

Plateforme Sécurisée de Requêtes Biomédicales Fédérées

GDD, LS2N, Institut du Thorax

Contexte & motivation

Relier des **données cliniques sensibles** (qui ne peuvent quitter l'hôpital) à des connaissances publiques (ex. UniProt) est crucial pour la médecine de précision et l'étude des maladies rares. La centralisation est exclue (RGPD, secret médical). SaFE-KG explore des applications fédérées permettant des requêtes sécurisées tout en garantissant souveraineté et confidentialité.

Problème : hétérogénéité des accès

Chaque organisation applique des politiques différentes (rôles, attributs, consentements). Il faut comprendre, traduire et composer ces règles sans dégrader les performances.

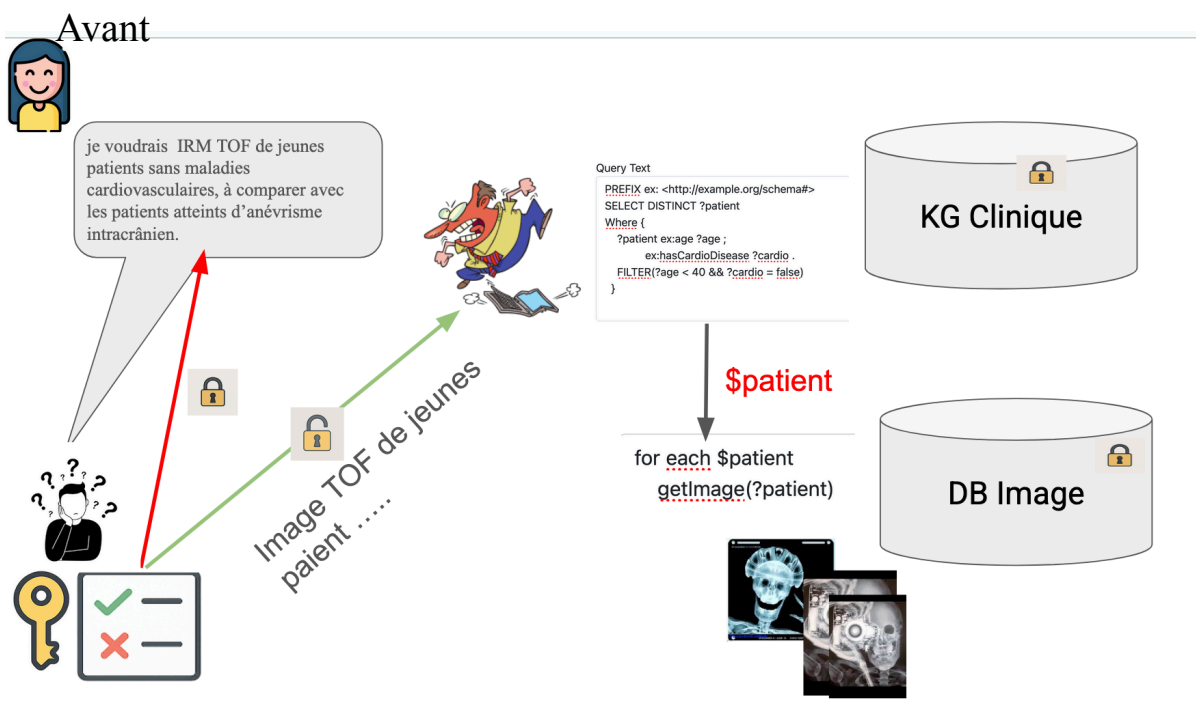
Scénarios

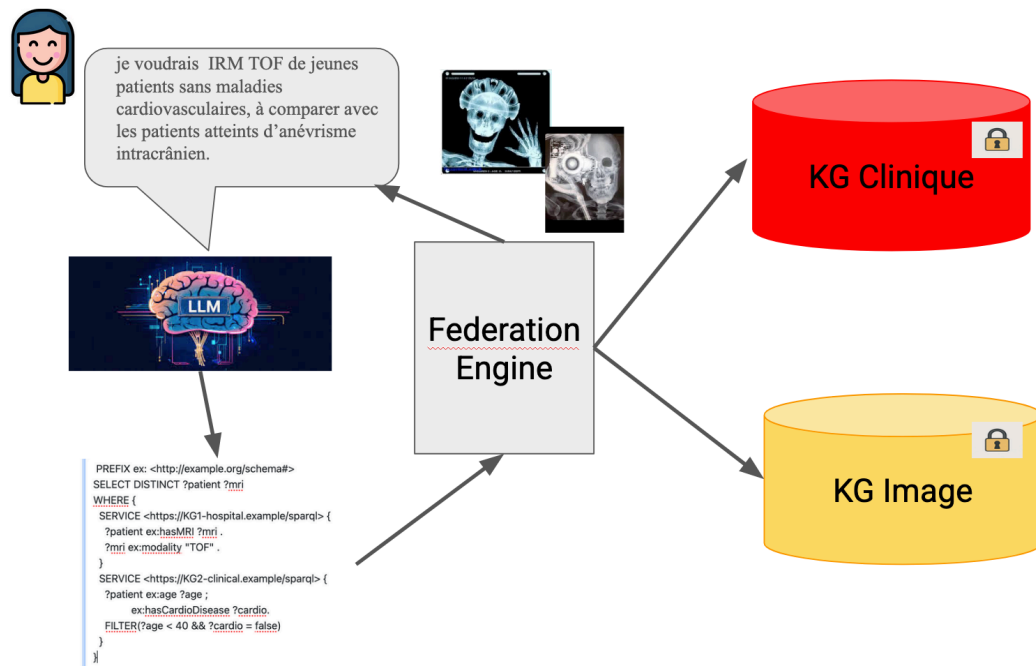
UC1 — Fédération intra-organisation

Deux projets internes (**même modèle de sécurité**).

- KG1 : métadonnées d'IRM (patients avec anévrisme intracrânien)
- KG2 : **clinique** (e.g. patients jeunes sans maladies cardiovasculaires, patients âgés avec facteurs de risque comme par exemple l'hypertension, le tabagisme)

Q1 : « Je voudrais IRM TOF de jeunes patients sans maladies cardiovasculaires, à comparer avec les patients atteints d'anévrisme intracrânien. »





Après; avec des fédérations sécurisées

Qui contrôle le droit d'accès de Q1 ?

- Côté engine (je fais confiance, c'est facile et efficace)
- **UC2 — Fédération inter-organisation**

Deux organisations avec modèles de sécurité différents :

- KG1 (registre imagerie anonymisée) : ABAC (ODRL/XACML), accès si `purpose=research` \wedge `project=SaFE-KG` \wedge `country` \in {FR,BE}, etc.
- KG2 (hôpital, clinique sensible) : RBAC (OIDC/Keycloak), rôles tels que `MedecinRadiologue`, `ChercheurAutorise`.

Q2 : « Donne-moi les IRM TOF de jeunes patients hypertendus. »

SPARQL (privacy-preserving : ?x ne sort pas de la source de données KG2)

```

PREFIX ex: <http://example.org/schema#>
SELECT DISTINCT ?y ?mri
WHERE {
  #KG1 : IRM TOF (pseudonyme patient = ?y), soumis à ABAC (purpose/project/etc.)
  SERVICE <https://KG1-imaging.example/sparql> {
    ?y ex:hasMRI ?mri .
    ?mri ex:modality "TOF" .
  }

  #KG2 : confirmation clinique côté hôpital (RBAC) sans exposer ?x ni l'âge
  SERVICE <https://KG2-clinical.example/sparql> {
    FILTER EXISTS {
      GRAPH <urn:protected> {

```

```
?x_local ex:pseudo ?y ;
      ex:age ?age_local ;
      ex:hasHypertension true .
FILTER(?age_local < 40)
}
}
}
}
```

Ici, seul l'identifiant des données d'imagerie anonymisée est renvoyé ; ni `?x_local` (sensible) ni `?age_local` (sensible) ne sortent de KG2.

Qui contrôle l'exécution de Q2 ?

- Côté endpoint (dans la clause SERVICE)

Objectifs du projet

- Plateforme démo couvrant UC1 & UC2.
- Fédération : déployer un moteur (Comunica, FedX, ou FedUP).
- Sécurité :
 - KG2 (RBAC) : Keycloak (OIDC), rôles, scopes ; graphe protégé urn:protected.
 - KG1 (ABAC) : règles ODRL/XACML (purpose/project/pays/temps).
 - Propagation du token : le client transmet le JWT à chaque SERVICE.
- Interface & API : REST + mini UI pour exécuter Q1/Q2, afficher provenance et latence.
- Observabilité : logs par source/graph, métriques latence & complétude.

Périmètre technique

- Architecture : diagrammes (modules endpoints, fédérateur, authN/Z, API/UI).
- Déploiement : 2 endpoints RDF (Fuseki/Comunica), datasets fournis (imagerie/clinique).
- Politiques : RBAC (Keycloak), ABAC (ODRL simple), option : règle SPARQLLM pour valider des droits.
- Évaluation : temps de réponse, taux de complétude, comparaison avec/ sans filtres de sécurité.

Livrables


- Code source (GitHub) et instructions de déploiement.
- Jeux de données RDF (nettoyés) .
- Rapport d'environ 10 pages : conception, implémentation, résultats, limites et perspectives.

Encadrements :

Hala Skaf-Molli <hala.skaf@univ-nantes.fr>, Pascal Molli <pascal.molli@univ-nantes.fr>, Alban Gaignard <alban.gaignard@univ-nantes.fr>, Gabriela Montoya, gabriela.montoya@univ-nantes.fr,

Références:

- [1] Molli, Pascal, Hala Skaf-Molli, Sébastien Ferré, Alban Gaignard, and Peggy Cellier. 2025. "SparqLLM: Retrieval-Augmented SPARQL Query Processing." Pp. 1–5 in *ESWC 2025 - 22nd European Semantic Web Conference*. Portoroz, Slovenia. [[pdf](#)] [[online poster](#)] SPARQLLM repository: <https://github.com/GDD-Nantes/SPARQLLM>

- [2] J. Aimonier–Davat, M.H. Dang, **Pascal Molli**, Brice Nédelec, **Hala Skaf–Molli**.  FedUP Querying Large Scale Federations of SPARQL Endpoints The ACM Web Conference (WWW '24), Singapore. [10.1145/3589334.3645704](https://doi.org/10.1145/3589334.3645704). [hal-04538238](https://hal.archives-ouvertes.fr/hal-04538238)
- [3] Bonatti, P. A., et al. [Access and Usage Control for Federations of Knowledge Graphs](#). Are Knowledge Graphs Ready for the Real World? Challenges and Perspective (2024): Dagstuhl, 14, 2, 2024
- [4] Kirrane, S., Mileo, A. & Decker, S. [Access control and the resource description framework: A survey](#). Semantic Web. 8, 311-352 (2017)
- [5] Schwarte, A. et al. [FedX: Optimization Techniques for Federated Query Processing on Linked Data](#). ISWC (1). 7031 pp. 601-616 (2011)