

La culture dessine son « graphe »

En soutenant des projets d'interconnexion de données et référentiels culturels, le ministère de la Culture et de la communication vise à constituer une base de connaissances, appelée « graphe culture », recourant aux technologies du web sémantique. Ce travail d'harmonisation et de modélisation permettra de décloisonner les différents secteurs concernés.

En lien avec la mission Etalab, créée en 2011, le ministère de la Culture et de la Communication (MCC) coordonne la politique d'ouverture et de partage des données culturelles. Une plateforme dédiée, data.culturecommunication.gouv.fr, a été lancée fin 2016, afin de valoriser les biens culturels et offrir à leurs producteurs des outils de réutilisation, tels que la cartographie ou la data-visualisation. Dans la lignée de cette politique, le MCC s'est très rapidement emparé de la question de l'application du web sémantique au domaine dont il a la responsabilité. Cet enjeu stratégique a conduit le ministère à sensibiliser ses services et établissements publics à ces technologies. Certains établissements publics, comme la Bibliothèque nationale de France (BnF) ou

services et établissements publics à produire, gérer, exposer et lier ses référentiels culturels en utilisant les technologies du web sémantique.

LE PROGRAMME HADOC : DÉCRIRE N'IMPORTE QUEL BIEN CULTUREL

La constitution du « graphe culture » s'inscrit dans la continuité du programme Hadoc, conduit par le ministère depuis 2008, qui vise à l'harmonisation de la production des données culturelles. Le ministère a lancé ce programme en partant du constat que plusieurs secteurs métiers, tels que l'inventaire du patrimoine, les monuments historiques ou encore les musées, des secteurs certes différents mais néanmoins proches, fonctionnaient totalement en

Des secteurs proches, comme l'inventaire du patrimoine, les monuments historiques ou les musées, fonctionnaient en silos et ne communiquaient pas entre eux

silos et ne communiquaient pas entre eux. Le modèle de données Hadoc a vocation à décrire n'importe quel bien culturel. À terme, l'ensemble des descriptions ainsi géné-

le Centre Pompidou, ont été pionniers en la matière. La BnF a lancé la plateforme data.bnf.fr dès 2011, dans une dynamique d'ouverture, de structuration et de liage des données.

UNE FEUILLE DE ROUTE FORMALISÉE EN NEUF AXES

Afin de définir une stratégie sur le web sémantique, globale à l'ensemble de ses services et établissements, le ministère a constitué en 2013 des groupes de travail réunissant des experts de chaque secteur (bibliothèques, archives, musées, musique, audiovisuel et spectacle vivant). Leurs réflexions communes ont abouti à la production d'une feuille de route¹, publiée en 2014 et formalisée selon neuf axes opérationnels et stratégiques, qui couvre les aspects tant techniques ou technologiques qu'organisationnels. Cette feuille de route permet de répondre aux enjeux actuels du web des données liées en définissant notamment un programme de formation des agents du MCC, un partenariat de R&D signé avec l'Inria¹ ou encore des preuves de concept permettant d'expérimenter et démontrer l'interopérabilité sémantique des données culturelles.

Un des neuf axes stratégiques et opérationnels de la feuille de route propose la constitution d'un « graphe culture ». L'objectif pour le ministère est d'inviter ses

rées est amené à constituer le référentiel des biens culturels. Cet important travail d'harmonisation et de modélisation a concerné la partie descriptive du bien culturel, mais également la partie « référentiel » qui l'alimente. Les services concernés ont donc travaillé à mettre en cohérence les thésaurus spécialisés utilisés par les différentes bases de données sectorielles.

L'APPLICATION GINCO : GÉRER LES VOCABULAIRES SPÉCIALISÉS

Ginco², acronyme de Gestion informatisée de nomenclatures collaboratives et ouvertes, a été développé par le MCC dans le cadre du programme Hadoc pour proposer un outil complet de création et de gestion de vocabulaires spécialisés. Ginco implémente la norme ISO-25964³ pour la production des vocabulaires. Cette norme, publiée en deux volets, en 2011 et en 2013, révisé les normes de conception de thésaurus et d'interopérabilité dans la perspective du web des données liées. L'ensemble des services du MCC collaborent à la création, l'harmonisation et l'alignement des référentiels par cet outil, qui permet ensuite d'exposer l'ensemble des vocabulaires scientifiques et techniques en SKOS. La plateforme de diffusion de Ginco, distincte de celle de production, permet une consultation en ligne, un téléchargement au format SKOS ou des requêtes complexes en SPARQL. Ces référentiels

[1] Le MCC et l'Inria ont signé une convention-cadre, le 12 décembre 2016, pour soutenir des projets de R&D culturels innovants.

[2] [culturecommunication.github.io/ginco/](https://github.io/ginco/)

[3] ISO 25964-1(2011) « Thésaurus pour la recherche documentaire » et ISO 25964-2(2013) « Interopérabilité avec d'autres vocabulaires ».

[4] Isni, code international normalisé des noms (en anglais *International Standard Name Identifier*), www.isni.org

[5] Viaf, fichier d'autorité international virtuel (en anglais *Virtual International Authority File*): <https://viaf.org>

du MCC sont alignés avec d'autres référentiels nationaux de la sphère culturelle, comme Rameau de la BnF, ou extérieurs et internationaux, comme celui du Getty ou Iconclass.

LE PROJET ONOMA : IDENTIFIER LES ACTEURS CULTURELS

Au même titre que les vocabulaires spécialisés sur les domaines, les sujets représentés ou les entités abstraites, le champ des acteurs historiques fait également l'objet d'un chantier important dans le cadre du programme Hadoc. Le projet Onoma (en cours) vise à créer un référentiel des acteurs, en s'appuyant sur les technologies du web sémantique. La notion d'« acteurs » recouvre les auteurs, les créateurs, les producteurs, ou toute personnalité intervenant dans le cycle de vie d'un bien culturel. La clé de voûte du projet repose sur l'attribution d'un identifiant pérenne et unique pour ces acteurs. Un des neuf axes stratégiques de la feuille de route propose notamment de déployer le dispositif Isni⁴. La base Isni, qui s'est constituée à partir de la base Viaf⁵, comporte plus de 7,5 millions d'identifiants sur des auteurs patrimoniaux et contemporains. La BnF est devenue une agence d'enregistrement en 2014 et assure l'attribution et la diffusion d'identifiants Isni. L'enjeu pour le MCC, à travers le projet Onoma, est de déployer largement les identifiants Isni pour les auteurs et acteurs historiques peu connus ne disposant pas d'identifiants et de diffuser et interconnecter les données relatives à ceux déjà identifiés. Cette démarche contribuera à la visibilité et au rayonnement des auteurs français à l'échelle nationale et internationale.

L'ensemble de ces référentiels culturels, formalisés selon les standards du web sémantique et interconnectés, participe à la constitution de la base de connaissances appelée le « graphe culture ». L'enjeu pour le MCC est de favoriser l'enrichissement de ce graphe en soutenant des projets d'interconnexion de données et référentiels culturels, mais aussi de veiller à sa maintenance et à son actualisation. La perspective de ce « graphe culture » marque l'engagement du MCC pour une coproduction des connaissances, destinée à désenclaver les bases de données institutionnelles et casser les fonctionnements en silos en misant sur les approches participatives.

MARIE-VÉRONIQUE LEROI

Chargée de mission au département
de l'innovation numérique
Ministère de la Culture et de la Communication
marie-veronique.leroi@culture.gouv.fr

© Library of Congress Prints and Photographs Division



Woman on tricycle,
followed by men on
penny-farthings,
Hy Sandham, 1887.
Une feuille de route
(<http://vu.fr/3r6>)
a été publiée par le MCC,
en 2014, pour répondre
aux enjeux actuels
du web de données.

● ● ● ARCHIPÉDIE, UNE ENCYCLOPÉDIE DÉDIÉE À L'ARCHITECTURE

archipédie.citedelarchitecture.fr

La Cité de l'architecture et du patrimoine (Capa) s'est engagée, ces dernières années, dans la refonte et l'harmonisation de ses sites Web et de ses bases de données. Archipédie, l'un des premiers résultats de cette démarche, se présente comme une encyclopédie numérique collaborative, consacrée à l'architecture moderne et contemporaine, et entend s'adresser au plus grand nombre. La connexion interinstitutionnelle de données a pu être expérimentée, dans la mesure où les notices d'architectes d'Archipédie sont liées par l'identifiant Isni aux notices correspondantes de la base Agorha, de l'Institut national de l'histoire de l'art (Inha). Ainsi une recherche sur un architecte permet de consulter la notice qui lui est consacrée, mais également celle sur ses œuvres ou les collectifs auxquels il a pris part.

● ● ● DOREMUS, UN PROJET DE REPRÉSENTATION SÉMANTIQUE D'ŒUVRES MUSICALES

www.doremus.org

Doremus est un projet financé par l'Agence nationale de recherche et réunissant la BnF, la Philharmonie de Paris, Radio France, et des partenaires universitaires et privés. Doremus s'attache à formaliser et décrire de la façon la plus complète une œuvre musicale. Ce projet s'inscrit pleinement dans la feuille de route pour le Web sémantique du MCC et participe à la constitution du « graphe culture ». Prévu pour être finalisé en 2018, il mettra à disposition des modèles et des référentiels pour la représentation sémantique d'œuvres musicales, afin de rendre interopérables les catalogues d'œuvres existants.