optimisation.data.edf.fr/RefOptim/referentiels/Site/Entity/HYDRAU+BAIGTS> skos:exactMatch <https://hydro.data.edf.fr/organisation/site/BAIGTH>
 - Etc...
- Le **métier** peut requêter !

Le Chat Bot'Hy

CHAT BOT GRAPHE REQUÊTE

```
1- PREFIX onto: <http://www.ontotext.com/>
2- PREFIX PPE_Property: <https://optimisation.data.edf.fr/RefOptim/referentiels/PPE/Property/>
3- PREFIX relation: <https://optimisation.data.edf.fr/RefOptim/referentiels/>
4- PREFIX Generation_Unit_Property:
  <https://optimisation.data.edf.fr/RefOptim/referentiels/Generation_Unit/Property/>
5- PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
6- PREFIX gh: <https://hydro.data.edf.fr/ontologie/grafhydro/>
7- PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
8
9- select ?geh_label ?site_hydro ?site_doaat ?ppe ?eic where {
10   ?ppe PPE_Property:CODE_ETC ?eic .
11   ?ppe relation:PPE-Site ?site_doaat .
12   ?site_doaat skos:exactMatch ?site_hydro .
13   ?gu gh:has_site ?site_hydro .
```

EXECUTER

geh_label	geh_label	geh_label	geh_label
AZUR ECRINS	AZUR ECRINS	AZUR ECRINS	AZUR ECRINS
site_hydro	site_hydro	site_hydro	site_hydro
https://hydro.data.edf.fr/organisation/site_doaat	https://hydro.data.edf.fr/organisation/site_doaat	https://hydro.data.edf.fr/organisation/site_doaat	https://hydro.data.edf.fr/organisation/site_doaat
https://optimisation.data.edf.fr/ppe	https://optimisation.data.edf.fr/ppe	https://optimisation.data.edf.fr/ppe	https://optimisation.data.edf.fr/ppe
https://optimisation.data.edf.fr/eic	https://optimisation.data.edf.fr/eic	https://optimisation.data.edf.fr/eic	https://optimisation.data.edf.fr/eic
172000000966285W	1720000000BREILHL	1720000000PAGANH4	1720000000SSDAIH7

geh_label	geh_label	geh_label	geh_label
AZUR ECRINS	AZUR ECRINS	AZUR ECRINS	AZUR ECRINS
site_hydro	site_hydro	site_hydro	site_hydro
https://hydro.data.edf.fr/organisation/site_doaat	https://hydro.data.edf.fr/organisation/site_doaat	https://hydro.data.edf.fr/organisation/site_doaat	https://hydro.data.edf.fr/organisation/site_doaat
https://optimisation.data.edf.fr/ppe	https://optimisation.data.edf.fr/ppe	https://optimisation.data.edf.fr/ppe	https://optimisation.data.edf.fr/ppe
https://optimisation.data.edf.fr/eic	https://optimisation.data.edf.fr/eic	https://optimisation.data.edf.fr/eic	https://optimisation.data.edf.fr/eic
1720000000BELVEHX	1720000000966287S	1720000000P-VARHK	1720000000VALABHA

Trois réflexions pour terminer

- **Un LLM et une base graphe sont des réseaux de données**
 - Poids d'un LLM = opaque / souvent propriétaire / jetable
 - Base graphe = explicable / souverain / capitalisable dans le temps
 - Mais IA probabiliste aide à construire la base graphe : [nazca](#) , [paraphrase MiniLM-L3-v2](#), [camembert](#) , LLM, etc...
- **Deux grandes barrières** pour une base graphe en entreprise :
 - Coût marginal décroissant => il faut attendre **une taille minimale pour capter de la valeur**
 - De **nombreuses appréhensions**
- Pour que ça fonctionne, **approche systémique** :
 - Des utilisateurs curieux
 - Des sponsors qui acceptent de prendre un peu de risque (argent, réputation, etc...)
 - Des « techos » bricolo
 - Des polymathes
 - Des connexions humaines
 - Mettre les mains dans le cambouis / tester
 - De la persévérance

Quelle est la généalogie des Skywalker ?

Copilot La recherche web est désactivée

La généalogie des Skywalker dans l'univers **Star Wars** est assez emblématique. Voici un aperçu simplifié :

Origines

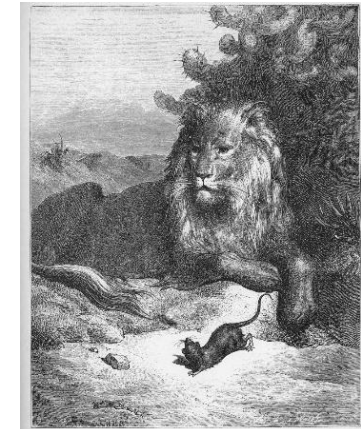
- **Shmi Skywalker** : Mère d'Anakin Skywalker. Elle vivait sur Tatooine. Anakin est né sans père biologique (conception mystérieuse liée à la Force).

Génération 1

- **Anakin Skywalker** (alias Dark Vador)
Jedi prodige, bascule du côté obscur et devient Sith.
Époux de Padmé Amidala.

Génération 2

- **Luke Skywalker** : Fils d'Anakin et Padmé. Jedi emblématique qui a vaincu l'Empire.





Merci