

FranceArchives et le Web Sémantique

Chloë Fize (Service interministériel des Archives de France), Elodie Thiéblin (Logilab)

Article sous licence CC-BY 4.0 rédigé à l'occasion des Journées du Logiciel Libre 2021

Présentation générale de FranceArchives

Qu'est-ce que c'est les archives?

Selon le code du patrimoine, les archives sont *l'ensemble des documents, y compris les données, quels que soient leur date, leur forme et leur support matériel, **produits ou reçus** par toute personne physique ou morale, et par tout organisme public ou privé, **dans l'exercice de leur activité**. Ces documents sont soit conservés par leurs créateurs ou leurs successeurs pour faire la preuve d'un droit ou d'un événement, soit transmis à l'institution d'archives compétente en raison de leur valeur historique...*



Elementaire non ?

Les archives sont plus simplement des documents, divers et variés !



Dans notre imaginaire, en général, les archives ne sont que de vieux papiers poussiéreux, rédigés dans des langues obscures et à première vue indéchiffrables, jalousement cachés au fond de sombres et froids placards... Et dans le pire des cas, elles sont cachées dans les sous-sols ou les greniers... Un petit peu comme ça :



Image libre de droit Pixabay

Mais détrompez-vous, les documents d'archives sont partout et peuvent être bien plus agréables à admirer que vous ne l'imaginez. En France, plus de **4 000 kilomètres linéaires** d'archives sont conservés dans plus de **500 services** d'archives nationales, régionales, départementales et municipales sans compter les services d'archives privés (entreprises, associations, etc). On regroupe les documents en **fonds**.

Voici des fonds, bien proprement rangés dans leurs cartons... C'est quand même plus sympathique ?



Photographie des grands dépôts des Archives nationales, site de Paris, magasin du Trésor des Chartes (série J) ©Chloe FIZE, 2017

Mais comment s'y retrouver ? Comment savoir que LE document que je recherche est bien dans cette boîte nommée simplement par des lettres et des chiffres ? Pour cela, il faut les décrire et ensuite les communiquer à qui veut les consulter. Car la vocation première des archives, c'est que tout le monde puisse en effet les consulter... **Oui, oui, y compris vous !**

Ressources en ligne des archives

Les archivistes ont toujours cherché à exploiter les technologies les plus en pointe pour communiquer à tous les publics les documents qu'ils conservent : microfilms, numérisation, site web... Et de fait, depuis plus de **20 ans**, les services d'archives mettent à disposition de tous des inventaires avec ou sans documents numérisés, consultables directement en ligne sur plus de **300 sites internet**.



Copie d'écran du site des archives départementales des Vosges, mars 2021.

La raison? Que tout le monde puisse y avoir **accès** ! Eh oui, les archives c'est comme la bonne humeur, ça se communique, et par tous les moyens ! ✨

C'est là toute la vocation du **portail FranceArchives** :

- Permettre aux chercheurs, étudiants, curieux, amateurs de généalogie ou qui que vous soyez, de repérer les ressources de nombreux services d'archives publics et privés pour, dans un second temps, les consulter sur les sites web ou dans les salles de lecture de ces services.
- Valoriser les fonds et services d'archives des quatre coins de la France.
- Mettre à disposition des ressources archivistiques professionnelles ou des textes de loi.

Comment y accéder?

L'accès et la recherche sur le portail sont construits pour être les plus intuitifs possible et pour mener le chercheur, amateur ou expert à trouver son bonheur dans cette caverne aux merveilles... Suivez le guide !

FranceArchives

FranceArchives : qu'est-ce que c'est ?

Le portail est porté par le Ministère de la Culture et a été mis en ligne au mois de **mars 2017**. Il est géré et maintenu par le service interministériel des archives de France (SIAF).

FranceArchives en chiffres :

Au mois de **mars 2021**, la **105ème** convention d'adhésion au portail FranceArchives a été signée. Vous pouvez donc consulter les fonds de 2 ministères, 4 services à compétences nationales (Archives nationales, Archives nationales du monde du travail, Archives nationales d'Outre-Mer et la Médiathèque de l'architecture et du patrimoine), 63 archives départementales, 19 archives municipales, 13 établissements publics, 4 associations ou entreprises.

Plus de **57 000** instruments de recherche sont consultables et réutilisables. Ils contiennent près de **13 000 000** de descriptions.

FranceArchives : comment ça fonctionne ?

Effectuer une requête simple



FranceArchives
PORTAIL NATIONAL DES ARCHIVES

ACCÈS PROFESSIONNELS

MENU

Partez à la découverte de 15 siècles d'archives

Que trouverez-vous sur FranceArchives ?

Georges Clemenceau

x | Q

?

Tout en s'aidant de l'autocomplétion

The screenshot shows a search bar with the text 'Georges Clemenceau' and a search icon. Below the search bar, there are two main categories: 'Lieu' and 'Personne ou institution'. The 'Lieu' category lists several locations with the number of documents associated with each. The 'Personne ou institution' category lists individuals and institutions related to Georges Clemenceau.

Lieu	
Georges-Clemenceau, avenue	54 documents
Nice (Alpes-Maritimes, France) -- Georges-Clemenceau (avenue)	32 documents
Clemenceau -- Quai Georges	18 documents
Perros-Guirec (Côtes-d'Armor) -- Boulevard Georges-Clemenceau	7 documents
Pau (Pyrénées-Atlantiques, France) -- Place Georges-Clemenceau	5 documents
Bel-Esbat (Saint-Vincent-sur-Jard) : maison de Georges Clemenceau	4 documents
Guingamp (Côtes-d'Armor) -- Boulevard Georges-Clemenceau	3 documents

Personne ou institution	
Clemenceau, Georges (1841-1929)	675 documents
Clemenceau, Georges Cloupet Max	1 document
Georges-Clemenceau (hôpital)	1 document
COLLEGE GEORGES CLEMENCEAU (LYON)	1 document

Et enfin affiner sa recherche grâce aux facettes

AFFINER VOTRE RECHERCHE :

The screenshot shows a search refinement interface with a search bar and several facets. The search bar contains the text 'Rechercher dans ces résultats'. The facets are: États, Numérisé, Types de ressources, Lieux de conservation, Période chronologique, and Producteurs. Each facet has a dropdown arrow and a list of options with their respective counts.

Rechercher dans ces résultats

États	
non	627
oui	47

Types de ressources	
Documents d'archives	674
Inventaires d'archives	1

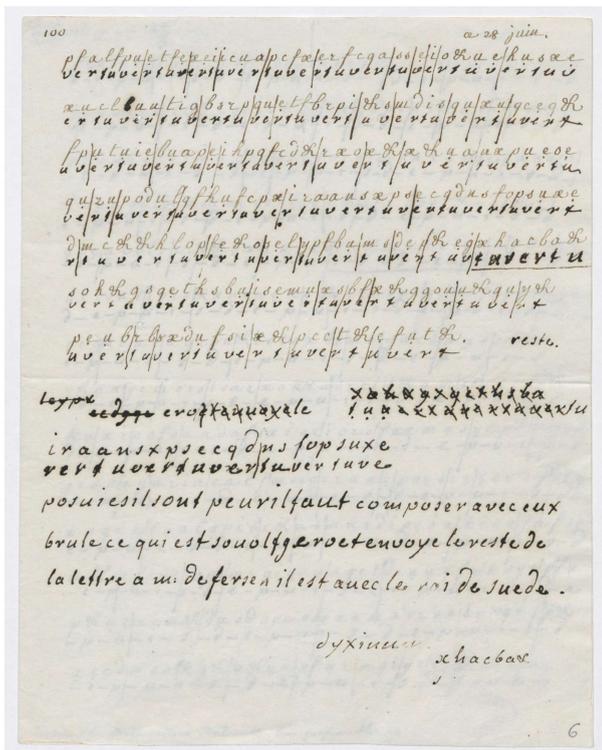
Lieux de conservation	
1700 - 1799	2
1800 - 1899	10
1900 - 1999	254

Producteurs	
-------------	--

Les archives et leurs métadonnées

Vous avez réussi à trouver votre bonheur ? Parfait ! Mais vous n'avez pas trouvé de documents d'archives numérisés ? C'est normal, seulement 5% des archives conservées en France sont numérisées. Le reste est accessible en ligne uniquement grâce aux descriptions ou métadonnées et doit être consulté dans les salles de lecture des services qui les conservent.

Le document d'archives numérisé est une photographie du document. Sur FranceArchives, vous pourrez trouver certains documents numérisés, à l'image du célèbre exemple qui suit :



Lettre cryptée échangée entre la reine Marie-Antoinette et le comte Axel de Fersen,

datée du 28 juin 1791 - AN440AP/1, dossier 1, pièce 6, page 1.

Mais vous ne trouverez la plupart du temps que des métadonnées, à savoir le contenu de la lettre, son auteur, son destinataire, sa date, des remarques sur sa forme et/ou son fond. C'est là qu'est la différence entre un document numérisé et une métadonnée numérique.

Vous voulez un exemple? Très bien, les métadonnées numériques c'est... ça :

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/
<responseDate>2021-02-25T13:24:04Z</responseDate>
<request verb="ListRecords" metadataPrefix="ead">https://archives.calvados.fr/oai-pmh</request>
<ListRecords>
  <record>
    <header>
      <identifier>32a5f38c-8fa5-41c9-9f67-9cac6211043e</identifier>
      <datestamp>2020-04-09T17:22:43Z</datestamp>
    </header>
    <metadata>
      <default:ead audience="external" xsi:schemaLocation="urn:isbn:1-931666-22-9 http://www.1
      <default:eadheader findaidstatus="unverified-partial-draft">
        <default:eadid>FRAD014_001654</default:eadid>
      <default:filedesc>
        <default:titlestmt>
          <default:titleproper>2683W - Université de Caen</default:titleproper>
          <default:subtitle>
            Reclassement de plusieurs versements de l'Université de Caen.
          </default:subtitle>
          <default:author>Didier Bouquin</default:author>
        </default:titlestmt>
      <default:publicationstmt>
        <default:publisher>Archives du Calvados</default:publisher>
      <default:address>
        <default:addressline>61, rue de Lion-sur-Mer</default:addressline>
        <default:addressline>14000 Caen</default:addressline>
      </default:address>
      <default:date normal="2017">2017</default:date>
      </default:publicationstmt>
    </default:filedesc>
  </record>
</ListRecords>
```

Avouez que ça vous fait rêver ! Eh bien cette belle lettre que vous avez vu précédemment peut être transformée en données et donc ressembler à... ces lignes en couleurs et comportant plein d'informations pouvant paraître incompréhensibles. N'ayez pas peur on va tout vous expliquer.

Mais alors d'où viennent les métadonnées et à quoi ressemblent-elles?

Les services d'archives décrivent leurs fonds dans des *instruments de recherche*. Ce sont ces instruments qui sont mis en ligne sur FranceArchives et consultables par tout un chacun. Ils ne donnent pas accès au document numérisé, comme nous l'évoquions plus haut, mais à sa *description*. Ces éléments permettent de décrire avec précision les documents que l'on va retrouver dans le fonds et donc de répondre à vos multiples questionnements sans sortir le document de son joli carton de protection. Quand vous consultez une notice sur le portail, il vous est ensuite possible d'accéder directement au site web du service qui conserve le document décrit grâce au bouton *Accéder au site*.

Alors où sont passées les données que nous avons vu plus haut ? Elles sont là, partout, juste sous vos yeux. Vous les voyez ? Regardez...

Voici ce que vous voyez lorsque vous requêtez FranceArchives :

Gard

Snr Cévennes, 1989. Haribo Ricqles zan, 1991. Solitech, 1992. Gravure d'azur, 1992. Seac Codognan, 1992....

 Date : 1971 - 1992

 Télécharger la cote

Titre : Gard
Snr Cévennes, 1989. Haribo Ricqles zan, 1991. Solitech, 1992.
Gravure d'azur, 1992. Seac Codognan, 1992.

Inventaire d'archives : [Premier ministre ; Délégation à l'aménagement du territoire et à l'action régionale \(DATAR\) - Délégation interministérielle à l'aménagement ...](#)

Sujet : [entreprise, politique économique](#)

Fonctions : [aide de l'etat, aménagement du territoire](#)

Institution de conservation :
Archives nationales

[Contact](#) [Accéder au site](#)



Voici ce que nous traitons :

```
<?xml version="1.0" encoding="UTF-8"?>
<ead xmlns:xlink="http://www.w3.org/1999/xlink">
  <eadheader dateencoding="iso8601" langencoding="iso639-2b" repositoryencoding="iso15511">
    <eadid url="https://www.siv.archives-nationales.culture.gouv.fr/siv/IR/FRAN_IR_019511">FRAN_IR_019511
  </eadid>
  <filedesc>
    <titlestmt>
      <titleproper>Premier ministre ; Délégation à l'aménagement du territoire et à l'action régionale
        (DATAR) - Délégation interministérielle à l'aménagement et à la compétitivité du territoire (DIACT)
        (1971-1992)</titleproper>
      <subtitle>Répertoire (20020491/1-20020491/52)</subtitle>
    </titlestmt>
    <publicationstmt>
      <publisher>Archives nationales</publisher>
      <date calendar="gregorian" era="oe" normal="2002-01-01/2002-12-31">2002</date>
    </publicationstmt>
  </filedesc>
  </eadheader>
  <unitid>
    <extptr href="https://www.siv.archives-nationales.culture.gouv.fr/siv/UD/FRAN_IR_019511/d_5_3"/>
  </unitid>
  <unittitle>
    <emph render="bold">Gard</emph>
  </unittitle>
  <unittitle>Snr Cévennes, 1989. Haribo Ricqles zan, 1991. Solitech, 1992. Gravure d'azur, 1992.
    Seac Codognan, 1992.</unittitle>
  </did>
```

Les métadonnées sont bien là. Cet intermédiaire invisible pour l'utilisateur permet de mettre en forme de façon lisible les informations.

Mais alors comment êtes-vous parvenu jusqu'à ce résultat ? Comment, parmi les milliers de résultats, les seuls qui vous ont été proposés ont-ils été choisis ?

Vous avez vu les nombreux liens cliquables en bleu et soulignés que vous trouvez un peu partout... Ce souvent des noms de lieux, de personnes ou des thèmes, eh bien tous ces

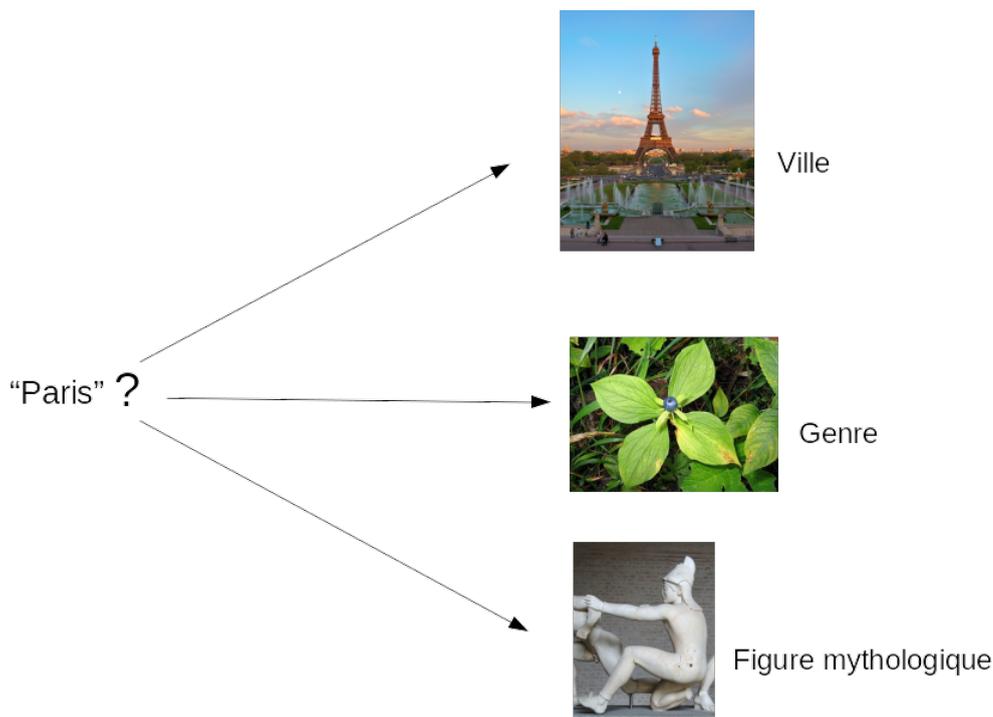
termes sont des **autorités** qui sont extraites des instruments de recherche pour être groupées avec leurs semblables et alignées sur de plus gros portails de données tels que Data.BnF ou Wikidata.

Quelle en est l'utilité ?

L'identification de ces ressources permet de lever l'ambiguïté sur un nom : être sûr qu'on parle bien de la même personne ; ou relier plusieurs noms à une même ressource.

Prenons un exemple : "**Paris**" c'est à la fois le nom de la capitale française, d'un genre de plante et d'une figure mythologique : 3 ressources différentes (donc 3 URI) portent le même nom. Paris a beau être une ville fleurie et mythique, il est tout de même important de pouvoir différencier tous ces éléments lors d'une requête.

Mais certains éléments peuvent présenter un cas inverse : l'autrice Dominique Aury est également connue sous plusieurs pseudonymes très différents les uns des autres : Anne Cécile Desclos et Pauline Réage. Ici il y a donc une seule ressource (1 URI) qui porte ces 3 noms. Pour que vous trouviez toujours le même résultat, ces 3 noms doivent être tous rattachés à la même personne et ne pas figurer comme étant 3 éléments différents et distincts.



Cette différenciation ou ce regroupement est réalisé grâce à l'URL (vous savez les liens incompréhensibles écrits dans votre barre de navigation... Eh bien en réalité ils ont un sens !)

L'utilisation d'URL pour identifier les ressources est la base du Web sémantique (ou Web de données).

- > <https://www.wikidata.org/wiki/Q90> (Capitale de France)
- > <https://www.wikidata.org/wiki/Q162121> (Genre de Plante)
- > <https://www.wikidata.org/wiki/Q167646> (Figure mythologique)

3 "liens" différents, pour 3 thèmes complètement différents, mais qui sont tous requêtables avec le même mot.

Le Web sémantique

Le Web sémantique a été inventé par Tim Berners-Lee, le fondateur du Web. Son idée est d'utiliser les technologies du Web pour y faire transiter non seulement des documents (pages Web, comme c'est le cas aujourd'hui) mais aussi des données.

Comme dans le Web que nous connaissons tous, le protocole HTTP visible dans l'URL (on vous avait dit que ça avait un sens!) est utilisé pour faire voyager les données. Les ressources quant à elles sont identifiées par des URL (*Uniform Resource Location*) appelées aussi URI pour mettre l'accent sur le côté identification (*Uniform Resource Identifier*).

Alors, cela étant dit, qu'est-ce que cela implique concrètement ? Nous y venons.

Pourquoi "sémantique" ?

Le Web sémantique, aussi appelé Web de données, porte ce nom car il permet aux machines de "comprendre" le contenu du Web (sémantique <-> sens).

Dans le Web de documents, nous (les humains) voyons et comprenons les informations suivantes :

en HTML

```
<h1>Les berlingots Eysséric</h1>
<p>La fabrique Eysséric produit des berlingots dans le
<a href="https://www.vaucluse.fr/">Vaucluse</a>.
</p>
```

l'ordinateur, lui, comprend :

```
<h1>??? ??????????</h1>
<p>?? ?????????? ?????????? ?????????? ??? ?????????????? ????? ??
<a href="https://www.vaucluse.fr/">?????????</a>.
</p>
```

Nous aimerions qu'il puisse comprendre:

Nom	Produit	Localisation
Fabrique Eysséric	berlingots	Vaucluse

Nous souhaiterions que l'ordinateur comprenne les relations entre les éléments et la nature de ces éléments, comme nous en somme.

Pour cela, les données transmises doivent être structurées et identifiées (nous l'avons vu plus haut, grâce aux URI).

Données structurées



Le W3C (*World Wide Web Consortium*) définit des standards pour le Web (encore un morceau de votre barre de navigation décrypté !).

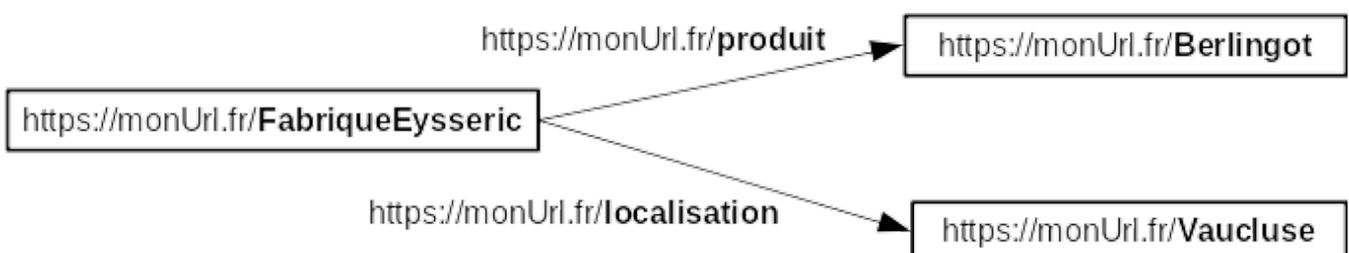
Pour représenter les données dans le Web sémantique, on utilise ces standards et le *Resource Description Format* (RDF). Ce dernier consiste à représenter les données sous forme de **triplets** utilisant des URI, comme des phrases très simples : sujet - prédicat (verbe) - objet.

Les données structurées de notre exemple deviennent alors :

`https://monUrl.fr/FabriqueEysseric https://monUrl.fr/produit https://monUrl.f`

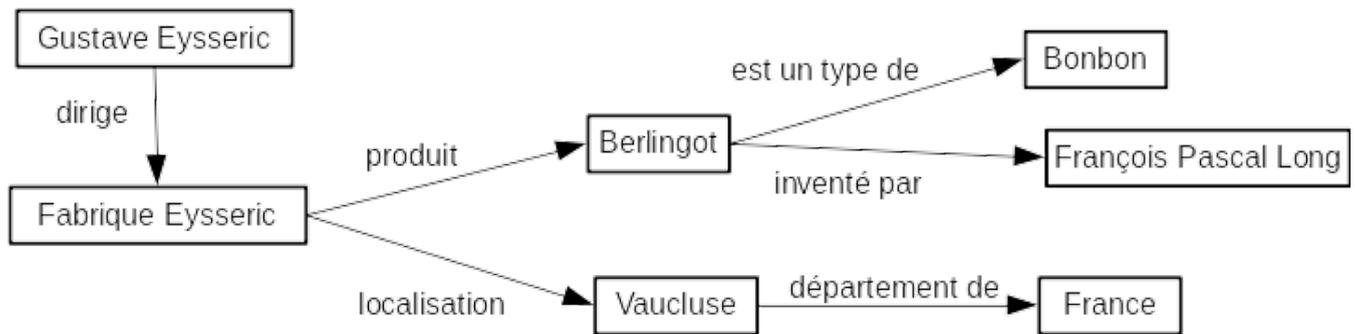
`https://monUrl.fr/FabriqueEysseric https://monUrl.fr/localisation https://mor`

Et parce qu'on sait bien qu'un joli dessin vaut mille mots, on peut aussi représenter les triplets bout à bout sous forme de graphe.



En récupérant les données relatives aux autres ressources du Web, on peut étendre le graphe de données, tant qu'il y a des données.

À l'instar du Web de documents où les documents sont interconnectés grâce aux liens hypertexte, les ressources sont reliées les unes aux autres dans le Web de données.



Pourquoi utiliser le Web sémantique dans FranceArchives ?

Besoin FranceArchives	Réponse Web sémantique
Désambiguïser les autorités	Utilisations d'URI comme identifiants
Données accessibles	Protocole HTTP
Référencement par moteurs généralistes	Contribution au google graph
Alignements référentiels nationaux	Ontologies, alignements
Enrichissement des données propres à FA	Geonames, data.bnf, wikidata
Limiter la responsabilité de maintenance des données	Décentralisation

Parmi les besoins de FranceArchives, nous avons déjà vu que l'utilisation d'URL comme identifiants (ce qui en fait des URI) répond au problème de désambiguïstation des autorités.

De même, le protocole HTTP, base du Web, permet de rendre les données disponibles sur le Web sans application tierce.

Maintenant que nous avons tous ces éléments, il ne reste plus qu'à chercher !

Référencement par les moteurs de recherche généralistes

Certains moteurs de recherche (dont le plus connu de tous) se mettent au RDF !

Ils utilisent des données en RDF insérées dans le code d'une page Web pour mieux comprendre de quel sujet elle traite.

Grâce à cette compréhension, ils peuvent afficher certains résultats sous des formes personnalisées... Exemple !

Le moteur de recherche utilise les triplets RDF pour afficher les recettes de pâte à crêpes

sous forme de petites cartes :

The screenshot shows a search engine interface with the query 'recette pate a crepe' in the search bar. Below the search bar, there are navigation options: 'Tous', 'Vidéos', 'Images', 'Shopping', 'Actualités', 'Plus', 'Paramètres', and 'Outils'. The search results indicate 'Environ 522 000 résultats (0,48 secondes)'. Under the 'Recettes' section, three recipe cards are displayed:

- Pâte à crêpes** by Marmiton, 4.1 stars (804 reviews), 30 min.
- Crêpes : la meilleure recette rapide** by 'Recette de cuisine - Le Journal...', 5.0 stars (2,1 k reviews), 10 min.
- Recette inratable des Crêpes moelleuses...** by Hervé Cuisine, 5.0 stars (185 reviews), 20 min.

Ainsi, vous n'avez même pas besoin de chercher LA meilleure recette de pâte à crêpes, votre ami le moteur de recherche l'a fait pour vous. Et comme il sait que vous n'aimez pas perdre votre temps et que vous aimez quand même bien quand il y a de jolies images qui vous mettent l'eau à la bouche, il vous propose de ne pas utiliser la molette de votre souris et de cliquer directement sur la carte que vous préférez.

Alignement vers des référentiels

Il y a plusieurs avantages à lier les données que nous publions sur le Web de données à des référentiels nationaux (ou internationaux).

Ontologie de référence

Une ontologie (ou un vocabulaire) est un ensemble d'URI que l'on va utiliser pour représenter les prédicats (ou flèches en version graphe) du RDF. L'ontologie définit les types de ressources présentes dans les données et les relations qui peuvent exister entre elles.

C'est un peu le schéma d'une base de données relationnelle ou la liste des noms de colonnes d'un tableur.

Si on compare les données au langage, l'ontologie serait la grammaire ainsi qu'une partie du vocabulaire.

Le fait d'utiliser des ontologies standards dans ces données RDF permet de se "brancher" plus facilement avec d'autres graphes de données.

Sources de données de référence

En liant ses données à d'autres bases de données sur le Web, FranceArchives y trouve plusieurs avantages.

Tout d'abord, cela lui permet d'**enrichir ses propres données**.

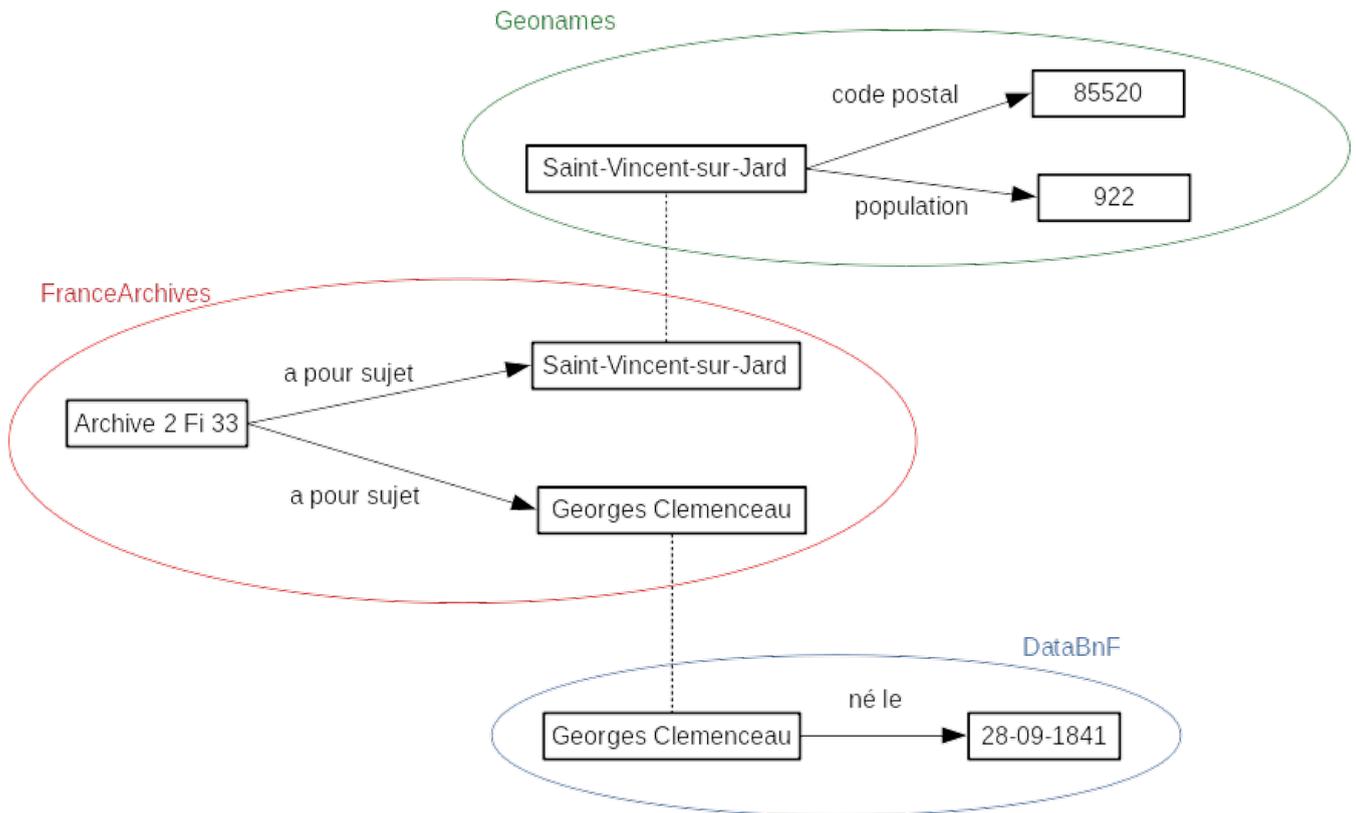
En effet, un document d'archives implique toujours des lieux et/ou des personnes. On tente de normaliser les pratiques de nommage de ces entités (dans quel sens on met quelle information) afin d'aider davantage au liage des données : **Charles, de Gaulle (1890-1970)** ou **de Gaulle, Charles** ou **Général de Gaulle (Charles, 1890-1970)**. De même pour les noms de lieux : **Sumène, Sumène (Gard - 30), Sumène (Gard)**, etc.

Dans les notices, seuls figurent généralement le nom du lieu (avec son département) et le nom de la personne. En liant les données de FranceArchives à d'autres bases, on peut ainsi étendre le graphe de données et enrichir les informations que nous avons au départ. Comme nous l'avons vu, plus il y a de mentions permettant de désambigüiser un terme (à placer au Scrabble), meilleure sera la qualité de la donnée et donc plus performant sera le schéma RDF et au final les résultats de recherche.

Ce schéma montre l'exemple de la notice Sur la plage de Saint-Vincent-sur-Jard...

(<https://francearchives.fr/fr/facomponent/e6c3ff90ff8dc0578c3fed6909b5e9968965913b>) dont la description RDF (<https://francearchives.fr/fr/facomponent/e6c3ff90ff8dc0578c3fed6909b5e9968965913b/rdf.xml>) peut être obtenue en ajoutant à l'url le suffixe `/rdf.xml` ou `/rdf.ttl`.

Cette notice parle notamment de Georges Clemenceau et de la commune de Saint-Vincent-Sur-Jard. Initialement, il y avait peu d'informations sur ces deux ressources. En liant Saint-Vincent-sur-Jard à son pendant dans Geonames (<https://www.geonames.org/2976583/saint-vincent-sur-jard.html>), une base de données regroupant des lieux, nous avons pu enrichir les données en récupérant notamment le code postal et la population. De la même manière, en liant Georges Clemenceau à son pendant dans DataBnF (https://data.bnf.fr/fr/11897013/georges_clemenceau/), nous avons pu enrichir les données en récupérant notamment sa date de naissance et quelques éléments biographiques notables.



Le second avantage qu'apporte l'alignement (c'est-à-dire le fait de lier sa base avec une autre) est de **limiter la maintenance des données** : moins on les manipule, mieux elles se portent.

FranceArchives peut profiter d'informations libres et ouvertes sur les personnes, les thèmes, les lieux pour valoriser ses données tout en se concentrant sur la publication et la maintenance des données d'archives uniquement.

data.bnf

Le projet data.bnf (<https://data.bnf.fr/fr/about>) a pour but de rendre les données de la BnF **utiles et exploitables sur le web**. Elles permettent notamment de rassembler des informations sur les ressources conservées au sein de la BnF : **documents, ouvrages, auteurs, thèmes, etc.** Les pages sont indexées par les moteurs de recherche : les **données disponibles et requêtables sont souvent invisibles lors d'une recherche classique** car enfouies dans les données et métadonnées des ressources BnF.

Sur FranceArchives, ces liens permettent d'ajouter des informations sur un sujet donné.

Wikidata

Wikidata (<https://www.wikidata.org/wiki/Wikidata:Introduction/fr>) est une base **open source, gratuite, collaborative** et qui, de la même manière que DataBnF, met à disposition des données compréhensibles aussi bien par les humains que par les machines. Cette base de données aide Wikipédia en facilitant la maintenance des fameuses boîtes d'informations

que nous consultons tous dès que nous cherchons des informations sur la célèbre encyclopédie.

De la même manière que la précédente, les renvois vers Wikidata ajoutent une plus-value aux données consultables sur FranceArchives.

Clemenceau, Georges (1841-1929)

Date de naissance :

28 septembre 1841

Date de décès :

24 novembre 1929

Description :

Né à Mouilleron-en-Pareds (Vendée), fils de médecin et médecin lui-même, Georges Clemenceau (1841-1929) entre dans la carrière politique au lendemain de la journée révolutionnaire du 4 septembre 1870 et devient maire de Montmartre (XVIII^e arrondissement de Paris) au moment où...

Activité :

docteur d'université ; député ; journaliste ; ministre ; ministre ; président du Conseil ; président du Conseil ; sénateur

Source :

[Archives nationales : Clemenceau, Georges \(1841-1929\)](#)

Autres ressources :

Wikidata : [Georges Clemenceau](#)

[data.bnf.fr : Georges Clemenceau \(1841-1929\)](#)



Geonames

Geonames (<https://www.geonames.org/about.html>) est une base de données libres et ouvertes sur les données géographiques.

DataCulture

DataCulture (<http://data.culture.fr/thesaurus/>) : le Ministère de la Culture publie un référentiels de sujets classés hiérarchiquement (en thésaurus). Les thèmes de FranceArchives sont alignés sur les ressources de DataCulture.

Axes futurs d'amélioration

FranceArchives utilise déjà des technologies du Web sémantique. Pour aller plus loin, les chantiers suivants sont envisagés.

Interrogation en SPARQL : SPARQL (oui il faut le lire comme un mot prononcé *SparKeul* et ne pas le jouer au scrabble celui-là sauf si on joue en anglais, car ça fait un jeu de mot pétillant avec *to sparkle*) est le langage d'interrogation du RDF. Rendre possible l'interrogation des données produites dans ce langage permet aux utilisateurs et utilisatrices de rechercher très précisément les informations voulues.

Utilisation de l'ontologie RiC-O (<https://www.ica.org/standards/RiC/ontology.html>) : cette ontologie (Records in Contexts - Ontology) est développée et maintenue par le Conseil International des Archives. Elle est en passe de devenir un standard pour le monde archivistique. L'utiliser pour décrire les données de FranceArchives permettra de se brancher plus facilement aux données d'autres services d'archives qui en font aussi usage.

I have a dream...

... *that one day* tout le monde pourra rechercher simplement et trouvera du premier coup toutes les informations désirées !

Dans le monde numérique, ce qui est bien c'est qu'on peut rêver, et rêver grand ! Alors que diriez-vous de pouvoir faire une requête telle que : *Je cherche les archives concernant le village de naissance du général de Gaulle et la période 1945-1962* et que le moteur de recherche vous remonte directement les documents qui traitent exactement de ce dont vous, humain, vous parlez ? Imaginez un monde où l'on pourrait interroger les bases de données en langage naturel.

Nous pouvons conclure cet article rédigé à l'occasion des Journées du Logiciel Libre 2021, sur le thème des *Utopies concrètes et accessibles* par cette proposition d'amélioration : un accès unique à toutes les données du web, requêtables en langage naturel et sans bruit documentaire... Un *International Knowledge Portal* !