

Numérisation des archives sonores de l'immigration à l'heure des digital humanities¹ : les métadonnées METS au service de la conservation du patrimoine kabyle

Sabine LOUPIEN

Université Paris 8 (Saint-Denis)

La croissance des collections numériques fait surgir de nouvelles interrogations. La masse des flux électroniques à organiser, conserver, partager et diffuser révèle des enjeux fondamentaux tant aux plans culturel et juridique que technique et organisationnel. La dématérialisation de l'information transforme les savoirs professionnels et ouvre de nouveaux champs de recherche. Au cœur de ces réflexions, les bibliothèques numériques sont, plus que jamais, confrontées aux impératifs de collecte, d'enrichissement, de diffusion et de pérennisation des corpus. Numérisation des collections, interopérabilité² des systèmes et conservation du patrimoine numérique constituent des défis manifestes pour les professionnels. La maîtrise des technologies et des systèmes d'information devient essentielle. Dans ce cadre, les métadonnées³ Metadata Encoding and Transmission Standard (METS) participent, via la construction d'ensembles numériques interopérables, à la mise en commun et à la transmission d'un patrimoine pérenne et cohérent.

Afin d'assurer l'accès du plus grand nombre à la mémoire sonore, écrite et iconographique, les programmes des grandes institutions patrimoniales et de leurs tutelles s'orientent vers la numérisation et l'organisation des corpus dans des formats ouverts et interopérables, et selon des normes et des protocoles internationaux. Qu'il s'agisse d'images fixes ou animées, textes publiés ou notes manuscrites, corpus sonores inédits et édités — comme les fonds de chansons kabyles de l'immigration — la diffusion et la conservation des collections (indépendamment du support, du contenu et du format des données) requièrent en effet le respect de politiques raisonnées de gestion de l'information.

L'OPEN ARCHIVAL INFORMATION SYSTEM : UN MODÈLE D'ORGANISATION DES DONNÉES NUMÉRIQUES

Dans cet esprit, modèles conceptuels et procédures appliquées sont mis en œuvre afin d'organiser et de conserver les données. Le modèle de référence Open

Archival Information System (OAIS) est devenu en 2003 la norme ISO 14721. Il permet une organisation des fonds d'archives et ce, dès leur entrée dans l'archive numérique ayant vocation à les conserver, jusqu'à leur diffusion, par le biais par exemple, de catalogues de bibliothèques. Ce modèle conçu en 1995 pour archiver des informations techniques et scientifiques, à la demande de l'ISO et à l'initiative du Consultative Committee for Space Data Systems (CCSDS)⁴, s'est depuis ouvert à d'autres métiers et notamment aux professions des archives et des bibliothèques. Il propose une organisation conceptuelle de l'archivage des objets physiques et numériques et vise la pérennisation des ressources électroniques. La fonction principale de l'OAIS est l'établissement de règles communes à l'ensemble des organisations chargées de la production, de la diffusion ou de la sauvegarde de données électroniques. Ses préconisations concourent à aider les décideurs à prendre des orientations en matière d'implémentation de systèmes d'information compatibles avec une politique d'archivage à très long terme.

L'OAIS reste toutefois conceptuel et ne définit pas une mise en œuvre spécifique des systèmes. C'est pourquoi des archives électroniques mutualisées, dédiées aux besoins des institutions et de leurs collections, sont créées par les organismes de recherche, les grandes bibliothèques et les Archives nationales. Plusieurs programmes d'archives pérennes, établis selon la norme OAIS et consistant en la conservation de documents numériques d'intérêt national ou international, sont ainsi conçus.

On note parmi eux, le Système de Préservation et d'Archivage Réparti⁵ (SPAR) lancé en 2009 par la Bibliothèque nationale de France. Son fondement est la conservation des documents numériques dans une double optique de sécurité et de pérennité de l'information.

Autre initiative, les Archives de France ont mis en production en 2010 une plate-forme PILote d'Archivage Électronique (PILAE). Dans le cadre du développement de l'administration électronique, PILAE est conçu pour recevoir, traiter, conserver et communiquer les archives numériques des services centraux de l'État. Il a vocation à intégrer le futur Système d'Information Archivistique (SIA) du centre des Archives nationales situé à Pierrefitte-sur-Seine.

LES MÉTADONNÉES METS : UNE IMPLÉMENTATION DE L'OAIS POUR LA CONSERVATION ET L'ÉCHANGE DES OBJETS NUMÉRIQUES

Le format de métadonnées METS est issu du projet Making of America II (MOA-2)⁶. Il a été élaboré à l'initiative de la Digital Library Federation en 2001 (version 1.1 en 2002) par un groupe de travail pour le développement d'un standard d'objet appliqué aux bibliothèques numériques. Maintenu par la Library of Congress, METS est un schéma eXtensible Markup Language (XML)⁷ défini pour permettre la conservation et l'échange de documents numériques répondant au modèle OAIS. Ce schéma non propriétaire et extensible se décline autour de sept sections arborescentes et de trois classes de métadonnées. Lors de sa création, et c'est l'atout de ce format, le document METS contient la description de la structure hiérarchique des objets électroniques qui forment la ressource. Il répertorie les noms ainsi que la localisation des fichiers et assemble les métadonnées administratives, structurelles ou descriptives associées aux documents. Le document METS peut alors être conservé et diffusé au sein d'une archive OAIS. Il comprend d'un côté la ressource numérique et de l'autre, les métadonnées associées.

SECTIONS ARBORESCENTES, MÉTADONNÉES, CARTES DE STRUCTURE ET HYPERLIENS POUR DÉCRIRE, ORGANISER ET DIFFUSER LE PATRIMOINE SONORE

De manière générale, les sept sections⁸ METS permettent la description et l'ordonnance des collections numériques. L'avantage de ce format est donc d'admettre l'archivage des objets hétérogènes simples, comme une photographie ou une page, ou la conservation des objets complexes⁹, à l'instar d'un livre composé de chapitres.

Prenons l'exemple de la numérisation d'un disque 45 tours de la chanteuse Hanifa. La préservation et l'échange de ce document associé à la photographie de l'artiste impliquent à la fois la rédaction d'informations descriptives (les notices de catalogue), la conservation des aspects techniques (les différents formats de fichiers) et l'enregistrement de données administra-



محمد الكمال

المغني المبتدع

MOHAMMED
EL KAMAL

LE CHANTEUR INNOVATEUR

اسطوانات باي



Mohamed El Kamal/ Disques Pathé.
Collection Electric and Musical Industries (EMI)/Génériques

tives (droits d'auteur et cycle de vie du document). De même, sont nécessaires la considération de la structure physique de l'objet (organisation des fichiers) et la connaissance de sa structure logique (agencement intellectuel des différentes parties).

Pour mener à bien la numérisation, les métadonnées de structure, fondamentales pour comprendre un corpus, permettent de connaître tous les fichiers (inventaire, emplacement, taille des fichiers) qui composent un document quels que soient leur format et leur nature. Elles retracent la relation physique (linéaire) entre ces documents, leur ordre d'affichage et indiquent le chemin vers les fichiers cibles. Mais elles décrivent également les relations logiques tissées entre les fichiers: délimitation des chapitres, index ou détail du contenu d'un objet (vue de détail de la pochette d'un disque 45 tours, par exemple)¹⁰.

C'est la carte de structure qui reconstitue la navigation, linéaire ou logique, dans le document. Ainsi, pour un fonds d'archives consacré à l'œuvre de Slimane Azem, les notes prises par le musicien lors de la composition des mélodies et sa documentation personnelle (coupures de presse ou écrits critiques à propos, par exemple, du contexte politique algérien) seront liées, au sein de la carte de structure, aux fichiers sonores par le biais des hyperliens METS.

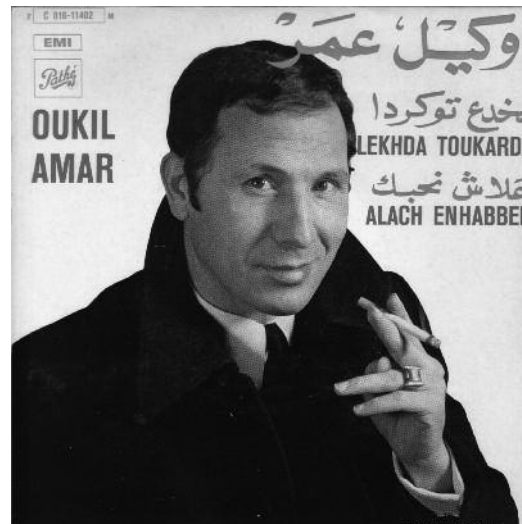
Les métadonnées METS permettent d'envisager la construction de cartes numériques en s'appuyant sur le cheminement intellectuel des collecteurs ou des artistes. Il devient alors possible d'organiser, échanger et préserver à long terme des collections sonores jusque-là inédites.

Dans le même esprit, un modèle de structuration commun aux fonds – inédits ou édités – de chansons kabyles de l'immigration, candidats à une conservation pérenne, permet d'établir une arborescence numérique de l'ensemble des documents: affiches, articles, photographies, contrats, notes manuscrites, pochette de disques vinyles, cassettes audio, etc. À l'inverse d'un schéma appliqué exclusivement à des fonds édités – qui suit la structuration éditoriale des objets publiés –, il convient de se fonder sur l'observation de l'ordonnement personnel des collecteurs et des artistes, pour proposer un modèle s'appliquant à des fonds sonores inédits. Matrice de l'organisation des fonds dans l'archive OASIS, ce modèle générique arborescent peut alors être associé à des métadonnées appliquées

et à des procédures spécifiques de gestion des fonds de chansons kabyles.

ILLUSTRE LA DÉMARCHE DES ARTISTES ET TRACER L'ORGANISATION MATÉRIELLE DU FONDS D'ARCHIVES: UNE APPROCHE À LA FOIS CONCEPTUELLE ET PRATIQUE

Sans nul doute, l'apport du format METS réside – notamment pour les fonds inédits – dans sa carte de structure, la diversité de ses liens et ses trois classes de métadonnées. Ainsi formulée, l'information de



Oukil Amar / Pathé Marconi.
Collection Electric and Musical Industries (EMI)/Génériques

contexte illustre et conserve le cheminement des artistes tout en décrivant très précisément l'organisation physique des documents (ordre des mélodies, incise dans la composition, suite des feuillets de texte). Toutefois, la conservation numérique du patrimoine sonore ne fait pas abstraction d'interrogations tant de forme que de fond.

Point cardinal du travail du chercheur comme du gestionnaire de l'information, la mise en contexte des fonds est essentielle à leur complète compréhension. Qu'elle soit effectuée au plan général (il peut s'agir alors du contexte politique et social d'une collecte ou

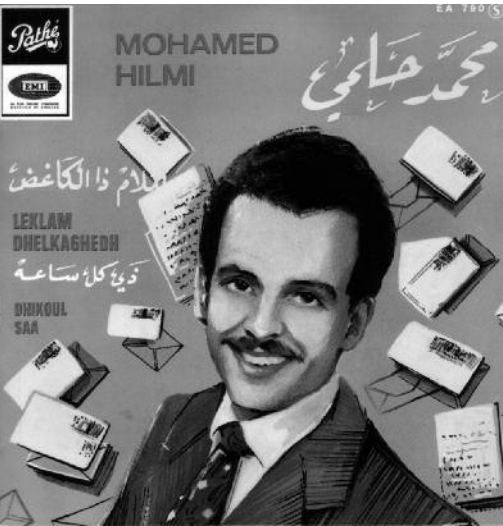
encore des approches musicologiques ou historiques d'une œuvre) ou particulier via le tissage de liens entre les éléments d'un fonds, cette contextualisation ne peut faire l'économie des outils innovants des technologies de l'information développés par les bibliothèques numériques. Précisément, afin de respecter l'interopérabilité des catalogues et le traitement homogène des corpus sonores, l'élaboration d'un modèle d'ordonnement commun s'impose. Il devient alors possible de verser dans de vastes ensembles numériques des fonds musicaux inédits, comme le corpus constitué par l'ethnomusicologue indien Deben Bhattacharya¹¹ ou des documents édités – tout ou partie –, tels les fonds de chansons kabyles de

En d'autres termes, à travers l'observation des relations logiques et sémantiques entre les fonds sonores et la diversité de la documentation qui les accompagne (catalogues personnels, notes, analyses, photographies, croquis), il convient de retrouver la démarche initiale du collecteur ou de l'artiste (interprète, compositeur, parolier). Si l'enregistrement sonore reste au centre de l'analyse, il importe de décrire l'approche et le parcours personnel des acteurs et de les restituer par le biais des hyperliens METS.

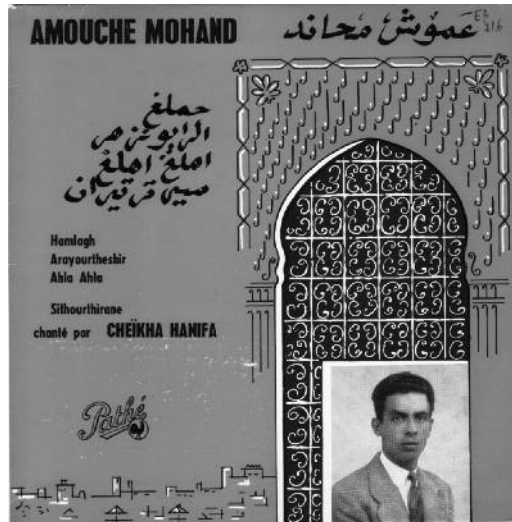
Mais dégager une organisation commune aux divers corpus observés fait surgir différentes interrogations en termes de méthode et d'orientation. L'orientation archivistique de cette démarche est confrontée au

principe d'homogénéité du traitement des fonds mais aussi au fondement déontologique du respect de ces derniers et à la pratique du traitement archivistique selon d'évidentes contraintes de temps et de moyens. En outre, afin d'œuvrer à une plus grande cohérence du travail et favoriser un traitement rapide du document sonore, textuel ou iconographique, l'établissement de procédures normalisées est fondamental. Mais comment mettre en œuvre un modèle unique d'organisation à partir de fonds dissemblables tant par leur contenu que par leur forme? Par ailleurs, en considérant les moyens

alloués et afin d'avoir une approche efficace et pertinente, quelle méthode employer pour obtenir une vision d'ensemble de fonds parfois très vastes ou difficiles d'accès? En d'autres termes, comment représenter un fonds aussi conséquent que celui du grand amateur de jazz Charles Delaunay¹², ou étudier les nombreuses collaborations et le vaste répertoire de l'auteur-compositeur-interprète Kamal Hamadi? De plus, selon le principe du respect des fonds constitués par les collecteurs, de quelle manière suivre l'agencement des documents établis par ces derniers et traduire, sous forme numé-



Mohamed Hilmi / Pathé Marconi.
Collection Electric and Musical Industries (EMI)/Génériques



Amouche Mohand / Pathé Marconi.
Collection Electric and Musical Industries (EMI)/Génériques

l'immigration. Mais comment structurer des ensembles sonores hétérogènes tant par leur forme que par leur fond? Ici, pas de normes bibliographiques ou éditoriales; l'enjeu est de découvrir la démarche du collecteur et, sans la trahir, de la traduire dans le cadre d'une arborescence numérique fidèle à l'esprit du langage XML et des métadonnées METS. Ainsi, il s'agit de comprendre en quoi le témoignage d'un chanteur kabyle rendant compte de la vie musicale parisienne dans les années trente est lié aux notes manuscrites de l'ethnomusicologue collectant ses propos, pour traduire alors ce lien dans l'architecture METS.

rique, leur parcours intellectuel ? Si pour les documents édités, les collections numériques suivent l'organisation des œuvres : chapitre, paragraphe, texte et fichier texte pour les monographies ou encore, collection de disques, disque à l'unité, face de disque puis fichier audio pour les corpus musicaux, la difficulté consiste dans le cadre de fonds inédits à retrouver l'ordonnement logique créé par le compositeur, le musicien instrumentiste ou le collecteur. En quoi ce document de 1947 émanant de la Société des auteurs est-il lié au premier disque de Cheikh El-Hasnaoui ? Quelle association existe entre ces notes manuscrites, ce contrat d'auteur tapuscrit et la chanson « Boulayoune thiberkanine »¹³ ? Comment traduire ce lien dans l'architecture METS ? Ici, pas de normes ou de chartes éditoriales, l'enjeu et non le moindre, est de comprendre la démarche de l'auteur et de la restituer dans le cadre d'une arborescence numérique.

Autre écueil, le niveau de l'objet numérique paraît essentiel. Quel objet choisir comme premier niveau de structuration du fonds ? Quels contours donner à cet objet numérique ? Pour un fonds inédit, la face d'un disque est-elle l'item¹⁴ ou est-ce plutôt le morceau et donc la plage ? Dans le cadre de témoignages oraux, que faire d'un lot de photographies sur le même témoin (musicien, locuteur kabyle) ? Celui-ci doit-il être considéré comme un objet ou comme un ensemble plus important ? Enfin, quelle place accorder à l'analyse documentaire ou scientifique faite par des spécialistes a posteriori, et de quelle manière organiser cette information avec le reste du fonds ?

CONCLUSION

La préservation à long terme des documents sonores et l'accès du plus grand nombre à un patrimoine souvent méconnu et dispersé constituent les apports fondamentaux de la mise aux normes numériques des corpus de l'immigration. Le format XML METS concourt à l'organisation et à la sauvegarde de l'information de contexte qui accompagne le document numérique. Qu'il s'agisse de retracer le cheminement intellectuel des auteurs ou de replacer les fonds dans un contexte culturel plus vaste, retrouver la démarche des collecteurs, des éditeurs ou des artistes est au cœur de cette approche.

Aujourd'hui, de nombreux projets de conservation et de diffusion des archives sonores voient le jour. Parmi eux, des portails ouverts et interopérables sont particulièrement innovants¹⁵. Associé au protocole de diffusion OAI-PMH¹⁶, METS qui permet l'empaquetage et la préservation des collections, s'inscrit dans ce mouvement. Ainsi, à l'heure des *digital humanities*, l'archivage numérique du patrimoine kabyle ouvre de nouvelles perspectives de conservation mais aussi de diffusion au plus grand nombre d'une mémoire sonore de l'immigration. ■

1. « Les *digital humanities* (humanités numériques) désignent une transdiscipline, porteuse des méthodes, des dispositifs et des perspectives heuristiques liés au numérique dans le domaine des Sciences humaines et sociales », in *Manifeste des digital humanities*, Paris, ThatCamp, mai 2010.

2. On désigne par interopérabilité la possibilité d'échanger des fichiers avec d'autres utilisateurs équipés de matériels ou de logiciels différents. Pour garantir l'interopérabilité, il faut veiller à utiliser des formats de fichiers ouverts, c'est-à-dire dont les spécifications sont connues et accessibles à tous.

3. Les métadonnées, données sur les données, permettent à la fois de décrire, d'organiser et d'accéder à l'information numérique. Très diverses, du Dublin Core à l'Encoded Archival Description (EAD) en passant par les métadonnées Learning Object Metadata (LOM) pour décrire les ressources de l'enseignement jusqu'aux métadonnées Moving Picture Experts Group (MPEG) pour les documents audiovisuels, orientées et recommandées selon leur usage, ces dernières ont une fonction essentielle dans le monde de la conservation et de l'information. Elles répondent à des protocoles d'échange des documents sur le Web ainsi qu'à des impératifs d'interopérabilité.

4. Le CCSDS est un groupe international formé par plusieurs agences spatiales. Il a pour fonction de développer des standards et de faire des recommandations dans le domaine des données scientifiques et techniques. Pour une explication détaillée du modèle OAIS voir CCSDS. *Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data System Practices*. Washington, D.C., Magenta Book, juin 2012.

5. Se reporter à la communication d'Emmanuelle Bermes, Isabelle Dussert Carbone, Thomas Ledoux et al. *La préservation numérique à la Bibliothèque nationale de France : présentation technique et organisationnelle*. In 74e congrès IFLA, Bibliothèques sans frontières : naviguer vers une compréhension globale, 10-14 August 2008, Québec.

6. Sous l'égide de la Digital Library Federation, le projet MOA-2 s'est consacré à l'élaboration de métadonnées d'objets numériques pour l'encodage, la recherche, l'affichage et la navigation.

7. Le langage XML est devenu un standard en 1998. Issu de la norme Standard Generalized Markup Language (SGML), XML

favorise l'organisation des éléments et de leur sémantique. Il sert de fondement pour créer des langages balisés spécialisés tels les langages Metadata Object Description Schema (MODS), Metadata Encoding and Transmission Standard (METS), Resource Description Framework (RDF) ou encore Really Simple Syndication (RSS).

8. Les sept sections METS sont les suivantes : 1- l'en-tête ou « Mets Header » ; 2- les métadonnées descriptives ou « Descriptive Metadata Section » ; 3- les métadonnées administratives ou « Administrative Metadata Section » ; 4- la liste des fichiers qui composent l'objet numérique : « File Section » ; 5- la carte de structure ou « Structural Map Section » ; 6- les liens entre les éléments de la carte de structure : « Structural Link section » ; 7- « Behavior Section » qui recense les programmes et les logiciels nécessaires au bon fonctionnement de l'objet conservé.

9. Un objet simple présente un niveau de structure simple alors que les objets complexes comptent plusieurs échelles d'accès. Ces échelles ou hauteurs d'accès sont appelées granularités. Ainsi, l'objet complexe « collection » peut comporter plusieurs niveaux de granularité : le document, le volume, le tome, la page puis la vue de détail de la page.

10. Dans le cadre du XML Pointer Language, un système de pointage interne relie les fichiers et les métadonnées d'un document METS. Dans l'exemple ci-dessus, le pointer *area* viserait un sous-segment du fichier c'est-à-dire un détail de la pochette.

11. Collectionneur passionné, Deben Bhattacharya (1921-2001) a parcouru l'Europe, l'Asie et l'Afrique afin de collecter les musiques de la vie quotidienne. De mariages en concerts, il a rassemblé sept cents heures de musique fixées sur bandes magnétiques. À côté des sources sonores, un important corpus documentaire enrichit le fonds conservé au Département de l'audiovisuel de la BnF : catalogue personnel décrivant et illustrant les enregistrements, annotations musicales, courriers, documents administratifs, contrats, etc.

12. Charles Delaunay (1911-1988), fils des peintres Sonia et Robert Delaunay, commence très tôt une collection de disques qu'il complète tout au long de sa vie. À la fois organisateur de concerts, producteur, imprésario et éditeur, il fonde avec Hugues Panassié la revue *Jazz Hot* en 1935, puis crée le label Swing, première compagnie musicale consacrée au jazz. Il lance le label Vogue avec Léon Cabat et dirige *Jazz Hot* jusqu'en 1980. En 1979, il fait don à la BnF de sa considérable collection de disques et de ses archives ; un corpus d'environ 22 000 disques, 11 000 78 tours et 11 000 33 tours ainsi que de nombreuses cassettes analogiques, films et bandes magnétiques ou encore des copies d'émissions radiophoniques.

13. La chanson « Boulayoune thiberkanine » – L'homme aux yeux noirs – fut éditée en 1949.

14. On qualifie d'item l'élément minimal d'un ensemble ordonné.

15. Citons par exemple les catalogues collectifs d'archives sonores et audiovisuelles, Telemeta et le Portail du patrimoine oral.

16. L'Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) est un protocole d'interrogation et d'échange de métadonnées fondé sur le langage XML et le format Dublin Core. Il fonctionne par moissonnage régulier d'entrepôts de données

structurés et organisés par chaque institution ; seules les métadonnées documentaires des catalogues OAI sont moissonnées.

BIBLIOGRAPHIE

ASSOCIATION INTERNATIONALE D'ARCHIVES SONORES ET AUDIOVISUELLES, Comité technique, [Dietrich Schüller, éd]. *Sauvegarde du patrimoine sonore : éthique, principes et stratégies de conservation. Version 3*. [En ligne]. IASA, décembre 2005. IASA-TC 03. Fichier PDF. Disponible sur : <http://www.iasa-web.org/downloads/publications/TC03_French.pdf> (consulté le 15 juillet 2012).

BAUDE Olivier (dir.), *Corpus oraux. Guide des bonnes pratiques*, Paris, CNRS Éditions, 2006, 203 p.

BEAUBIEN Richard, PRESTON Cecilia, et al., *METS. Metadata encoding and transmission standard. Primer and reference manual*, Washington, D.C., Digital Library Federation, 2007, 128 p.

BERMES Emmanuelle, DUSSERT CARBONE Isabelle, LEDOUX Thomas, et al, *La préservation numérique à la Bibliothèque nationale de France : présentation technique et organisationnelle*, in IFLA. Bibliothèques sans frontières : naviguer vers une compréhension globale, 10-14 Août 2008, Québec, IFLA, 2008. [En ligne]. Disponible sur : <http://archive.ifla.org/IV/ifla74/papers/084-Bermes_Carbhone_Ledoux_Lupovici-trans-fr.pdf> (consulté le 10 juillet 2012).

CALDERAN Lisette, HIDOINE Bernard, MILLET Jacques, *Métadonnées. Mutations et perspectives*. Séminaire INRIA, 29 septembre-3 octobre 2008, Dijon. Paris, ADBS, 2008, 282 p.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, *L'archive ouverte : rôle d'une archive ouverte*, in Centre de ressources pour la description de l'oral, 2012. [En ligne]. Disponible sur : <<http://crdo.risc.cnrs.fr/exist/crdo/>> (consulté le 25 janvier 2012).

COLLARD Claude (dir.), *Des sources pour l'histoire de l'immigration en France de 1830 à nos jours*, Paris, Bibliothèque nationale de France, 2006, 427 p.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS), *Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data System Practices*. [En ligne]. Washington, D.C., Magenta Book, June 2012. Disponible sur : <<http://public.ccsds.org/publications/archive/650xom2.pdf>> (consulté le 20 juillet 2012).

INTERNATIONAL ASSOCIATION OF SOUND AND AUDIOVISUAL ARCHIVES (IASA), *Ethical Principles for Sound and Audiovisual Archives*, IASA Special Publication, n°6, 2010.

MAHFOUFI Mehenna, « La chanson kabyle en immigration : une rétrospective », in *Hommes et Migrations*, septembre 1994, n° 1179, pp. 32-39.

MOKHTARI Rachid, *La chanson de l'exil. Les voix natales (1939-1969)*, Alger, Casbah Éditions, 2001, 157 p.

THE LIBRARY OF CONGRESS, *METS Implementation Registry*, in METS. Metadata Encoding and Transmission Standard. Official Web Site. [En ligne]. Disponible sur : <<http://www.loc.gov/standards/mets/mets-registry.html>> (consulté le 10 juillet 2012).