

Presentation francearchives.fr : les archives de France et le Web sémantique

04 juillet 2016

V_1.1

- *titre* : **francearchives.fr : les archives de France et le Web sémantique**

- *description en moins de 400 mots* :

Les archives de France diffusent déjà des informations sur le Web sémantique, notamment le thésaurus d'indexation des archives qui est disponible en RDF (SKOS) et interrogeable en SPARQL. <http://data.culture.fr/thesaurus/sparql>

Elles s'engagent maintenant dans un projet numérique national, projet d'ouverture des archives auprès du plus grand nombre. Ce type de service existe déjà dans d'autres pays européens, notamment au Royaume-Uni avec le portail Discovery. Ces portails sont capables d'orienter les internautes dans leurs recherches, quelles qu'elles soient : scientifiques, généalogiques, administratives. Ce sera la fonction de Francearchives.fr, qui regroupera de nombreux contenus éditoriaux et les inventaires et données de plusieurs centaines de services d'archives, mis à jour et traités.

Francearchives.fr est actuellement en phase de développement. Le programme est piloté par les Archives de France en partenariat avec les ministères des Affaires étrangères et de la Défense. Des représentants du réseau territorial y sont bien sûr associés. Les travaux de développement du portail ont démarré fin 2015 avec la société Logilab.

Une précision : le portail francearchives ne remplace pas les sites existants. Au contraire, il met en valeur leur richesse et renvoie les internautes vers les images d'archives numérisées qui restent hébergées sur les sites des services d'archives participants. Il améliore leur visibilité auprès d'un public conçu d'emblée comme le plus large possible.

Cette simplicité apparente implique en fait des traitements importants et complexes sur les données numériques. Un internaute cherchant dans les archives le testament de Voltaire doit en fait trouver celui de François-Marie Arouet. Nous le savons, mais la machine doit apprendre à relier ces deux noms, ce que permettent aujourd'hui les outils du Web dit "sémantique", pour faire émerger les informations concernant les lieux, les dates et les thèmes notamment.

Un portail fédérateur doit désormais assurer la diffusion maximale des données, et favoriser ce qu'on appelle la « réutilisation des données publiques », intégrée dans le mouvement *Open data*. Francearchives offre les conditions juridiques et techniques permettant la diffusion et la réutilisation des données : c'est la licence ouverte de l'État, dite « licence Etalab » qui a été choisie pour atteindre cet objectif.

L'intervention proposée présentera le projet dans sa dimension stratégique : les besoins auxquels il répond, les solutions mises en place, et les raisons pour lesquelles Web sémantique et Open data sont indissociables dans le cadre des données d'archives.

- auteur présenté en quelques phrases :

Romain Wenz est conservateur du patrimoine, chef de projet au Service interministériel des archives de France depuis 2015. Il a notamment travaillé de 2009 à 2013 à la Bibliothèque Nationale de France comme expert métadonnées et chef du projet « Web sémantique » data.bnf.fr. Conservateur des bibliothèques et archiviste-paléographe de formation, Romain Wenz a soutenu en 2007 une thèse d'École Nationale des Chartes sur le port d'armes en France à la fin du Moyen Âge.