

Tags: Archiv Zugang,

Archival linked open data – le projet suisse aLOD

L'accès facilité du public aux inventaires et aux données est l'un des défis des archivistes qui doivent proposer des solutions innovantes pour compléter l'approche métier traditionnelle, encore trop orientée pour les spécialistes.

Origine du projet

Il est en effet important que les archives en ligne s'ouvrent à un public plus large que le public habituel des chercheurs. Les processus d'accès sont ainsi à repenser. Le Web des données (abrégé LD pour *linked data*) est l'une des pistes à explorer.

Suite à des discussions passionnées, un groupe d'archivistes s'est formé en 2014 pour tenter d'expérimenter ce Web des données appliqué aux inventaires d'archives. Un projet, au départ informel, car ne dépendant d'aucune structure, a reçu l'appui des directions des plusieurs institutions: Archives fédérales suisses, Archives des cantons de Neuchâtel, du Valais, de Genève, de Bâle-Ville et de la Ville de Baden, accompagnées par la société Zazuko, spécialisée dans le Web des données. Tous les échelons des administrations publiques suisses sont ainsi représentés.

«Le Web des données est une initiative [...] visant à favoriser la publication de données structurées sur le Web, non pas sous la forme de silos de données isolés les uns des autres, mais en les reliant entre elles pour constituer un réseau global d'informations. Il s'appuie sur les standards du Web [...] mais plutôt qu'utiliser ces standards uniquement pour faciliter la navigation par les êtres humains, le Web des données les étend pour partager l'information également entre machines. Cela permet d'interroger automatiquement les données, quels que soient leurs lieux de stockage, et sans avoir à les dupliquer.» (définition Wikipédia)

Objectifs du projet

Il a fallu se mettre d'accord au sein du groupe de travail sur les priorités, car deux tendances se dessinaient. La première souhaitait mettre l'accent sur la partie technologique avec la mise sur pied d'une infrastructure complète et documentée qui soit adaptées aux besoins spécifiques des archives. La seconde insistait sur un résultat permettant une présentation exemplifiée et pédagogique de cette technologie, afin de la partager avec la communauté archivistique. Pour conjuguer ces deux attentes, il a été convenu de produire une infrastructure pilote qui ait un résultat visuel parlant.

Deux étapes ont été définies: l'approche technologique par la réalisation d'une infrastructure nourrie avec les données RDF (le modèle de description des données en LD) créées à partir de métadonnées descriptives disparates, et l'approche utilisateur par le développement d'une interface d'accès qui propose d'autres modes de présentation de ces données aux utilisateurs.

Choix de jeux de données sur une thématique commune

Il a fallu choisir un thème, afin que les différentes institutions disposent de jeux de données qui aient un rapport entre eux. Des thématiques telles que les finances, les routes ou l'image de la Suisse dans le monde ont été envisagées, mais elles ne permettaient pas aux institutions de pointer sur un fonds ou un ensemble de fonds constitué. Il a donc été décidé de se concentrer sur une période historique qui a produit des fonds se retrouvant dans toutes les institutions partenaires. Ainsi le choix s'est porté sur la période de la *République helvétique* et celle de la *Médiation* (1798–1813).



Faux assignats de la Période française (Département du Léman) (cote CH AEG ADL L 9)

Approche de la technologie LD: réalisation et résultats

Chaque institution a dû commencer par mettre concrètement les mains dans le cambouis afin de constituer des données en RDF à partir d'une extraction de données de son système d'information archivistique. Le fichier de départ était donc constitué d'un set de métadonnées descriptives, d'un format différent selon les systèmes d'information: XML-EAD, CSV ou Excel. Ces jeux de données ont été transformés en RDF avec l'aide de la société Zazuko.

Un jeu de données en RDF est un ensemble de triplets, qui établissent des relations sous la forme de sujet-prédicat-objet. Une bonne introduction à ces notions se trouve sur le lien suivant.

Une plate-forme technique (*triplestore*) a été mise sur pied pour le stockage et l'accès aux données RDF. Les jeux de données RDF ont été ensuite importés dans le triplestore. La provenance des données est garantie par le choix des noms de domaines dans les URLs. Ainsi, «data.ge.alod.ch» sert à identifier les données de Genève, «data.staatsarchiv-bs.ch» celles de Bâle-Ville, etc. Ce modèle permet que chacun soit responsable de ses propres données, ce qui correspond à l'approche fédéraliste.

Cette étape technique a eu pour livrable concret la colocalisation, au sein d'une même infrastructure, des données qui proviennent de différents systèmes, sur lesquelles des recherches peuvent être effectuées.

Pour effectuer des recherches sur des données RDF, on peut naviguer entre les URLs des entités qui ont été définies et découvrir les prédicats et objets/valeurs qui leur sont associées sous forme de pages Web. Ce n'est pas très impressionnant, mais c'est la preuve basique de l'existence de ce Web des données pour un humain.

Approche utilisateurs: réalisation et résultats

Notre principe de départ établit que c'est aux institutions d'aller vers l'utilisateur et non le contraire. Une analyse sur la base d'interviews des besoins de l'utilisateur d'archives a été effectuée.

Ces interviews ont été réalisées sur deux groupes d'utilisateurs définis au préalable: les chercheurs spécialisés ou professionnels et les non-spécialistes. Ils ont permis de définir un concept d'interface et un cahier des charges. Les points les plus importants sont les suivants:

- Un champ unique de recherche est présent en haut de la page. Lorsqu'il est vide, la liste de tous les résultats est par défaut affichée. Cette liste se présente sans pagination et est déroulée en scrollant, quelle que soit la taille de la liste, permettant d'explorer les données sans avoir de mots-clés, selon le principe «les archives s'ouvrent». Il n'est pas fait de différence entre la recherche simple et avancée.
- Dans une première étape, l'utilisateur entre un mot-clé et la liste se réduit par filtrage.
- Le lot résultat est composé de la liste des notices descriptives. Un code couleur indique sur chaque résultat son niveau dans la tectonique.
- Une timeline fixe est affichée en bas de l'écran et indique visuellement où les documents du lot résultat se situent dans la fourchette temporelle couverte par les archives.
- Chaque résultat a un marqueur qui le situe dans le temps en reportant sa position dans la timeline.

Un prototype d'une nouvelle génération d'interface utilisateur a ainsi été développé par étapes pour répondre au cahier des charges. Il est accessible (sans garantie) à l'URL www.alod.ch/search.

Conclusion provisoire

À ce stade du projet, la mise en commun de données provenant de différents systèmes au sein d'une infrastructure dans laquelle des recherches «brutes» peuvent être effectuées a été réalisée. Nous avons pu ainsi répondre au premier défi technologique. Un cahier des charges pour le développement de l'interface utilisateur a été rédigé; cette interface existe en tant que démonstrateur.

En d'autres termes, la partie immergée de l'iceberg a été réalisée et la partie visible pour les utilisateurs disponible. Il s'agit maintenant de définir des *best practices* pour une implémentation à large échelle afin de simplifier le travail des institutions qui souhaitent adopter cette solution.

D'une approche pragmatique, consensuelle et expérimentale, le projet aLOD a pu mettre sur pied une solution prototype qui permet de présenter les avantages et le potentiel du Web des données. Ce prototype est aussi un excellent banc d'essai pour toute une série d'expérimentations futures que nous souhaitons conduire autour de la nouvelle norme [Records in Context](#), ou de la production participative pour l'enrichissement des données. Une autre direction possible est de faire évoluer cette infrastructure vers une solution productive, fiable et hautement redondante.

Ont participé au projet aLOD: Jean-Luc Cochard, Frédéric Noyer, Alain Dubois, Lambert Kansy, Anouk Dunant Gonzenbach, Michaël Luggen, Tobias Wildi, Urs Meyer, Adrian Gschwend. Qu'ils en soient tous ici chaudement remerciés.



Anouk Dunant Gonzenbach

Anouk Dunant Gonzenbach travaille depuis 15 ans comme archiviste aux Archives d'Etat de Genève, où elle est responsable des projets numériques: archivage électronique; gouvernance des documents et données électroniques; outils numériques et médiation numérique. Elle est membre du groupe Records Management-archivage définitif de l'Association des archivistes suisses. Elle tient un blog professionnel: www.hieretdemain.ch

.

Abstract

Français

Le web des données (*linked data* LD) est une piste intéressante en matière d'accès facilité du public aux inventaires et aux données d'archives. Le groupe de travail aLOD (*archival linked open data*), dont les membres font partie d'institutions d'archives publiques suisses, développe un projet pour explorer les potentiels de cette technologie. Ainsi, une infrastructure pilote nourrie à partir de métadonnées descriptives disparates a été créée. Une interface d'accès novatrice a également été développée.

Deutsch

Linked Data (LD) stellen eine interessante Möglichkeit dar, um der Öffentlichkeit den Zugang zu den Beständen und Daten von Archiven zu erleichtern. Die Mitglieder der Arbeitsgruppe aLOD (*archival linked open data*) stammen aus öffentlichen Archiven in der Schweiz. Sie haben es sich zum Ziel gesetzt, in einem Projekt die Potenziale der Linked Data-Technologie für Archive genauer zu untersuchen. Das Projekt *aLod* hat eine Pilot-Infrastruktur aufgebaut, basierend auf deskriptiven Metadaten aus verschiedenen Quellen. Auch eine innovative Schnittstelle für den Zugriff auf diese Daten wurde durch die Arbeitsgruppe entwickelt.