



HAL
open science

INTEROPERABILITE ET CONVERGENCE DES PRATIQUES DE DESCRIPTION A LA BIBLIOTHEQUE NATIONALE DE FRANCE: LA VALORISATION DE LA COLLECTION CHARLES CROS

Marianne Deraze, Xavier Loyant

► **To cite this version:**

Marianne Deraze, Xavier Loyant. INTEROPERABILITE ET CONVERGENCE DES PRATIQUES DE DESCRIPTION A LA BIBLIOTHEQUE NATIONALE DE FRANCE: LA VALORISATION DE LA COLLECTION CHARLES CROS. Communication [Information Médias Théories]: revue québécoise des recherches et des pratiques en communication et information, 2016, Interopérabilité culturelle, 34 (1), 10.4000/communication.6618 . hal-01417692

HAL Id: hal-01417692

<https://bnf.hal.science/hal-01417692v1>

Submitted on 15 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

INTEROPERABILITE ET CONVERGENCE DES PRATIQUES DE DESCRIPTION A LA BIBLIOTHEQUE NATIONALE DE FRANCE : LA VALORISATION DE LA COLLECTION CHARLES CROS

Marianne Deraze, Xavier Loyant

Résumé

L'interopérabilité des données, permise par le développement du web sémantique, semble amorcer un mouvement de convergence des pratiques de description des bibliothèques, centres d'archives et musées (LAM). La stratégie de valorisation de la collection Charles Cros à la Bibliothèque nationale de France est un exemple particulièrement révélateur de ce mouvement et de ses limites, en même temps qu'elle illustre parfaitement l'évolution nécessaire des pratiques professionnelles au sein des LAM.

Mots-clé

Interopérabilité sémantique

Collection Charles Cros

Bibliothèque nationale de France (BnF)

LAM

CIDOC CRM

Depuis une dizaine d'années, le concept de LAM pour « Library, Archive, Museum » s'est développé au sein de la communauté professionnelle de la culture et du patrimoine¹. Derrière cet acronyme, se trouve le point de départ d'une réflexion sur la convergence des problématiques auxquelles sont confrontés ces trois types d'institutions dans un contexte de transformation profonde des usages et des pratiques de leurs publics : à l'heure du numérique, la multiplication et la disponibilité supposée immédiate de l'information auraient transformé

¹ Zorich, 2008 ;Bermès, 2011.

les usagers des bibliothèques, centres d'archives et musées en consommateurs d'information². La bibliothéconomie, l'archivistique et la muséologie sont donc interrogées dans leurs méthodes de traitement et leurs pratiques de description, et appelées ainsi à faire dialoguer les données des LAM et à les rendre accessibles de la manière la plus immédiate possible pour le public. Les plus prudents tentent d'établir des passerelles entre ces trois cultures professionnelles quand les plus ambitieux cherchent à définir un modèle conceptuel de données commun et universel et au sein duquel toutes les données devraient être interopérables, c'est-à-dire lisibles par différents outils et réutilisables par toute institution ou tout particulier. Cela suppose à la fois l'utilisation d'un langage commun, compréhensible par tous, c'est-à-dire le choix d'une norme de structuration des données, et une infrastructure technologique capable d'analyser et restituer cette structuration aux utilisateurs. Constitutifs du web sémantique, ces deux aspects sont indispensables à la visibilité des données sur Internet et donc, à la diffusion des connaissances.

Le département de l'Audiovisuel de la bibliothèque nationale de France (BnF) conserve une collection d'appareils de lecture et d'enregistrement de documents sonores, vidéo et multimédia, dite collection Charles Cros, du nom du concepteur français, et oublié depuis, de ce qu'Edison commercialisa au même moment sous le nom de phonographe³. Riche de plus de 1400 objets, la collection continue de s'accroître mais souffre d'un défaut de visibilité : elle n'est signalée ni décrite dans aucun catalogue et n'est accessible qu'à un public spécialisé, sur rendez-vous. Grâce au soutien d'un programme de recherche, le département de l'Audiovisuel a pu entreprendre, depuis 2013, de remédier à cette situation en développant deux axes de valorisation : la description de la collection dans un catalogue public et la présentation d'une sélection d'objets dans une exposition virtuelle. Cette entreprise de médiatisation d'une collection de type muséal dans une bibliothèque bénéficie d'un contexte favorable au moment où les travaux scientifiques en sciences de l'information se multiplient sur le concept de médiation des savoirs à l'ère numérique et où les grandes institutions patrimoniales voient leurs pratiques converger à la faveur du développement d'outils qui ont permis l'émergence de la notion d'humanités numériques, ou « digital humanities. »

² Badillo, 2015.

³ Charles Cros décrit en 1877 un système qu'il nomme « paléophone » et qui permettrait selon lui d'enregistrer et restituer les sons, sans toutefois produire de prototype. Quelques mois plus tard, Edison présente son « phonographe », fondé sur les mêmes bases techniques que le paléophone rêvé par Charles Cros.

L'objectif de cette étude est de montrer, en nous appuyant sur l'exemple du traitement de la collection Charles Cros, comment la mise en œuvre de l'interopérabilité culturelle repose moins sur la convergence des LAM vers un modèle de description unique que sur la mutualisation, sur la mise en commun des pratiques et des modèles propres à chaque culture professionnelle et désormais utilisables par toutes.

LES INSTITUTIONS PATRIMONIALES FACE AU DEFI DE LA MEDIATION NUMERIQUE

Le concept de médiation des savoirs défini dans les années 1970 comme discours permettant la diffusion des connaissances et favorisant l'interactivité s'est trouvé de nouveau interrogé avec l'apparition et le développement d'Internet, comme en témoignent notamment les travaux de Dominique Cotte⁴ et de Jacques Bonnet⁵, qui explorent le concept de médiation des savoirs à l'ère numérique et en décrivent les évolutions, les enjeux pour les professionnels de l'information et les conséquences sur la sociologie des institutions. Ces travaux montrent l'importance des notions de réception, d'interopérabilité et d'interdisciplinarité en termes de communication numérique. Repris par la communauté des sciences humaines, ils permettent une riche réflexion sur la notion d'humanités numériques⁶ et les évolutions que ce concept sous-entend pour le travail des chercheurs et des professionnels du patrimoine culturel⁷. L'intérêt suscité par le développement de la réflexion sur les humanités numériques en tant que « champ multidisciplinaire à l'intersection des technologies digitales et des sciences humaines⁸ » a conduit chercheurs et professionnels à approfondir la définition de l'environnement technique nécessaire au développement de ce champ. Le web sémantique, défini par Tim Berners-Lee comme « extension du Web actuel, dans lequel l'information [est] fournie avec des significations bien définies, davantage en mesure d'habiliter des personnes et des ordinateurs à travailler en coopération »⁹, est apparu d'une part comme l'environnement naturel des *digital humanities* et, d'autre part, comme un objectif pour les institutions culturelles. Cet objectif ne pourra être atteint que par la mise en œuvre d'une interopérabilité

⁴ Cotte, 2013.

⁵ Bonnet, 2014.

⁶ Pouyllau, 2013.

⁷ Favier, 2015.

⁸ Id, p. 84.

⁹ Berners-Lee, 2001.

technique, qui permettra aux différents systèmes d'exploitation de communiquer entre eux ; d'une interopérabilité sémantique, qui associera à chaque donnée une définition et des relations avec d'autres données permettant aux humains comme aux machines de comprendre son sens ; d'une interopérabilité de gestion, qui verra converger les processus de travail eux-mêmes¹⁰. Processus qui ne saura faire l'économie d'une adaptation des métiers de l'information et de la documentation autour de trois missions : structurer, normaliser et qualifier l'information¹¹.

Ces travaux ont fait prendre conscience aux professionnels de l'information et de la communication de l'intérêt que revêtait une bonne présentation des données qu'ils publient sur Internet. Cette bonne présentation impose d'adapter la structure des données produites pour les rendre accessibles : alors que la bibliothéconomie s'était fondée sur la constitution de catalogues, bases et répertoires qu'il fallait interroger successivement pour en explorer le contenu, Internet impose de refonder ces pratiques pour que des catalogues propres à chaque institution puissent être interrogés par les moteurs de recherche les plus communs sur le web. Il s'agit d'ouvrir les données pour les rendre visibles, et de faire dialoguer les différents réservoirs de données. Se dessine ainsi la perspective d'un web au sein duquel les données produites par les institutions culturelles sont facilement accessibles, mais également facilement récupérables et réutilisables par tous.

Ainsi, les professionnels des bibliothèques n'ont pu que dresser le constat de l'inadaptation de leurs pratiques et de leurs outils au web sémantique : « Parce qu'elles ne sont pas structurées selon les standards du Web, la plupart de notices bibliographiques restent enfouies dans les profondeurs du Web invisible¹². » Forte de ce constat, la Bibliothèque nationale de France a entrepris depuis 2010 une vaste « transition bibliographique », long processus par lequel, progressivement, les données des différents catalogues et bases de la BnF sont convertis dans des formats interopérables. Au cœur de cette transition, le programme de valorisation de la collection Charles Cros s'est trouvé confronté à ces problématiques d'interopérabilité technique, sémantique et de gestion, ce qui a nous a conduits à nous intéresser à un aspect moins développé dans la littérature scientifique, à savoir les conditions, enjeux, et forces à mobiliser pour mettre en œuvre l'interopérabilité sémantique au sein d'une communauté

¹⁰ Bachimont, 2011.

¹¹ Cotte, 2013, *op. cit.*

¹² Illien, 2013.

professionnelle. Le projet de valorisation de la collection Charles Cros, que nous attacherons ici à décrire et analyser, en est un exemple pertinent en ce qu'il montre comment une collection de type muséal et dont la constitution s'apparente à celle d'un fonds d'archives peut être traitée avec des outils bibliothéconomiques garantissant l'interopérabilité des données produites. Pour le démontrer, nous relaterons d'abord brièvement l'histoire de cette collection, ce qui nous permettra de comprendre en quoi les objets qui la composent sont à la fois uniques et liés au contexte dans lequel ils sont entrés dans la collection et y ont été conservés. Cette analyse nous permettra ensuite de comprendre nos besoins en termes de traitement et donc de justifier les choix que nous avons faits dans notre entreprise de production de données interopérables (définition du modèle de données, choix de l'outil de production du catalogue). Enfin nous montrerons, en nous appuyant sur cette expérience, comment l'utilisation des outils du web sémantique contribuent à renouveler, au-delà des compétences spécialisées, les pratiques professionnelles en ce qu'ils favorisent le travail en réseau, renforcent la communauté professionnelle des bibliothèques et dessinent les contours d'une nouvelle communauté rassemblant professionnels des bibliothèques, des musées et des archives.

CONCEPTION D'UNE METHODE DE TRAITEMENT DE LA COLLECTION CHARLES CROS

L'origine de la collection Charles Cros remonte aux origines mêmes du département de l'Audiovisuel, les Archives de la parole, laboratoire fondé en 1911 au sein de la Sorbonne dans le but de constituer un centre d'archives orales : le linguiste Ferdinand Brunot avait notamment pour objectif la constitution d'un atlas sonore des langues¹³. La nécessité de disposer d'appareils d'enregistrement et de lecture (à l'époque, des phonographes) est donc consubstantielle à l'idée même d'archives de la parole. Dans les années 1920, les Archives de la parole deviennent un Musée de la parole et du geste. Les premiers appareils utilisés sont exposés dans les espaces de l'institution et un premier musée du phonographe voit ainsi le jour : c'est l'embryon de l'actuelle collection Charles Cros. Le projet initial change à la fois d'échelle et de nature lorsque le dépôt légal des documents sonores est institué, à partir de 1938 : le législateur crée une institution publique, la Phonothèque nationale, qui se voit confier la charge de collecter le dépôt légal des phonogrammes et de mettre ces documents à la disposition du public. La Phonothèque, dans laquelle le Musée de la parole et du geste finit

¹³ Cordereix, 2005.

par se confondre, poursuit une politique d'acquisitions de phonographes anciens en même temps qu'elle suscite des dons destinés à enrichir l'ancien musée du phonographe.

Au milieu des années 1970, le législateur ajoute au champ du dépôt légal les documents vidéo et multimédia. La charge en est confiée à la Bibliothèque nationale, à laquelle la Phonothèque est rattachée dans un nouveau département de la Phonothèque nationale et de l'audiovisuel. Les appareils constituant la salle rétrospective de la Phonothèque sont alors remisés en attendant l'aménagement et l'ouverture d'un musée qui doit leur être consacré et prendre à cette occasion le nom de Charles Cros. Ce qui n'empêche pas la collection de continuer à s'accroître, d'autant que le département doit à présent faire l'acquisition des appareils permettant de communiquer au public les documents vidéo et multimédia, appareils qui rejoignent à leur tour la collection Charles Cros une fois que les services qui les utilisent n'en ont plus besoin. A partir de 1992, le département reçoit la charge du dépôt légal des documents électroniques, et la collection Charles Cros s'accroît également d'ordinateurs et consoles de jeu. Elle témoigne non plus seulement des missions passées du département, la collecte d'archives orales, mais également de l'étendue de ses collections documentaires, qui couvrent toutes les technologies sonores, vidéo et multimédia, des premiers phonographes à feuille d'étain aux lecteurs numériques les plus récents, et ainsi toute l'édition audiovisuelle française.

L'objectif du projet de catalogage et valorisation de la collection peut se résumer en un principe : mettre en relation les ressources et connaissances dont nous disposons et le public qui pourrait en avoir l'usage. Une mise en relation très peu aisée jusque récemment car non seulement la collection n'était encore décrite dans aucun catalogue public, mais encore parce que le public visé par notre programme (chercheurs, amateurs, spécialistes et curieux) ne penserait pas immédiatement à chercher des informations sur les pratiques et les techniques audiovisuelles dans une bibliothèque, ni à y trouver une collection d'appareils d'enregistrements et de lecture.

Le signalement de la collection sur différentes interfaces est ainsi un préalable à toute possibilité de valorisation ; il nous faut donc produire des données interopérables, échangeables et réutilisables aussi bien par la BnF que par les autres institutions qui possèdent des données sur des domaines similaires. Il a fallu pour cela conceptualiser la collection Charles Cros, dont l'inventaire s'était développé de manière empirique. Nous avons ainsi eu à considérer la collection dans son ensemble et à en définir l'architecture avant de pouvoir nous

intéresser aux appareils eux-mêmes et à la façon dont nous allons décrire chaque objet, analyse préalable à la définition d'un modèle de données puis au choix de l'outil de production. Il s'est donc agi de considérer la collection à une échelle en quelque sorte macroscopique avant de focaliser notre analyse sur les objets pris individuellement, afin d'identifier les types d'information requis pour une description aussi complète que possible de la collection.

Le choix de l'outil de production du catalogue dépendait en effet de la façon dont nous allons présenter la collection Charles Cros : il s'agissait d'en produire un classement qui dépasse les logiques successives de constitution et d'accroissement. Le rattachement de la Phonothèque nationale à la Bibliothèque nationale en 1975 marque en effet une rupture dans la logique d'enrichissement de cette collection jusqu'alors conçue comme un musée. Les appareils, objets désormais atypiques au sein de la Bibliothèque nationale, sont remisés en magasin et la logique d'accroissement s'infléchit. Si l'illustration des techniques audiovisuelles est encore une logique d'enrichissement, la collection se structure à présent autour de l'histoire particulière du département de l'Audiovisuel et non plus des seuls documents sonores : les appareils sont rassemblés en fonction des trois grands domaines médiatiques que constituent documents sonores, vidéo et multimédia. Aujourd'hui, la politique d'accroissement de la collection tente de combiner ces deux logiques en collectant des appareils qui ont servi ou auraient pu servir à consulter les documents diffusés en salle de lecture, et qui sont un reflet de ces collections documentaires, tout en cherchant à rendre compte des techniques et de leur évolution.

Ainsi, si nous avons choisi de montrer comment la collection Charles Cros aujourd'hui prolonge le musée du phonographe d'autrefois, axé sur la technique, nous aurions pu classer la collection Charles Cros selon de grands domaines techniques et définir autant de catégories que nous pouvons distinguer de médias représentés dans cette collection : son, image fixe, image animée, informatique, jeu vidéo. Chaque catégorie aurait ensuite été subdivisée en autant de sous-ensembles correspondant aux évolutions techniques apportées : pour les appareils de lecture et d'enregistrement du son par exemple, un ensemble consacré aux phonographes aurait lui-même été subdivisé en sous-ensembles correspondant aux différents types de phonographes (à feuille d'étain, à cylindre, à disque).

Si nous avons choisi au contraire de présenter la collection selon l'organisation du travail au sein du département de l'Audiovisuel, nous aurions regroupé ces appareils en fonction des

médias auxquels ils se rattachent : médias sonores, médias vidéo et multimédias. Ces ensembles obéissent en effet aujourd'hui à des stratégies d'accroissement distinctes. Chaque catégorie aurait à son tour été divisée en ensembles rappelant la politique d'accroissement menée par chaque institution successive. Cette collection pourrait en effet être traitée comme un fonds d'archives dont les grandes séries auraient été produites l'une après l'autre par ces institutions. La série du matériel technique des Archives de la Parole (1911-1924) serait suivie de la série du Musée de la Parole et du Geste (1924-1938), et ainsi de suite. Les fonds du service des Documents sonores du département de l'Audiovisuel (1975-...), celui du service de l'Images animée (1975-...) et celui du service Multimédia (1992-...) permettraient de rendre compte de la logique d'accroissement propre à chaque producteur.

Cette analyse de la collection a ainsi permis de mettre en évidence son caractère double, entre collection muséale et fonds quasiment archivistique : les objets qui la constituent, s'ils possèdent bien une valeur intrinsèque, ne se comprennent véritablement que reliés les uns aux autres, comme faisant partie d'ensembles et de sous-ensembles eux-mêmes liés les uns aux autres. Notre classement devait donc tenir compte de ces multiples liens et notre modèle de données, permettre de les expliciter.

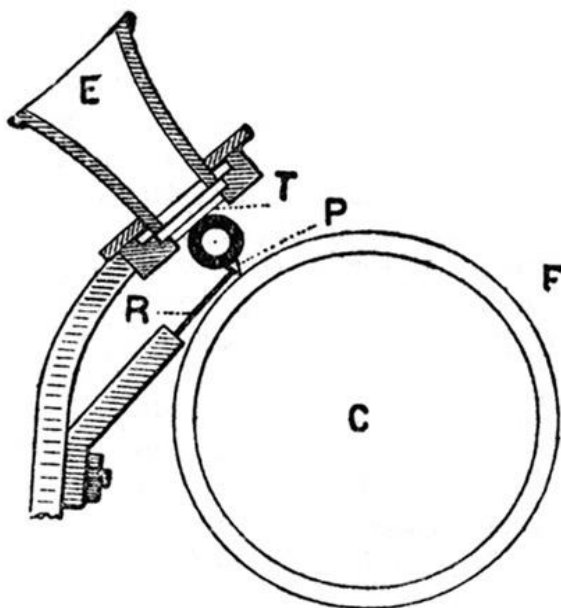
DEFINITION DE L'ARCHITECTURE DE DONNEES

Pour définir une architecture de données propre à rendre compte de la complexité de la collection, restait à combiner cette analyse d'ensemble avec une réflexion sur les données nécessaires à la modélisation de chaque objet, c'est-à-dire sur le type d'informations que nous souhaitons renseigner pour chaque objet à décrire. Il a fallu pour cela tenir compte à la fois du caractère unique de ces objets dans la collection, mais aussi, puisqu'il s'agit pour la quasi-totalité d'entre eux d'objets industriels, de leur caractère sériel. Or, la pratique bibliothéconomique voudrait que nous décrivions les objets pièce à pièce, tels qu'ils se présentent. Mais comment rendre compte de leur caractère industriel, sériel, des ensembles auxquels ils se rattachent et de la complexité même de la collection entière ?

Nous avons répondu à cette question en identifiant trois niveaux de description : le type d'objet, le modèle industriel, l'item présent dans la collection. Nous allons décrire ces niveaux en nous appuyant sur le phonographe Edison Class M, le premier phonographe à cylindre doté d'un moteur électrique, conçu par Thomas A. Edison en 1888, dix ans après l'invention du premier phonographe à feuille d'étain. La collection Charles Cros en possède trois

exemplaires ; l'un d'entre eux provient des Archives de la Parole et les deux autres, de la Phonothèque nationale.

On identifie dans un premier niveau de description les *types* d'objets conservés dans la collection. Ces types sont identifiés par leurs caractéristiques techniques et les composants nécessaires à leur bon fonctionnement. Par exemple, un phonographe à cylindre est caractérisé par son procédé technique, la transcription mécanique d'une onde sonore sur un cylindre, mais encore par un certain nombre d'éléments : présence d'un système d'entraînement, qu'il soit électrique ou mécanique, d'un système d'enregistrement composé d'un diaphragme et d'un pavillon, d'un système de lecture composé d'un diaphragme de lecture ainsi que d'un système d'amplification. Cette définition inclut tous les appareils appelés « phonographes à cylindre » et exclut tous les autres. Le phonographe Edison Class M fait donc partie du type d'objet « phonographes à cylindre » puisqu'il rassemble toutes les caractéristiques techniques (voir Figure 1, Coupe transversale et schématique d'un phonographe).



Coupe transversale et schématique d'un phonographe : C, cylindre enregistreur ; F, manchon de cire ; E, embouchure ; T, membrane vibrante ; R, ressort maintenant l'aiguille P.

Figure 1. Coupe transversale et schématique d'un phonographe à cylindre.

Dictionnaire du phono [En ligne].

<http://www.hervedavid.fr/francais/phono/dico%20phono.htm>.

Page consultée le 29 septembre 2015.

Un second niveau de description est dédié à l'objet industriel, c'est à dire au *modèle*, à la *série* de production. Un objet industriel n'est, par essence, pas unique : produit en grand nombre, seul le numéro de série permet de distinguer un exemplaire d'un autre. Ici, ce sont donc les caractéristiques communes à l'ensemble de ces objets d'une même série ou d'un même modèle qui doivent être décrites. Les caractéristiques communes à l'ensemble des phonographes Class M produits par Edison sont donc renseignées dans ce niveau de description, consacré au modèle industriel ou, en termes bibliothéconomiques, à l'*objet conceptuel* : ces objets, construits à partir des mêmes plans (voir Figure 2, Schéma du Class M), ont un même concepteur, un même producteur, portent la même marque commerciale et sont vendus sous la même appellation. Le Class M a ainsi été créé par Thomas A. Edison et commercialisé sous la marque *Edison Electric Phonograph* par la *National Phonograph Company* entre 1888 à 1911. L'objet industriel a connu des modifications considérées comme mineures, parce qu'elles ne touchaient ni au fonctionnement technique, ni aux composants nécessaires ni à l'appellation commerciale de l'objet, qui sont donc considérées comme des versions différentes d'un même objet industriel.

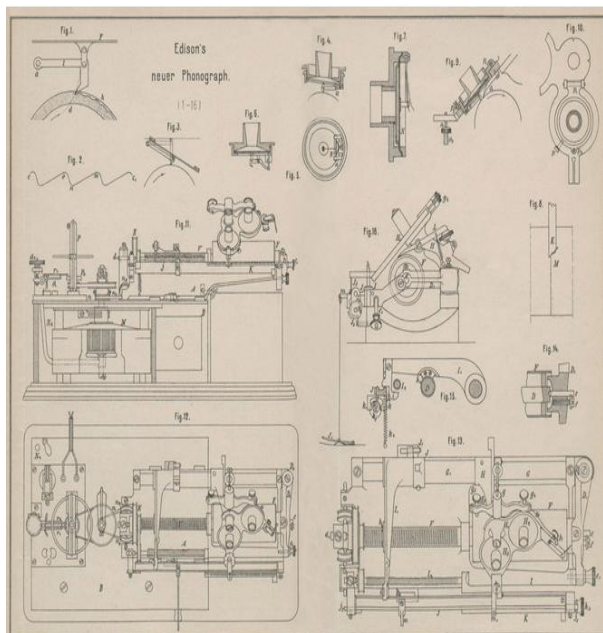


Figure 2. Schéma du phonographe Edison Class M.

Phonorama [En ligne]. <http://www.phonorama.fr/phonographes-edison.html>. Page consultée le 29 septembre 2015.

Le dernier niveau de description est dédié à l'objet physique tel que conservé dans la collection Charles Cros, ou *item*. Ici, c'est l'objet tel qu'il nous est parvenu qui nous intéresse,

et les différents événements qui ont marqué son histoire et lui donnent son caractère unique. Le Class M n° 61 à l'inventaire de la collection Charles Cros (voir Figure 3, Phonographe à cylindre Edison Class M, collection Charles Cros n° 61) se distingue ainsi de tous les autres Class M par son numéro d'inventaire, son numéro de série, ses spécificités de production, et les modifications qui lui ont été apportées, que ces modifications soient techniques, comme celles que peut apporter un propriétaire à un objet qu'il personnalise (remplacement de certaines pièces par exemple), ou bien le fruit de dégradations ou restaurations.

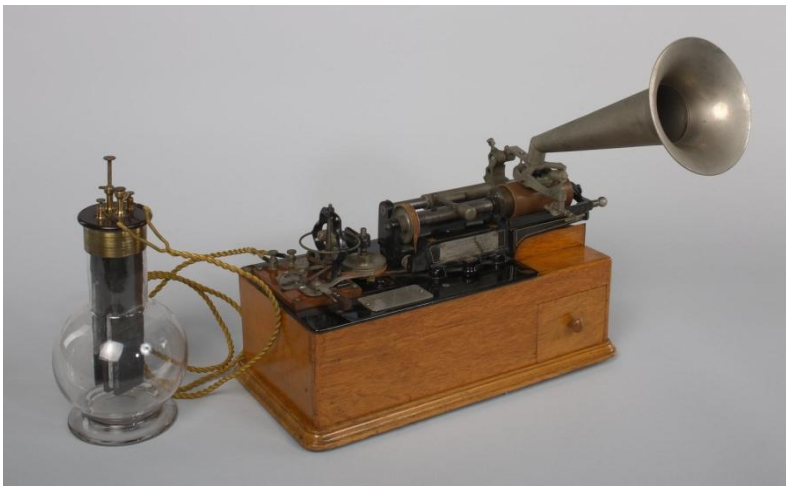


Figure 3. Phonographe à cylindre Edison Class M, collection Charles Cros n° 61 [En ligne]. <http://gallica.bnf.fr/ark:/12148/btv1b55009433w.r=edison.langFR>. Page consulté le 29 septembre 2015.

L'analyse de la collection nous a donc conduits à nous interroger sur les données essentielles à la modélisation non seulement d'une série d'objets, mais également d'un ensemble cohérent où les liens entre chaque item sont signifiants, formant une collection elle-même liée aux autres ensembles patrimoniaux que sont les collections audiovisuelles de la BnF.

La mise à plat de ces informations nous a permis de construire une architecture de données qu'il fallait ensuite mettre en œuvre en utilisant les outils développés par la BnF. Nous avons pour cela étudié les modèles conceptuels tirés des logiques bibliothéconomique, archivistique et muséologique, ainsi que les outils de production disponibles. C'est en comparant les vertus respectives de chacun d'entre eux que nous avons pu choisir la solution la mieux adaptée à nos besoins. Notre choix a ici été conditionné par la manière dont nous voulions présenter et

décrire la collection Charles Cros et par les types d'institutions identifiées comme possédant des collections proches par leur nature des objets constituant cette collection¹⁴.

DEFINITION DU MODELE CONCEPTUEL

UNE OPTION BIBLIOTHECONOMIQUE : LE CATALOGUE GENERAL DE LA BNF ET LES FRBR

Le catalogue général est l'outil de production de données le plus ancien et le plus utilisé à la BnF. Il a longtemps reposé sur une logique de description des objets pièce à pièce et sur les fondements de la bibliothéconomie telle qu'elle s'est élaborée au début du XX^e siècle, pour la description des éditions. Il est de ce fait adapté au traitement des monographies et périodiques édités, généralement reçus au titre du dépôt légal. Depuis 2010, la BnF a entrepris d'en réformer la structure pour y intégrer le modèle FRBR¹⁵, développé et utilisé par les professionnels des bibliothèques comme un modèle commun de production de données interopérables¹⁶. Les données, structurées selon les FRBR, sont ainsi liées d'une manière nouvelle.

La pratique de catalogage traditionnelle décrit chaque édition d'un document dans une notice bibliographique¹⁷ comportant plusieurs champs de données. Le document n'est accessible qu'à partir de sa notice bibliographique et aucune de ces données n'est réutilisable. Les FRBR permettent au contraire de décrire les différents aspects d'un document (depuis son caractère unique, matériel jusqu'à sa signification intellectuelle), de lier toutes les données contenues dans les notices bibliographiques entre elles, et de faciliter leur recherche et leur réutilisation. Ainsi le chercheur accède en seul point d'accès, la page dédiée à l'œuvre, à l'ensemble des

¹⁴ Le travail de modélisation de la collection a été élaboré en collaboration avec le département de l'Information bibliographique et numérique (IBN), avec pour fil conducteur de garantir à la fois la justesse et la finesse de la description, mais aussi l'interopérabilité sémantique des données produites.

¹⁵ FRBR : *Functional Requirements for Bibliographic Records* ; en français, Fonctionnalités requises des notices bibliographiques.

¹⁶ Tillet, 2004.

¹⁷ La notice bibliographique rassemble les données concernant une édition présente dans la bibliothèque et les « données locales », correspondant aux informations de gestion utiles pour retrouver et communiquer les exemplaires de cette édition.

éditions de cette œuvre et, pour chaque édition, à l'ensemble des exemplaires de cette édition présents dans les collections de la BnF (voir figure 4, Les quatre niveaux de description du modèle FRBR). Appliqués au signalement d'un exemplaire des *Misérables*, de Victor Hugo, les FRBR décrivent dans un premier niveau les caractéristiques de l'exemplaire que le lecteur a en main, comme sa cote, son adresse de rangement ou son état de conservation. Le second niveau d'information décrit les caractéristiques de l'édition, caractéristiques communes à tous les exemplaires de cette édition comme la date, le nom de l'éditeur ou encore le nombre de pages. Un troisième niveau décrit le contenu intellectuel de l'exemplaire : c'est le texte d'Hugo qui est décrit, en signalant par exemple sa langue originale et les langues dans lesquelles il a été traduit. Enfin, un quatrième niveau, plus générique encore, décrit la création abstraite, l'œuvre à laquelle se rattache le contenu : ici, « *Les Misérables* de Victor Hugo » ; toutes les éditions et traductions de cette œuvre sont signalées sur la page qui lui est consacrée, ce qui les rend bien plus aisément repérables pour l'utilisateur.

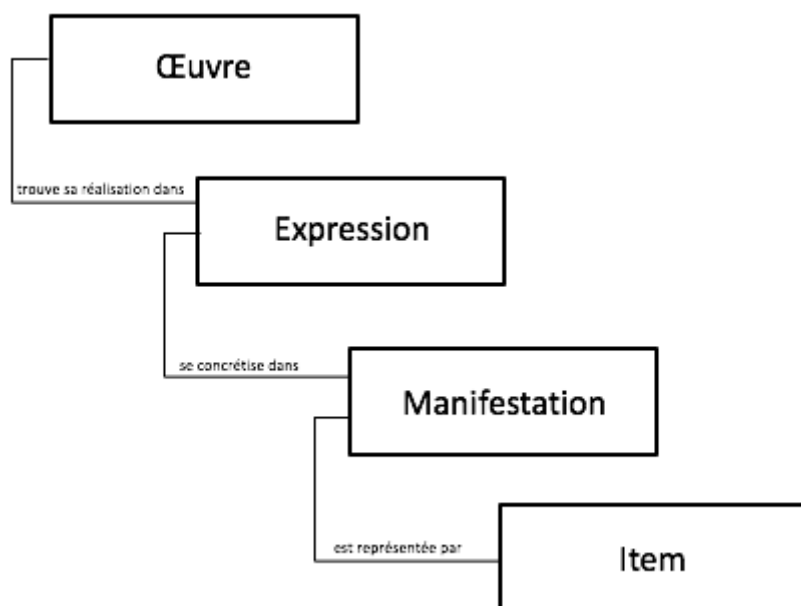


Figure 4. Les quatre niveaux de description du modèle FRBR : œuvre, expression, manifestation et item. D'après : IFLA, *Les Fonctionnalités requises des notices bibliographiques*, deuxième édition française établie par la Bibliothèque nationale de France, p. 14 [En ligne].

http://www.bnf.fr/documents/frbr_rapport_final.pdf Page consultée le 29 septembre 2015.

On pourrait tenter d'appliquer le même traitement à un appareil de la collection Charles, considérant non plus un exemplaire d'une édition des *Misérables*, mais un exemplaire d'un

modèle donné de phonographe, comme un des trois phonographes Edison Class M présents dans la collection (voir Figure 5, Schéma simplifié des correspondances entre la structure des données définie pour la collection Charles Cros, le modèle FRBR et le catalogue général de la BnF). Chacun de ces trois Class M se distingue par de légères différences de forme, qui s'expliquent par leur histoire individuelle, et par leur appartenance à des sous-ensembles différents à l'intérieur de la collection (fonds des Archives de la parole pour l'un, de la Phonothèque nationale pour les deux autres).

Le niveau de description correspondant au *modèle* pour la collection Charles Cros rassemble des informations contenues en FRBR dans les niveaux de description de l'œuvre (son nom, le Class M) et de son expression (caractéristiques techniques et accessoires communs par exemple). Dans le catalogue général de la BnF, ces deux types d'informations (œuvre et expression) ne sont pas contenus dans la notice bibliographique mais dans la notice d'autorité de titre conventionnel (TIC), à laquelle les notices bibliographiques sont liées.

Le niveau de description correspondant à l'*item* dans le modèle défini pour la collection Charles Cros rassemble quant à lui des informations contenues en FRBR dans les niveaux de description de la manifestation (comme la date de production/édition ou la version du modèle décrit) et de l'item (accessoires présents, pièces remplacées ou encore lacunes). Dans le catalogue général de la BnF, les données sur la manifestation et les données sur l'item sont contenues dans la notice bibliographique.

Reste à signaler le *type* d'objet, en l'occurrence un phonographe à cylindre. On pourrait envisager de le faire en créant des notices d'autorité dans RAMEAU, le répertoire d'autorités-matières encyclopédique et alphabétique unifié, utilisé par la BnF pour l'indexation par sujet : une notice consacrée au phonographe à cylindre décrirait les caractéristiques communes à l'ensemble des phonographes à cylindre, à laquelle on rattacherait toutes les notices de modèles de phonographes à cylindre, dont la notice du modèle Class M.

Enfin, l'appartenance de ces objets à l'ensemble qu'est la collection Charles Cros pourrait être indiquée par un lien de la notice d'autorité de titre « Phonographe Class M d'Edison » vers une notice d'autorité titre parente « Collection Charles Cros ».

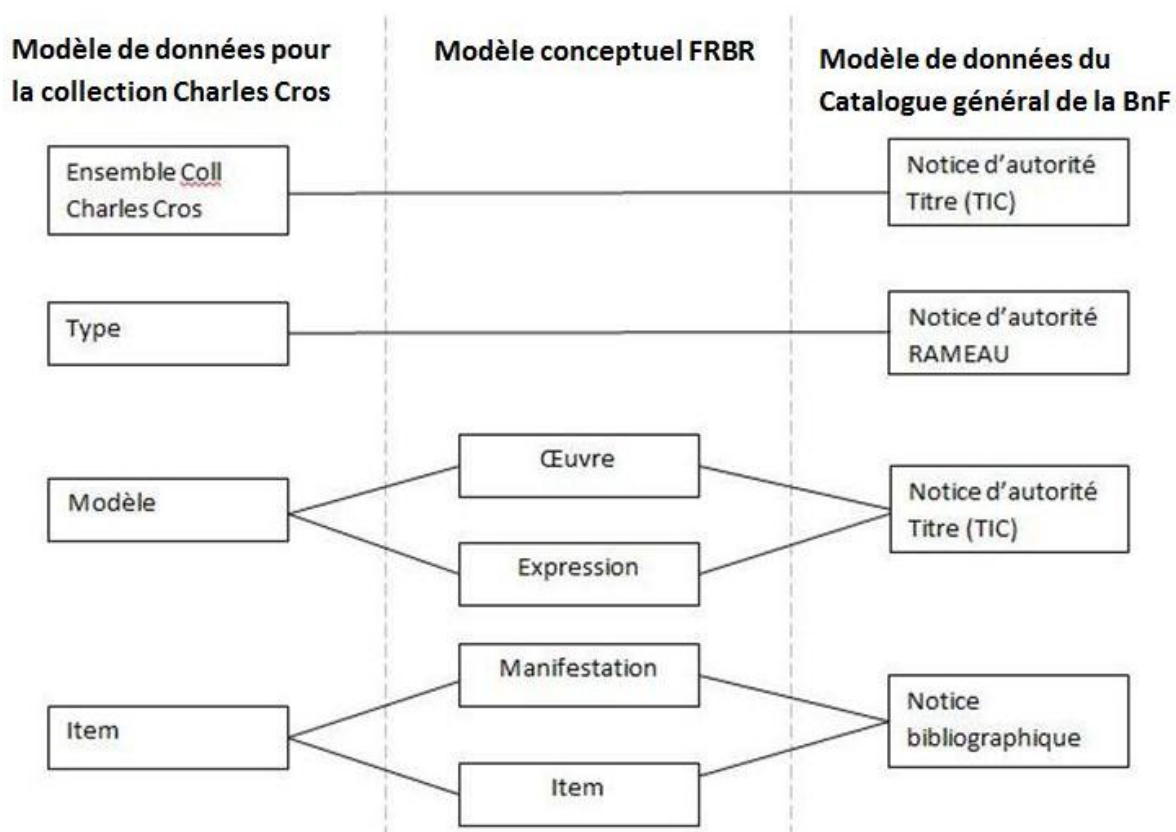


Figure 5. Schéma simplifié des correspondances entre la structure des données définie pour la collection Charles Cros, le modèle FRBR et le catalogue général de la BnF.

Or, on le voit, si les FRBR et le catalogue général nous permettent de décrire efficacement les objets selon les trois niveaux d'informations que nous avons caractérisés, ils ne nous permettent pas, en revanche, d'explicitier les liens qu'entretiennent ces objets les uns avec les autres ni de les contextualiser. Autrement dit, si nous avons décrit les trois Class M de la collection Charles Cros dans le catalogue général de la BnF, nous n'aurions pas pu décrire les sous-ensembles différents dont ils font partie. La mention de leur appartenance au fonds des Archives de la Parole ou à celui de la Phonothèque nationale ne serait apparue que dans le niveau consacré à la description de l'item, sans qu'on puisse lier les appareils appartenant au même sous-ensemble.

De plus, les correspondances effectuées entre l'entité « Œuvre » du modèle FRBR et le « modèle industriel » restent approximatives et ne rendent pas compte de certaines distinctions fondamentales, par exemple entre *l'auteur* d'une *création* originale et le *concepteur* d'un *objet* industriel. Comment qualifier, dès lors, l'inventeur du procédé technique qui est à l'origine de l'objet ? Il ne s'agit pas d'un auteur à proprement parler, et son invention ne saurait être assimilée à une création originale. Les définitions, même si elles

peuvent sembler proches, sont différentes y compris sous l'angle juridique car le droit distingue ces deux types de création.

Décrire la collection Charles Cros selon le modèle FRBR du catalogue général de la BnF nous aurait donc confrontés à deux limites. D'abord, le modèle FRBR, développé par la communauté des bibliothèques, ne propose pas de structure directement applicable aux objets industriels et au modèle défini pour la collection Charles Cros. D'autre part, le catalogue général a pour logique la description des documents pièce à pièce. Or les particularités de la collection Charles Cros résident précisément dans ses logiques successives d'accroissement et dans les multiples liens entre les objets qui la composent et que nous souhaitons expliciter.

UN CHOIX MUSEOLOGIQUE : LE CIDOC CRM

Nous nous sommes donc tournés vers un autre modèle de données, plus propice à témoigner de la richesse sémantique et de la complexité d'une collection d'objets puisqu'il a été développé par la communauté professionnelle des musées¹⁸. Il s'agit du CIDOC CRM, pour « Conceptual reference model » ou « Modèle conceptuel de référence », publié par le Comité international pour la documentation (CIDOC) au sein de l'ICOM, le Conseil international des musées. Il « inclut la description détaillée d'objets individuels qui font partie des collections, ainsi que des groupes d'objets et des collections dans leur ensemble¹⁹, » soit exactement ce que ne nous permettent pas de réaliser les FRBR dans le catalogue général de la BnF. Le CIDOC CRM a un autre avantage : c'est « un modèle souple qui peut être utilisé partiellement ou augmenté d'ajouts²⁰ ». Autrement dit, l'utilisation de ce modèle nous permettrait une grande flexibilité, indispensable pour décrire avec justesse des objets très différents les uns des autres.

¹⁸ La communauté des musées et celle des bibliothèques a collaboré pour établir une extension du CIDOC CRM appliquée aux objets dans les bibliothèques. C'est le modèle FRBRoo, pour « orienté objet. ». Ce modèle permet de décrire des objets dans une optique et une pratique bibliothéconomique. Les objets de la collection Charles Cros ayant été dès l'origine décrits dans une optique non pas bibliothéconomique mais bien muséologique, et la collection ayant été considérée dès l'origine comme un musée et non une collection de bibliothèque, le modèle FRBRoo n'est pas apparu adapté à la description de la collection.

¹⁹ Szabados, 2012

²⁰ *Ibid*

Le modèle de données que nous avons défini est donc en partie dérivé du CIDOC CRM, qui prend pour valeurs fondamentales le temps, la durée, l'évènement, autant d'éléments qui constituent l'histoire d'un objet de musée. Le CIDOC CRM distingue ainsi deux entités essentielles : les entités temporelles, qui se définissent par un début et une fin : un état de conservation, une période historique, un évènement, une activité ; les entités persistantes (ou les *endurants*) qui influent sur les entités temporelles : personnes, choses matérielles, objets conceptuels, propositionnels ou symboliques. Ces entités sont définies dans ce modèle par leurs relations entre elles, grâce au système de triplet qui organise aujourd'hui le web sémantique et permet aux outils informatiques de définir le sens d'une donnée (les relations qu'elle a avec d'autres données) au moyen d'algorithmes. Le modèle de triplet permet ainsi une automatisation croissante des processus de production de données.

Reprenons l'exemple du phonographe Edison Class M (voir figure 6, Architecture de données de la collection Charles Cros appliquée au phonographe Edison Class M n°61 et sa représentation dans le modèle CIDOC CRM). Cet objet est un phonographe à cylindre. C'est aussi un objet industriel appelé Class M ainsi qu'un exemplaire unique. Il fait partie de la collection Charles Cros et du sous-ensemble que représente le fonds des archives de la Parole. Le modèle CIDOC CRM permet de rendre compte de toutes ces strates d'informations : l'*item* est associé à une entité « objet matériel », tandis que les niveaux *type* et *modèle* entrent dans la catégorie d'entités appelées « type ». La distinction entre ces deux notions est faite par une relation entre les deux entités : le type d'objet est le terme générique du modèle. De plus, la logique du CIDOC CRM reposant sur la notion de temps, chaque objet peut être inscrit dans un temps long (on décrira ses caractéristiques permanentes) ou court (on décrira des événements, comme une date d'acquisition, une date d'exposition, ou de dégradation).

Le CIDOC CRM a également l'avantage de faire apparaître les relations entre l'objet décrit et son environnement. Le Class M n° 61 à l'inventaire de la collection Cros fait référence, en tant qu'objet *matériel*, à la description de la collection Charles Cros et de sa gestion. L'objet *conceptuel* fait quant à lui référence à la description du fonds des appareils des Archives de la Parole dont il fait partie. Ce fonds est lui-même décrit comme un objet conceptuel.

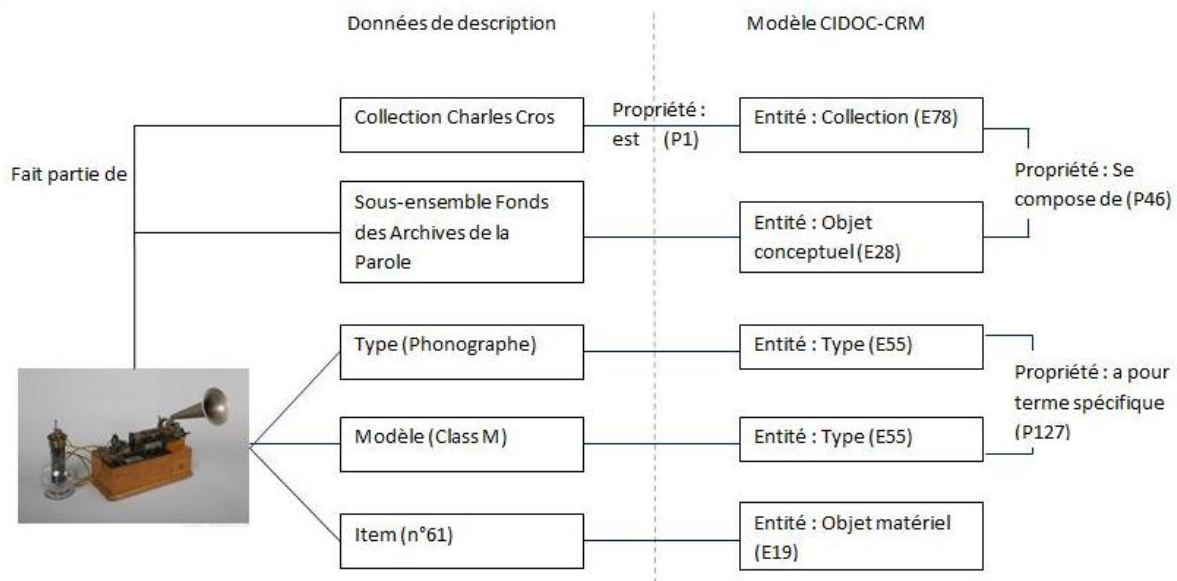


Figure 6. Architecture de données de la collection Charles Cros appliquée au phonographe Edison Class M n° 61 et sa représentation dans le modèle CIDOC CRM.

En prenant comme modèle le CIDOC CRM pour construire notre architecture de données, nous garantissons que notre catalogue en ligne sera fondé sur une logique muséologique largement approuvée par la communauté des musées. Nous préparons la voie à une éventuelle réutilisation de nos données par des institutions similaires ou parentes, ce que permettra également le langage de description que nous avons choisi.

LA BASE BNF ARCHIVES ET MANUSCRITS, OUTIL FLEXIBLE

Le modèle de données que nous avons élaboré pour la collection Charles Cros a ensuite été traduit dans un outil modulable. Développée à partir de 2002 pour répondre à un besoin très précis, celui de pouvoir décrire les manuscrits, la base BnF archives et manuscrits (BAM) est finalement peu à peu devenue l'équivalent d'un catalogue des fonds spécialisés de la Bibliothèque nationale de France. Contrairement au catalogue général, la description dans BAM obéit à une logique archivistique. Elle utilise le langage informatique XML et un vocabulaire de balises normalisé, l'EAD (*Encoded Archival Description*) pour mettre en œuvre la norme internationale archivistique ISAD(G)²¹ (*International Standard Archival Description-General*, en français : Norme générale et internationale de description archivistique). La base permet ainsi de décrire des fonds organiques en faisant apparaître leur plan de classement, structuré hiérarchiquement autour des notions de fonds, séries et

²¹ International Council on Archives, 2000

documents. Elle permet donc de montrer les relations existant entre différents objets décrits individuellement. Nous avons donc choisi cette base comme outil de production car elle permet de traiter la collection Charles Cros comme une collection organiquement constituée, tout en rendant compte de la typologie technique des appareils évoquée plus haut. Il est ainsi possible d'y décrire l'ensemble des niveaux constitutifs de la collection : type d'objet, modèle et item, ainsi que les sous-ensembles auxquels ils se rattachent.

De plus, BAM est une des principales plateformes de mise en œuvre de l'interopérabilité technique et sémantique de la Bibliothèque, et ce, grâce au standard d'encodage, ou vocabulaire de balises normalisé utilisé pour y décrire les fonds, l'EAD.

UN LANGAGE FORMALISE POUR DES DONNEES ENRICHIES

C'est précisément l'utilisation de ce standard qui nous donne la possibilité d'irriguer les autres outils de production et de diffusion de la BnF avec nos données. L'EAD utilise en effet le langage de balisage XML, qui a pour principal atout sa relative simplicité, ce qui facilite sa prise en main par des producteurs de données qui ne sont pas nécessairement familiers avec les langages informatiques : sa structure se veut en effet proche de la structure du langage humain et repose sur une hiérarchie bien identifiée de composants ainsi que sur le principe de subsidiarité²². C'est ce qui permet de décrire des collections organiquement constituées, et pas uniquement des objets à la pièce²³. Les bibliothèques ont rapidement compris quel intérêt elles pouvaient en tirer, notamment pour le traitement des collections spécialisées. Aujourd'hui, le nombre de balises EAD, c'est-à-dire les types d'informations disponibles pour décrire un fonds, s'élève à plus de 145. Il est ainsi possible de décrire de manière très fine tous types de documents. De sorte que, si nous avons finalement choisi de classer la collection Charles Cros selon une typologie technique, la finesse de description du langage EAD nous permet de procéder également à un classement par type de médias : chaque appareil se verra attribuer une balise « type d'objet » et une balise « type de média. » On pourra accéder à l'ensemble des appareils relevant d'un même type de média ou d'un même type d'objet en interrogeant les index, listes regroupant les données d'autorité du catalogue général de la BnF et réutilisées dans BAM qui permettent un accès spécifique : par personne physique, morale,

²² Cela signifie qu'un composant peut contenir plusieurs sous-composants, mais qu'aucun de ces sous-composants ne peut faire partie de plusieurs composants.

²³ Dalbin, 2011

par type de document, par titre, etc. Ainsi notre description de la collection Charles Cros rend-elle compte des différentes logiques qui continuent de présider à son accroissement.

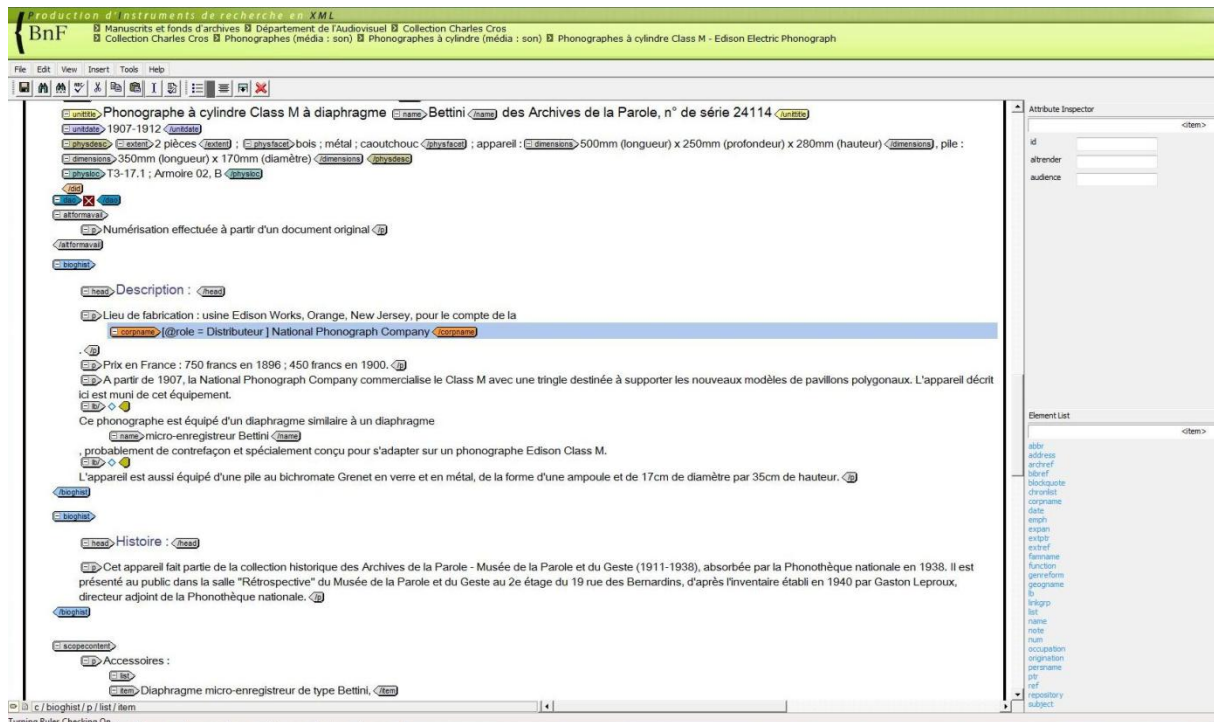


Figure 7. Fiche descriptive du phonographe Edison Class M, collection Charles Cros n° 61, encodée en EAD (capture d'écran des premiers éléments de description de l'item dans l'interface professionnelle).

C'est également sur le langage EAD que repose la construction de passerelles garantes d'interopérabilité entre les différents catalogues de la Bibliothèque nationale de France, prérequis indispensable à notre projet de valorisation de la collection dans une exposition virtuelle.

Cette exposition virtuelle doit se développer dans Gallica, la bibliothèque numérique de la BnF, qui récupère les données des notices des différents catalogues et bases de manière automatique : les informations produites dans BAM sont reportées dans Gallica et associées à chaque objet numérisé. Cette opération est rendue possible par un *mapping*, c'est-à-dire une cartographie des correspondances entre les métadonnées typées dans des balises spécifiques : balises en XML EAD dans BAM, balises en Dublin Core dans Gallica. Mais ce mapping reste partiel car seules 15 balises EAD sont reportées dans Gallica, fournissant les éléments essentiels de la description : identifiant pérenne de la notice, titre, date, type de document, format, provenance, description sommaire et droits de propriété intellectuelle attachés. La

notice Gallica ne livre donc aucun élément de contextualisation, ni aucune information sur le traitement du fonds dont sont issus les documents ou objets numérisés(voir figure 7, Schéma des correspondances entre l'architecture de données définie pour la collection Cros en EAD dans BAM et les données telles qu'elles se présentent en Dublin Core dans Gallica.). Pour plus d'informations, le « gallicanaute » est renvoyé à la notice complète de l'objet dans le catalogue où il a été décrit.

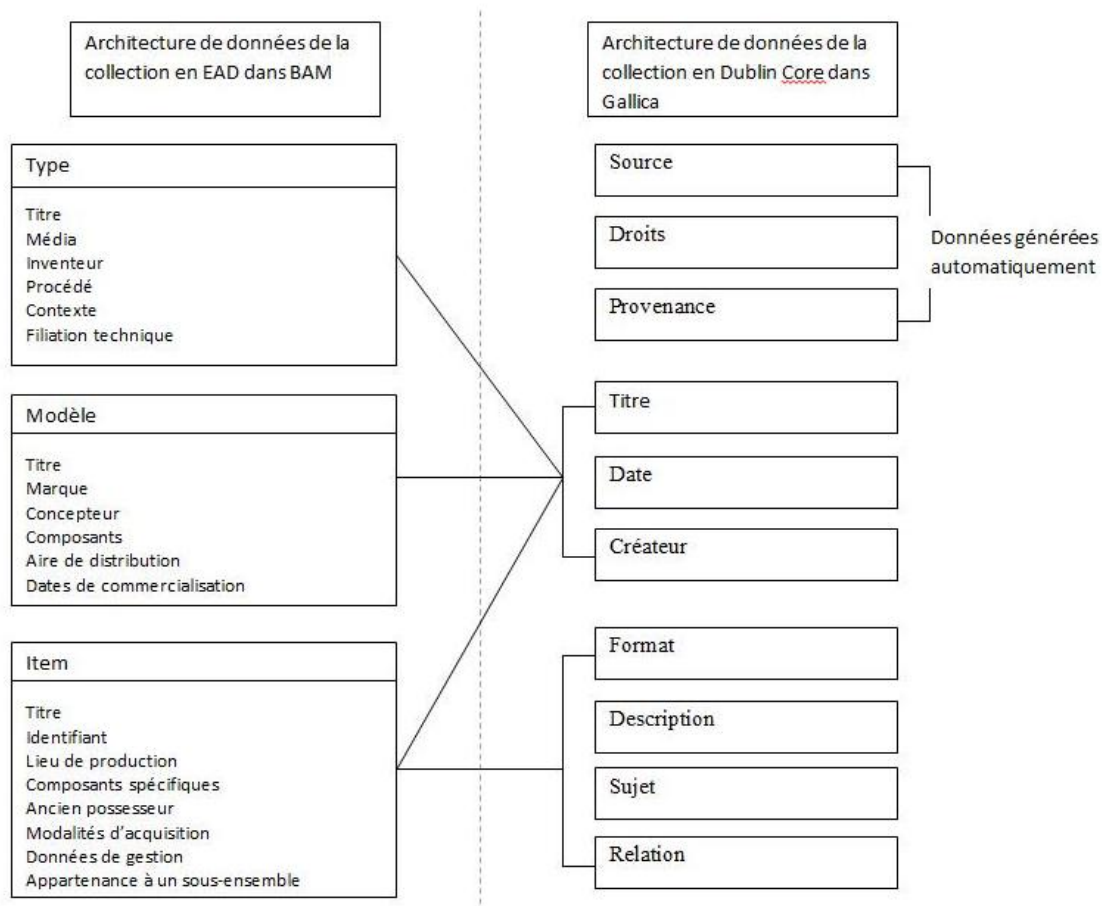


Figure 8. Schéma des correspondances entre l'architecture de données définie pour la collection Cros en EAD dans BnF archives et manuscrits et les données telles qu'elles se présentent en Dublin Core dans Gallica.

L'interopérabilité entre ces deux outils est donc assurée, d'un point de vue sémantique, par une formalisation du langage et des mots utilisés pour décrire, tels qu'ils sont partagés par tous les producteurs de données d'un même outil de production. D'un point de vue technique, par une passerelle : le *mapping* profite de la formalisation du langage pour automatiser le processus de correspondances. L'information, saisie une seule fois, est accessible en deux points différents (BAM et Gallica) destinés à deux usages et deux publics

différents(chercheurs et grand public), et qui correspondent au moyen de liens renvoyant l'utilisateur de l'un à l'autre. L'outil de production que nous avons adopté permet donc de rendre compte de la complexité de la collection Charles Cros tout en proposant différents niveaux de granularité de l'information. Nous avons ainsi élaboré une structure de données à partir d'un modèle conceptuel proposé par la communauté professionnelle possédant les collections et les données les plus similaires à la collection Charles Cros, celle des musées. Nous avons enfin veillé à la formalisation de ces données dans un langage largement utilisé par notre communauté professionnelle. BAM est un point d'entrée dans le réseau des outils de la BnF en ce que cette base permet de reverser les données dans les autres réservoirs de la BnF, notamment Gallica, la bibliothèque numérique de la BnF et data.bnf.fr, agrégateur de données d'autorité.

En présentant l'outil que nous avons conçu pour produire et diffuser les connaissances sur la collection Charles Cros, en détaillant nos choix conceptuels et nos réalisations, nous avons cherché à mettre en valeur les principaux enjeux que soulève la mise en œuvre de l'interopérabilité sémantique et technique pour les producteurs de données et les institutions culturelles au sein desquelles ils exercent. Ces enjeux sont autant internes, car les producteurs de données ont aujourd'hui besoin d'outils de production connectés les uns aux autres, qu'externes, car ils ont également besoin de mettre en place des pratiques d'indexation collaboratives pour optimiser l'usage de ces outils connectés. L'interopérabilité des outils de production de données invite ainsi les professionnels des institutions patrimoniales à réfléchir à l'évolution de leurs propres pratiques. Comme le souligne Emmanuelle Bermès²⁴, « la démarche de l'internaute n'est pas centrée sur les institutions mais sur les contenus. » Pour l'internaute, peu importe finalement d'où provient l'information, pour peu qu'il la trouve et qu'elle lui soit utile. En retour, les professionnels du patrimoine culturel ne sont-ils pas invités à converger vers un modèle conceptuel commun et applicable à la description de tous les objets culturels ? Peut-on revendiquer la libre circulation des données culturelles, des connaissances, véhiculées par un ou des modèles communs, sans voir s'émousser les spécificités propres à chaque culture professionnelle ?

Le cas de la Bibliothèque nationale de France, développé à travers l'étude de la valorisation en ligne de la collection Charles Cros, apporte quelques éléments de réponse sur le

²⁴ Bermès, 2011, *op. cit.*

développement de l'interopérabilité sémantique, ses enjeux et ses conséquences pour les professionnels.

MISE EN ŒUVRE DE L'INTEROPERABILITE SEMANTIQUE A LA BnF

DES OUTILS DE PRODUCTION CONNECTES

La BnF s'est dotée d'outils connectés les uns aux autres, qui allègent le travail des producteurs de données, garantissent une interopérabilité technique, et dessinent l'amorce d'une interopérabilité sémantique en matière d'indexation.

En effet, nous l'avons vu, BAM s'appuie sur un langage formalisé pour le web de données. Il prend pour ce faire en compte les évolutions les plus récentes au sein des communautés professionnelles des archives et des bibliothèques. Le catalogue général de la BnF et notamment ses fichiers d'autorité (personnes, organisations, noms communs), remontent à la fondation des principes bibliothéconomiques définis au début du XX^e siècle. Un des enjeux majeurs pour les producteurs de données est de pouvoir réutiliser et enrichir les données d'autorités de la BnF quel que soit l'outil sur lequel ils travaillent. Les concepteurs de BAM ont donc développé un outil permettant l'intégration automatique des données du catalogue général dans les métadonnées d'une notice EAD. Ce développement a notamment permis de résoudre une des limites de l'interface publique BAM, qui réside dans l'impossibilité de proposer des instanciations multiples. Par exemple, s'il est impossible de faire figurer dans les balises EAD que le type d'objet « phonographe à disque » peut aussi s'appeler, tout aussi légitimement, « gramophone », terme généralement utilisé dans les sources anglo-saxonnes, la base de production permet d'établir un lien avec les données d'autorité gérées par la BnF et affichées dans son catalogue général. Le terme de phonographe à disque, contenu dans une balise EAD, sera associé à l'identifiant pérenne donné à la notice d'autorité renvoyant à ce terme dans RAMEAU²⁵, la liste d'autorités normalisées coordonnée par la BnF. Cette notice d'autorité consacrée au phonographe à disque signalera dans les termes équivalents celui de « gramophone », et donnera accès à la totalité des documents du catalogue général ayant indexé le terme de « phonographe à disque ». Ainsi, l'internaute cherchant des informations sur les gramophones sera redirigé vers les informations détenues sur les phonographes à disque.

²⁵Répertoire d'autorité-matière encyclopédique et alphabétique unifié

Mais pour que ces différents outils soient utilisés au maximum de leurs possibilités, il est indispensable que les agents qui les prennent en main en connaissent parfaitement la logique et le fonctionnement. Par exemple, pour décrire la collection Charles Cros dans BAM, le catalogueur doit également connaître la logique et les possibilités du catalogue général et de RAMEAU : si le catalogue général propose des notices d'autorité *personne, organisation* et *marque* qui peuvent être réutilisées dans BAM, c'est en revanche le répertoire RAMEAU qui permet l'indexation des types d'objet dans BAM. C'est encore au catalogueur d'effectuer une recherche manuelle dans ces différents outils pour garantir une indexation fine et détaillée de la collection. C'est encore à lui de maîtriser le *mapping* qui permet d'établir des correspondances entre les notices produites dans BAM et celles générées dans Gallica : pour que les notices générées dans Gallica contiennent les informations les plus essentielles, il est indispensable de savoir quelles balises EAD sont converties en Dublin Core, pour savoir quelles balises utiliser pour décrire les éléments qu'on souhaite voir figurer dans la notice Gallica. Par exemple, Gallica affiche les informations concernant l'auteur d'un document. Mais le concepteur d'un objet industriel n'est pas considéré comme tel. Les objets de la collection Cros présentés dans Gallica n'auront donc, dans un premier temps du moins, aucune mention de propriété intellectuelle. Des développements sont en cours pour adapter le mapping à de plus nombreux types d'objets présentés dans Gallica.

L'interopérabilité sémantique des outils de la BnF repose donc encore en partie sur les producteurs eux-mêmes, qui doivent désormais être capables d'appréhender l'ensemble de leur environnement de travail et les principes sur lesquels repose la communication entre les différents outils qu'ils ont à leur disposition. Les producteurs de données sont ainsi les premiers acteurs de l'interopérabilité : c'est la manière dont ils suivent le modèle conceptuel et l'architecture de données choisis qui conditionne la qualité des données diffusées.

PRATIQUES COLLABORATIVES DE DESCRIPTION

Mais la mise en œuvre de l'interopérabilité sémantique n'a pas seulement conduit les institutions à faire évoluer leurs outils ou langages de description et d'indexation, elle suscite également des pratiques professionnelles nouvelles fondées sur la collaboration entre professionnels, non plus seulement à l'intérieur des institutions, mais encore entre institutions. C'est ainsi que, pour accompagner les producteurs de données dans la transition bibliographique, le groupe de réflexion national sur l'EAD en bibliothèque a mis au point un guide des bonnes pratiques constitué d'éléments normatifs et d'un forum sur lequel les

utilisateurs peuvent dialoguer pour résoudre leurs problèmes sans nécessairement passer par les instances prescriptrices. Ce guide est un outil très efficace d'aide au catalogage, grâce notamment aux exemples tirés de situations réelles. Pour autant, même s'il essaie d'agréger des ressources sur l'EAD et sur des domaines connexes, les producteurs de données ne le utilisent pas encore au maximum de ses possibilités. Par exemple, le forum de discussion qui l'accompagne et qui pourrait servir à l'ensemble de la communauté, reste peu employé. C'est que les bonnes pratiques ne sont pas encore le fruit de l'expérience commune des producteurs, mais demeurent bien, pour l'instant, le fruit de la prescription des experts des formats de description. Ce constat explique en partie pourquoi les pratiques de description sont encore très divergentes dans les faits, alors que les modèles de description convergent eux continuellement. Ces différences sont sans doute dues entre autres, au manque de recul des producteurs de données eux-mêmes sur ces évolutions d'autant que les normes qu'on leur demande de mettre en œuvre sont bien souvent définies à un échelon international et politique qui ne se répercute qu'en sourdine dans leur travail quotidien. Les normes internationales, conçues pour une utilisation universelle, ne peuvent tenir compte des particularités propres à chaque institution ou chaque collection et il appartient donc aux producteurs de données de faire entrer ces particularités dans le cadre général de la norme. Il en résulte des modèles de données multiples, fondés sur des modèles communs transdisciplinaires, dont celui défini pour la collection Charles Cros est un exemple parmi d'autres.

AGREGATION ET CONVERGENCE DES DONNEES

Les objectifs fixés à la réalisation de data.bnf.fr résument quant à eux l'essentiel de la politique de la Bibliothèque nationale de France en matière d'outils connectés. Data.bnf.fr est un agrégateur de données internes, qui met à la disposition de tous des données d'autorité issues de différents réservoirs, mais de manière transparente pour l'internaute. Sur data.bnf.fr, les données du catalogue général, de BAM et de Gallica sont agrégées sur une même page, donnant à voir toutes les informations disponibles à la BnF sur le sujet. Ainsi, si l'on effectue dans data.bnf.fr une recherche sur Thomas A. Edison, l'inventeur du phonographe et concepteur du Class M, la page de résultats présente tous les documents et objets dont Edison est signalé comme auteur ou sujet, mais également tous les auteurs et sujets qui lui sont liés, et chaque document signalé ramène à la source à laquelle il a été moissonné.

Data.bnf.fr, qui fonctionne conformément aux recommandations du modèle FRBR, a pour objectif d'inciter le public à réutiliser les données de l'institution, en permettant de télécharger des données dans de nombreux langages informatiques, du pdf au JSON en passant par RDF.

Surtout, les données accessibles par data.bnf.fr sont des données authentifiées, validées par une institution de référence. Car dans cette toile où l'information, véridique comme erronée, vérifiée ou non, circule à foison, les institutions culturelles que sont les bibliothèques, archives et musées font aujourd'hui le pari que les internautes cherchent des sources d'information fiables, et c'est ce que data.bnf.fr leur garantit²⁶. Ainsi, chaque donnée d'autorité de la BnF est authentifiée humainement par vérification de sa redondance avec les données d'autorité d'autres institutions, et informatiquement au moyen d'un identifiant pérenne unique, l'ark²⁷ : une url²⁸ à laquelle sont associées des métadonnées qui permettent à une machine d'associer une adresse Internet avec une entité unique. De même pour la collection Charles Cros, les données produites dans BAM, diffusées sur Gallica et data.bnf.fr, sont à la disposition des institutions comme du public intéressé.

DES ENJEUX PROFESSIONNELS QUI DEPASSENT LE CADRE NATIONAL

Ce mouvement d'ouverture des données est aujourd'hui partagé par un très grand nombre d'institutions de par le monde. L'authentification des données est facilitée par le Fichier d'autorité international et virtuel ou VIAF, pour *Virtual International Authority File* : VIAF agrège l'ensemble des notices d'autorités produites par des établissements considérés comme institutions de référence, parmi lesquelles la plupart des grandes bibliothèques nationales mais également Wikipedia²⁹. Par institutions de référence, on entend des institutions reconnues par leurs pairs sur la foi de la richesse et de l'authenticité de leurs données autant que sur leur popularité auprès du public.

Ainsi, la mobilisation d'une communauté internationale facilite d'une part les échanges et permet d'autre part à tous les professionnels de suivre et de comprendre les réflexions en cours dans les domaines qui les touchent, même quand leur institution n'est pas spécialiste du domaine. Pour le traitement de la collection Charles Cros en l'occurrence, la confrontation au modèle du CIDOC CRM et les travaux des grands établissements qui y travaillent ont mis en lumière la nécessité d'intégrer des éléments de description de durée et d'évènement à

²⁶ Duhamel, 2014.

²⁷ *Archival Research Key*

²⁸ *Uniform Resource Locator*

²⁹ Bourdon, 2011.

l'architecture de données choisie pour sa description dans BAM, et qui sont absents de la norme archivistique ISAD(G).

Ce sont donc les établissements qui ont obtenu le statut d'institution de référence qui ont aujourd'hui un rôle dans l'évolution future des pratiques professionnelles au sein des bibliothèques, centres d'archives et musées. Obtenir cette reconnaissance est un enjeu de pouvoir qui se joue au niveau mondial. Les normes internationales qui modèlent les pratiques futures dans le domaine des LAM³⁰ sont ainsi le fruit de collaborations, influences et confrontations entre ces institutions de référence. Les institutions culturelles d'envergure internationale ont en effet rapidement compris, dans un contexte où l'interopérabilité technique et sémantique est devenue un enjeu majeur, l'importance de participer à la formalisation d'une norme. Participer à la décision, c'est avant tout être sûr que les spécificités de ses collections ou de ses traditions de description soient prises en compte. Il est ensuite plus facile, d'une part, de conformer ses propres données à la norme et, d'autre part, de se positionner en pôle de référence. C'est pourquoi il existe aujourd'hui des groupes de travail internationaux pour chaque métier de la culture : sur la norme archivistique internationale ISAD(G), sur le modèle conceptuel FRBR et sur celui du CIDOC CRM. Ces groupes de travail coopèrent et se nourrissent de leurs réflexions mutuelles, mais ils cherchent également à affirmer leur position dans leur domaine d'expertise.

CONCLUSION

La convergence des modèles et pratiques de description ne semble donc pas devoir aboutir à la fondation d'un modèle universel qui verrait s'abolir les spécificités propres à chaque culture professionnelle constituant les LAM. Il semble bien, au contraire, que ce soit précisément parce que les outils du web sémantique permettent de faire converger des données produites selon des modalités pratiques et dans des optiques diverses vers des points d'accès communs, que les LAM pourront à l'avenir voir leur fondements respectifs renforcés : chaque institution pourra décrire de la manière la mieux adaptée des collections qui ne relèvent pas de son champ d'expertise traditionnel, comme la Bibliothèque nationale de France peut à présent promouvoir en ligne une collection longtemps considérée comme un musée et d'ailleurs décrite selon une approche muséologique. Plutôt qu'à l'émergence d'un modèle commun, on assiste donc au renforcement et surtout à la médiatisation des modèles

³⁰ Ilien, 2013, *op. cit.* p. 26.

propres à chaque domaine des LAM : une bibliothèque peut désormais s'emparer des modèles de description muséologique et archivistique, comme un musée peut utiliser plus aisément les outils bibliothéconomique et archivistique.

Le traitement de la collection Charles Cros de la Bibliothèque nationale de France a ainsi permis de montrer comment les outils du web sémantique redéfinissent les pratiques au sein de la communauté professionnelle des bibliothèques, processus particulièrement prégnant dans le domaine de la description de contenus. En agrégeant les données qu'elles produisent et en les rendant accessibles par les moteurs de recherche et réutilisables par quiconque, les institutions culturelles bénéficient de moyens de promotion et de mise en relation de leurs contenus jusqu'à présent inédits. L'interopérabilité apparaît dans ce contexte comme une chance à saisir pour les institutions culturelles, en même temps qu'elle est devenue un enjeu sinon de pouvoir, au moins d'influence : les grandes institutions de demain seront celles qui seront parvenues à intégrer ces outils à leurs pratiques et qui auront réussi la promotion la plus étendue de leurs données sur le web. De la même manière les communautés professionnelles les plus influentes seront celles qui réuniront le plus grand nombre d'institutions autour d'un modèle ou d'une norme commune de description. Pour les filières métier, au cœur des institutions, cette évolution se vit au quotidien, dans l'adaptation nécessaire à des pratiques et des outils de description en constant développement, dans la réactivité indispensable des équipes, et dans l'accompagnement dont elles doivent bénéficier pour négocier au mieux cette « transition bibliographique. » Car l'interopérabilité culturelle ne donnera pleine mesure de ces formidables capacités que si l'ensemble de la communauté professionnelle se saisit de cet enjeu et met en œuvre au quotidien les outils qui la garantissent, tout en développant les compétences des personnels pour leur donner les moyens d'utiliser ces outils.

BIBLIOGRAPHIE

BACHIMONT, Bruno, Fabien GANDON, Gautier POUPEAU, Bernard VATANT, Raphaël TRONCY, Stéphane POUYLLAU, ... et Manuel ZACKLAD (2011), « Enjeux et technologies : des données au sens », *Documentaliste – Sciences de l'Information*, 4 : 24-41.
DOI : [10.3917/docsi.484.0024](https://doi.org/10.3917/docsi.484.0024)

BADILLO Patrick-Yves et Nicolas PÉLISSIER (2015), « Usages et usagers de l'information numérique », *Revue française des sciences de l'information et de la communication*, 6 : 1-11.

BERMES, Emmanuelle (2011), « Bibliothèques, archives et musées : l'enjeu de la convergence des données du patrimoine culturel », *Documentaliste – Sciences de l'Information*, 48(4) : 45-48.

BERNERS-LEE, James HENDLER et Ora LASSILA (2001), « The semantic web », *Scientific American* : 29-37. [En ligne].
http://www.minervaeurope.org/structure/workinggroups/userneeds/prototipo/progproto/intero_perabilita_f.html. Page consultée le 4 décembre 2015.
DOI : [10.1038/scientificamerican0501-34](https://doi.org/10.1038/scientificamerican0501-34)

BONNET, Jacques, Rosette BONNET et Daniel RAICHVARG (dir.) (2014), *Communication et intelligence du social, Acteurs, auteurs ou spectateurs des savoirs, de l'éducation et de la culture*, T2, Paris, éditions l'Harmattan.

BOURDON, Françoise et Vincent BOULET (2011), « VIAF : un pivot pour l'accès multilingue à diverses collections », Actes de la conférence internationale de l'IFLA. [En ligne]. <http://conference.ifla.org/past-wlic/2011/79-bourdon-fr.pdf>. Page consulté le 4 décembre 2015.

CORDEREIX, Pascal (2005), « Les fonds sonores du département de l'Audiovisuel de la Bibliothèque nationale de France », *Le Temps des médias*, 5(2) : 253-264

COTTE, Dominique, Dominique LAHARY, Françoise GENOVA, Fabien GANDON, Jean DUFOUR, Christophe WILLAERT... Charles HUOT (2013), « Les métiers, entre traditions et modernité », *Documentaliste – Sciences de l'Information*, 3 : 42-59.
DOI : [10.3917/docsi.503.0042](https://doi.org/10.3917/docsi.503.0042)

DALBIN, Sylvie, Emmanuelle BERMES, Antoine ISAAC, Romain WENZ, Yann NICOLAS, Tayeb MERABTI ... Dominique COTTE (2011), « Approches documentaires : priorité aux contenus », *Documentaliste – Sciences de l'Information*, 2011, 4 : 42-59
DOI : [10.3917/docsi.484.0042](https://doi.org/10.3917/docsi.484.0042)

DUHAMEL, Benjamin (2014), Les technologies du web sémantique et du record linkage au service de data.bnf.fr et du Linked Open Data culturel : étude sur les nouveaux paradigmes informationnels, Mémoire de master en sciences de l'information et de la communication sous la direction de Gildas ILIEN et Gérald KEMBELLEC, GERIICO, Université Lille III.

FAVIER, Laurence et Widad Mustafa El HADI (dir.) (2015), « L'archivage numérique des savoirs : perspectives européennes », *Les Cahiers du numérique*, 11 : 83-102.
DOI : [10.3166/lcn.11.1.9-13](https://doi.org/10.3166/lcn.11.1.9-13)

IFLA, Les Fonctionnalités requises des notices bibliographiques, rapport final, deuxième édition française établie par la Bibliothèque nationale de France, p. 14. [En ligne].
http://www.bnf.fr/documents/frbr_rapport_final.pdf. Page consultée le 29 septembre 2015.

ILLIEN, Gildas, Odile HOLOGNE, Stéphane POUYLLAU, Gilles ALFONSI, Jean-Pierre TROEIRA, Jean DELAHOUSSE ... Charles HUOT (2013), « Enjeux professionnels », *Documentaliste-Sciences de l'Information*, 3 : 26-41.

INTERNATIONAL COUNCIL ON ARCHIVES (2000), ISAD(G): *Norme générale et internationale de description archivistique*, Ottawa, ICA.

POUYLLAU, Stéphane (2013), « Web de données, big data, open data, quels rôles pour les documentalistes ? », *Documentaliste – Sciences de l'Information*, 3 : 32-33.

TILLET, Barbara (2004), « FRBR, qu'est-ce que c'est ? Un modèle conceptuel pour l'univers bibliographique », traduit de l'anglais par la Bibliothèque nationale de France, Library of Congress. [En ligne]. <http://www.loc.gov/cds/downloads/FRBR.PDF>. Page consultée le 4 décembre 2015.

SZABADOS, Anne-Violaine, Katell BRIATTE et Rosemonde LETRICOT (dir.) (2012), « Utiliser l'ontologie CIDOC CRM pour l'information relative au patrimoine culturel », *THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques*, Paris, Éditions de la Maison des sciences de l'homme.

ZORICH Diane, Günter WAIBEL et Ricky ERWAY (2008), « Beyond the silos of the LAMs: Collaboration among libraries, archives, museums », Dublin Ohio, OCLC Research, p. 5-8. [En ligne]. <http://www.oclc.org/content/dam/research/publications/library/2008/2008-05.pdf>. Page consultée le 4 décembre 2015.