

Le portail FranceArchives : de l'agrégation d'instruments de recherche à la constitution d'un réservoir de données

28 janvier 2020 - Les métadonnées archivistiques en transition

Manonmani Restif

Camille Monnier

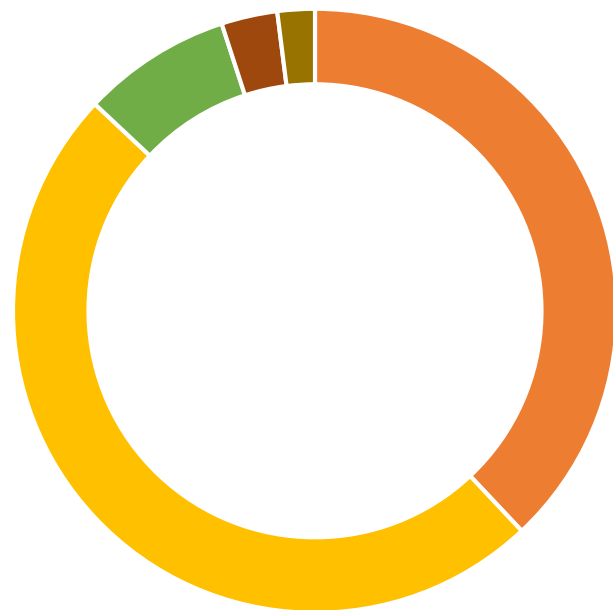
Service interministériel des Archives de France

FranceArchives?

Portail interministériel, agrégateur des métadonnées archivistiques françaises.



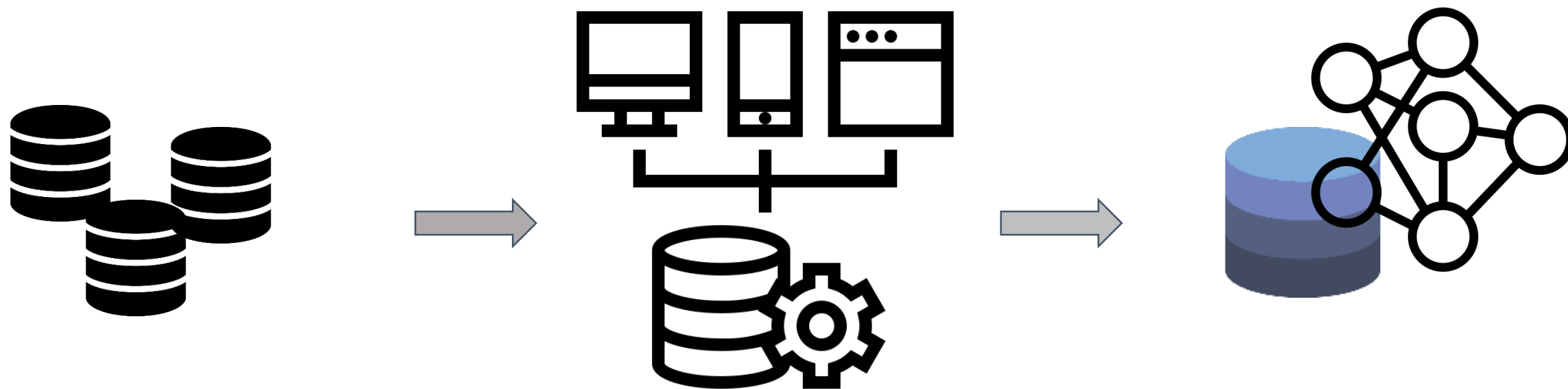
Répartition des données par type de service



- 69 services
- 47 000 inventaires
- 9 million d'unités de description



De l'agrégation au réservoir de données

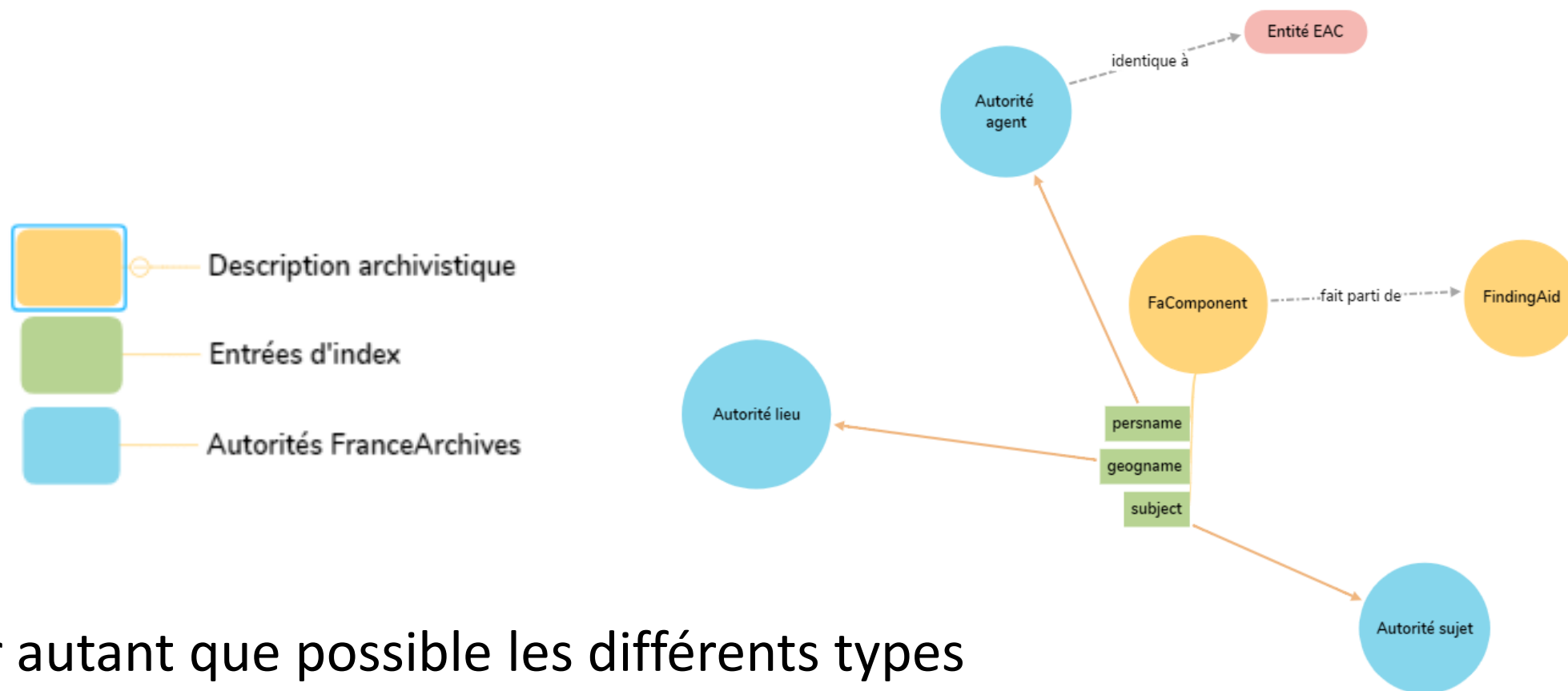


Permettre l'interrogation complexe et profonde, la connexion de services web, la réutilisation des métadonnées archivistiques françaises.

Plan

1. Quel modèle de données actuellement?
2. Traitement des agents, des lieux, des thèmes

Modèle de données actuel



Éclater autant que possible les différents types d'informations, vers un modèle plus orienté entité relation.

Opportunités des alignements

- A l'échelle du portail : construire, nettoyer, rationaliser des points d'accès pour mieux les valoriser
- A l'échelle du web de données : entités qui sont les points de jonction entre les données des diverses bases de connaissances

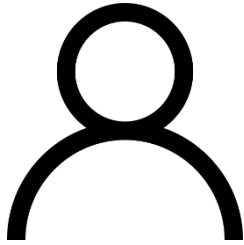
{ BnF



- persname
- geogname
- subject



Scripts d'alignements

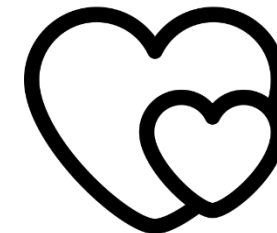


Normes exploitée : NF Z 44-061
Autorités alignées : 14%



Norme exploitée : NF Z 44-081
Autorités alignées : 50%

Conclusion alignements et normalisation =



Des URI pérennes?



<https://francearchives.fr/fr/findingaid/68e919893b41489542908f1c95647e633f9f13fe>

- Identification des unités de description archives : pérennité tributaire des identifiants délivrés par les services partenaires, producteurs des données.
- Identification des autorités : pérennité des URI produites sur le portail, lors de la fusion de deux autorités, une redirection http est appliquée.

Vers une montée en qualité des données?



Mutualisation de traitements, reversement des données traitées, construction d'un réservoir de données cohérent.

Perspectives

- Lancement d'un groupe de travail description/indexation



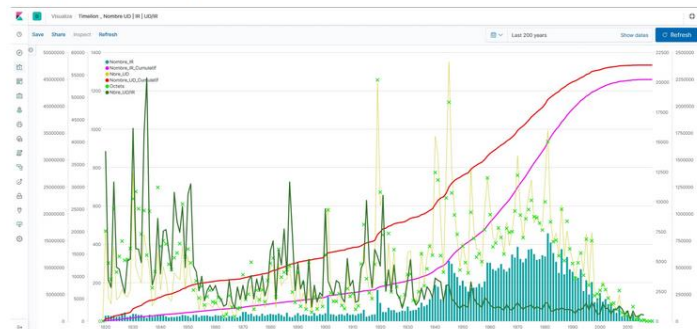
Le pilotage par la métadonnée

La visualisation améliore notre compréhension du stock d'information. Dès lors, elle peut être envisagée comme outil d'audit qualité, de traçabilité des flux ou de pilotage de l'activité.

Par exemple, l'archiviste peut chercher les notices qui ne sont pas associées au référentiel des « types de document » en isolant celles pour lesquelles la valeur de l'indicateur est « VIDE » :



Une analyse statistique permet d'observer des tendances sur les flux entrants.



- OpenRefine : compléter les alignements
- Etude en cours : implémentation de Kibana sur le portail (outil de visualisation et d'exploration de données basé sur Elastic Search)

Merci !

Manonmani Restif

Camille Monnier

Service interministériel des archives de France

Nous contacter :

Francearchives.siaf@culture.gouv.fr

Source pictogrammes : Flat Icon <https://www.flaticon.com>