

GARANCE : une solution pour la diffusion des référentiels sémantisés des Archives nationales de France SPARNA

Florence Clavaud (Lab des AnF) et Thomas Francart (Sparna)

Objectifs du projet :

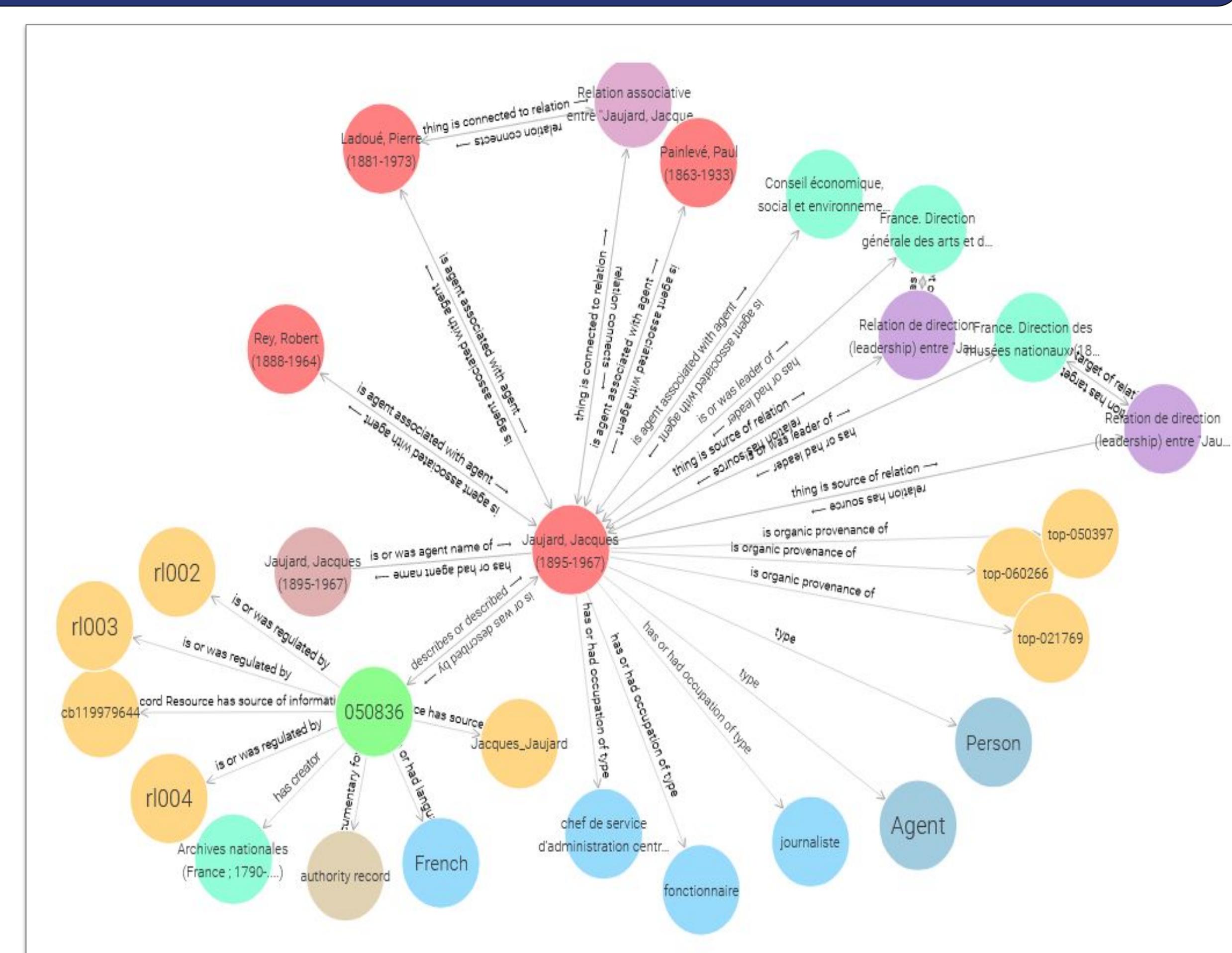
- Générer un unique graphe de connaissances RiC-O 1.1. à partir des données de référence produites et maintenues depuis des décennies par les AnF sur les entités de contexte des archives que l'institution conserve
- Doté ces données d'une interface de consultation et de recherche performante, ergonomique et facile à maintenir ; les rendre accessibles aux machines. Bref, FAIRiser ces données !

La génération du graphe



Dans les 16019 notices de producteurs validées au 21 janvier 2025	
Relations encodées entre les notices d'autorité (entre les entités décrites)	Nombre de relations
Relations hiérarchiques	9956
Relations temporelles	4585
Relations associatives	6901
Relations familiales	606
Total	22048
Relations d'identité (avec des notices extérieures aux AnF, à la BnF ou dans la base ISNI)	3438
Relations avec des instruments de recherche des AnF	37090

Une relation qui existe entre deux entités données est comptée une seule fois



Le point de départ : des fichiers XML de structure variable, parfois très pauvre ; souvent des listes de noms ; des redondances ; pas de liens explicites entre lieux.

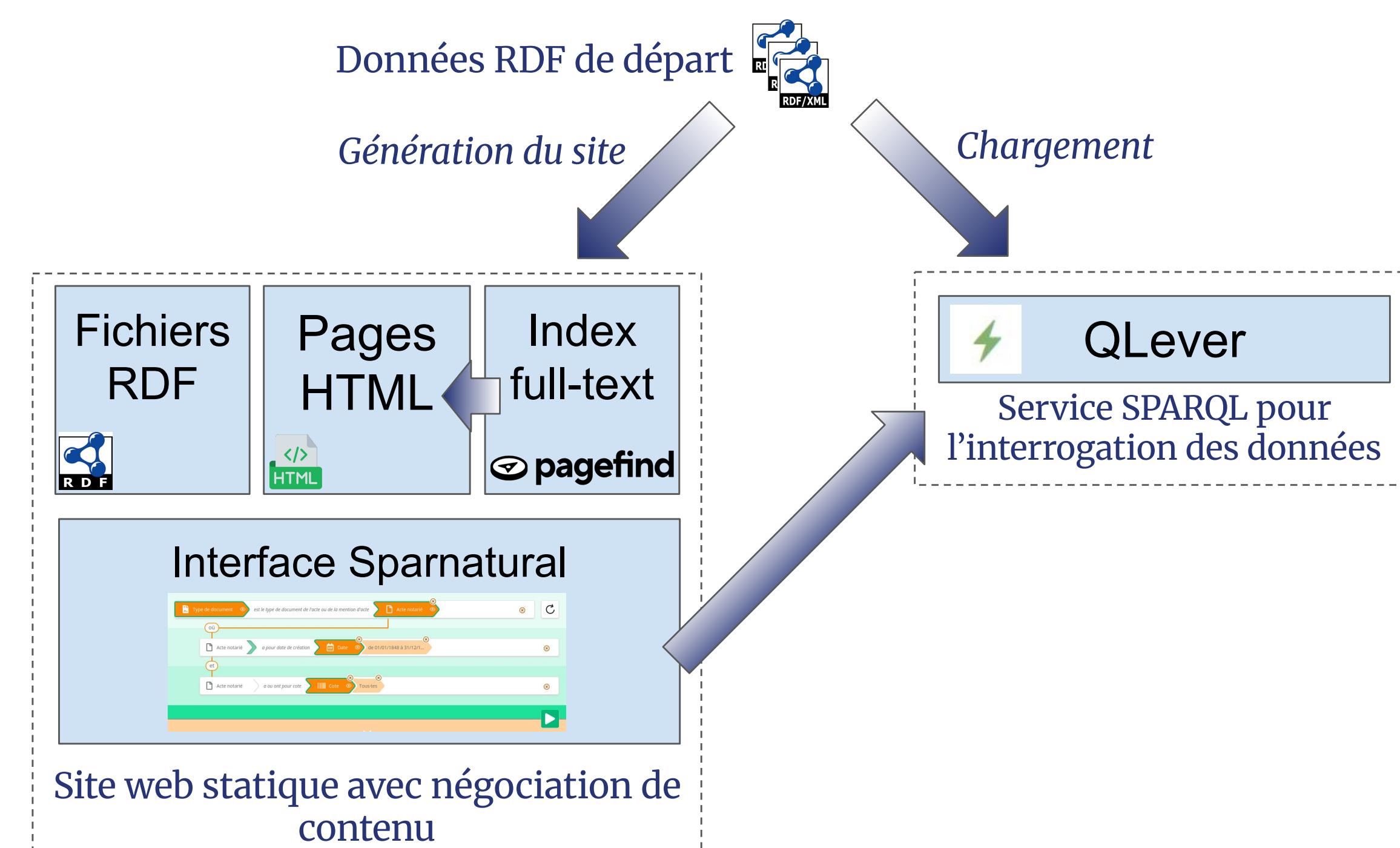
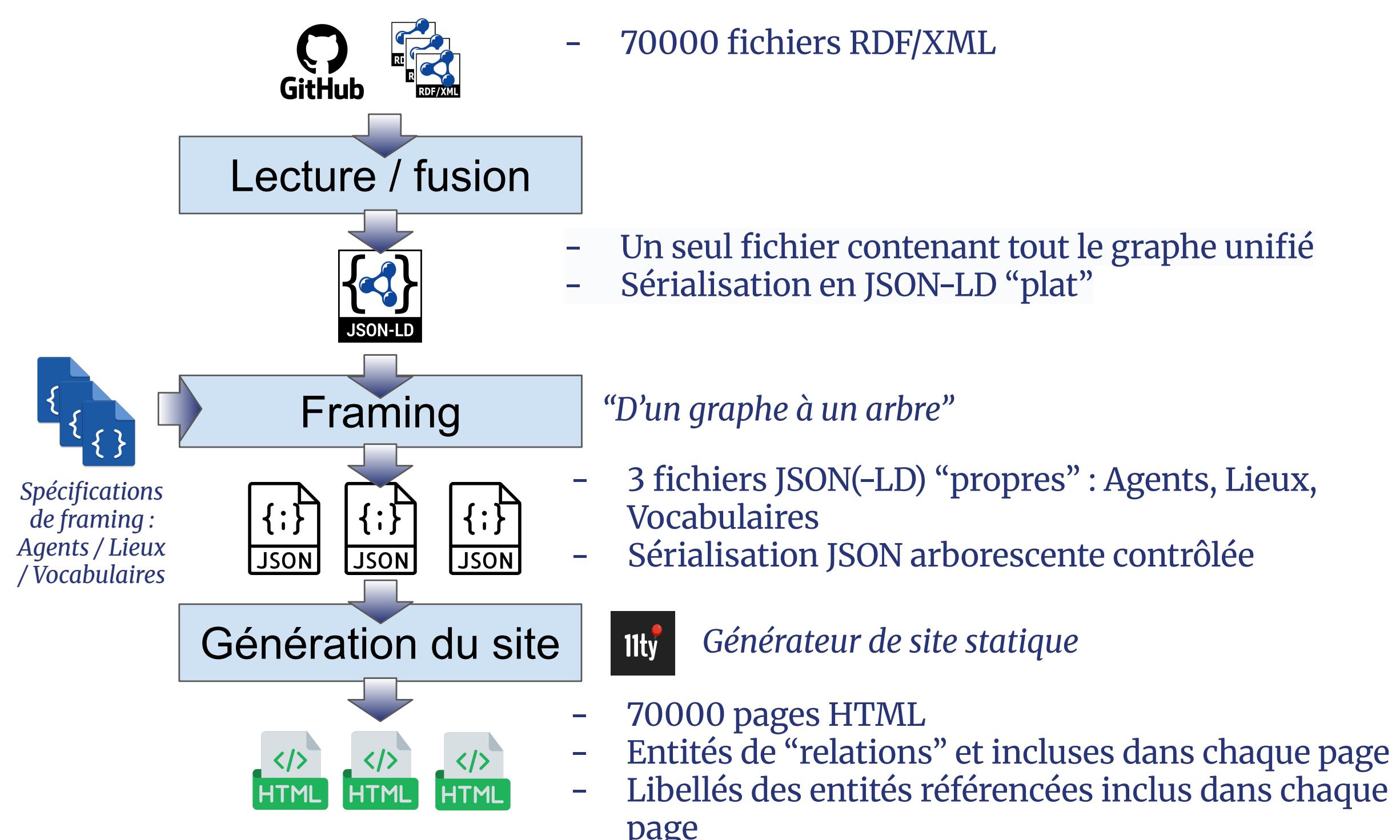
Mais en fait, déjà un graphe sous-jacent !

- Travail commencé en 2021 (première release sur GitHub en mai 2022 : <https://github.com/ArchivesNationalesFR/Referentiels>)
- Nouvelle phase démarrée en 2023
- Plusieurs workflows de transformation, pour tirer le meilleur parti de l'existant et l'enrichir

Beaucoup de changements :

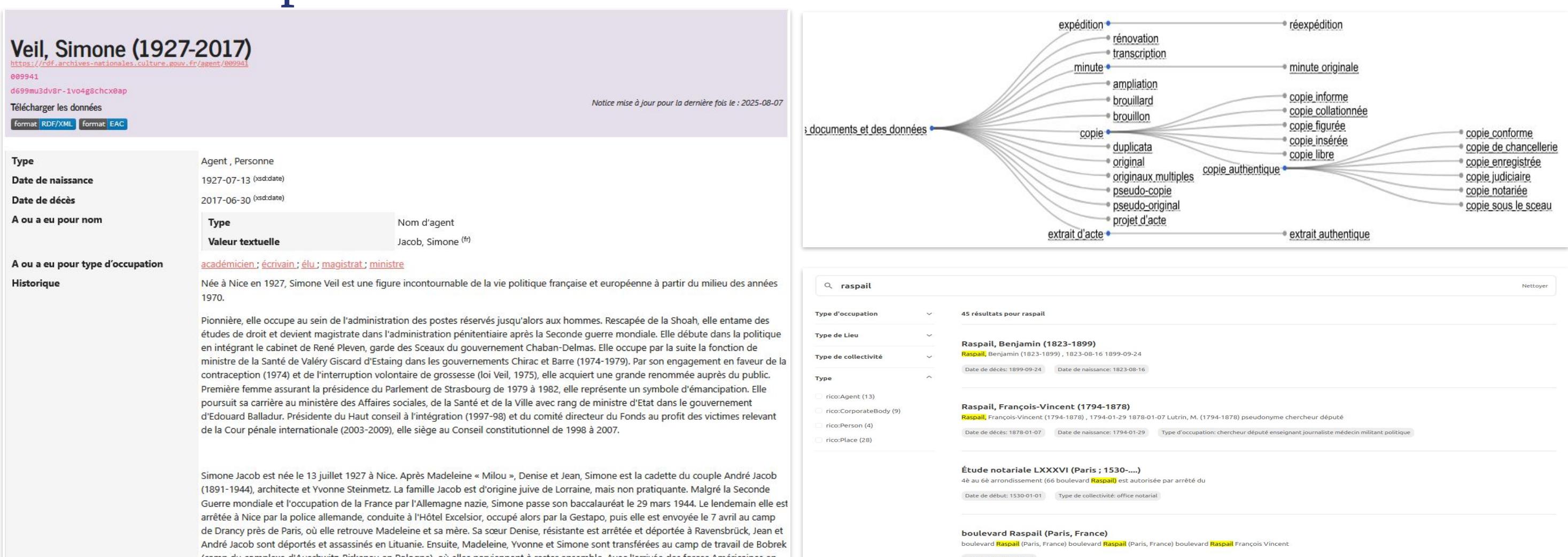
- Vocabulaires : conversion en SKOS/RiC-O, avec métadonnées et nouveaux concepts
- Agents : 255 notices dédoublonnées
- Lieux : beaucoup de travail ! circonscriptions administratives françaises et voies de Paris alignées avec les données IGN, INSEE et Mairie de Paris ; des descriptions enrichies manuellement → typage des lieux, liens entre lieux, coordonnées géographiques, notes historiques, listes d'événements...

L'application web GARANCE : architecture



Premiers résultats

La partie statique du site web est quasiment achevée : pages pour les entités, URIS cliquables, liens de téléchargement, fonctionnalité de recherche rapide.



Feuille de route fin 2025-2026

- Sur les données :
 - Addition de concepts aux vocabulaires
 - Mise à jour des données sur les circonscriptions administratives françaises
 - Fin du travail sur les voies de Paris
 - Ajout de liens de localisation pour les édifices
 - Et bien sûr, la mise à jour de toutes les données à partir d'un nouvel export du SI

Sans oublier :

- Un travail sur les fichiers source (pour corriger des problèmes récurrents de qualité, autant que possible)
- Plus tard, l'ajout de nouveaux fichiers d'agents

- Sur l'interface :
 - Configuration complète de la fonctionnalité de recherche rapide
 - Import des données dans une base de graphes (très probablement QLever, un outil open source récent très performant)
 - Intégration d'une interface de recherche avancée permettant d'explorer l'ensemble du graphe, à l'aide de l'éditeur visuel de requêtes SPARQL Sparnatural
 - Fonctionnalités d'accès pour les machines : SPARQL end point, négociation de contenu
 - Documentation sur la structure du graphe, produite à partir de son profil d'application spécifié en SHACL
 - Un peu de data visualisation et si possible, intégration de Sparnatural AI

→ Mise en ligne officielle du site prévue fin 2025 ou début 2026