
Une Preuve de concept pour la sémantisation et la visualisation orientée utilisateur de données archivistiques

Florence Clavaud

florence.clavaud@culture.gouv.fr

Archives Nationales, France

Emmanuel Château-Dutier

emmanuel.chateau.dutier@umontreal.ca

Université de Montréal, Canada

Notre présentation concerne un travail exploratoire aux objectifs qualitatifs, qui ambitionne de fournir des pistes de réflexion pour avancer vers la sémantisation des métadonnées des documents d'archives et la représentation à l'écran de ces données RDF, en vues multidimensionnelles, porteuses de sens et manipulables.

Trois organismes français (les Archives nationales de France, la Bibliothèque nationale de France, le Service interministériel des Archives de France) et un laboratoire de recherche canadien (le Département d'histoire de l'art et des études cinématographiques de l'Université de Montréal) se sont associés en 2015 pour réaliser une preuve de concept visant à démontrer qu'il est possible :

- de représenter en RDF, en veillant à la précision, à l'exactitude et à l'utilisabilité des triplets obtenus, des métadonnées archivistiques produites de différentes manières et selon diverses perspectives (celles d'institutions patrimoniales, celles d'un chercheur) ;
- d'enrichir les triplets obtenus en créant de nouveaux triplets, qu'il s'agisse de procéder à des alignements ou d'établir de nouvelles relations par inférence ;
- de produire une interface de recherche et d'exploration analytique et graphique qui soit dynamique, ergonomique et signifiante,

sans sacrifier la granularité informationnelle ni la lisibilité.

Il s'agit donc de réaliser un démonstrateur, sous la forme d'une application web construite en utilisant des composants libres, dont les sources seront placées sous licence libre et déposées dans un entrepôt public.

Cette opération est relativement complexe. Il n'y a pas de réel précédent dans le domaine des archives. De plus, jusqu'à ce jour, il n'existait pas d'ontologie générique du domaine. Enfin, il n'existe pas a priori de librairie ou logiciel satisfaisant directement la totalité des besoins, en particulier pour ce qui concerne la visualisation des jeux de données.

Les bibliothèques, comme la Bibliothèque nationale de France (avec *data.bnf.fr*), et plusieurs musées, se sont lancés dans la conversion en RDF de leurs métadonnées et dans la réalisation d'interfaces de présentation de ces données, dès lors que des modèles conceptuels et ontologies appropriés ont été disponibles pour représenter leurs collections (CIDOC-CRM et FRBR). Par contre, si divers projets spécifiques ont déjà, soit produit des ontologies pour la description des archives et mis en œuvre ces ontologies (par exemple le projet LOCAH, dont les résultats sont consultables via *The Archives Hub Linked Data*), soit cherché à produire des représentations analytiques ou graphiques de ces descriptions, jusqu'ici ces réalisations n'avaient pas en même temps les ambitions de la généralité, de la complétude et d'une grande précision.

Il a fallu attendre fin 2016 pour la publication d'un modèle conceptuel global pour la description des archives (*Records In Contexts-Conceptual Model* ou RiC-CM) ; l'ontologie OWL correspondante, *Records In Contexts-Ontology* ou RiC-O), qui est l'ontologie de référence pour notre projet, sera publiée en 2017.

En outre, si la data visualisation est aujourd'hui devenue, en s'appuyant sur des technologies de plus en plus performantes, un important domaine d'action et d'innovation pour les humanités numériques, il est à notre connaissance encore très rare de trouver articulés la dimension temporelle, très forte en archivistique et en histoire, et une représentation précise, en graphe, d'objets historiques et des différents éléments de leur contexte, le tout étant susceptible de former un réseau très dense. Ainsi, les divers projets réalisés jusqu'ici présentent uniquement de façon relativement simple les agents et les documents (bibliothèque publique de la ville de New York), ou encore uniquement des réseaux d'agents (SNAC), ou encore des agents, leur histoire et leurs relations dans une perspective diachronique,

mais pas les ressources documentaires liées (*Kindred Britain*).

Dans le cas de notre projet, nous avons commencé par choisir un périmètre pour les jeux de métadonnées source et préparer ces jeux de métadonnées, chaque partenaire travaillant de son côté, selon sa propre perspective, tout en respectant quelques règles élaborées en commun.

Les jeux de métadonnées sont composés de notices descriptives d'organismes et personnes physiques acteurs dans deux domaines fonctionnels de l'administration française (la gestion des monuments historiques et bâtiments civils des années 1795 à nos jours ; la gestion des bibliothèques publiques et de la lecture du 19e siècle à nos jours) et d'instruments de recherche archivistiques décrivant les ensembles documentaires produits par ces entités. Ils sont conformes aux standards archivistiques actuels que sont les normes ISAD(G) et ISAAR(CPF) et leurs transpositions techniques, la DTD EAD 2002 et le schéma XML/EAC-CPF. Nous avons choisi de procéder nous-mêmes à la sémantisation de ces jeux de métadonnées. Un vocabulaire des fonctions des entités encodé en SKOS/RDF, le seul à avoir été produit collectivement, s'ajoute à ce corpus.

Un cahier des charges a été élaboré, une consultation lancée et une société de services a été choisie fin octobre 2016 pour concevoir et réaliser le démonstrateur. Elle réalisera le travail selon une approche agile, en interaction forte avec l'équipe projet, qu'il s'agisse d'analyser l'ontologie et les triplets RDF fournis en entrée, de bâtir des scénarios de recherche, d'alignement, d'enrichissement ou de visualisation, ou de tester les versions successives du logiciel.

Comme le projet sera achevé fin octobre 2017, une version quasiment définitive du prototype devrait être disponible en août 2017. Un bilan critique, méthodologique et prospectif sera ensuite réalisé et publié par les entités impliquées. En avant-première, après une présentation rapide des objectifs, des enjeux et de l'historique du projet, nous nous attacherons à évoquer deux de ses aspects principaux :

- la mise en œuvre, pour la représentation en RDF des métadonnées archivistiques retenues, de l'ontologie RiC-O : présentation de l'ontologie et de ses principes de conception, discussion des choix d'adaptation qui ont été faits pour les besoins du projet, présentation des résultats obtenus, évaluation de ces résultats en ce qui

concerne la granularité d'expression et les possibilités de raisonnement induites ;

- la conception de l'interface de recherche et de visualisation : présentation de la méthode suivie, des choix faits et des résultats obtenus, de leur intérêt et de leurs limites pour l'utilisateur final, que celui-ci soit un chercheur averti ou un amateur moins connaisseur des concepts archivistiques.

Bibliographie

Bertin, J. (1977). *La graphique et le traitement graphique de l'information*. Paris : Flammarion.

Bibliothèque nationale de France (2011-). *Data.bnf.fr*. <http://data.bnf.fr>

Drucker, J. (2014). *Graphesis: visual forms of knowledge production*. MetaLABprojects. ISBN 9780674724938

Fekete, J-D. (2010). "Visualiser l'information pour la comprendre vite et bien". Dans *L'Usager numérique*. Séminaire INRIA, 27 septembre-1er octobre 2010. ADBS, pp. 161-194.

Few, S. (2013). "Data Visualization for Human Perception". Dans Soegaard, M. et Friis Dam, R. (éds.), *The Encyclopedia of Human-Computer Interaction*, 2e édition. En ligne : <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception>

Friendly, M. (2009). *Milestones in the history of thematic cartography, statistical graphics, and data visualization*. Version PDF en ligne : <http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf>

Gibbs, F. W. (2016). "New Forms of History: Critiquing Data and Its Representations". *The American Historian*, 7. En ligne : <http://tah.oah.org/february-2016/new-forms-of-history-critiquing-data-and-its-representations/>

Grandjean, M. (2014). "La connaissance est un réseau". *Les Cahiers du Numérique*, 10/3 : 37-54. DOI :10.3166/LCN.10.3.37-54.

Grandjean, M. (2015). "Introduction à la visualisation de données, l'analyse de réseau en histoire". *Geschichte und Informatik*, 18-19 : 109-128. Version PDF en ligne : <http://www.martingrandjean.ch/wp-content/uploads/2015/09/Grandjean2015.pdf>

- Januals, B. et Minel, J-L.** (2016). " La construction d'un espace patrimonial partagé dans le Web de données ouvert ". *Communication*, 34/1. En ligne : <http://communication.revues.org/6650>. DOI : 10.4000/communication.6650.
- Jenkins, N., Meeks, E. et Murray, S.** (2013). *Kindred Britain*. Stanford University Libraries. <http://kindred.stanford.edu/>
- Manovich, L.** (2010). What is Visualization? En ligne : http://manovich.net/content/04-projects/064-what-is-visualization/61_article_2010.pdf
- Natale, E., Sibille, Ch., Chachereau, N., Kammerer, P., Hiestand, M. et al** (2015). " La visualisation des données en histoire / Visualisierung von Daten in der Geschichtswissenschaft ". *Geschichte und Informatik*, 18/19.
- Krempel, L.** (2011). " Network visualization ". Dans Scott, John G. et Carrington, P. J. (éds.), *The SAGE handbook of social network analysis*. Pp. 558-577.
- International Council on Archives. Experts Group on Archival Description** (2016). *Records In Contexts: a conceptual model for archival description*. Consultation Draft v0.1. En ligne : <http://www.ica.org/fr/publication-de-records-contexts-par-legad>
- International Council on Museums (ICOM). International Committee for Documentation (CIDOC)** (2011). CIDOC Conceptual Reference Model. Version 5.0.4. En ligne : <http://www.cidoc-crm.org/get-last-official-release>
- International Federation of Libraries Associations and Institutions** (2008). *Functional Requirements for Bibliographic Records* (FRBR). En ligne : http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf (en) ; http://www.bnf.fr/documents/frbr_rapport_final.pdf (fr).
- International Federation of Libraries Associations and Institutions** (2016). *FRBR-Library Reference Model (FRBR-LRM)*. Draft for world-wide review. En ligne : http://www.ifla.org/files/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf
- International Working Group on FRBR and CIDOC CRM Harmonisation** (2008). *FRBR object-oriented definition and mapping to FRBRer*. Version 0.9 draft. En ligne http://archive.ifla.org/VII/s13/wgfrbr/FRBRoo_V9.1_P R.pdf.
- The New York Public Library.** (n. d.) Archives and manuscripts: beta tools & experiments. <http://archives.nypl.org/tools>
- University of Liverpool** (2010-2013). *The Archives* hub Linked Data. <http://data.archiveshub.ac.uk/>
- University of Virginia Institute for Advanced Technology in the Humanities.** (n.d.) *Social Networks and Archival Contexts* (SNAC) prototype. http://socialarchive.iath.virginia.edu/home_prototype.html
- Yau, N.** (2013). *Data Points: Visualization That Means Something*. John Wiley & Sons. ISBN 978-1-118-46219-5