
Les humanités numériques pour valoriser le patrimoine minier

Eric Kergosien^{*1}, Amélie Daloz², Bernard Jacquemin³, and Stéphane Chaudiron³

¹Groupe d'études et de recherche interdisciplinaire en information et communication (GERiiCO, EA 4073) – Université de Lille, Sciences Humaines et Sociales – France

²Groupe d'études et de Recherche Interdisciplinaire en Information et COmmunication (GERiiCO) - EA 4073 – Université de Lille : EA4073 – France

³Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication (GERiiCO) - EA 4073 (GERIICO) – Université de Lille, Sciences Humaines et Sociales : EA4073 – Domaine Universitaire du Pont de Bois – BP 60149 F-59653 Villeneuve d'Ascq CEDEX, France

Résumé

Les travaux présentés dans cette communication sont menés dans le cadre du projet ANR MémoMines[1], dont l'objectif est d'apporter des outils et méthodes scientifiques pour aider à la sauvegarde et à la mise en valeur du patrimoine minier des Hauts-de-France. En effet, un enjeu social important dans le domaine du patrimoine culturel et notamment celui des mines est lié à la collecte, l'analyse, la publication et la mise en valeur de l'histoire et de la mémoire collectives des acteurs du domaine, à l'oral comme à l'écrit. Dans ce sens, la formalisation de l'information sur le patrimoine culturel représente un véritable défi en raison de la diversité et de l'incomplétude des données. De plus, ces données sont hétérogènes et peuvent être trouvées dans différentes sources, en ligne ou hors ligne : bases de données, bibliothèques, musées, dossiers de presse, expertise des intervenants, etc. Cette diversité des ressources pose de nombreux défis tels que la documentation des données, la représentation, l'intégration et l'interopérabilité au sein d'une même base de connaissances.

Pour répondre à ces problématiques, nous proposons une méthodologie pour la construction d'une base de connaissances, que l'on nomme également ontologie de domaine du patrimoine minier (exploitation du charbon) circonscrit au territoire des départements du Nord et du Pas-de-Calais en France et sur une période allant du début de l'exploitation (vers 1750) jusqu'à nos jours avec par exemple la réhabilitation des anciens sites de production et des habitats liés à la mine. Nous souhaitons dans ce sens structurer et formaliser des connaissances caractérisant les acteurs, les lieux, les entités temporelles et les thématiques présentes dans les corpus documentaires oraux et écrits analysés, et l'ontologie produite doit alors nous permettre d'en faciliter l'accès.

La méthodologie que nous présenterons dans notre communication se décompose en 3 étapes, la première étant la formalisation des entités nommées (lieux, acteurs, entités temporelles) et d'un vocabulaire structuré décrivant le domaine des mines en une première ontologie du domaine des mines. Les entités temporelles correspondent aux références calendaires (dates et années). Concernant les entités nommées acteurs et lieux, nous nous appuyons sur les ressources produites par la Mission Bassin Minier, partenaire du projet qui gère pour l'état le label UNESCO. Concernant le vocabulaire du domaine, nous nous appuyons sur le

*Intervenant

thésaurus ThesoMines défini dans le cadre du projet (Daloz et Chaudiron, 2019), validé par des experts du domaine et disponible dans sa version 1.0 en accès libre[2] sur la plateforme OpenTheso mise à disposition sur les serveurs de la Très Grande Infrastructure de Recherche HumaNum[3]. Le thésaurus, défini selon le standard SKOS (*Simple Knowledge Organization System*), est composé de 558 descripteurs et 764 non-descripteurs, et est structuré en quatre domaines (culture éducation, économie, environnement naturel, histoire) et cinq facettes (activité, bâti, gisement, matériel, opération) afin de couvrir aussi bien le patrimoine matériel qu’immatériel, selon la définition qu’en donne l’UNESCO.

Au sujet des formalismes, la plupart des tentatives pour résoudre les problèmes d’interopérabilité sémantique se concentrent sur la standardisation, voir la normalisation de formalismes communs tels que FRBR (Le Boeuf, 2005), FRBRoo (Doerr *et al.*, 2013), CIDOC CRM (Le Boeuf *et al.*, 2015), etc. Au regard de l’état de l’art et de nos premières expérimentations (Kergosien *et al.*, 2019), nous faisons le choix de formaliser l’ontologie produite en CIDOC CRM, modèle conceptuel de référence conçu pour modéliser les domaines du patrimoine culturel. Nous détaillerons à partir d’exemples concrets les choix de formalisation des connaissances provenant des ressources sélectionnées (thésaurus, liste de lieux, etc.).

La seconde étape consiste à décrire le corpus de documents oraux et écrits à partir de la première ontologie produite. Cette seconde étape correspond à l’instanciation (ou peuplement) du modèle ontologique. Nous inscrivons ici nos travaux dans la politique d’ouverture des données qui a pris de l’ampleur au cours de la dernière décennie. Dans la continuité de la loi pour une République numérique (7 octobre 2016) renforçant la politique d’ouverture des données publiques, le ministère de la Culture s’est engagé dans une politique d’ouverture et de partage des données publiques, notamment pour permettre aux citoyens de participer à l’amélioration de la qualité des données. C’est ainsi que les Archives nationales, en lien étroit avec le service interministériel des Archives de France, viennent de publier un nouveau régime de réutilisation des informations publiques qu’elles conservent, qu’il s’agisse des inventaires et des corpus d’archives. Il s’agit de documents librement communicables à tous et sur lesquels des tiers ne détiennent pas de droits de propriété intellectuelle. Les informations doivent être librement communicables au sens de l’article L.213-1 du Code du patrimoine et elles ne doivent pas avoir été communiquées par autorisation ou par dérogation. C’est également dans ce cadre que les Archives nationales ont publié plusieurs corpus librement réutilisables sur la plateforme de données ouvertes du ministère : data.culturecommunication.gouv.fr. En 2019, le ministère de la Culture a dans la continuité de son programme lancé la plateforme ouverte du patrimoine POP, nouvel outil pour la gestion des données culturelles du ministère, comportant deux volets : (1) une plate-forme de production, qui permet le versement de métadonnées (images, photographies, informations descriptives ou réglementaires) ; (2) une plate-forme de diffusion de ces données, accessible au grand public[4]. Dans le cadre du projet, nous travaillons avec le partenaire INA pour intégrer notre ontologie de domaine sur la plateforme OKAPI[5] dédiée à l’indexation et l’analyse et la publication de corpus de données audiovisuelles. Afin d’explicitier cette deuxième étape d’instanciation, nous nous appuyons sur une analyse fine d’un premier entretien mené auprès d’un ancien mineur, pour montrer à la fois l’intérêt de notre démarche, et également les difficultés rencontrées liées à l’indexation, et notamment au choix du niveau de granularité de description retenu.

La troisième étape consiste à enrichir le premier modèle ontologique produit à partir de nouvelles connaissances présentes notamment dans les entretiens menés auprès des anciens mineurs (nouveaux termes, nouvelles entités nommées, etc.), puis de valider la base de connaissances obtenue avec les experts du domaine. Nous étayerons cette dernière étape à travers différents exemples tirés de notre travail.

L’ontologie produite sera prochainement rendue disponible dans le format OWL CIDOC CRM, afin de la rendre accessible au plus grand nombre. Notre démarche peut en effet être généralisée à d’autres patrimoines en danger et, dans ce sens, une documentation associée intégrant les schémas d’identification des entités que nous sommes en mesure de marquer à partir d’un document sera également proposée, pour tous ceux qui souhaiteraient utiliser le CIDOC CRM sur un cas concret lié à la valorisation du patrimoine.

<https://memomines.hypotheses.org>

Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International

<https://opentheso.huma-num.fr/opentheso/api/theso/MemoMines>

<https://www.pop.culture.gouv.fr/>

Open Knowledge Application Programming Interface, <https://okapi.inalco.fr/mediascape>

Bibliographie :

Daloz, Amélie et Chaudiron Stéphane. 2019. Méthodologie de conception d'un thésaurus du domaine minier. In B. Jacquemin, M. Ghenima (Éds.), *La numérisation info-documentaire. Actes du 21e Colloque international sur le Document Numérique* (p. 11-21). Europia

Doerr, Martin, Gradmann Stefan, LeBoeuf Patrick, Aalberg Trond, Bailly Rodolphe and Olensky Marlies. 2013. *Final Report on EDM – FRBRoo Application Profile Task Force* [Final Report]. Europeana. <https://pro.europeana.eu/files/Europeana.Professional/EuropeanaTech/EuropeanaTech.FRBRoo.pdf>

Kergosien, Eric, Daloz Amélie, Chaudiron Stéphane, Jacquemin Bernard. Le CIDOC CRM, un modèle ontologique pour représenter les connaissances du bassin minier des Hauts de France. Premiers résultats. *6th Hyperheritage International Symposium (HIS6) "La fabrique du patrimoine à l'ère numérique"*, Nov 2019, Paris, France.

Le Boeuf Patrick. 2005.. *Functional Requirements for Bibliographic Records (FRBR) : Hype or Cure-All ?*. The Haworth Information Press (Binghamtown, NY), pp.316.

Le Boeuf, Patrick., Doerr Martin, Ore Christian-Emil and Stead Stephen Derek (Éds.). 2015. *Definition of the CIDOC Conceptual Reference Model (Version 6.2.1)*. ICOM/CIDOC Documentation Standards Group, CIDOC CRM Special Interest Group. <http://www.cidoc-crm.org/Version/version-6.2.1>