



HAL
open science

Éditorialisation d'un fonds audio documentaire, stratégie et enjeux

Boris Blanckemane

► **To cite this version:**

Boris Blanckemane. Éditorialisation d'un fonds audio documentaire, stratégie et enjeux. domain_shs.info.docu. 2016. mem_01476034

HAL Id: mem_01476034

https://memsic.ccsd.cnrs.fr/mem_01476034

Submitted on 24 Feb 2017

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

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

Ecole Management et Société-Département CITS

INTD

Éditorialisation d'un fonds documentaire audio sur le web: stratégies et enjeux

Présenté et soutenu le 2 décembre 2016 par *Boris Blanckemane*

MEMOIRE pour obtenir le Titre enregistré au RNCP
"Chef de projet en ingénierie documentaire"
Niveau I

Promotion 46



Paternité pas d'utilisation commerciale

"J'ai voulu que les voix aimées
Soient un bien que l'on garde à jamais,
Et puissent répéter le rêve
Musical de l'heure trop brève"

Charles Cros, *Le Collier de griffes*, 1908

Remerciements

Merci à Evelyne Broudoux pour ses conseils, ses remarques enrichissantes et ses pistes de lectures.

Je tiens à remercier tout particulièrement Claude Mussou, l'équipe de l'Inathèque et celle du Dépôt Légal du Web de l'INA pour leur accueil, leur sympathie et leur confiance.

Merci à Vincent Rappeneau, mon fidèle binôme durant cette année de formation à l'INTD pour sa complicité, son efficacité dans nos projets d'étude et pour m'avoir ouvert aux écrits de P.Murray, J-C Michéa et G.Orwell.

Ce mémoire est dédié à ma maman qui m'a transmis l'insatiable appétit des mots et de leur musicalité.

Résumé

Ce mémoire, rédigé dans le cadre du titre professionnel de niveau 1 à l'INTD présente une étude des stratégies et des enjeux de l'éditorialisation des phonogrammes. Après avoir défini et contextualisé les notions polysémiques de phonogramme et d'éditorialisation, seront présentées les différentes techniques d'éditorialisation que sont l'ajout de métadonnées, l'ajout de mots-clés, le chapitrage et les commentaires, tant du point de vue de la méthode que des avantages et des inconvénients attenants. Nous verrons si et comment ces techniques sont utilisées via une étude comparative de plusieurs services d'écoute en ligne mettant à disposition des phonogrammes de différentes natures à destination de différents publics. Enfin, ces techniques seront appliquées via un cas d'étude concret de mise en ligne d'un corpus de phonogrammes.

Descripteurs : phonogrammes, métadonnées, audio, éditorialisation, valorisation, stratégie, chapitrage, mot-clés, commentaires.

Abstract

This essay, written as part of the course "Titre Professionnel de niveau 1" in INTD presents a study of strategies and issues of the editorial content of phonograms. Having defined and contextualised polysemous notions of phonogram and editorial content, the various editorialisation approach will be presented: adding metadata, adding keywords, adding chapters and reviews, both from the point of view the method and from the advantages and disadvantages. We'll see whether and how these methods are used via a comparative study of several streaming websites offering content based on phonograms of various kinds and to different public. Finally, these methods will be applied through a concrete study case of putting on-line a corpus of phonograms.

Keywords: phonograms, editorial content, valuation, strategy, chaptering, keywords, comments.

Table des matières

| | |
|---|-----------|
| INTRODUCTION | 12 |
| 1. DEFINITION D'UN FONDS DOCUMENTAIRE AUDIO | 13 |
| 1.1 ASPECTS HISTORIQUES | 13 |
| 1.1.1 <i>Du paléophone au mp3 : évolution du support</i> | 13 |
| 1.1.2 <i>Du Muzak à Spotify: évolution du streaming et des conditions d'écoutes</i> | 15 |
| 1.2 ASPECTS TECHNIQUES | 16 |
| 1.2.1 <i>La compression audio</i> | 16 |
| 1.2.2 <i>Les protocoles</i> | 16 |
| 1.3 ASPECTS ECONOMIQUES | 17 |
| 1.4 ASPECTS JURIDIQUES | 18 |
| 2. EDITORIALISATION D'UN FONDS DOCUMENTAIRE AUDIO | 20 |
| 2.1 EDITORIALISATION | 20 |
| 2.2 TECHNIQUES D'EDITORIALISATION | 21 |
| 2.2.1 <i>Les métadonnées : quand nécessité rime avec pluralité</i> | 22 |
| 2.2.1.1 Métadonnées de propriétés | 22 |
| 2.2.1.2 Métadonnées de gestion | 23 |
| 2.2.1.3 Métadonnées des phonogrammes matérialisés..... | 23 |
| 2.2.1.4 Métadonnées des phonogrammes dématérialisés | 24 |
| 2.2.1.5 Métadonnées des phonogrammes hébergées | 25 |
| 2.2.1.6 Le PBCore : l'approche nord-américaine..... | 26 |
| 2.2.1.7 L'approche de la FAMDT..... | 27 |
| 2.2.2 <i>Les mot-clés : dilemne entre efficacité et subjectivité d'un outil</i> | 30 |
| 2.2.3 <i>Les commentaires : l'éditorialisation à l'heure du Web 2.0</i> | 32 |
| 2.2.4 <i>Le chapitrage : diviser pour mieux éditorialiser</i> | 33 |
| 2.2.4.1 Approche théorique | 33 |
| 2.2.4.2 Approche pratique..... | 34 |
| 3. COMPARAISON DE STRATEGIES D'EDITORIALISATION DE SITES WEB | 35 |
| 3.1 PREAMBULE..... | 35 |
| 3.2 PRESENTATION DES SITES..... | 36 |
| 3.2.1 <i>Spotify et Deezer : deux acteurs clés de l'offre musicale</i> | 36 |
| 3.2.2 <i>Soundcloud et Jamendo: les réseaux sociaux sonores</i> | 37 |
| 3.2.3 <i>France Inter et France Culture : deux radios du service public</i> | 37 |
| 3.2.4 <i>Les sites de la FAMDT et de la bibliothèque sonore de Paris</i> | 38 |

| | | |
|-----------|--|------------|
| 3.3 | GRILLES DE COMPARAISON | 39 |
| 3.4 | PRESENTATION DES RESULTATS ET LIMITES DE LA COMPARAISON..... | 40 |
| 3.5 | RESULTATS DETAILLES | 41 |
| 3.5.1 | <i>Deezer</i> | 41 |
| 3.5.2 | <i>Spotify</i> | 43 |
| 3.5.3 | <i>Soundcloud</i> | 45 |
| 3.5.4 | <i>Jamendo</i> | 48 |
| 3.5.5 | <i>France Culture</i> | 49 |
| 3.5.6 | <i>France Inter</i> | 51 |
| 3.5.7 | <i>Bibliothèque sonore de Paris</i> | 53 |
| 3.5.8 | <i>Portail du patrimoine oral</i> | 56 |
| 3.6 | LE CAS DE L'INSTITUT NATIONAL DE L'AUDIOVISUEL..... | 59 |
| 3.6.1 | <i>L'INA: quand l'archivage devient une mission d'état</i> | 59 |
| 3.6.2 | <i>Le poste de consultation multimédia</i> | 59 |
| 3.7 | RECAPITULATIF DES RESULTATS | 62 |
| 3.8 | CONCLUSION DE LA COMPARAISON..... | 63 |
| 4. | MISE EN APPLICATION : LE BLOG DES ATELIERS DL WEB | 64 |
| 4.1 | CONTEXTE | 64 |
| 4.2 | ENJEU | 64 |
| 4.3 | STRATEGIE ADOPTEE | 65 |
| 4.3.1 | <i>Traitement du fonds documentaire</i> | 65 |
| 4.3.2 | <i>Editorialisation du fond</i> | 68 |
| 4.3.3 | <i>Valorisation</i> | 70 |
| 4.3.4 | <i>Résultats</i> | 72 |
| | CONCLUSION | 72 |
| | ANNEXES | 77 |
| | BIBLIOGRAPHIE | 95 |
| | SITOGRAPHIE | 104 |
| | GLOSSAIRE | 105 |

Table des illustrations

| | |
|---|----|
| Figure 1: Évolution des moyens et du prix du stockage informatique | 14 |
| Figure 2: Fonctionnement des éléments du PBCore 2.1 | 26 |
| Figure 4: Résultats de la recherche Deezer..... | 41 |
| Figure 3: Moteur de recherche Deezer | 41 |
| Figure 5: Deezer: page d'un artiste | 42 |
| Figure 6: Deezer: zone commentaires d'une page artiste | 42 |
| Figure 7: Spotify: résultats de recherche dans la barre latérale..... | 43 |
| Figure 8: Spotify: résultats détaillés de recherche | 43 |
| Figure 9: Spotify: détail d'une page artiste..... | 44 |
| Figure 10: Soundcloud: moteur de recherche | 45 |
| Figure 11: Soundcloud: résultats de la recherche | 45 |
| Figure 13: Soundcloud: présentation d'un phonogramme..... | 46 |
| Figure 12: Soundcloud: page d'un contributeur audio | 46 |
| Figure 14: Soundcloud: détails d'un phonogramme | 47 |
| Figure 15: Soundcloud: exemple de commentaires..... | 47 |
| Figure 16: Jammendo: moteur de recherche | 48 |
| Figure 17: Jammendo: page d'un artiste | 48 |
| Figure 18: Jammendo: page d'un titre | 48 |
| Figure 19: France Culture: moteur de recherche | 49 |
| Figure 20: France Culture: limite de la suggestion à la saisie | 49 |
| Figure 21: France Culture: page de résultats d'une recherche..... | 49 |
| Figure 22: France Culture: page d'une émission..... | 50 |
| Figure 24: France Inter: résultats d'une recherche | 51 |
| Figure 23: France Inter: moteur de recherche | 51 |
| Figure 25: France Inter: page d'une émission | 52 |
| Figure 27: Bibliothèque Sonore de Paris: listes d'auteurs et de genres..... | 53 |
| Figure 26: Bibliothèque Sonore de Paris: moteur de recherche | 53 |
| Figure 28: Bibliothèque Sonore de Paris: l'accessibilité relative de la recherche avancée | 54 |
| Figure 29: Moteur de recherche de l'édition adaptée..... | 54 |
| Figure 30: Bibliothèque Sonore de Paris: exemple de recherche..... | 55 |
| Figure 31: Bibliothèque sonore de Paris: détail d'un phonogramme | 55 |

| | |
|---|----|
| Figure 32: Portail du patrimoine oral: moteur de recherche | 56 |
| Figure 33: Portail du patrimoine oral: moteur de recherche avancée..... | 56 |
| Figure 34: Portail du patrimoine oral: présentation des résultats..... | 57 |
| Figure 35: Portail du patrimoine oral: détail d'un phonogramme | 58 |
| Figure 37: Poste de Consultation Multimédia de l'INA: exemple d'aide | 60 |
| Figure 36: Poste de Consultation Multimédia de l'INA: moteur de recherche | 60 |
| Figure 38: Poste de Consultation Multimédia de l'INA: résultats de recherche | 61 |
| Figure 39: Poste de Consultation Multimédia de l'INA: détail d'un phonogramme | 61 |
| Figure 40: Extrait de la page "Intervenants" du blog des Ateliers du dépôt légal du Web | 71 |
| Figure 41: Page web d'un atelier avant éditorialisation et valorisation..... | 73 |
| Figure 42: Page web d'un atelier après éditorialisation et valorisation | 74 |

Tables des tableaux

| | |
|---|----|
| Tableau 1: Répartition des zones dans l'ID3 v1 | 24 |
| Tableau 2: Correspondance métadonnées FAMDT et Dublin Core | 28 |
| Tableau 3: Bloc de métadonnée selon la FAMDT..... | 29 |
| Tableau 4: Avantages et Inconvénients de deux méthodes de chapitrage..... | 33 |
| Tableau 5: Offres d'abonnement Spotify | 36 |
| Tableau 6: Offres d'abonnement Deezer | 36 |
| Tableau 7: Récapitulatif des résultats de la comparaison de portails | 62 |
| Tableau 8: Tableau de description technique d'un fonds documentaire audio | 66 |
| Tableau 9: Tableau de description thématique d'un fonds documentaire audio ... | 67 |

INTRODUCTION

Le langage serait apparu entre 350 000 et 150 000 ans avant Jésus-Christ [1] faisant ainsi de l'oralité l'un des premiers vecteurs de transmission de la Connaissance utilisé par l'homme. De l'ode à la diffusion d'interviews, de la musique en ligne en passant par la chanson de geste, l'histoire de la tradition orale est intimement liée à l'histoire de l'humanité.

Littérature orale, opéras, émissions radiophoniques, discours, les documents sonores représentent un corpus dense, varié aussi bien en terme de contenant que de contenu.

Devant la profusion de ces documents et la pluralité de leur format, il convient de se demander comment est orienté l'auditeur vers ce qu'il recherche : quelles sont les techniques utilisées pour valoriser, mettre à disposition d'un public un fonds documentaire sonore et quels sont les enjeux sous-jacents à cette valorisation?

Après une première partie dans laquelle le terme de document sonore sera défini et précisé, et dans laquelle ses aspects techniques, économiques et juridiques seront présentés, une deuxième partie permettra de clarifier la notion d'éditorialisation et d'en présenter quelques unes des méthodes et techniques.

Une troisième partie proposera une présentation, puis une comparaison des méthodes et choix en matière d'éditorialisation d'un panel de sites mettant des documents sonores à disposition de différents publics pour différents usages.

Enfin, une dernière partie exposera la mise en œuvre d'une stratégie d'éditorialisation via le cas concret de la valorisation d'un fond mis en ligne par le service du dépôt légal du Web de l'Institut National de l'Audiovisuel.

1. Définition d'un fonds documentaire audio

Le terme « document audio » est une appellation ambiguë qui nécessite d'être précisée. Sa définition pourra dépendre du contexte : juridique, économique, scientifique....

Selon l'AFNOR, le terme normatif est « phonogramme » et correspond à un « document constitué par toute fixation exclusivement sonore de sons provenant d'une exécution musicale ou d'autres sons, quelle qu'en soit l'origine » [2].

Le phonogramme est défini par le CPI (Code de la Propriété intellectuelle) comme « la fixation d'une séquence de son » (Loi 92-597 du 1^{er} Juillet 1992 (chap. V, Art. L 215-1)).

Dans le contexte actuel de la généralisation de la dématérialisation documentaire, il peut être intéressant de reconsidérer ces définitions.

En effet, se distinguent les phonogrammes dit *matérialisés* (enregistrement magnétique ou optique de la voix et du son, disque, cassette-audio, disque-compact) de ceux dits *dématérialisés* (flux audio (ou *streaming*), enregistrements numériques sous forme de fichier informatique).

Dans la suite de ce mémoire et pour en faciliter la lecture, l'aspect support et l'aspect contenu sous-jacent au terme de phonogramme seront confondus et ce terme ainsi que ceux de *fichier audio* et *document sonore* seront utilisés sans distinction.

1.1 Aspects historiques

1.1.1 Du paléophone au mp3 : évolution du support

Le phonographe de Léon Scott de Martinville en 1857, le paléophone de Charles Cros et le phonographe de Thomas Edison en 1877 furent les premiers appareils à l'origine des phonogrammes [4].

Les supports initialement analogiques reposent sur le principe de création du son par la variation continue d'une grandeur physique. Suite à la création 1937 de la technologie numérique PCM (*Pulse Code Modulation pour Modulation par impulsion et Codage*) par Alec Reeves, ils deviennent numériques dès les années 1970, la création du son s'opère alors par la variation discrète d'une grandeur binaire.

Partie 1 : Définition d'un fonds documentaire audio

La problématique rencontrée par ce nouveau support est la taille extrêmement importante occupée par un fichier, taille qui est à mettre en regards des capacités de stockage de l'époque et du prix de ce stockage (cf. figure 1)

Il faudra attendre 1989 pour que la première méthode de compression audionumérique apparaisse avec la création et le dépôt du brevet du codec MP3 (ou *MPEG Layer 3*) par l'institut allemand Fraunhofer [5].

Sans rentrer dans des considérations trop techniques ayant trait au traitement du signal, le principe de la compression des données audio, compression dite « destructive », est d'exploiter les défauts de l'audition humaine pour créer des algorithmes permettant de supprimer les informations les moins perceptibles.

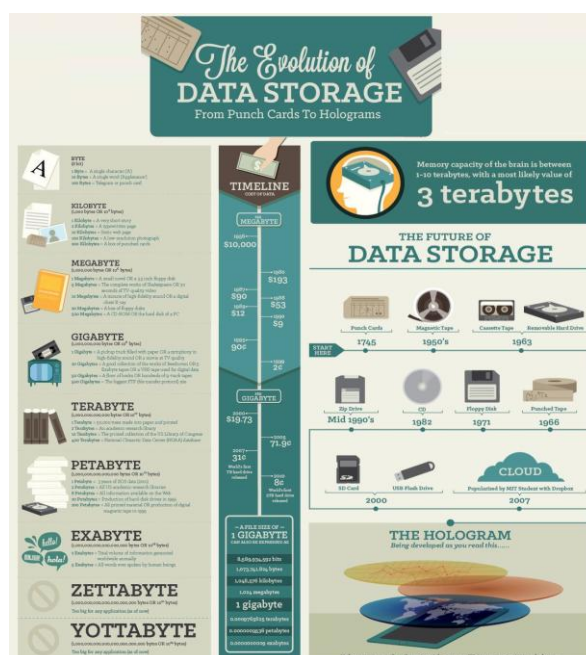


Figure 1: Évolution des moyens et du prix du stockage informatique (Source :WepageFX)

Au-delà des considérations audiophile et de son format libre, l'avantage de la méthode de compression mp3 est de réduire massivement (d'un facteur 5 à 10) la taille d'un fichier audio permettant ainsi son envoi, son partage et sa diffusion via le réseau Internet tout en garantissant une qualité d'écoute au moins acceptable.

Les fichiers audio sont ainsi partagés, légalement ou non, via des réseaux comme IRC, Usenet ou encore Napster qui, dès 1999, a participé significativement à la démocratisation de l'utilisation du format MP3¹ sur le Web.

Depuis les années 2000, de nombreux autres formats de compression sont apparus, propriétaires ou non : *vorbis*, *flac*, *ape*. Au-delà des simples considérations de stockage, c'est dorénavant un consensus entre la taille du

¹ On confond souvent le format d'un fichier multimédia audionumérique avec son codec, son conteneur ou son extension.

fichier, la qualité audio et l'usage qui oriente le choix d'un format de compression plutôt qu'un autre.

Aujourd'hui, les évolutions et recherches sur les phonogrammes concernent principalement les techniques de diffusion visant à proposer de nouvelles techniques de spatialisations. Ainsi, depuis la diffusion en stéréophonie, les techniques ont évoluées vers des modes dit « multicanaux » à l'instar des technologies Dolby®, puis Dolby Surround®. Aujourd'hui, les innovations et la recherche s'orientent vers le son binaural et la diffusion holophonique (son 3D) qui permettent, sinon promettent, une immersion plus complète de l'auditeur [6].

1.1.2 Du Muzak à Spotify: évolution du streaming et des conditions d'écoutes

Le premier système de transmission de signaux (donc par extensions de signaux sonores) est le Muzak dont le brevet est déposé par George O. Squier en 1922 [7].

Cependant, il faudra attendre la création du réseau Internet, le développement des échanges via ce dernier et l'amélioration de la bande passante (capacité et stabilité) pour permettre l'essor du streaming. Ainsi, en 1993, *Severe Tire Damage* est le premier groupe de musique à proposer un concert diffusé en direct sur le web, utilisant et permettant ainsi d'éprouver une nouvelle technologie (*MBONE*) [8].

En 1994, WXYV sera la première station de radio à proposer une diffusion de ses programmes sur Internet. En 1995, Word Magazine, sera le premier magazine en ligne qui mettra à disposition de ces auditeurs des morceaux écoutables en streaming. En 1996, Virgin Radio sera la première radio européenne à proposer ses programmes sous 3 formats : sur la bande FM, en direct en streaming et en podcast [9].

En 2001, Rhapsody est lancée comme la première plate-forme d'achat de fichier audio en ligne suivi en 2003 de MySpace qui permettra aux utilisateurs de mettre en ligne leur propres créations sonores puis Spotify en 2008, première plateforme à proposer un service *freemium* [10].

Depuis, les acteurs et les offres se sont multipliés et offrent aux particuliers, aux institutions et aux entreprises de larges possibilités d'écoutes et de diffusions.

1.2 Aspects techniques

1.2.1 La compression audio

Un phonogramme numérique peut se présenter sous plusieurs formats. Originellement, le format utilisé était le PCM, encapsulé au sein d'un conteneur créé conjointement en 1991 par Microsoft et IBM : le *Wave*.

Si ce format assurait une excellente qualité sonore, il restait très volumineux : un fichier correspondant à une capture audio de 5 minutes représentait environ 75 méga-octets de données.

A une époque où la vitesse de connexion au réseau Internet était de l'ordre de 56 kbit/s, l'échange de tels fichiers représentait une limite technique au streaming.

Le développement de codec de compression dans les années 90 a permis de réduire (par 10 environ) la taille de ces fichiers à qualité d'écoute égale et leur échange sur le réseau Internet devint ainsi plus envisageable.

Une vulgarisation du principe de compression est disponible en Annexe 1.

Le développement des vitesses de connexion à Internet ayant aussi décuplé, il a été possible d'assurer le streaming de phonogrammes encodés avec des qualités toujours meilleures.

1.2.2 Les protocoles

Le streaming a pu se démocratiser grâce aux développements de protocoles réseaux adaptés comme le RTSP (*Real-Time Streaming Protocol*) et l'UDP (*User Datagram Protocol*) qui permettent de découper un flux de données en paquets de taille moindre, ou grâce au multicast qui rend possible la diffusion d'un seul et unique fichier depuis un poste unique (serveur) vers différents postes clients.

1.3 Aspects économiques

Dès qu'il fût possible de le mettre sur support, le document sonore musical a pu faire l'objet d'un marché de plus en plus important, en perpétuel développement et qui a montré sa capacité à se réinventer pour s'adapter aussi bien aux supports qu'aux habitudes de consommation. Au premier semestre 2016, un tiers de la population française écoute la musique en streaming et a écouté 12,8 milliards de titres. Cet engouement de la population pour ce nouveau mode de consommation de la musique (représentant 36% du marché et ayant enregistré 44% de progression) a entraîné une croissance semestrielle de 6%, soit la meilleure performance depuis 6 ans [11].

Le document audio non musical fait lui aussi l'objet d'un marché, certes moins important mais qui ne doit pas être négligé.

Parmi ces documents, on distingue notamment le livre audio. En effet, celui-ci représente 1% du marché du livre papier en France mais 10% aux USA [12].

En France, entre 2008 et 2013, selon IPSOS, le nombre des ventes de livres audio a augmenté de 156 % [13], la production est de plus en plus importante bien qu'elle soit une production de niche réservée majoritairement aux secteurs jeunesse et aux malvoyants.

Le marché de l'archive sonore est principalement détenu par l'Institut National de l'Audiovisuel qui propose à ses utilisateurs (particuliers, entreprises, institutions, chercheurs) un fonds documentaire sonore de plusieurs milliers d'heures.

1.4 Aspects juridiques

De par leur nature variées en terme de contenus (œuvre musicale, entretien, conférence) et de contenants (cassette, CD, fichiers numériques), les questions relatives au droit des phonogrammes sont complexes et délicates à appréhender pour le non juriste.

L'idée dans le présent mémoire est de rappeler brièvement quelques notions clés à retenir pour comprendre les enjeux juridiques révélés par la gestion d'une phonothèque ouverte au public.

Dans un premier temps, il convient de rappeler que c'est l'article L213-1 alinéa 1^{er} du Code de la propriété intellectuelle qui définit les phonogrammes comme « toute fixation exclusivement sonore de sons provenant d'une exécution ou d'autres sons » [14].

Plutôt que de parler du droit d'une œuvre, il conviendrait de parler des droits d'une œuvre.

On distingue donc :

- le droit moral attribué à l'auteur de l'œuvre qui se distingue par :
 - ses caractères :
 - éternel : il survit à la mort du créateur et 50 ans après,
 - inaliénable : la cession est juridiquement impossible et irrecevable,
 - ses attributs régis par :
 - le droit de divulgation et de retrait: l'auteur a les pleins pouvoirs quant au choix de divulguer/retirer son œuvre,
 - le droit de paternité : l'auteur peut revendiquer son statut d'auteur,
 - le droit au respect : l'auteur a la garantie du respect de son nom et de celui de son œuvre.
 - le droit de représentation : l'auteur est assuré d'être informé, crédité et éventuellement rétribué pour chaque exploitation de son œuvre,
 - le droit de reproduction : toute reproduction totale ou partielle de l'œuvre ne peut se faire sans autorisation de l'auteur sauf cas d'utilisation particulier (citations, parodie...),

- les droits voisins : qui assure la couverture juridiques de ceux qui exploiteront l'œuvre (interprète, journalistes, producteurs...).

Comme le rappelle Fabien Muller dans son mémoire très complet sur ces problématiques juridiques : « la personne à l'initiative de la fixation détient des droits voisins sur le phonogramme sans qu'il ne soit tenu compte de l'originalité et de la créativité de cette fixation, et ces droits existeront dès lors que l'acte matériel de captation sonore sera, pour la première fois, réalisé. Occurrence rare mais possible, un phonogramme peut donc ne pas être une œuvre, et ainsi être source d'un droit voisin sans pour autant être source d'un droit d'auteur » [15].

Il est donc impératif lors des opérations de valorisation et d'éditorialisation d'un fonds documentaire audio de garder à l'esprit les questions relatives aux respects des droits de l'œuvre et de son(s) auteur(s).

2. Editorialisation d'un fonds documentaire audio

2.1 Editorialisation

L'éditorialisation est un concept récent, multi-facettes, polysémique dont la définition dépendra du contexte.

C'est Brigitte Guyot qui est à l'origine de l'interprétation actuelle de ce terme qu'elle définit dès 2006 comme un complément du travail auctorial, « une projection vers un horizon d'interprétation, plus ou moins large et lointain, vers des personnes proches, connues, ou au contraire anonymes pouvant être amenées à l'utiliser » [16]. L'éditorialisation est donc une étape de raffinement du document permettant ainsi d'en faciliter l'exploitation et la diffusion. Le document devient alors une coproduction entre son auteur et son éditorialisateur.

Inspiré par sa consœur, Bruno Bachimont, directeur de la recherche de l'Université de technologie de Compiègne, présente l'éditorialisation comme le processus qui permet de produire un document numérique à partir de plusieurs autres documents [30] [20].

Marcello Vitali-Rosati, professeur au département des littératures de langue française de l'Université de Montréal, spécialiste de l'éditorialisation, développe cette idée. Il définit dans un premier temps l'éditorialisation comme la pratique qui regroupe toutes pratiques éditoriales qui ne sont pas du ressort de l'édition [17].

Selon Stéphane Crozat, l'éditorialisation est un processus s'articulant autour « d'étapes permettant aux différents savoir-faire de s'exprimer et de s'opérationnaliser avec les niveaux requis d'exigence et de qualité » [25]. L'éditorialisateur devra donc être polyvalent, s'adapter au corpus cible, le contextualiser afin de créer des parcours entre les différents documents [26].

Il apparaît que la nature du simple document est dépassée. Ce dernier s'inscrit dans une dualité en étant à la fois tout et partie du tout. Ce peut-être une étape et/ou une destination en fonction l'adéquation entre le niveau d'éditorialisation produit par l'éditorialisateur et celui attendu par le lecteur-auditeur.

L'éditorialisation correspond donc à un processus ouvert – par opposition à l'édition qui est un processus fermé (dans la mesure où une fois ce processus terminé, son évolution, son remaniement est limité et impactant) – reposant sur des moyens techniques et technologiques, permettant de diffuser, commenter, partager et donc valoriser un savoir, un contenu [22].

C'est un processus nouveau, complémentaire de l'édition, qui fait interagir des contenus (ou des ressources), un environnement technique (le réseau, les serveurs, les plateformes, les CMS, les algorithmes des moteurs de recherche), des structures et formats (l'hypertexte, le multimédia, les métadonnées) et enfin des pratiques (l'annotation, les commentaires, les recommandations via réseaux sociaux).

Pour résumer, dans le cadre d'une approche orientée Web, l'éditorialisation consiste à structurer et organiser l'information issue d'un document. Cette structuration peut s'opérer de façon thématique et/ou chronologique.

2.2 Techniques d'éditorialisation

S'il existe de nombreuses techniques d'éditorialisation, manuelles ou informatisées (prises de notes versus transcription automatique), intuitives ou déductives (mots-clés versus boussole sémantique), ouvertes ou fermées (commentaires versus notices) [27], les quatre qui seront présentées dans les paragraphes qui suivent correspondent à celles le plus en adéquation avec le processus d'éditorialisation d'un phonogramme :

- l'attribution des métadonnées, étape obligée pour tout document, phonogramme inclus et dont on ne peut faire l'économie dans une approche documentaire de l'éditorialisation. On pourra parler d'**éditorialisation documentaire**.
- les mots-clés qui, par leur format court et intuitif, permettent une compréhension rapide du sujet et passent un processus d'analyse du contenu du phonogramme. On pourra parler **éditorialisation exégétique**.
- la prise en charge des commentaires des auditeurs afin d'inscrire l'éditorialisation dans le Web 2.0. On pourra parler d'**éditorialisation participative**.

- le chapitrage qui permet de segmenter un phonogramme en plusieurs permettant une navigation plus rapide et plus facile au sein du contenu. On parlera d'**éditorialisation ergonomique**.

2.2.1 Les métadonnées : quand nécessité rime avec pluralité

A l'instar de tout autre type de document, la gestion correcte et l'utilisation optimale de phonogrammes ne peut faire l'économie des métadonnées.

Les métadonnées donneront ainsi plus de sens à un phonogramme, optimiseront son référencement et permettront ainsi une meilleure indexation par les moteurs de recherches.

Les métadonnées devront aussi faire l'objet de mise à jour en cas de besoin [23].

On distingue plusieurs types de métadonnées [24]:

- les métadonnées de propriétés,
- les métadonnées de gestion,
- les métadonnées de description.

2.2.1.1 Métadonnées de propriétés

Elles se basent sur deux standards : l'ISRC et l'ISWC.

L'*International Standard Recording Code* est au phonogramme ce que l'ISBN est au livre. En effet, ce système permet d'attribuer à un enregistrement un identifiant unique, indépendant du support et qui coexistera avec les autres métadonnées [18].

Il a été reconnu par l'ISO sous la norme ISO 3901.

Le code ISRC est composé de 12 caractères alphanumériques séparés par un tiret et précédé de la mention "ISRC" tels que :

- 2 caractères correspondent au pays du producteur du phonogramme (reprenant l'ISO 3166-1 alpha-2),
- 3 caractères correspondent au code unique assigné au producteur du phonogramme,
- 2 caractères correspondent aux 2 derniers chiffres de l'année de dépôt du phonogramme,
- 5 caractères correspondent au code assigné au phonogramme.

Exemple : ISRC FR-AB5-07-12345.

L'*International Standard Musical Work Code* permet l'identification d'une œuvre musicale en tant que création unique et intangible.

Il est composé de trois parties:

- 1 caractère correspondant au préfixe
- 9 caractères correspondant à l'identifiant
- 1 caractère de vérification.

Exemple: T-000.000.001-0

Ces deux métadonnées sont complémentaires. L'ISRC se focalise sur l'aspect enregistrement de l'œuvre quand l'ISWC se focalise sur l'aspect composition [19]. Il peut donc y avoir plusieurs ISRC pour un même ISWC.

L'ISRC est déterminé par le producteur de l'enregistrement (maison de disque ou syndicat auquel l'artiste est affilié (en France, la Société Civile des Producteurs Phonographiques ou la Société civile des Producteurs de Phonogrammes). L'ISWC est déterminé par l'*International Federation of the Phonographic Industry*.

2.2.1.2 Métadonnées de gestion

Ce sont les métadonnées qui permettent de gérer le phonogramme au sein d'un catalogue, d'une collection, d'une offre de vente. Il peut s'agir du prix du phonogramme, de ses éventuelles identifiants issus d'un système de gestion interne (base de données dédiée), des avis et commentaires des auditeurs, et de toutes informations jugées pertinentes par le gestionnaire de collection lors de son travail de promotion, d'éditorialisation, de valorisation et de diffusion de l'œuvre. Il s'agit donc de métadonnées qui par extension deviennent des métadonnées à forte valeur juridique [28].

2.2.1.3 Métadonnées des phonogrammes matérialisés

Seuls les supports optiques permettent l'encapsulation de métadonnées à l'inverse des supports analogiques qu'ils soient magnétiques (cassette audio, bande master, DAT) ou mécanique (microsillons).

i. Encapsulation des données au sein d'un CD : le *CD-Text*

Mise en place en 1996, cette méthode utilise pour stocker les métadonnées soit :

- la zone d'amorce d'un CD qui dispose de 5 ko d'espace mémoire,

- les *subchannels* d'un CD (zone d'un CD permettant de stocker des données non audio) permettant de disposer de 31 Mo d'espace mémoire.

Dans tous les cas, ce système doit répondre aux exigences de la norme IEC 61866:1997 (*Audiovisualsystems - Interactive text transmission system (ITTS)*)

ii. Encapsulation des données dans les DVD-A et les SACD

Bien que peu utilisés, les DVD-Audio (DVD-A) et les Super-Audio CD (SACD) utilisent la technologie DSD (*Direct Stream Digital*) et un encodage dit de qualité supérieure plébiscité par les audiophiles. Le DSD repose sur deux types d'encodages : le DFF et le DSF. Seul ce dernier permet l'encapsulation des métadonnées au format ID3 (cf. paragraphe suivant).

2.2.1.4 Métadonnées des phonogrammes dématérialisés

Dans le cas des fichiers dématérialisés, les métadonnées peuvent être intégrées directement dans le fichier audio grâce aux ID3.

Conçu par Eric Kemp en 1996, dans sa première version, il s'agit d'un espace de 128 octets situés en fin de fichier, codés en ISO 8859-1 et répartis comme suit :

| | |
|------------------|---------------|
| Titre du morceau | 30 caractères |
| Artiste | 30 caractères |
| Album | 30 caractères |
| Année | 30 caractères |
| Commentaire | 30 caractères |
| Genre | 8 bit |

Tableau 1: Répartition des zones dans l'ID3 v1

Il peut être édité via :

- un logiciel dédié (MP3Tag, Tag&Rename...),
- un lecteur audio (Foobar, Winamp...),
- lors de la création du fichier via des logiciels de conversion supportant cette technologie.

Cette version jugée insuffisante en termes de métadonnées évoluera en 1998 en une version 2 adaptée au streaming, non limitative en nombre de champs (seule la taille du tag est limitée à 256 Mo), permettant d'illustrer le phonogramme par

une image et supportant les caractères UNICODE. Les informations sont alors placées en début de fichier [29].

Les champs supportés par le système ID3 V2 sont listés en [Annexe 2](#).

L'ID3 est principalement utilisé pour les fichiers de type MP3. Cependant, il existe d'autres systèmes :

- l'*ApeTAG*. Moins populaire que l'ID3, ce système stocke les tags à la fin du fichier. Les métadonnées ainsi stockées sont disponibles en [Annexe 3](#).
- le *vorbis comment* utilisé majoritairement pour les formats de fichiers FLAC, Vorbis [30]. Les longueurs de champs sont limités à 4 294 967 295 octets. Il supporte l'UTF-8 et contient généralement les champs détaillés en [Annexe 4](#).

2.2.1.5 Métadonnées des phonogrammes hébergées

Les métadonnées peuvent être associés aux fichiers via le catalogue qui les contient et affichées sur la page qui les héberge.

2.2.1.6 Le PBCore : l'approche nord-américaine

Mis au point aux Etats-Unis, le *Public Broadcasting Metadata Dictionary* a été créé en 2001 par les diffuseurs de programmes et les institutions en charge de fonds documentaires audio-visuels (bibliothèques, entreprises privées...).

Ouvert, gratuit, sous licence *Creative Commons*, et basé sur le Dublin-Core, le PBCore repose sur la même syntaxe programmatique que le langage XML. Chaque média peut être décrit via 4 classes de contenus (*content classes*), 15 conteneurs (*containers*), and 82 éléments structurés comme indiqué dans la figure 2 [22].

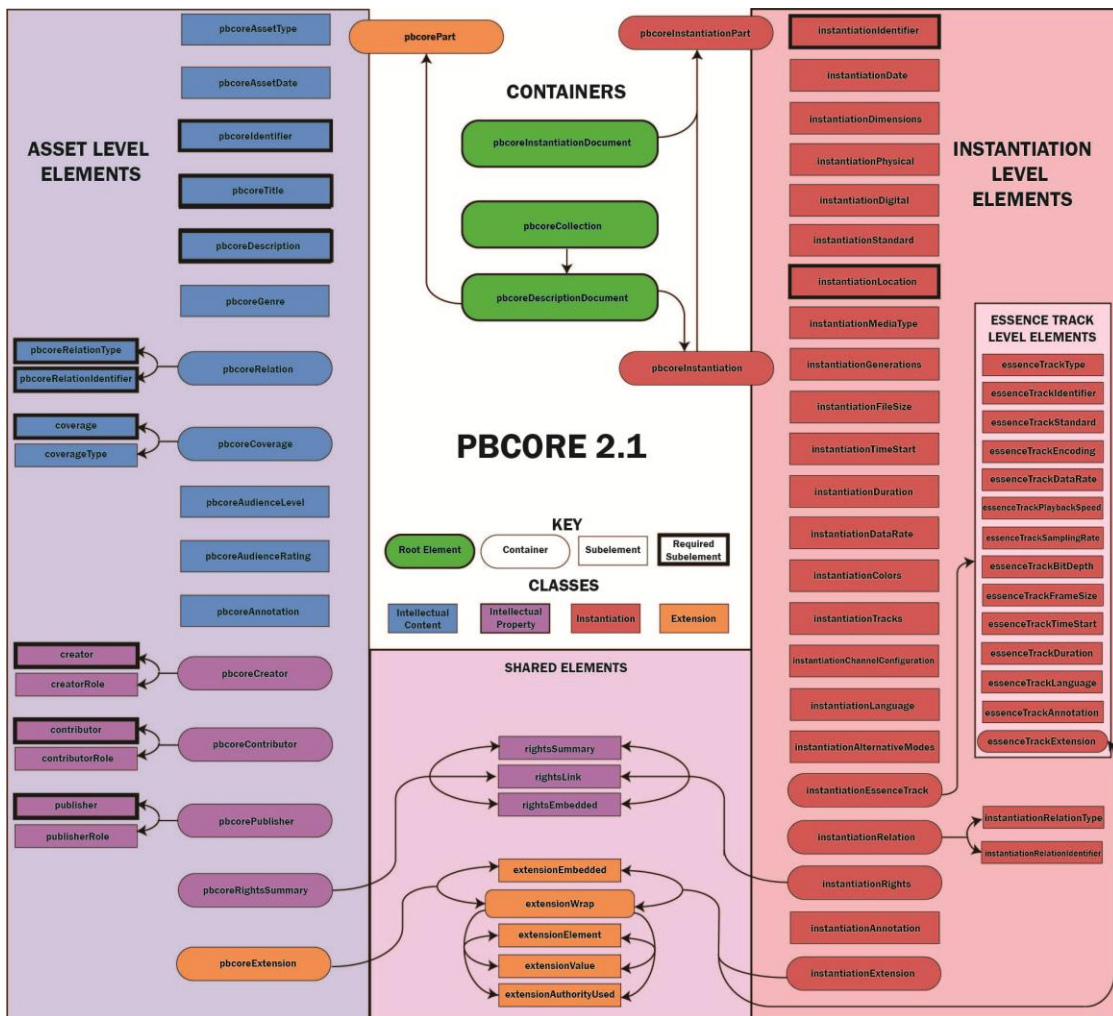


Figure 2: Fonctionnement des éléments du PBCore 2.1

Des exemples sont disponibles en [Annexe 5](#).

2.2.1.7 L'approche de la FAMDT

Proposé par la Fédération des Associations de Musiques et Danses Traditionnelles depuis 1994 et régulièrement mis à jour, le *Guide d'analyse documentaire du son* contient les recommandations quant à la gestion des métadonnées d'un fonds documentaire audio [31].

Le principe repose sur l'analyse des documents à la fois dans leur globalité et dans leur spécificité.

Chaque document peut être ainsi caractérisé par des métadonnées regroupées en onze blocs comme décrit dans le tableau 3.

De même, afin de rendre possible l'exploitation de ces données via le protocole OAI-PMH et donc d'en assurer l'interopérabilité, la FAMDT propose une correspondance des champs fondamentaux vers les champs du Dublin-Core (cf. tableau 3 ci-après)

| Champs | Élément Dublin Core correspondant (français/anglais) |
|--|---|
| Titre | Title/Titre |
| Intervenant(s) Fonction(s) | Creator/Créateur |
| Mots-clés Noms cités Périodes citées Lieux cités Classification du conte ou de la chanson Catalogues - Titres Catalogues - Numéros Type de danses Nom des instruments de musique | Subject/Sujet ou mots-clés |
| Résumé | Description/description |
| Nom du catalogue collectif | Publisher/Editeur |
| Signature de la notice | Contributor/Contributeur |
| Date d'enregistrement | Date/Date |
| Type de document Genre du document Nature du document | Type/Type de ressource |
| Support du document Durée du document | Format/Format |
| URL de la notice | Identifier/Identifiant de ressource |
| Numéro d'inventaire | Source/Source |
| Langue | Language/Langue |
| Fonds Corpus | Relation/Relation |

| Champs | Élément Dublin Core correspondant (français/anglais) |
|----------------------------------|---|
| Enquête Contient | |
| Lieu d'enregistrement | Coverage/Couverture |
| Droits Niveau de consultation | Rights/Droits |

Tableau 2: Correspondance métadonnées FAMDT et Dublin Core

| Intitulé du bloc | Données du bloc |
|----------------------------|---|
| Identification | Numéro d'inventaire Identifiant pérenne Fonds Corpus Enquête Lieu(x) de consultation Localisation de l'original Côte Autre(s) côte(s) Signature de la notice |
| Conditions de consultation | Droits Déposant Mode d'obtention Date d'entrée Niveau de consultation |
| Désignation | Type (support physique) Nature Genre |
| Contexte de création | Intervenants Fonctions Date d'enregistrement Lieu(x) d'enregistrement Latitude Longitude Langue(s) Code des langues Commentaires sur la langue |
| Titre | Titre |
| Description Technique | Support Durée Qualité du son Autres caractéristiques |

| Intitulé du bloc | Données du bloc |
|----------------------------------|--|
| | techniques Matériel d'accompagnement |
| Analyse | Résumé Transcription Noms cités Périodes cités Lieu(x) cités Mots-clés |
| Relation | Contient Exploitation URL Fichier attaché Voir aussi |
| Spécialité musique – œuvre orale | Incipit littéraire Type de formation instrumentale ou vocale Type de voix Timbre de l'air Refrain Effectif des interprètes Forme de la chanson Coupe de la chanson Classification du conte ou de la chanson Catalogues – Titres Catalogues – Numéro Formulette d'introduction au conte Formulette de conclusion au conte Formulette interne au conte Type de danse |

| Intitulé du bloc | Données du bloc |
|---|--|
| | Nom des instruments de musique Indice des instruments de musique |
| Catalogage d'une émission radiophonique | Lieu(x) de production Date(s) de diffusion Producteur(s) |
| Gestion | Niveau de catalogage Date de numérisation Date de saisie / mise à jour de la fiche |

Tableau 3: Bloc de métadonnée selon la FAMDT

2.2.2 Les mots-clés : dilemme entre efficacité et subjectivité d'un outil
Selon la définition de la norme AFNOR NF Z 47-102, on appelle *tag* ou *mots-clés* un « mot ou groupe de mots choisi soit dans le titre ou le texte d'un document, soit dans une demande de recherche documentaire, pour en caractériser le contenu ». Il s'agit donc de termes qui tiennent le rôle de marqueur sémantique ou lexical très utilisés sur les sites Web notamment avec le développement du Web 2.0. En effet, ils sont facilement interprétables par l'utilisateur, permettent une navigation intuitive entre les différents documents d'un corpus et présentent en outre des intérêts pour le référencement et l'indexation des documents par des moteurs de recherche.

Les tags sont généralement déterminés à partir de la méthode fréquentielle (Lee et Baik, 2004) via laquelle on calcule la fréquence d'apparition de chaque mot clés à laquelle on associe une méthode de sélection des mots les plus représentatifs du contenu sémantique du document [32] [33] [34].

Les tags peuvent être regroupés au sein :

- de taxonomie ou de folksonomie : une classification des mots-clés selon la réflexion et des méthodes élaborées par un ou plusieurs auteurs, spécialistes ou non;
- d'un *tag-cloud* ou nuage de mots clés : une représentation visuelle dans laquelle la taille de la police d'écriture d'un terme sera d'autant plus importante que le terme est représentatif du corpus
- de *tag-thunder* ou tonnerre de mots clés qui est à l'oreille ce que le *tag-Cloud* est à l'œil [24].

Si des tags peuvent être proposés de façon automatique par les CMS et les DAM (à partir du contenu du document et/ou de ses métadonnées), ils sont le plus souvent déterminés par l'utilisateur à partir de son appréciation et sa compréhension du contenu. Il y a donc une part importante de subjectivité dans leur choix.

Cependant, l'utilisation d'outils de fouille de textes permet de lister les termes les plus utilisés ce qui peut faciliter l'opération de taggage et la rendre plus pertinente.

Ce genre de logiciels permet de plus le regroupement des mêmes mots sous un terme générique via des processus de lemmatisation (exemple : « Internet », « Web », « toile » peuvent être regroupé sous le seul terme de « Web »).

Dans le cas d'un fichier audio, il faut le transcrire au format texte au préalable. Bien que cette opération soit couteuse en terme de temps, elle peut représenter un véritable intérêt dans la mesure où le document audio est complété par un document de transcription qui augmentera sensiblement la pertinence de l'opération de taggage.

En fonction de la nature du fichier audio (qualité de la captation, qualité d'élocution), cette transcription pourra être plus ou moins fidèlement effectuée par des logiciels de transcription automatique comme VoxSigma proposé par la société Vocapia.

Une transcription effectuée par un humain sera plus pertinente mais plus chronophage.

Le choix de la méthode dépendra donc de la nature du fond, des ressources humaines et matérielles mises à disposition et du temps alloué à cette étape.

Dans le cadre d'un processus d'éditorialisation d'un fonds documentaire audio, l'utilisation de tags :

- permettra:
 - de caractériser le contenu sémantique des phonogrammes,
 - de relier différents phonogrammes entre eux et permettre à l'utilisateur une navigation par mots clés au sein d'un ou de plusieurs corpus,
- sera tributaire:
 - du temps alloué à l'opération,
 - de la subjectivité dans le choix des mots-clés.

2.2.3 Les commentaires : l'éditorialisation à l'heure du Web 2.0

Il serait dommage de penser que le seul référenceur (dans le sens d'auteur, éditeur, documentaliste, webmaster,...) du document pourra caractériser pleinement et justement ce dernier. En effet, comme évoqué précédemment, l'attribution de métadonnées et surtout de tags fait appel à une certaine subjectivité.

De fait, faire reposer l'éditorialisation d'un document uniquement sur la subjectivité et l'appréciation d'une personne reviendrait donc à réduire le contenu informationnel de ce document à la compréhension et à la valeur qu'il a pour cette personne.

Ainsi, de la même manière que les folksonomie participent à la valorisation d'un document, il peut être intéressant de laisser aux lecteurs/auditeurs la possibilité de poster des commentaires et ainsi de participer à l'éditorialisation d'un phonogramme en y ajoutant de l'information : précision, complément bibliographique....

Le commentaire est donc une méthode d'éditorialisation particulièrement adaptée au paradigme du Web 2.0 dans lequel c'est la communauté d'utilisateurs qui donne de la valeur aux objets du web qu'elle consulte en se positionnant à la fois en tant que consommateurs, producteurs et, donc, éditorialisateurs.

Le choix d'ouvrir l'éditorialisation aux commentaires dépendra :

- du document: quelle est sa nature ? quel est son public ? quels sont les droits associés au document et à respecter ?
- de la charte d'éditorialisation adoptée,
- de l'espace sur lequel le document est mis à disposition : espace public, espace privé, réseau d'entreprise, site institutionnel...

Enfin, l'ouverture d'un document à la commentarisation suppose :

- un travail de modération de la part du responsable du corpus afin d'éviter les commentaires redondants, non-pertinents, qui contreviendraient aux droits associés au document,
- de disposer d'une base de données permettant de stocker ces commentaires.

2.2.4 Le chapitrage : diviser pour mieux éditorialiser

2.2.4.1 Approche théorique

En plus de l'ajout de métadonnée, de mots-clés, de commentaire, l'éditorialisation d'un contenu peut s'effectuer en donnant une valeur chronologique à un contenu sonore.

Au-delà du simple critère de date, il peut être intéressant dans le cadre d'un document audio correspondant à un entretien, une conférence ou un séminaire d'en chapitrer le contenu.

Chaque chapitre est alors défini par un *time code* de début et de fin, et une caractérisation de chapitrage : l'intervenant, la thématique, le lieu, l'heure, la date (dans le cadre d'une captation s'étalant sur plusieurs journées)...

Le tableau ci-dessous décrit les deux principales voies de chapitrage qui peuvent être adoptées ainsi que leurs avantages et inconvénients respectifs :

| Type de chapitrage | Principe | Avantages | Inconvénients |
|--------------------|---|--|---|
| Par intervenant | à chaque intervenant correspondra un chapitre | →adapté aux longues prises de paroles le découpage →ne nécessite pas d'écouter l'intégralité du document | →inadapté lors de dialogues très dynamiques de type « Ping-pong » |
| Par thématique | à chaque nouvelle thématique correspondra un chapitre | →adapté aux documents contenant plusieurs thématiques différentes et correctement identifiées | →le découpage ne nécessite d'écouter l'intégralité du document |

Tableau 4: Avantages et Inconvénients de deux méthodes de chapitrage

Il est à noter que le chapitrage transforme un document unitaire (a fortiori primaire) en autant de sous-documents (a fortiori secondaires) qu'il y aura de chapitres. Il induit donc une augmentation non négligeable du nombre de documents à gérer et doit faire l'objet d'une réflexion préalable, en particulier lorsque l'éditorialisation s'opère dans un contexte technologique présentant des limites en terme du nombre de fichiers supportés.

2.2.4.2 Approche pratique

D'un point de vue pratique, le chapitrage nécessite des logiciels d'édition audio qui permettront de lire et de découper le document original en sous-documents/chapitres.

A partir de l'écoute du document, mais aussi en détectant les changements d'interlocuteurs (par exemple à partir des creux dans la forme d'onde du document audio), les time codes de début et de fin de chaque chapitre sont relevés. A partir de ces derniers, on génère une *cuesheet* qui permet de découper le fichier.

Le guide présent en Annexe 6 décrit les différentes étapes du processus et une méthode pour industrialiser le procédé.

Une fois découpé en chapitre, le document peut être mis à disposition de l'auditeur sous plusieurs formes :

- intégralement et sous forme originale : via un lecteur audio contenant le phonogramme
- intégralement et sous forme découpée : via une liste de lecture dans laquelle l'auditeur pourra naviguer selon ses besoins
- partiellement : via un ou plusieurs chapitres qui correspondront aux choix de mise en ligne du responsable du corpus ou aux résultats fournis par un moteur de recherche lors de la recherche effectuée par un utilisateur...

A noter que de la même manière qu'il est possible de chapitrer une vidéo (comme par exemple cela est fait sur les films au format DVD ou Blue-Ray), il existe des conteneurs audio (mp4 et mka) qui supportent le chapitrage et permettent donc de maintenir l'intégrité du document. Cependant, leur mise en place est techniquement plus contraignante qu'un chapitrage par découpage du document, et leur lecture sur le web s'appuie sur des outils qui nécessitent des connaissances en langages de développements Web.

Une fois chapitrée, chaque sous-document peut être considéré comme un nouveau document qu'il est possible d'editorialiser à son tour (ajout de métadonnées, tags, commentaires...).

3. Comparaison de stratégies d'éditorialisation de sites web

3.1 Préambule

Pour cette partie, huit acteurs répartis selon 4 vocations seront comparés quant à leur stratégie respective d'éditorialisation :

- deux sites à vocations commerciales : Spotify et Deezer,
- deux sites à vocation communautaires : Soundcloud et Jamendo,
- deux sites à vocations journalistiques radiophoniques : France Culture et France Inter,
- deux sites à vocations associatives : le site de la Fédération des Associations de Musiques et de danses traditionnelles, le site des bibliothèques sonores de Paris.

Les sites de l'étude ont été choisis afin de disposer d'acteurs dont le contenu, l'offre et le public cible soient les plus variés possibles.

Pour chaque type de vocation, l'offre est pléthorique et de nombreux acteurs se créent ou disparaissent régulièrement [35]. Les exemples choisis ici ne l'ont pas nécessairement été parmi les acteurs les plus importants (en terme de volume d'écoute, de ventes, de visites...) mais parmi les plus représentatifs de leur secteur et/ou offrant des stratégies d'éditorialisation qui présentent un intérêt dans le cadre de ce mémoire.

3.2 Présentation des sites

3.2.1 Spotify et Deezer : deux acteurs clés de l'offre musicale

Spotify est une plateforme suédoise qui offre la possibilité à ses utilisateurs d'écouter des morceaux de musiques en streaming. Cette plateforme repose sur une technologie dite de *Pair-à-Pair* ainsi que l'utilisation de DRM.

Créée en 2006, elle revendique, 10 ans après, 5 milliards de morceaux diffusés, un catalogue de 30 millions de titres et 100 millions d'utilisateurs [36].

Son modèle économique repose sur une offre Freemium décrite dans le tableau 5:

| Nom | Prix | Publicité | Temps d'écoute par mois | Qualité audio | Écoute hors-ligne | Contenu exclusif |
|-----------------|-------------------|-----------|-------------------------|-----------------|-------------------|------------------|
| Spotify | Gratuit | Oui | Illimité | 96 à 160 kbit/s | Non | Non |
| Spotify Premium | 9,99 EUR par mois | Non | Illimité | 96 à 320 kbit/s | Oui | Oui |

Tableau 5: Offres d'abonnement Spotify

Deezer est une plateforme française créée en 2007 qui propose le même type de service que Spotify. Elle compte 16 millions d'utilisateurs, environ 7 millions d'abonnés et 40 millions de titres. Son modèle économique est à la fois basé sur le système Freemium mais s'agrément de partenariats notamment avec des opérateurs téléphoniques afin de profiter d'une plus large diffusion auprès des utilisateurs [37].

De même que Spotify, son modèle économique repose sur une offre Freemium décrite dans le tableau 6:

| Nom | Prix | Publicité | Musique illimitée sur tablette et ordinateur | Musique illimitée sur mobile | Qualité audio | Musique accessible en hors connexion |
|---------|-------------|-----------|--|------------------------------|--------------------------|--------------------------------------|
| Free | Gratuit | Oui | Oui | Non | Standard (128 kbps) | Non |
| Premium | 9€99 / mois | Non | Oui | Oui | Haute Qualité (320 kbps) | Oui |

Tableau 6: Offres d'abonnement Deezer

3.2.2 Soundcloud et Jamendo: les réseaux sociaux sonores

Soundcloud est une plateforme suédoise créée en 2007. Principalement destinée aux artistes amateurs, on y trouve aussi des comptes d'entreprises privées (maisons de disques), de radio du service public (Radio France), d'institutions (INA) et d'artistes sous contrats.

Cette plateforme revendique 200 millions de visiteurs uniques par mois et 40 millions d'inscrits.

Elle propose 6 offres dont 3 offres pour les diffuseurs. Pour les auditeurs, le site est totalement gratuit (l'inscription n'est pas nécessaire pour profiter des contenus), sans publicité et illimité. Le téléchargement des œuvres est à la discrétion des diffuseurs.

Pendant la rédaction de ce mémoire, Soundcloud a lancé une offre Soundcloud Go offrant un certain nombre d'option agrémentant l'expérience auditeur (possibilité d'écoutes hors-connexion, catalogue enrichi suite à des accords avec des maisons de disques). Cependant, les conditions exactes de l'offre ne sont pas encore clairement détaillées sur l'interface en français du site, seules les informations tarifaires sont disponibles: l'offre Soundcloud Go sera vendu 9,99 € sur Android et 12,99 € sur iOS.

Jamendo a été créé en 2004 au Luxembourg et compte 2 millions de membres, 400 000 titres, 30 000 artistes.

A l'origine, l'intégralité de son catalogue était sous licence *Creative Commons*. En 2008, un service de licences payantes basées sur la licence *CC Plus* est mis en place. Il permet aux artistes intéressés de vendre des licences d'utilisations de leurs œuvres à des entreprises commerciales. Cependant, pour les particuliers, l'intégralité du catalogue est à disposition de façon illimitée, sans publicité, et ce en streaming ou en téléchargement.

3.2.3 France Inter et France Culture : deux radios du service public

France Inter est une station de radio publique créée en 1947 sous le nom du *Club d'Essai* qui changera de nom plusieurs fois : *Paris-Inter* (1947-1957), *France I* (1957-1963), *RTF Inter* (1963) pour devenir *France Inter* le 8 décembre 1963.

Sa grille de programmation est généraliste et regroupe à la fois des journaux, des émissions d'informations, des reportages mais aussi des émissions culturelles et de sociétés ainsi que des programmes musicaux.

En juin 2016, France Inter cumule 10,7% d'audience, soit 5,8 millions d'auditeurs quotidiens et un record de 10,9% de part d'audience. Elle est en première position en France sur le créneau matinal (7h-9h) avec 3,9 millions d'auditeurs et également première radio sur les autres horaires stratégiques de l'information (12h-14h et 18h-20h).

France Culture est une station de radio publique créée en 1963. Ses programmes sont orientés sur le savoir humain : les Arts, l'Histoire, les Sciences ainsi que des émissions religieuses. Elle est l'une des seules à proposer des fictions radiophoniques.

De même que France Inter, hormis certaines fictions soumises aux contraintes de leurs éditeurs, elle propose ses programmes en streaming et en Podcast via un site web dont la nouvelle version devrait, selon Sandrine Treiner la directrice de la radio, proposer une « bibliothèque infinie » [38].

Elle représente 2,9% d'audience cumulée pour la période d'avril-juin 2016 parmi le paysage des radios mesurées par Médiamétrie et 19,4% d'audience cumulée sur le panel « Radios Thématiques » pour la même période [39].

3.2.4 Les sites de la FAMDT et de la bibliothèque sonore de Paris

Créée en 1985 à l'initiative du ministère de la culture, la Fédération des Associations de Musiques et de Danses Traditionnelles réunit 90 associations du secteur des musiques et danses traditionnelles parmi lesquelles 10 centres de musiques traditionnelles en région.

Elle représente 10 000 adhérents, 417 000 spectateurs, 500 enseignants, 5 000 élèves et est soutenue par le Ministère de la Culture et de la Communication, au niveau de la Direction Générale de la Création Artistique et par le Centre National des Variétés.

Ses missions sont de :

- permettre le développement d'un réseau entre les acteurs des musiques et des danses traditionnelles,
- représenter et de valoriser ces acteurs auprès des médias et des institutions
- les accompagner dans leurs démarches administratives, juridiques, et documentaires ainsi que dans leurs travaux de recherches.

Elle s'appuie sur un fonds documentaire de 100 000 heures d'enregistrements sonores, 20 centres de documentation mis à disposition dans un catalogue collectif d'archives sonores sur les traditions orales sur le portail www.portaildupatrimoineoral.org.

A l'initiative de la BNF, ce portail a été lancé en 2011 [40].

La bibliothèque sonore de Paris est une initiative créée en 1972 par l'ophtalmologiste Charles-Paul Wannebroucq, et rattachée à l'Association des Donneurs de Voix reconnue d'utilité publique.

Ses 4600 membres s'organisent autour de cent-vingt établissements, répartis dans toute la France, qui proposent gratuitement leurs services à 15 000 non-voyants et malvoyants.

Originellement sur cassette audio, son fonds documentaire audio de 350 000 ouvrages est désormais disponible au format CD, et MP3 sur support CD, clé USB et carte SD [40].

3.3 Grilles de comparaison

Pour chaque site, seront évalués dans l'ordre :

- le moteur de recherche de la phonothèque mis à disposition des auditeurs,
- la gestion des mots clés notamment la possibilité pour l'auditeur d'en ajouter,
- la présence de commentaires,
- la présence de chapitrage.

L'idée sous-jacente à cet ordre d'évaluation est de se positionner comme un auditeur type qui, dans un premier temps, recherche un contenu audio puis peut être amené à profiter ou à participer, consciemment ou non, à l'éditorialisation du phonogramme consulté via les outils exposés dans le deuxième paragraphe de ce mémoire.

3.4 Présentation des résultats et limites de la comparaison

Pour chaque cas d'étude, une revue complète sera présentée. A la suite de ces revues et pour chaque critère, un tableau récapitulatif sera proposé.

Il est à noter que l'étude qui sera menée par la suite présente les limites suivantes :

- seules les versions pour ordinateurs et gratuites des sites seront évaluées ; les versions mobiles, les applications mobiles ou logicielles et les possibilités offertes des abonnements payants ne seront pas prises en compte,
- la comparaison ne traitera que de l'aspect *front-office* de la phonothèque mise à disposition de l'utilisateur. La dimension *back-office* ne sera pas considérée.
- l'évaluation de la stratégie d'éditorialisation à partir d'indicateurs (nombres de visites, avis des utilisateurs...) ne sera pas abordée. Il s'agit ici de donner un état des lieux des stratégies adoptées, en aucun cas il s'agit de juger les retombées de ces choix de stratégies.

L'étude sera donc effectuée en tant qu'utilisateur du service plutôt qu'en tant qu'auditeur-consultant.

3.5 Résultats détaillés

3.5.1 Deezer

Deezer propose un panneau latéral dans lequel un moteur de recherche est mis à disposition (Figure 3).

L'interface utilisateur propose une prévisualisation des premiers résultats obtenus suite à une recherche avec un regroupement par facettes (Artistes, Albums, Titres, *Playlists*).

Cependant, le fonctionnement du moteur de recherche est très opaque (absence d'aide, de *FAQ*).

En outre, il n'y a pas de possibilités de recherche avancée.

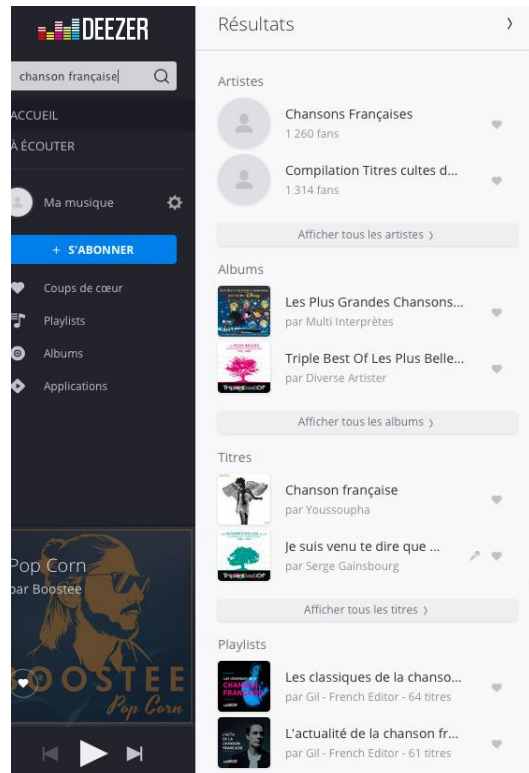


Figure 3: Moteur de recherche Deezer

Dans l'affichage complet des résultats de recherche, le moteur propose les mêmes facettes que dans l'affichage des résultats via le panneau latéral (avec une comptabilisation du nombre de résultats par facette) ainsi que les facettes *mix*, *podcast* et *profils*. L'utilisateur a aussi la possibilité de trier les résultats selon les métadonnées suivantes: titre, artiste, album, durée et popularité) (Figure 4).

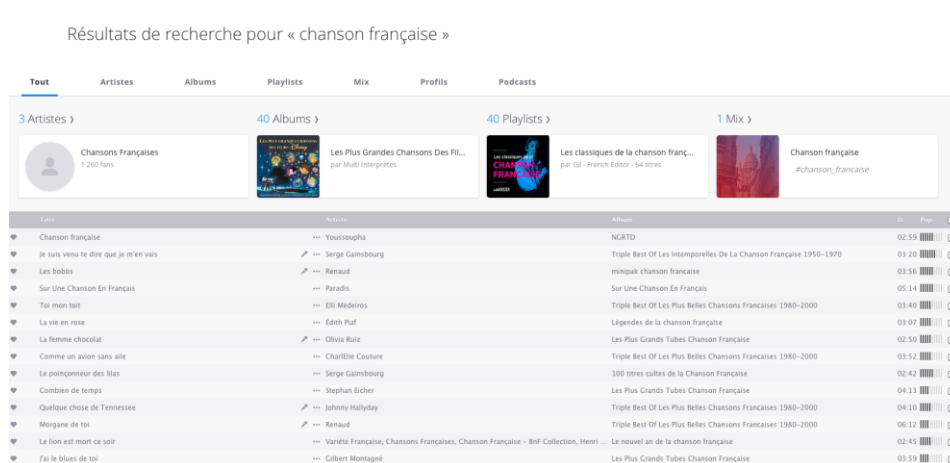


Figure 4: Résultats de la recherche Deezer

En choisissant un artiste de la liste, l'utilisateur accède à la page "Artiste" sur laquelle il dispose (Figure 5):

- d'une discographie,
- d'une biographie,
- des titres les plus populaires ou mis en avant par l'artiste (top-titres),
- une liste d'artistes similaires,
- des listes de lectures relatives à l'artiste,
- des commentaires sur l'artiste rédigés par d'autres utilisateurs et qui peuvent être partagés sur les réseaux sociaux (Figure 6).

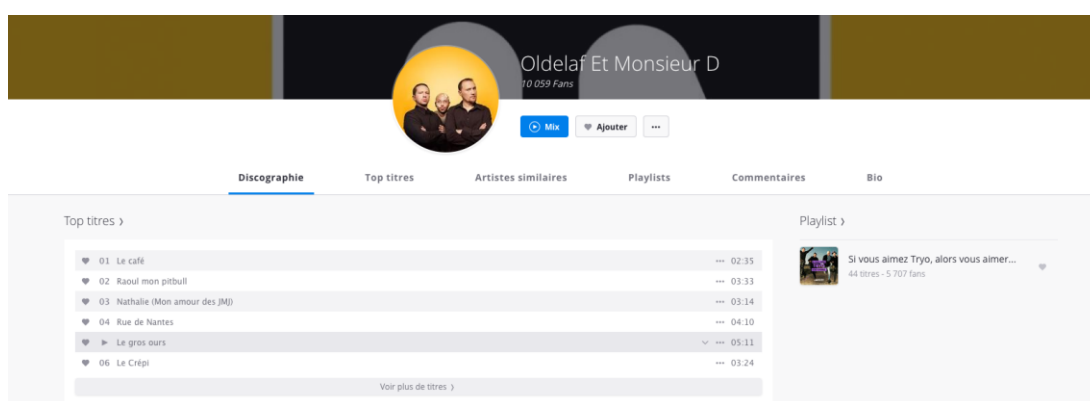


Figure 5: Deezer: page d'un artiste

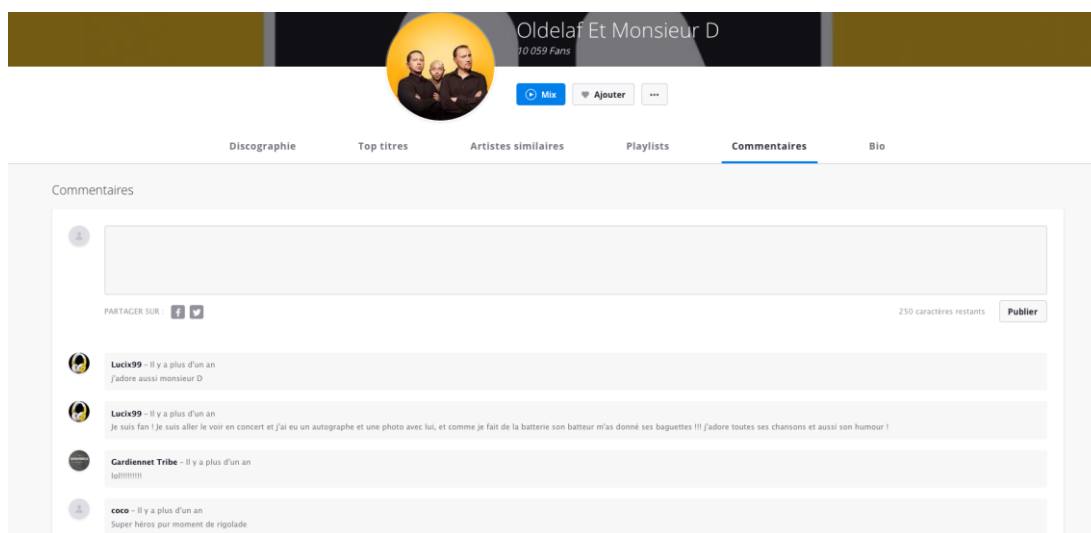


Figure 6: Deezer: zone commentaires d'une page artiste

3.5.2 Spotify

De façon quasiment identique à Deezer, Spotify propose un moteur de recherche accessible depuis un panneau latéral. L'affichage des résultats s'articule aussi en deux temps :

- affichage pendant la saisie des premiers résultats dans le panneau latéral (figure 7),
- présentation des résultats détaillés (figure 8).

On retrouve certaines mêmes facettes que sur Deezer: artiste, album, playlist ainsi que l'affichage des métadonnées artiste, titre, album, durée, popularité. Cependant, il n'est pas possible de trier les résultats.

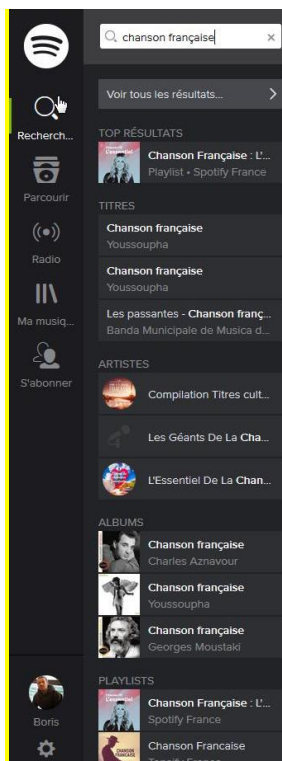


Figure 7: Spotify: résultats de recherche dans la barre latérale

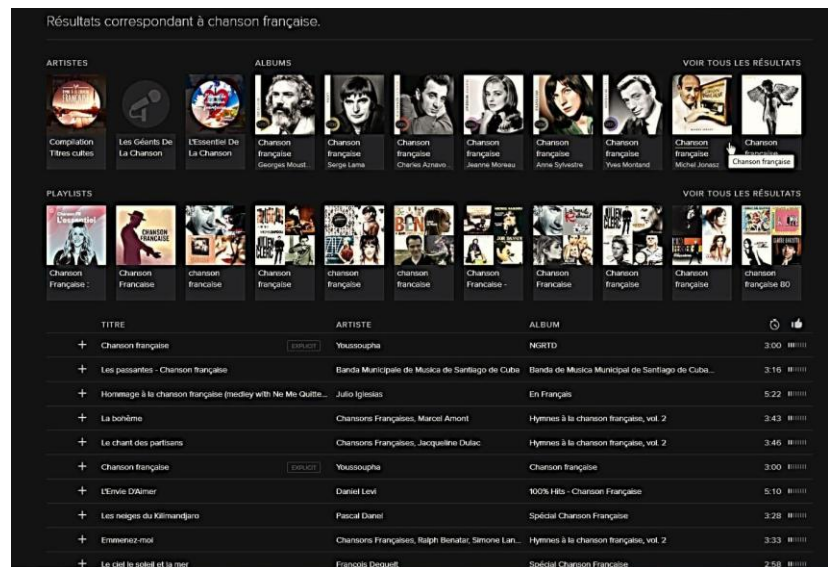
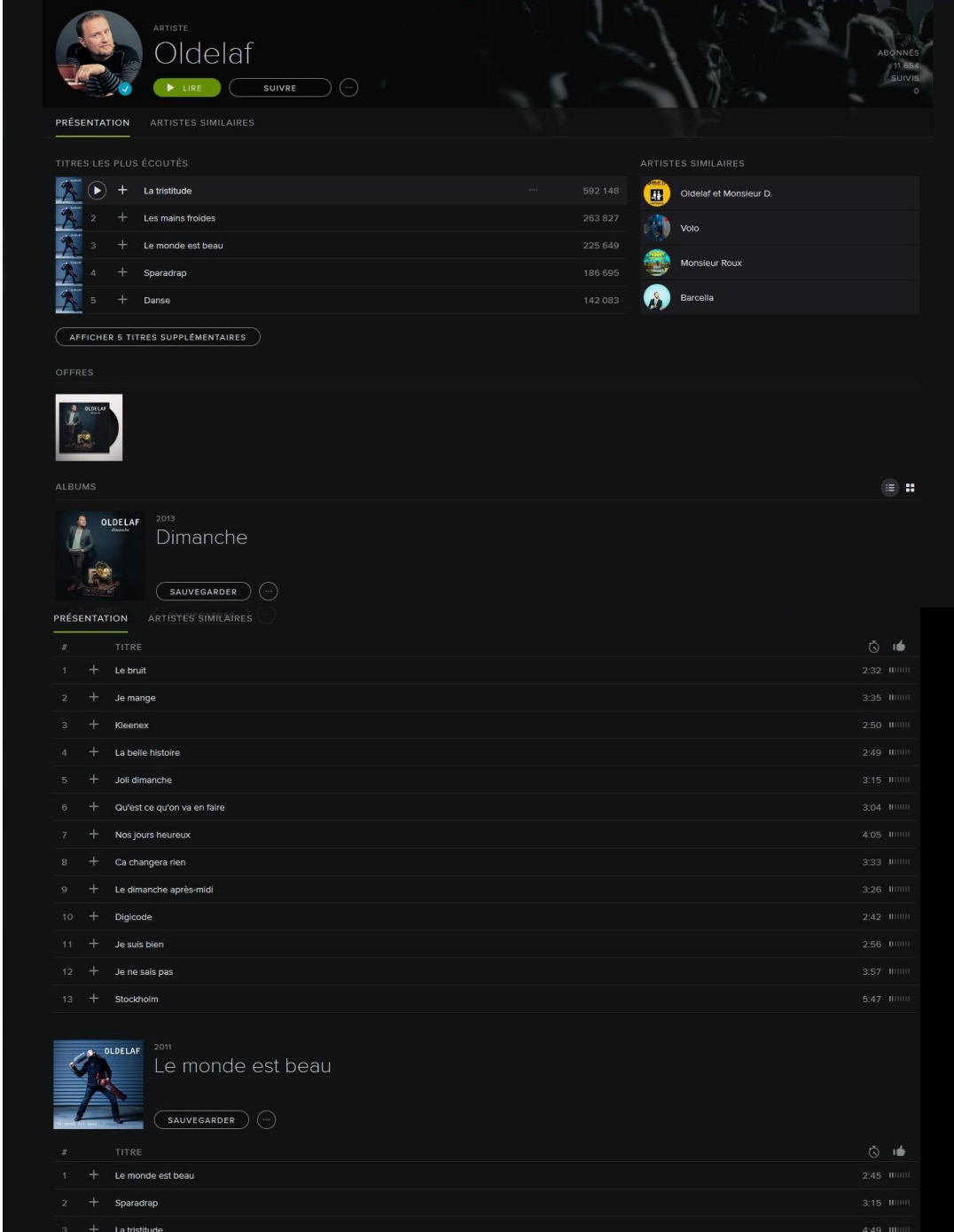


Figure 8: Spotify: résultats détaillés de recherche

Au niveau de la page artiste, Spotify ne propose pas de biographie ou de commentaires, uniquement les titres les plus écoutés, la discographie et des artistes similaires. On retrouve les mêmes métadonnées que dans la page de résultats. Il n'y a ni possibilité d'ajouter des commentaires, ni de tagger l'artiste mais seulement de *liker* et/ou partager un lien sur les réseaux sociaux vers la page artiste, un album ou un titre (figure 9).



The screenshot displays the Spotify profile for the artist Oldelaf. At the top, there is a profile picture, the name 'Oldelaf', and buttons for 'LIRE' and 'SUIVRE'. Below this, the page is divided into sections: 'PRÉSENTATION' and 'ARTISTES SIMILAIRES'. The 'TITRES LES PLUS ÉCOUTÉS' section lists five tracks with their respective play counts. The 'ARTISTES SIMILAIRES' section lists four other artists. Below these are sections for 'OFFRES' and 'ALBUMS'. Two albums are shown: 'Dimanche' (2013) and 'Le monde est beau' (2011). Each album has a 'SAUVEGARDER' button and a tracklist with track numbers, titles, and durations.

| Rang | Titre | Nombre de lectures |
|------|-------------------|--------------------|
| 1 | La tristitude | 592 148 |
| 2 | Les mains froides | 263 827 |
| 3 | Le monde est beau | 225 649 |
| 4 | Sparadrap | 186 695 |
| 5 | Danse | 142 083 |

| # | TITRE | Durée |
|----|-----------------------------|-------|
| 1 | Le bruit | 2:32 |
| 2 | Je mange | 3:35 |
| 3 | Kleenex | 2:50 |
| 4 | La belle histoire | 2:49 |
| 5 | Joli dimanche | 3:15 |
| 6 | Qu'est ce qu'on va en faire | 3:04 |
| 7 | Nos jours heureux | 4:05 |
| 8 | Ca changera rien | 3:33 |
| 9 | Le dimanche après-midi | 3:26 |
| 10 | Digicode | 2:42 |
| 11 | Je suis bien | 2:56 |
| 12 | Je ne sais pas | 3:57 |
| 13 | Stockholm | 5:47 |

| # | TITRE | Durée |
|---|-------------------|-------|
| 1 | Le monde est beau | 2:45 |
| 2 | Sparadrap | 3:15 |
| 3 | La tristitude | 4:49 |

Figure 9: Spotify: détail d'une page artiste

3.5.3 Soundcloud

Sur Soundcloud, le moteur de recherche est basique, n'offre pas de recherche avancée, et pas d'aide ou de FAQ. En revanche, il affiche les premiers résultats de recherche au cours de la saisie de l'utilisateur (Figure 10).

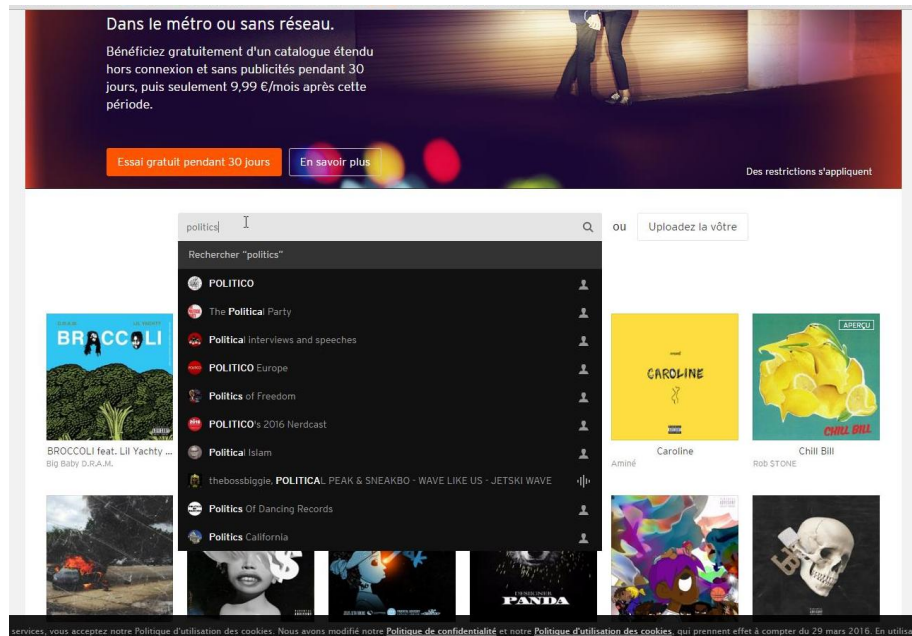


Figure 10: Soundcloud: moteur de recherche

Les résultats obtenus peuvent être précisés selon les facettes suivantes: Titres, Titres Soundcloud Go, Personnes, Albums, Playlist. Ils peuvent aussi être filtrés par date et par tags.

Il est impossible de trier les résultats par pertinence, nombre d'écoutes, nombre de like (Figure 11).

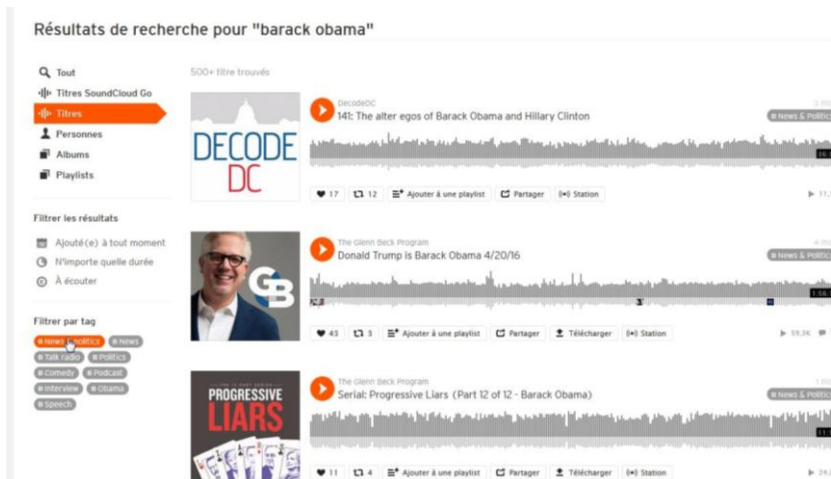


Figure 11: Soundcloud: résultats de la recherche

En accédant à la page d'un compte "Artiste", l'utilisateur peut visualiser (et à terme écouter) l'ensemble des titres mis en ligne. Pour chaque titre, l'utilisateur dispose des métadonnées suivantes : le titre, la durée, le nombre d'écoutes, de *likes* et de commentaires, une image, le mot clé principal du titre.

Enfin, l'utilisateur accède à des informations sur le compte de l'éditeur consulté : nombre d'abonné(e)s, nombre d'abonnements, nombre de titres, un descriptif du compte et différents liens vers des réseaux sociaux de l'éditeur (Figure 12).

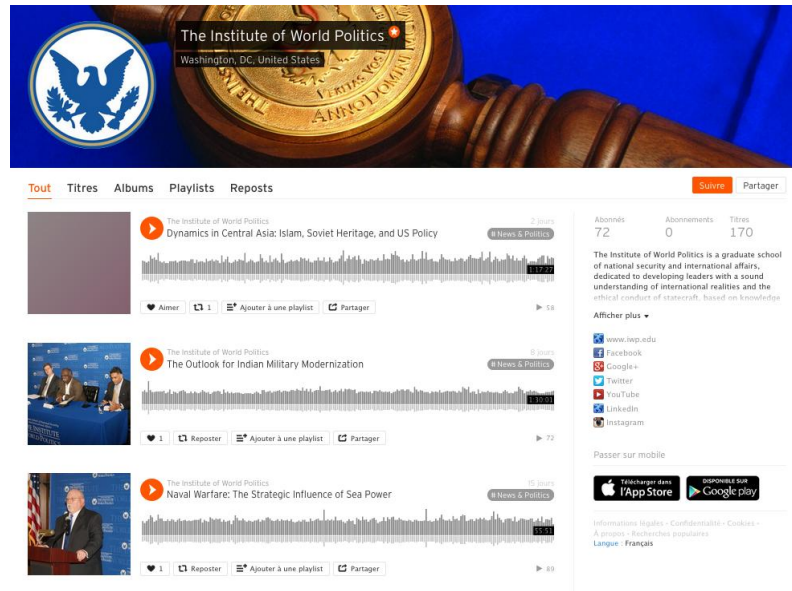


Figure 12: Soundcloud: page d'un contributeur audio

Soundcloud permet aussi à l'utilisateur de partager le titre sur les réseaux sociaux, et sous réserve qu'il soit inscrit, de le conserver dans ses favoris, le reposer sur sa page utilisateur et l'ajouter à une *playlist* de son choix (Figure 13).



Figure 13: Soundcloud: présentation d'un phonogramme

En cliquant sur un titre, l'auditeur dispose de nouvelles informations: un descriptif complet du titre, la liste des commentaires, et les différents tags du titre.

Il n'est pas possible à l'auditeur d'ajouter ses propres tags.

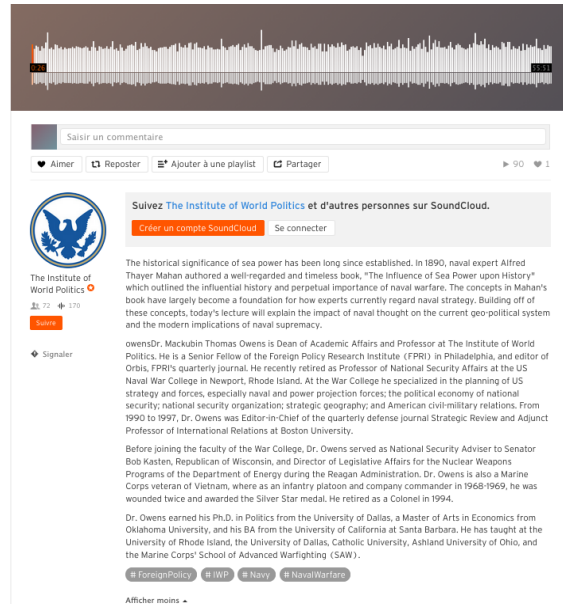


Figure 14: Soundcloud: détails d'un phonogramme

Il est à noter que sur Soundcloud, les commentaires sont associés à un minutage précis sur le titre. C'est l'un des très rares sites à offrir cette fonctionnalité.

En outre, il est possible à un auditeur de répondre aux commentaires des autres

auditeurs. Enfin, l'utilisateur qui a mis le titre en ligne peut supprimer ces commentaires (Figure 15).

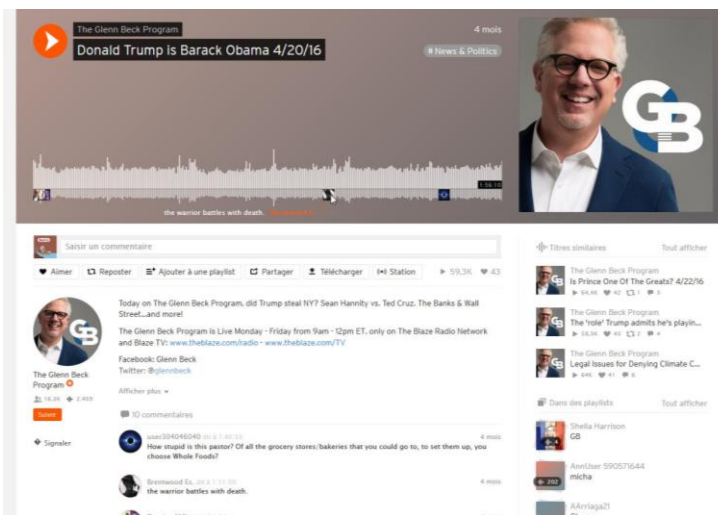


Figure 15: Soundcloud: exemple de commentaires

3.5.4 Jamendo

Jamendo propose un moteur de recherche prenant en charge la suggestion de recherche lors de la saisie mais ni options de recherche avancée, ni aide.

Pour chaque titre présent dans la fenêtre des résultats proposés, les tags associés et la durée sont affichés.

L'écoute, le téléchargement, l'ajout aux favoris ou dans une liste de lecture et le partage sont possibles (figure 16).

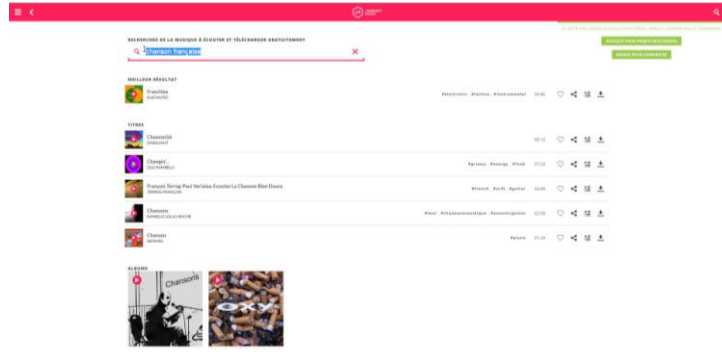
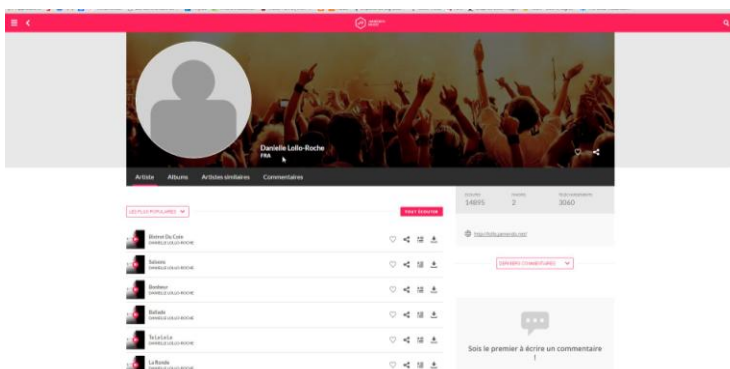


Figure 16: Jammendo: moteur de recherche

Un clic sur le nom de l'artiste affiche la page qui lui est associée sur laquelle sont indiqués le nombre d'écoutes, de téléchargements, la liste complète des titres



ainsi qu'une zone de commentaires. Un système d'onglets permet d'accéder aux albums et à une liste d'artistes similaires (figure 17).

Figure 17: Jammendo: page d'un artiste

La page d'un titre permet d'avoir accès aux paroles, au type de licence sous laquelle est déposée le titre, à une suggestion de titres supplémentaires et aux commentaires associés au titre et issus d'autres utilisateurs (figure 18).

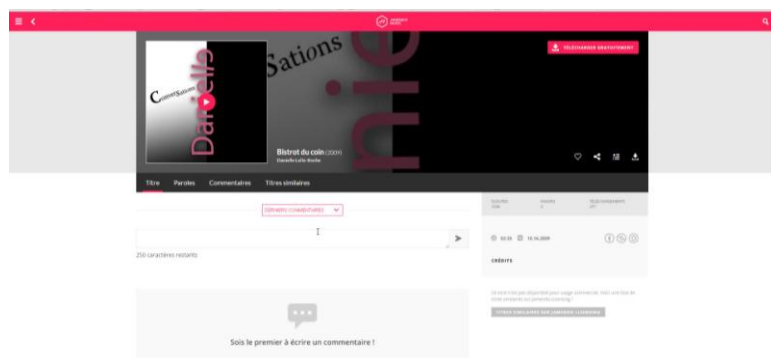


Figure 18: Jammendo: page d'un titre

3.5.5 France Culture

Le moteur de recherche prend en charge la recherche la suggestion de recherche pendant la saisie. Cependant, son fonctionnement semble erratique. En tapant le nom de l'émission *Des papous dans la tête*, on obtient beaucoup de bruit dans les résultats (figure 19). Il est préférable de saisir un mot clé plus discriminant comme *papous* (figure 20) pour obtenir le résultat escompté.

En outre, il n'y a pas de possibilité de recherche avancée ou d'aide à propos du moteur utilisé.

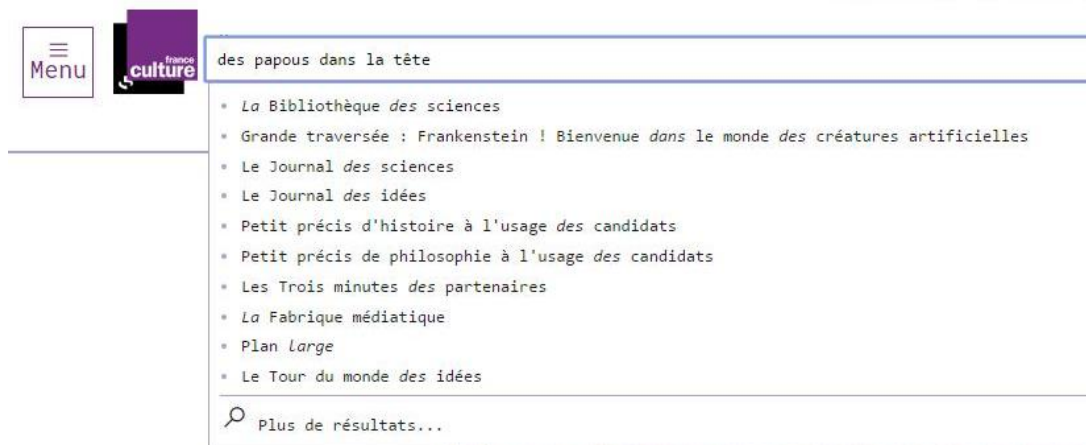


Figure 19: France Culture: moteur de recherche



Figure 20: France Culture: limite de la suggestion à la saisie

L'affichage des résultats détaillés n'offre pas plus de possibilités à l'utilisateur. Les résultats sont présentés de manière brute, sans tri particulier (figure 21).

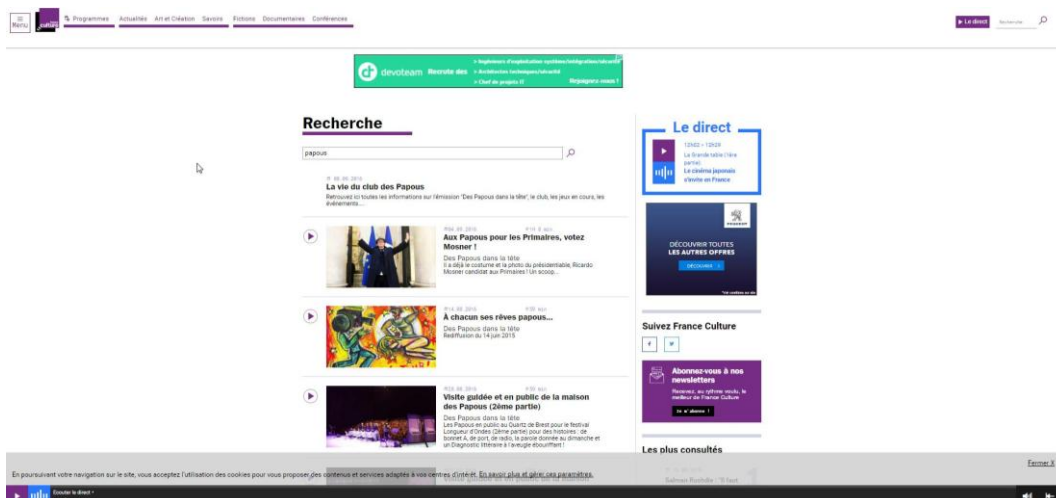
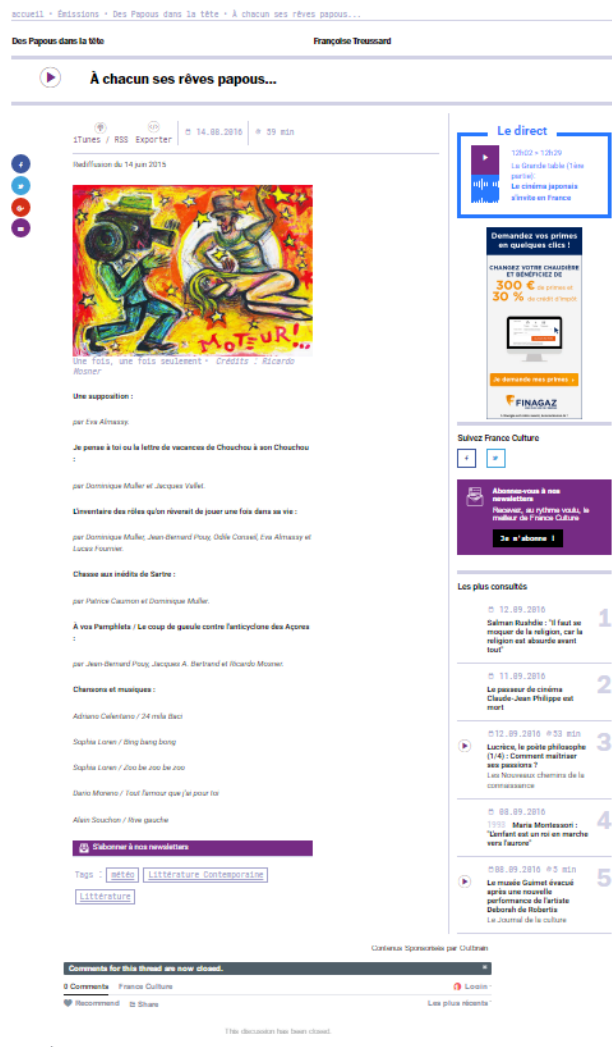


Figure 21: France Culture: page de résultats d'une recherche

En cliquant sur un résultat correspondant au titre souhaité, l'utilisateur accède à la page du phonogramme et dispose d'un chapitrage simple de l'émission, de sa durée, de la date de diffusion ainsi que d'une liste de mots clés. Il lui est aussi possible d'exporter le code HTML du média pour l'héberger sur un site de son choix (blog, réseaux sociaux...) et un lien vers le podcast du média.

Comme cela est souvent pratiqué, des boutons de partages sur les réseaux sociaux sont disponibles.

En revanche, il n'a pas la possibilité de participer à l'éditorialisation du document via des commentaires (Figure 22).



The screenshot shows a podcast page on the France Culture website. The main title is "À chacun ses rêves papous..." with a duration of 39 minutes and a broadcast date of 14.09.2016. Below the title is a colorful illustration of a man and a woman. The page lists several articles related to the podcast, such as "Une fois, une fois seulement" and "Chasse aux inédits de Sartre". There are social media sharing buttons and a "Salonner à nos newsletters" button. The page also features a "Le direct" section with a video player and a "Les plus consultés" section with a list of popular content.

Figure 22: France Culture: page d'une émission

3.5.6 France Inter

Le moteur de recherche est identique à celui utilisé sur France Culture: une barre de recherche simple, la prise en charge de suggestion à la saisie et l'absence de recherche avancée ou d'aide.

Ici, une recherche sur le mot clé "Matinale" renvoie des résultats dont la pertinence semble assez faible (figure 23).

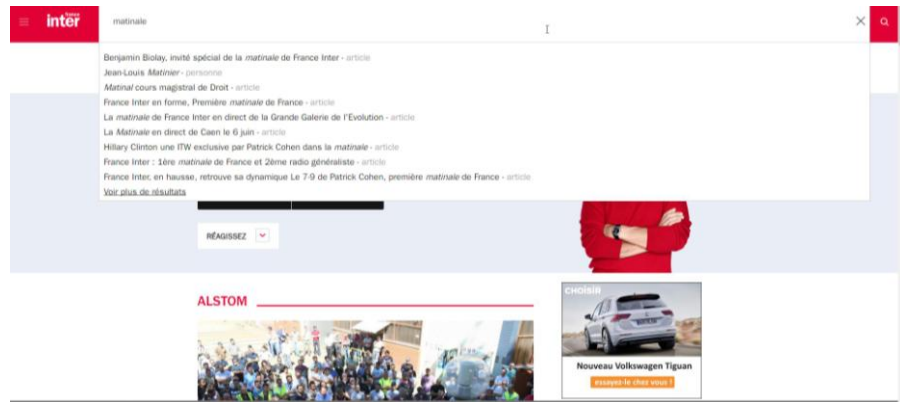


Figure 23: France Inter: moteur de recherche

L'affichage des résultats de la recherche met peu d'informations à disposition: titre de l'émission, présentateur et la catégorie. En revanche, aucune information quant à la durée ou à la date de diffusion (figure 24).

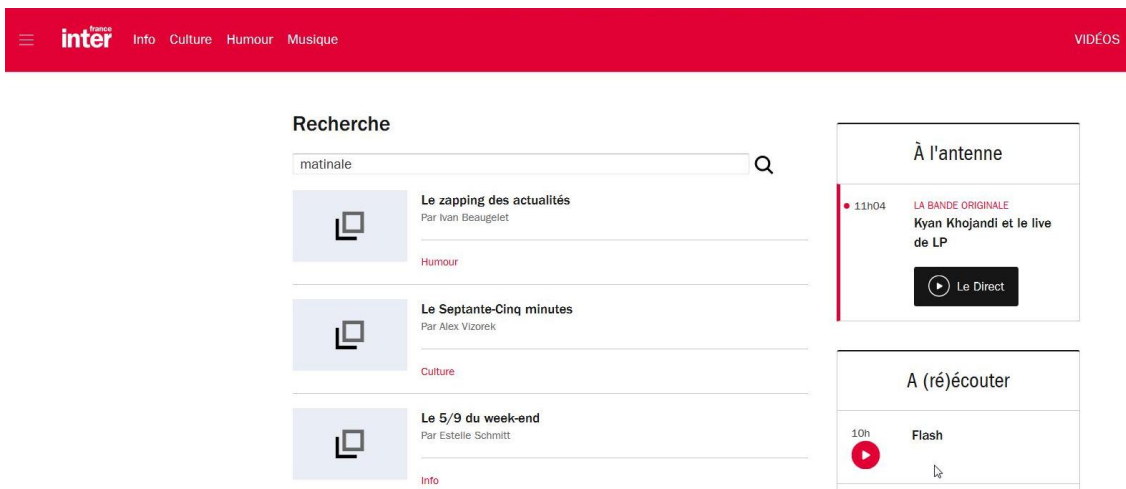


Figure 24: France Inter: résultats d'une recherche

Au niveau de la page correspondant au phonogramme choisi, de nombreuses

informations sont disponibles. En plus du titre, de la date de diffusion, l'auditeur dispose d'un chapitrage complet (*time code*, intitulé du chapitre correspondant à la chronique ainsi que le nom du chroniqueur) dans lequel chaque chapitre est considéré comme un flux à part entière, donc directement accessible.

En fin de page, on retrouve le nom de l'invité, un descriptif, les différentes références utilisées pendant l'émission, ainsi que l'équipe complète (nom des intervenants et leur(s) fonction(s)).

Enfin, des mots-clés associés à l'émission peuvent conduire l'auditeur à d'autres contenus sur le(s) même(s) thème(s).

A noter qu'il existe une zone commentaire mais qu'elle est fermée dans l'exemple ci-contre (figure 25).

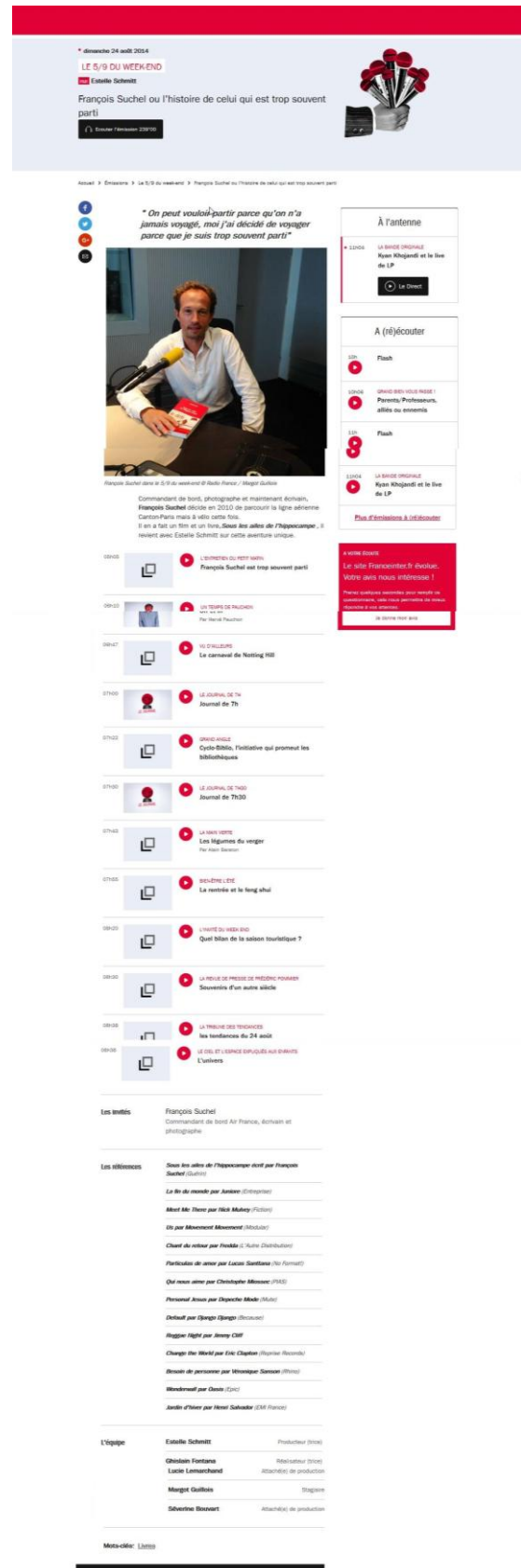


Figure 25: France Inter: page d'une émission

3.5.7 Bibliothèque sonore de Paris

Le site de la bibliothèque sonore de Paris propose au choix une navigation parmi l'intégralité du corpus mais met aussi à disposition un moteur qui permet de faire une recherche par mots-clés (titre, auteur) ainsi que de rechercher dans un corpus correspondant à un auteur ou à un genre (figure 26)



Figure 26: Bibliothèque Sonore de Paris: moteur de recherche

Ces deux dernières possibilités sont intéressantes mais les listes proposées ont un très (trop?) grand nombre d'entrées (figure 27).



Figure 27: Bibliothèque Sonore de Paris: listes d'auteurs et de genres

Il existe une recherche alphabétique par titre mais le lien est seulement disponible en bas de page (Figure 28). On peut imaginer que cela ne présentera pas de problème aux utilisateurs (malvoyants ou non-voyants) qui passent généralement par des navigateurs spécialisés qui lisent les différents liens de chaque page visités.



Figure 28: Bibliothèque Sonore de Paris: l'accessibilité relative de la recherche avancée

A noter qu'il existe un autre moteur de recherche qui permet d'interroger le catalogue nationale de l'édition adaptée. En plus des mêmes champs de recherche que le moteur précédent, il dispose d'une aide à l'utilisateur et d'une recherche par bibliothèque (figure 29)

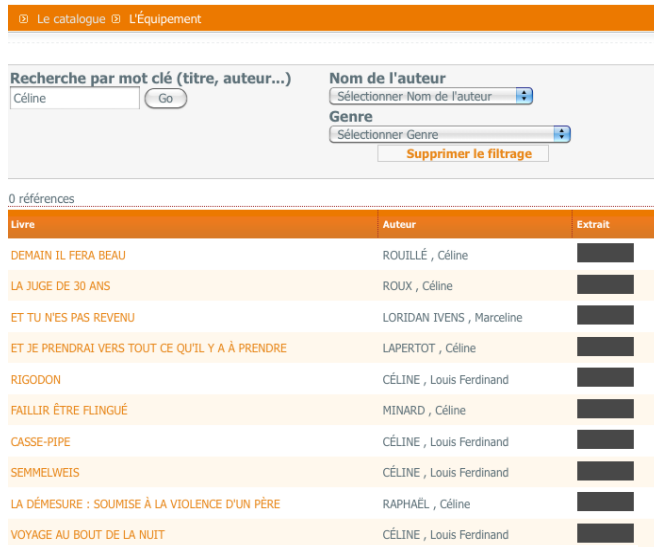


Figure 29: Moteur de recherche de l'édition adaptée

Ici, en effectuant une recherche sur le roman de Louis Ferdinand Céline *Voyage au bout de la nuit* à partir du simple nom de l'auteur, le moteur de recherche retourne les résultats suivants.

On note qu'il n'y a pas de possibilité de trier, de filtrer ou d'indicateur de pertinence.

En revanche, un extrait audio est écoutable directement depuis la page des résultats (figure 30).



Le catalogue L'Équipement

Recherche par mot clé (titre, auteur...)
Céline

Nom de l'auteur
Sélectionner Nom de l'auteur

Genre
Sélectionner Genre

0 références

| Livre | Auteur | Extrait |
|---|--------------------------|---------|
| DEMAIN IL FERA BEAU | ROUILLÉ, Céline | |
| LA JUGE DE 30 ANS | ROUX, Céline | |
| ET TU N'ES PAS REVENU | LORIDAN IVENS, Marceline | |
| ET JE PRENDRAI VERS TOUT CE QU'IL Y A À PRENDRE | LAPERTOT, Céline | |
| RIGODON | CÉLINE, Louis Ferdinand | |
| FAILLIR ÊTRE FLINGUÉ | MINARD, Céline | |
| CASSE-PIPE | CÉLINE, Louis Ferdinand | |
| SEMMELWEIS | CÉLINE, Louis Ferdinand | |
| LA DÉMESURE : SOUMISE À LA VIOLENCE D'UN PÈRE | RAPHAËL, Céline | |
| VOYAGE AU BOUT DE LA NUIT | CÉLINE, Louis Ferdinand | |

Figure 30: Bibliothèque Sonore de Paris: exemple de recherche

En choisissant le titre recherché, l'utilisateur accède aux métadonnées suivantes : auteur, éditeur, disponibilité, genre, référence et durée du fichier. Un extrait sonore est aussi proposé. Un champ commentaire est présent mais son contenu semble destiné aux gestionnaires de collection et aux administrateurs du site plus qu'aux utilisateurs (figure 31).

Il n'est pas possible de partager, commenter ou d'ajouter des mots-clés sur l'ouvrage.

Détail de l'ouvrage

| | |
|------------------|--------------------------------------|
| Titre | VOYAGE AU BOUT DE LA NUIT |
| Commentaires | 78W-a |
| Extrait sonore | |
| Auteur | CÉLINE, Louis Ferdinand |
| Nom de l'Editeur | Editions DENOËL |
| Disponibilité | disponible |
| Genre | Romans, Nouvelles, Contes, Aventures |
| Référence | 80078 |
| Duree | 19 heures 20 minutes |

Figure 31: Bibliothèque sonore de Paris: détail d'un phonogramme

3.5.8 Portail du patrimoine oral

Le portail du patrimoine oral offre deux interfaces de recherche.

La première, un simple champ de recherche, est accessible directement depuis la page d'accueil du site (figure 32).

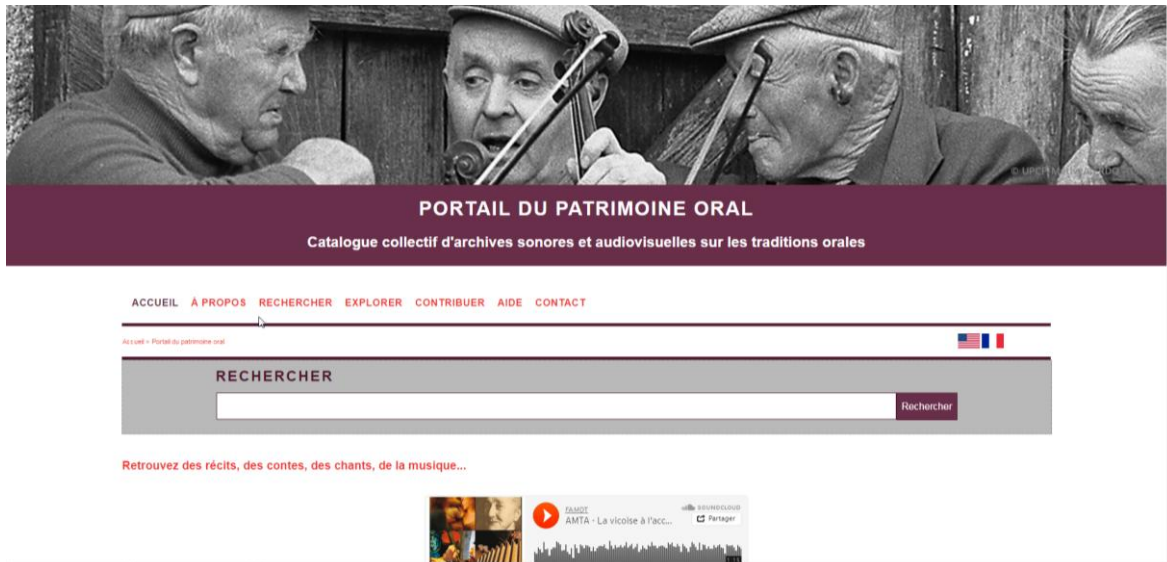


Figure 32: Portail du patrimoine oral: moteur de recherche

La deuxième, accessible via un lien dédié, offre plus de champs de recherche: mots-clés, contributeur, lieu, auteur(s), langue(s), côte ou numéro d'inventaire, sujet, mot(s) du titre (figure 33).



Figure 33: Portail du patrimoine oral: moteur de recherche avancée

Dans les deux cas, la recherche en cours de saisie n'est pas implémentée. Enfin, une page d'aide donne à l'utilisateur quelques renseignements sur le fonctionnement de la recherche au sein du site.

Les résultats de la recherche sont présentés sous forme d'une liste paginée limitée à 1000 éléments. Pour chaque élément, sont affichées les métadonnées correspondant au titres, aux contributeurs, la date ainsi qu'une accès direct à la notice (élaborée le portail oral) et à la notice originale (élaborée par le contributeur). En revanche, les critères de recherche ne sont pas rappelés dans la fenêtre de résultat. Il n'y a pas de possibilité ni de trier, ni de filtrer les résultats obtenus (figure 34).

1 - 25 de 1000 éléments 1 2 3 4 5 6 7 8 9 10 > >>

Explications sur la transmission des contes et récit de cinq histoires par un...
MMSH - Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme PACA - Provence - Alpes - Côtes d'Azur
 Castell, Claudette - enquêteur, Coulomb, Nicole - enquêteur, Volpilière, Marcel - informateur, Volpilière, Marinette - informateur
 1984-02-01
[VOIR LA NOTICE](#) | [VOIR LA NOTICE ORIGINALE](#) 🔊

Répertoire de contes traditionnels cévenols d'un ancien paysan cévenol
MMSH - Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme PACA - Provence - Alpes - Côtes d'Azur
 Castell, Claudette - enquêteur, Coulomb, Nicole - enquêteur, Volpilière, Marcel - informateur, Volpilière, Marinette - informateur
 1983-09-24
[VOIR LA NOTICE](#) | [VOIR LA NOTICE ORIGINALE](#) 🔊

Recherches de terrain en Cévennes de Jean-Noël Pelen
MMSH - Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme PACA - Provence - Alpes - Côtes d'Azur
 Pelen, Jean-Noël - enquêteur
 -1-11-30
[VOIR LA NOTICE](#) | [VOIR LA NOTICE ORIGINALE](#) 🔊

Contes facétieux et merveilleux racontés par une habitante de Vaison-la-Romaine
MMSH - Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme PACA - Provence - Alpes - Côtes d'Azur
 Ramel, Jean-Louis - enquêteur, 0844 - informateur
 1990-08-01
[VOIR LA NOTICE](#) | [VOIR LA NOTICE ORIGINALE](#) 🔊

A propos de l'origine des contes racontés par Sidonie et l'implantation de sa famille...
UFCP-Métive Parthenay
 participant : Rousselot, Micheline, informateur : Giraud, Sidonie, enquêteur : Rouger, Jany, enquêteur : Landry, Jean-Michel
 1978-09-13
[VOIR LA NOTICE](#) | [VOIR LA NOTICE ORIGINALE](#) 🔊

Contes facétieux et anecdotes d'un habitant de Mérindol-les Oliviers
MMSH - Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme PACA - Provence - Alpes - Côtes d'Azur
 Ramel, Jean-Louis - enquêteur, 0849 - informateur, 0859 - informateur
 1990-10-10
[VOIR LA NOTICE](#) | [VOIR LA NOTICE ORIGINALE](#) 🔊

Discussion sur les contes et récits fantastiques avec un couple de vieux cévenols
MMSH - Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme PACA - Provence - Alpes - Côtes d'Azur
 Castell, Claudette - enquêteur, Coulomb, Nicole - enquêteur, Volpilière, Marcel - informateur, Volpilière, Marinette - informateur
 1983-09-24

Figure 34: Portail du patrimoine oral: présentation des résultats

Une fois le document sélectionné, la page du document affiche un nombre plus conséquent d'informations et un lien vers la page du document sur le site du contributeur (figure 35).

Explications sur la transmission des contes et récit de cinq histoires par un agriculteur à la retraite en Cévennes

MMSH - Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme PACA - Provence - Alpes - Côtes d'Azur
VOIR LES INFOS D'ENTRÉE

| CHAMP | VALEUR |
|-------------------------|--|
| Mot(s) du Titre | Explications sur la transmission des contes et récit de cinq histoires par un agriculteur à la retraite en Cévennes |
| Auteur(s) | Castell, Claudette - enquêteur Coulomb, Nicole - enquêteur Voipilière, Marcel - informateur Voipilière, Marinette - informateur |
| Sujets | conte facéteux chien relation jeunesse-vieillesse patron relation de travail berger élection maire transmission des contes Celui qui voulait faire le chien La raie impeccable Le berger qui n'était pas d'accord avec son patron Les trois chevilles Le choix |
| Mot du résumé | L'entretien commence par les demandes de précisions des enquêtrés sur certains contes déjà répertoriés par elles. L'informateur se souvient de certaines origines des contes et cite les noms des personnes auprès de qui il les a appris. La suite de l'entretien est une série de citation des titres de contes et les explications sur les personnes les ayant racontés ou les lieux. Certains personnages d'anciens conteurs émergent. L'informateur raconte cinq contes en entier comme "Celui qui voulait faire le chien", "La raie impeccable", "Le berger qui n'était pas d'accord avec son patron", "Les trois chevilles" et "Le choix du Maire". La discussion continue sur le père de l'informateur qui connaissait beaucoup de contes issus ou non de l'Aimaniach. Enfin, la conversation est assez générale en fin d'enregistrement sur d'anciens conteurs, le réseau de connaissances des informateurs et des enquêtrés. |
| Editeur | Portail du Patrimoine Oral |
| Date | 1984-02-01 |
| Type(s) de document | archives sonores conte-légende-récit enquête sound |
| Format / durée | 1 bde 47min |
| Notice originale | http://phonothèque.mmssh.univ-aix.fr/Record.htm?record=19110288124919384809 |
| Cote ou n° d'inventaire | 421 |
| Langue(s) | français |
| Contenu dans | Mémoire orale du Mont-Lozère |
| Lieu | Alès |
| Droits | Contrat de dépôt signé en 2000. |

Figure 35: Portail du patrimoine oral: détail d'un phonogramme

Dans le cas où le fichier est accessible en ligne, il est disponible uniquement sur le site du contributeur, et au format asx (format de streaming audio supportés par un grand nombre de lecteurs multimédia).

Il n'y aucune possibilité d'éditorialisation par l'utilisateur que ce soit via des commentaires ou des tags. De même, aucun outil facilitant le partage des documents (sur des réseaux sociaux par exemple) n'est mis à disposition.

3.6 Le cas de l'Institut National de l'Audiovisuel

3.6.1 L'INA: quand l'archivage devient une mission d'état

L'Institut National de l'Audiovisuel est un établissement public à caractère industriel et commercial français créé le 6 janvier 1975.

Sa mission consiste à sauvegarder, valoriser et transmettre le patrimoine audiovisuel français.

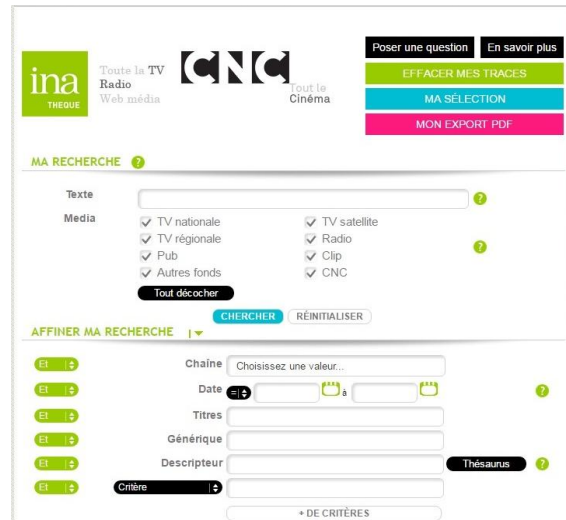
Son fond représente près de 14 700 000 heures de documents télévisuels et radiophoniques, 14 000 sites web médias, 1 200 000 photos, 123 chaînes de télé et de radio captées au titre du dépôt légal, et 1 463 881 heures sauvegardées et numérisées (fin 2015) [41].

3.6.2 Le poste de consultation multimédia

Afin de mettre à disposition son fonds documentaire, l'INA a mis en place des Poste de Consultation Multimédia. Installés dans les Bibliothèques et Médiathèques municipales à vocation régionale, ils permettent de consulter, l'intégralité des fonds audiovisuels via une interface dédiée. Pour les professionnel(le)s, cette consultation peut se faire avec le soutien et les conseils des équipes de documentalistes et de technicien(ne)s.

L'interface de recherche offre de nombreuses possibilités. Au delà du champ de recherche, il est possible de choisir le corpus de recherche, la chaîne, une plage de date, le titre, le générique, les descripteurs. Chaque critère (figure 37) peut être choisi inclusivement ou exclusivement à l'aide d'opérateur de recherche (*et, ou, sauf*).

En outre, un thésaurus est accessible afin d'utiliser avec plus de pertinence le champ descripteur. Le moteur offre aussi un historique de recherche chaque recherche est éditable et les résultats consultables ultérieurement.



The screenshot shows the 'MA RECHERCHE' section of the INA search engine. It includes a search bar, a 'Media' section with checkboxes for TV nationale, TV régionale, Pub, Autres fonds, TV satellite, Radio, Clip, and CNC. Below this is the 'AFFINER MA RECHERCHE' section with dropdown menus for 'Et' (operator), and input fields for 'Chaîne', 'Date', 'Titres', 'Générique', 'Descripteur', and 'Critère'. A 'Thésaurus' button is also visible next to the 'Descripteur' field.

Figure 36: Poste de Consultation Multimédia de l'INA: moteur de recherche

Enfin, chaque champ bénéficie d'une zone d'aide affichée dans la même fenêtre sous forme d'info-bulles (figure 38).



Figure 37: Poste de Consultation Multimédia de l'INA: exemple d'aide

L'affichage des résultats se fait dans une nouvelle fenêtre juxtaposée au formulaire de recherche. Pour chaque résultat, l'utilisateur dispose de la date du fichier, du logo de la chaîne correspondante, du titre, et de la durée. En sélectionnant un fichier, de nouvelles métadonnées sont affichées ainsi que le lecteur (figure 39 - cadre bleu). Les métadonnées sont exportables au format pdf (figure 39 - cadre orange).

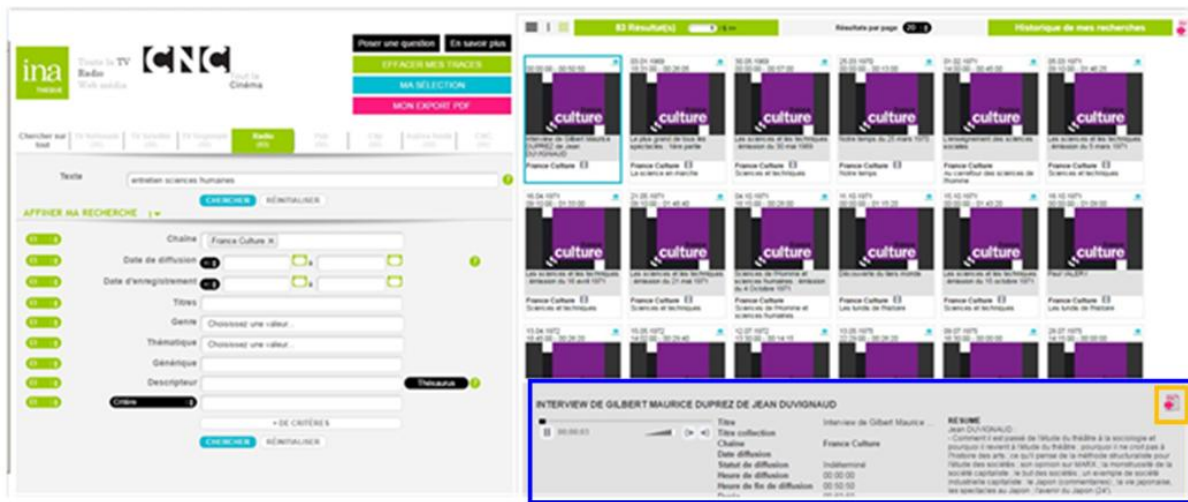


Figure 38: Poste de Consultation Multimédia de l'INA: résultats de recherche

En choisissant le fichier, l'utilisateur accède à la notice détaillée dans laquelle il peut trouver en plus du lecteur toutes les métadonnées associées au média notamment un résumé, le générique, la thématique...

Il est possible d'ajouter la notice à une sélection qui pourra ultérieurement faire l'objet d'export au format pdf.

En revanche, l'éditorialisation n'est pas ouverte: il n'est pas possible d'ajouter des notes, des commentaires ou des mots-clés, et ce même pour être consultés par l'utilisateur seul.

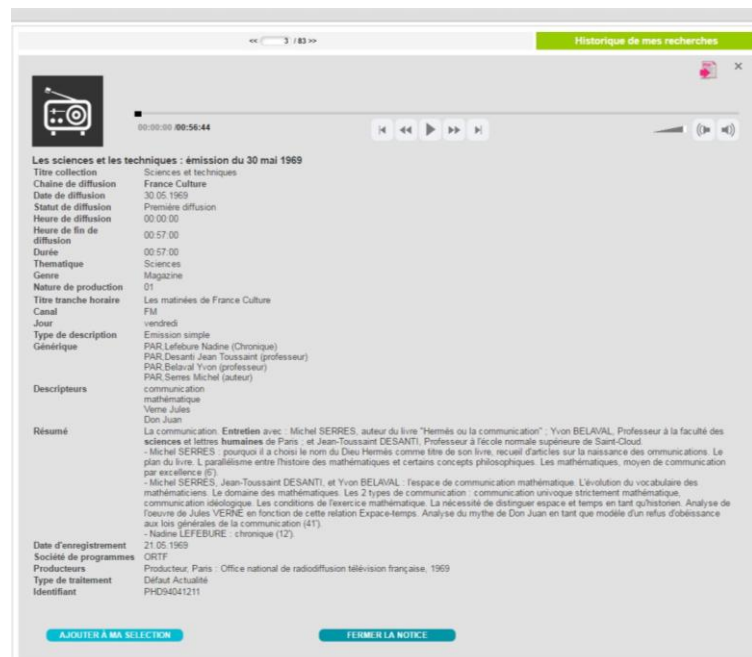


Figure 39: Poste de Consultation Multimédia de l'INA: détail d'un phonogramme

3.7 Récapitulatif des résultats

Le tableau 7 propose un résumé des résultats.

Pour chaque fonctionnalités, trois notes sont possibles:

- -1: la fonctionnalité est absente
- 0: la fonctionnalité est proposée mais de façon très sommaire
- 1: la fonctionnalité est présente

| Sites \ Critères | Moteur de recherche | | | Affichage des résultats | | Valorisation et éditorialisation | | | | |
|------------------------------|-------------------------------|-------------------|----------------|---------------------------|---------------|----------------------------------|----------------------|------------|--|--------------|
| | Moteur de recherche dynamique | Recherche Avancée | Aide du moteur | Métadonnées des résultats | Tri / Facette | Présence de mots clés | Métadonnées du titre | Chapitrage | Possibilité de réutilisation du titre ² | Commentaires |
| Spotify | 1 | -1 | -1 | 1 | 1 | -1 | 0 | -1 | 1 | -1 |
| Deezer | 1 | -1 | -1 | 1 | -1 | -1 | 0 | -1 | 1 | 1 |
| Soundcloud | 1 | -1 | -1 | 1 | 0 | 1 | 0 | -1 | 1 | 1 |
| Jamendo | 1 | -1 | -1 | 0 | -1 | 0 | 1 | -1 | 1 | 1 |
| France Inter | 1 | -1 | -1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| France Culture | 1 | -1 | -1 | 0 | -1 | 1 | 1 | 1 | 1 | 1 |
| Bibliothèque Sonore de Paris | -1 | 0 | -1 | 0 | -1 | 0 | 1 | -1 | -1 | -1 |
| Portail du Patrimoine Oral | -1 | 0 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | -1 |
| INA | -1 | 1 | 1 | 1 | -1 | 1 ³ | 1 | -4 | 0 | -1 |

Tableau 7: Récapitulatif des résultats de la comparaison de portails

² Export HTML, boutons de partage sur les réseaux sociaux

³ Les descripteurs du document sont ici considérés comme des mots-clés

3.8 Conclusion de la comparaison

Si dans le tableau précédent, il aurait été possible d'additionner pour chaque site les notes obtenues pour les différents critères afin d'obtenir une note globale, cela n'aurait pas été pertinent.

En effet, les fonctionnalités et les choix éditoriaux d'un site web proposant des phonogrammes sont à mettre en regard des utilisateurs de ce site.

L'utilisateur d'un service de musique en ligne n'aura pas les mêmes attentes vis à vis des métadonnées qu'un professionnel de l'information ou un chercheur.

De même, un utilisateur qui cherche à écouter ou réécouter une émission radiophonique n'aura pas besoin de fonctionnalités poussées au niveau du moteur de recherche.

Enfin, un site destiné aux malvoyants ne s'encombrera pas d'une interface trop riche d'information textuelle.

On peut néanmoins dégager des tendances:

- les moteurs de recherche sont généralement simples et de fait n'ont pas besoin ni ne mettent à disposition une aide quant à leur utilisation,
- le chapitrage est une fonctionnalité rarement utilisée. En effet, comme vu dans le paragraphe 2.2.4, il s'agit d'un processus long à mettre en place nécessitant des ressources techniques et humaines et pouvant multiplier significativement le nombre de documents secondaires obtenus à partir d'un document primaire,
- le partage sur les réseaux sociaux est largement plébiscité. Les gestionnaires de portails inscrivent donc complètement leur portail dans le web 2.0,
- les mots-clés sont très largement utilisés, ce qui confirme qu'au delà de leur pertinence, ils permettent une prise de connaissance rapide et sans ambiguïté du contenu d'un phonogramme.

4. Mise en application : le blog des ateliers DL Web

4.1 Contexte

Depuis 2006, l'INA a la mission d'archiver une liste de sites en rapport avec son activité patrimoniale d'archivage des programmes radiophoniques et télévisuels.

Afin d'étudier et promouvoir les usages du web et de ses archives à des fins d'étude et de recherche, des ateliers mensuels de recherches méthodologiques du dépôt légal du web à l'Ina ont été créés en novembre 2009.

Ils s'adressent aux étudiants, chercheurs, professionnels des médias et abordent des problématiques liées aux enjeux de la migration des contenus en ligne et à l'évolution des usages pour leur production, réception et conservation.

Chaque atelier a fait l'objet:

- d'une captation audio intégrale,
- d'une prise de notes.

4.2 Enjeu

Les différentes informations relatives à ces séances sont disponibles en ligne via un site Wordpress (<http://atelier-dlweb.fr/blog/>), à raison d'une page par atelier.

L'objet de cette mise en application est donc de:

- prendre connaissance du fonds documentaire,
- l'éditorialiser,
- le valoriser au sein du blog.

Comme évoqué précédemment, il est impératif de s'interroger quant aux publics visés qui se composent majoritairement:

- des participants aux ateliers (intervenants et spectateurs),
- des chercheurs intéressés par la thématique abordée lors de l'atelier.

4.3 Stratégie adoptée

4.3.1 Traitement du fonds documentaire

Le fonds documentaire était disponible au format numérique, sur support CD. Chaque CD a donc été encodé et stocké sur disque dur au format mp3.

Ce choix permettait:

- une meilleure garantie de conservation du fond: la durée de vie d'un disque dur représente un peu plus du double de celle d'un CD. En outre, les fichiers étant stockés sur des serveurs d'hébergement professionnels, leurs durée de vie est quasiment infinie.
- un format, mp3, adapté à la fois pour le web (taille moindre donc temps de chargement des fichiers réduits) et aux traitements audio éventuels.

Dans le cas où la captation s'articulait sur plusieurs fichiers (par exemple en cas de coupure inopinée de la captation), ces derniers ont été fusionnés afin de satisfaire l'approche documentaire suivante:

1 atelier = 1 CD = 1 fichier numérisé = 1 document primaire

Une fois numérisé, le fond a été décrit selon deux approches:

- une approche descriptive technique (tableau 8) qui permet de décrire le contenant du document (format, durée, taille, emplacement) à raison d'une ligne par fichier numérisé,
- une approche descriptive thématique (tableau 9) qui permet de décrire le contenu du document à raison d'une ligne par triplet {atelier, conférence, conférencier}. En outre, cette approche sera utile pour le travail de chapitrage.

Pour ces deux approches, il est vivement recommandé d'utiliser un tableur logiciel (Microsoft Office, OpenOffice Calc...) qui facilite d'une part la lecture mais aussi certaines opérations de gestion du fond (calcul de la durée totale du fond, calcul de la taille totale des fichiers du fond,

insertion de nouveaux documents plus pratique, possibilité d'effectuer différents tri des fichiers selon des métadonnées différentes).

| Nom du fichier original | Format Original | Format Après Numérisation | Taille du fichier numérisé (Mo) | Nom du fichier numérisé | Durée | Répertoire Réseau | Référence Atelier |
|-------------------------|-----------------|---------------------------|---------------------------------|---------------------------|----------|-------------------|-------------------|
| T001_20100115 | wav | mp3 | 239 | S1-a3_- _captation.mp3 | 02:04:28 | /2010-01-15 | S1-3 |
| T002_20100115 | wav | mp3 | 60,4 | S1-a3_- _captation.mp3 | 0:31:27 | /2010-01-15 | S1-3 |
| T001_20100212 | wav | mp3 | 204,4 | S1-a4_- _captation.mp3 | 01:46:26 | /2010-02-12 | S1-4 |
| T002_20100212 | wav | mp3 | 115,4 | S1-a4_- _captation.mp3 | 01:00:05 | /2010-02-12 | S1-4 |
| T001_20100312 | wav | mp3 | 392,2 | S1-a5_- _captation.mp3 | 03:23:45 | /2010-03-12 | S1-5 |
| T001_20100521 | wav | mp3 | 0,16 | S1-a7_- _captation.mp3 | 00:00:04 | /2010-05-21 | S1-7 |

Tableau 8: Tableau de description technique d'un fonds documentaire audio

| Titre de l'atelier | Référence atelier | Nom Fichier | Time code début | Time code fin | Intitulé du Chapitre | Intervenant |
|--|-------------------|-----------------------|-----------------|---------------|---|--------------------|
| Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte | S1-a3 | S1-a3_-_captation.mp3 | 00:00:00 | 00:07:43 | Présentation | Audrey Baneyx |
| Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte | S1-a3 | S1-a3_-_captation.mp3 | 00:07:44 | 00:18:00 | Équipe DL Web , La méthodologie utilisée à l'INA , intérêts et limites | Claude Mussou |
| Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte | S1-a3 | S1-a3_-_captation.mp3 | 00:18:01 | 00:32:21 | Équipe DL Web , La méthodologie utilisée à l'INA , intérêts et limites | Thomas Drugeon |
| Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte | S1-a3 | S1-a3_-_captation.mp3 | 00:41:26 | 01:00:52 | L'espace-temps d'un corpus web , problématiques du e-Diasporas Atlas | Mathieu Jacomy |
| Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte | S1-a3 | S1-a3_-_captation.mp3 | 01:17:40 | 01:50:18 | Projet Cartec, Une nouvelle approche pour l'archivage du Web basée sur l'analyse visuelle des pages | Stéphane Gançarski |
| Analyse de l'archive web : une dimension multimodale, Identification et « traçage » d'objets pertinents dans le texte, les images et les vidéos. | S1-a4 | S1-a4_-_captation.mp3 | 00:00:00 | 00:09:15 | Présentation du contexte et des problématiques | Audrey Baneyx |
| Analyse de l'archive web : une dimension multimodale, Identification et « traçage » d'objets pertinents dans le texte, les images et les vidéos. | S1-a4 | S1-a4_-_captation.mp3 | 00:09:16 | 00:22:25 | Introduction | Équipe DL Web |

Tableau 9: Tableau de description thématique d'un fonds documentaire audio

4.3.2 Editorialisation du fond

Une fois le fond numérisé et étudié, se posait la question de l'éditorialisation. Pour l'aborder le plus pertinemment possible, il a fallu répondre aux problématiques "quoi, pour qui, et où?" (version réduite de la méthode du QQOPQC).

Du fait qu'il s'agisse d'ateliers suivant généralement un déroulé chronologique de type [prise de parole du conférencier n°1, question du public, prise de parole du conférencier n°2, question du public], la mise en place d'un chapitrage se présentait comme un choix d'éditorialisation nécessaire et relativement simple à mettre en place.

En outre, les interventions ayant des thématiques précises, clairement identifiées, et destinées à être réécoutées par des chercheurs, le chapitrage présentait une véritable valeur éditoriale.

Enfin, l'éditorialisation du fond n'ayant pas de contrainte forte quant à la date de mise en ligne, il a été possible d'allouer du temps pour cette tâche.

Le blog disposant d'un espace de stockage conséquent, il a été décidé d'héberger les fichiers audio résultants de ce découpage directement sur le blog.

La constitution de listes de lecture à partir des différents chapitres d'un atelier (une liste de lecture correspondant donc à un atelier) a été rendue possible par l'utilisation d'un plugin dédié. En revanche, il n'a pas été possible de trouver un plugin permettant à l'utilisateur de commenter un phonogramme directement sur sa ligne de temps comme cela est possible sur Soundcloud.

Enfin, le fond étant destiné à être mis en ligne sur un outil de type CMS, l'utilisation de mots-clés s'imposait comme nécessaire pour:

- permettre à l'internaute de parcourir les contenus du corpus plus facilement et amener de la sérendipité lors de la navigation,
- permettre de regrouper des interventions de différentes conférences, mais ayant des thématiques communes et

proches, au sein des résultats d'une seule recherche (effectuée à l'aide du moteur de recherche intégré au blog).

En outre, Wordpress facilite la gestion des mots-clés et permet de distinguer les mots-clés associés à un média, et les mots-clés associés à la page contenant le média. Il a donc été possible de mettre en place un double taggage.

En général, les CMS recommandent entre 5 et 15 mots-clés par contenu (document, média, page, article...).

Dans notre cas, le média « Tommaso Venturini - Les sources du web pour la recherche en Sciences Humaines et Sociales » a pour mot-clé *internet;outil;accès;science politique;méthode;recherche*. L'article Wordpress contenant le média, relative à la conférence dans laquelle a eu lieu l'intervention de Tommaso Venturini a pour mot-clé *corpus; encyclopédie; gouvernance; media; méthode; opinion; outil; PEACe; projet; régulation; savoir; science*. L'utilisateur peut donc accéder au média depuis des mots-clés correspondant au média ou depuis des mots-clés correspondant à la thématique de l'ensemble de la conférence. Ce système de double taggage favorise donc la sérendipité.

Enfin, la difficulté de l'opération de taggage réside dans le fait que les mots-clés de chaque média devaient être choisis :

- en permettant de caractériser le plus singulièrement possible le média,
- en s'inscrivant dans la liste de mots-clés choisis pour les autres médias pour offrir à l'utilisateur la possibilité de naviguer entre différents médias via des mêmes mots-clés.

Pour le corpus considéré, il a fallu trouver la juste mesure entre un trop grand nombre de mots-clés (conduisant à une multiplicité de parcours de type *impasse*) et un nombre trop faible de mots-clés (conduisant à des parcours de type *sens unique*) afin de créer une navigation ouverte, thématique et connective (parcours de type *toile d'araignée*).

Cette opération nécessite donc d'être effectuée avec un suivi, idéalement dans un tableau et il n'est pas exclu que la liste des mots-clés associés à un média tagué soit repensée, remise en question lors du taggage du reste du corpus et évolue.

Le blog étant majoritairement consulté par des chercheurs, ou des professionnels, la commentarisation a été activée. En outre, le CMS utilisé (Wordpress) offre des *plugins* permettant de gérer les *spams* sans intervention humaine et de façon pertinente.

4.3.3 Valorisation

Une fois le chapitrage effectué:

- chaque document primaire a été découpé en documents secondaires (un document secondaire par triplet {atelier, conférence, conférencier},
- chaque document secondaire a été transcrit textuellement à l'aide d'un logiciel (Vocapia),
- chaque transcription a été analysée à l'aide du logiciel Tropes pour en extraire des mots clés.

De plus, certains intervenants ayant mis à disposition leur présentation au format numérique (pdf), le fonds documentaire audio pouvait être complété par ces dernières.

L'intervenant a aussi fait l'objet d'un travail d'éditorialisation et de valorisation afin qu'il soit plus qu'un simple triplet {nom de l'intervenant, fonction de l'intervenant, nom de la présentation de l'intervenant} et qu'il participe ainsi à la valorisation du corpus.

Ainsi, au sein de chaque page relative à un atelier et pour chaque intervenant, un lien html a été mis en place et conduit à une ancre⁴ vers la mention qui est faite de l'intervenant au sein d'une page intitulée "Intervenants". Cette page, qui liste l'ensemble des intervenants, et disponible en figure 40, fournit les informations suivantes:

- le nom et le prénom de l'intervenant,
- sa fonction,
- sa biographie succincte (rédigée à partir d'informations trouvées sur le site officiel de l'intervenant, sa page sur l'institution à laquelle il appartient ou son profil public sur un réseau social)

⁴ Une ancre est un outil ergonomique. C'est un lien HTML au sein d'une page HTML qui permet de se positionner précisément à un endroit voulu dans la page.

- la liste des ateliers auxquels il a participé avec des liens vers ses ateliers,
 - des liens vers son ou ses sites web, sa page linkedIn, son compte Twitter
- **Audrey Baneux** (INSERM) était, en date du 12/03/2010, docteur en informatique de l'université paris 6, spécialisée en ingénierie des connaissances et en text mining. A participé aux ateliers Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte ? (Saison 1 - atelier 3), Analyse de l'archive web : une dimension multimodale (Saison 1 - atelier 4), Quelle appropriation de l'archive par l'utilisateur ? (Saison 1 - atelier 5), Web social et intelligence collective : nouveaux territoires de la recherche (Saison 1 - atelier 7), Bilan et perspectives (Saison 1 - atelier 8), Publics et logiques de diffusion (Saison 2 - atelier 4), Digital Humanities, normes, passerelles et langages communs (Saison 2 - atelier 7) et Grands nombres, clics & boutons : sera-t-il, demain, encore possible d'archiver le web ? (Saison 3 - atelier 1)




 - **Christine BARATS** (Université Paris Descartes, CEDITEC) était, en date du 03/12/2010, maître de conférences en sciences de l'information et de la communication université paris v - rené descartes (iut - département gea), membre du ceditec (centre d'étude des discours textes écrits et communications équipe d'accueil 3119 de l'université paris-xii <http://www.univ-paris12.fr/www/labos/ceditec>) elle poursuit des recherches axées sur les technologies de l'information et l'enseignement supérieur et s'intéresse aux modalités de représentation et de transmission d'informations par les technologies de la communication. A participé à l'atelier Les observables du web (Saison 2 - atelier 2)



 - **Christophe Batier** (Université Claude Bernard Lyon 1) était, en date du 26/06/2015, directeur technique, chargé des plateformes d'e-learning au sein de l'université, a participé à l'atelier Le web en mode labs (fablabs, makers etc.) (Saison 6 - atelier 5)





Figure 40: Extrait de la page "Intervenants" du blog des Ateliers du dépôt légal du Web

Enfin, concernant la problématique des droits, chaque intervenant a été contacté afin de lui demander sa permission pour la diffusion en ligne de sa prestation. Les intervenants ont soit répondu positivement, soit n'ont pas répondu. Par défaut et du fait des sujets abordés (non soumis à des contraintes commerciales et/ou stratégiques) et des fonctions des intervenants (chercheurs), il a été décidé de diffuser les contenus pour lesquels aucune autorisation n'a été reçue et de les retirer en cas de litiges ultérieurs.

Enfin, les phonogrammes ont été mis en téléchargement et sous licence Creative Commons.

4.3.4 Résultats

Les figures 41 et 42 (ci-après) présentent les page web du blog correspondant à un atelier avant et après le travail d'éditorialisation.

Pour chaque atelier, la page correspondante contient rigoureusement la même structure, cette unité apportant ainsi de l'ergonomie dans la navigation et facilite le travail de développement du site et a fortiori les mises à jours éventuelles.

La page après éditorialisation contient un plus grand nombre de liens (mots-clés, participants) offrant donc à l'internaute la possibilité de naviguer sur le site (et par extension au sein du corpus documentaire) selon plusieurs parcours. La liste de lecture permet de naviguer au sein de la captation d'une séance et donc d'accéder plus rapidement au contenu désiré, contenu qui pourra être complété par des documents tiers (présentation au format pdf). La zone commentaire permet à l'internaute de réagir ou d'interagir et/ou de compléter la thématique de l'atelier avec d'autres sources, documents....

Cependant, il peut être intéressant de réfléchir à d'éventuelles voies d'amélioration de l'éditorialisation qui n'ont pas pu être mises en place du fait des contraintes de temps et/ou de moyens:

- ajout ou développement d'un plugin permettant à un utilisateur de commenter un phonogramme directement sur sa ligne de temps comme cela est possible sur Soundcloud,
- ajout d'une carte de navigation sémantique qui suggère à l'internaute d'autres contenus du blog à partir du parcours qu'il a déjà effectué (à la manière des recommandations d'achat que l'on trouve sur les sites marchands),
- rédiger et mettre à disposition des notices complètes pour chaque phonogramme.

The screenshot shows a web page for a workshop. At the top, there is a banner with the text 'Dicen IDF' and 'ina THEQUE'. Below the banner is a navigation menu with items: 'A propos', 'Dépôt légal du web', 'Les saisons d'Ateliers', 'Intervenants', 'Médiagraphie', and 'Archive Actus'. The main content area features the workshop title 'Séance N°3 : Web et audiovisuel : quelles frontières ? – vendredi 14 janvier, 14h30 – 17h30' and a list of sessions with their respective speakers and topics. A search bar is located on the right side of the page. Below the main content, there is a section for 'Tweets de @nadiweb' featuring a word cloud with terms like 'web', 'audiovisuel', 'dépot légal', and 'archive'. At the bottom, there is a 'Laisser un commentaire' section with a login status and a comment input field.

Séance N°3 : Web et audiovisuel : quelles frontières ? – vendredi 14 janvier, 14h30 – 17h30
Publié le janvier 21, 2011 par admin

14h30-14h40]
Réception des participants

[14h45-15h15]
Equipe DL web – Thomas Drugeon : *Archiver les contenus audiovisuels du web, contraintes et perspectives*

[15h15-15h45]
Alain Carou, responsable du service Images animées et coordinateur dépôt légal web au département de l'Audiovisuel de la Bibliothèque nationale de France : *Vidéos et sons dans l'archivage du web à la BnF*

[15h45-16h15]
Serge Gordey, producteur de documentaires : *Documentaires de télévision, webdocumentaires : expériences de producteur dans un monde hybride*

[16h15-16h30]
Pause

[16h30-17h00]
Josiane Jouet, professeur, Carism/IFP, Université Paris 2 et Coralie Le Caroff, doctorante, Carism/IFP, Université Paris 2 : *Observation en ligne – Le Jeu de la mort sur le forum de France 2 et Lepost.fr*

[17h00-17h30]
Dominique Cardon, France Télécom R&D, EHESS-CNRS : *Les médias de la critique des médias*

Recherche
Terme(s) recherché(s)
Catégorie
Toutes
Mot-clé
Tous les mots-clés
Filtrer par type
Tous
Go

Tweets de @nadiweb

Intégrer Voir sur Twitter

oeuvre acces archivage
archive audience
audiovisuel copie dépôt
légal document
documentaires droit d'auteur forum
galerie hadoop histoire Impaire
média médiamétrie médias maproduce
Musées numérique open data
patrimoine projet propriété

Web Attitude

- 20 ans de web
- FRING
- France Culture Place de la Toile

Laisser un commentaire
Connecté en tant que admin. Se déconnecter ?
Commentaire

Figure 41: Page web d'un atelier avant éditorialisation et valorisation

CONCLUSION

Si l'écrit et la parole existent tout deux depuis des centaines d'années, les documents issus de l'oralité sont beaucoup plus récents. De même que les écrits, ils présentent des aspects documentaire, économique et juridique qui s'articulent et se meuvent dans des paradigmes sociétaux, de consommation et de communication sans cesse renouvelés.

S'il existe certaines méthodes éprouvées pour les éditorialiser et donc les valoriser, le choix de celles utilisées dépendra d'une part de la nature du corpus mais aussi du public cible.

L'éditorialisation d'un fond entrainera donc sa valorisation si et seulement si les attentes du public sont préalablement et justement déterminées et la nature du fond connu et maîtrisé.

De plus, meilleure est la valorisation, meilleure sera l'utilisation du phonogramme en tant que source documentaire. Cette qualité aura des répercussions sur l'ensemble du contenu, génèrera du flux (augmentation du nombre de visiteurs, d'abonnés pour les services payants, de professionnels pour les sites institutionnels publics ou à l'accès limité), de l'utilisation qui permettront à terme de justifier le poste de coût et de temps nécessaire à une bonne éditorialisation.

Dans le cas d'un portail/site de mise à disposition de fonds documentaires audio, il est donc impératif de fixer dès le début des objectifs en terme d'éditorialisation afin d'être reconnu pour la valorisation qui est faite des corpus et ainsi de maximiser le succès rencontré.

Enfin, on est en mesure de se demander si l'éditorialisation d'un phonogramme ne s'inscrirait pas dans la dynamique du Web 2.0 en étant effectuée en dehors du site qui l'héberge. En effet, lorsqu'un internaute partage sur un réseau social un lien vers un phonogramme, ce lien (et par transposition le fichier associé) va pouvoir être commenté et complété par d'autres internautes.

Le phonogramme étant ainsi éditorialisé hors du site sur lequel il est hébergé, cette éditorialisation communautaire et extérieure repose sur la pertinence des ajouts éditoriaux effectués par les internautes, ajouts qui peuvent devenir hors du contrôle des gestionnaires et/ou ayant-droits du phonogramme.

Le développement et l'amélioration des sites mettant à disposition des phonogrammes ne pourront faire l'économie du travail des éditorialisateurs. Pluridisciplinaire, ce dernier conjugue à la fois des techniques documentaires, des connaissances en marketing et en droit, un bagage technique mais aussi un travail de veille.

ANNEXES

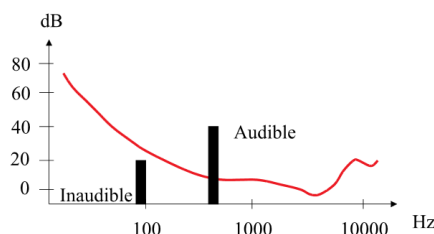
ANNEXE 1 : Principe de compression audio - l'exemple du MP3⁵

"Lorsque vous regardez le soleil et qu'un oiseau passe dans son axe, vous ne le voyez pas car la lumière provenant du soleil est trop importante. En acoustique, c'est similaire. Lorsqu'il y a des sons de fort volume sonore, vous n'entendez pas les sons faibles. Prenez l'exemple d'un orgue: lorsqu'un organiste ne joue pas, vous entendez le souffle dans les tuyaux, et quand il joue, vous ne l'entendez plus car il est masqué. Il n'est donc pas essentiel d'enregistrer tous les sons, c'est la première propriété utilisée par le format MP3 pour gagner de la place."

Gabriel Bouvigne, développeur de l'encodeur audio mp3 LAME.

La compression MP3 se base sur trois phénomènes physiques:

- le seuil de perception de l'oreille: toutes les fréquences ne sont pas entendues de la même façon par l'oreille humaine (figure ci-dessous)



- l'effet de masquage: certaines fréquences seront masquées par d'autres et donc inaudibles,
- l'effet de masquage temporelle: l'oreille ayant un temps de résolution de 1 ms, certaines fréquences en masqueront d'autres situées avant et après.

L'encodeur MP3 va donc, à l'aide d'un système de filtres:

- découper le fichier en zone fréquentielles,

⁵ Sources:

<http://docplayer.fr/13515766-Compression-d-un-signal-codec-audio-video.html>

<http://www-igm.univ-mlv.fr/~dr/XPOSE2002/compressionAugert/son.htm>

http://www.montefiore.ulg.ac.be/services/acous/STSI/file/ITAV_2010_Chap5.pdf

- appliquer à chacune de ces zones fréquentielles des algorithmes permettant de supprimer les fréquences inaudibles.

Annexe 2: ID3-V2, métadonnées disponibles

| Description | Frame | |
|-----------------------------------|-------|-------|
| | v2.3 | v2.4 |
| Audio encryption | AENC | |
| Audio seek point index | N/A | ASPI |
| Attachedpicture | APIC | |
| Comments | COMM | |
| Commercial frame | COMR | |
| Encryptionmethod registration | ENCR | |
| Equalization | EQUA | EQU2 |
| Event timing codes | ETCO | |
| General encapsulatedobject | GEOB | |
| Group identification registration | GRID | |
| Involved people list | IPLS | TIPLa |
| Linked information | LINK | |
| Music CD identifier | MCDI | |
| MPEG location lookup table | MLLT | |
| Ownership frame | OWNE | |

| Description | Frame | |
|------------------------------------|-------|-------|
| | v2.3 | v2.4 |
| Composer | TCOMc | |
| Content type | TCON | |
| Copyright message | TCOP | |
| Date | TDAT | TDRCb |
| Encoding time | N/A | TDEN |
| Playlist delay | TDLY | |
| Recording time | N/A | TDRCb |
| Release time | N/A | TDRL |
| Taggagge time | N/A | TDTG |
| Encoded by | TENC | |
| Lyricist/Textwriter | TEXTc | |
| File type | TFLT | |
| Time | TIME | TDRCb |
| Content group description | TIT1 | |
| Title/songname/content description | TIT2 | |

| Description | Frame | |
|--|-------|---------|
| | v2.3 | v2.4 |
| Conductor/performer refinement | TPE3 | |
| Interpreted, remixed, or otherwise modified by | TPE4 | |
| Part of a set | TPOS | |
| Produced notice | N/A | TPRO |
| Publisher | TPUB | |
| Track number/Position in set | TRCK | |
| Recording dates | TRDA | TDRCb |
| Internet radio station name | TRSN | |
| Internet radio station owner | TRSO | |
| Size | TSIZ | Dropped |
| Album sort order | N/A | TSOA |
| Performer sort order | N/A | TSOP |
| Title sort order | N/A | TSOT |
| International Standard Recording Code (ISRC) | TSRC | |
| Software/Hardware and settings used for encoding | TSSE | |

| Description | Frame | |
|--------------------------------|-------|------|
| Private frame | PRIV | |
| Play counter | PCNT | |
| Popularimeter | POPM | |
| Position synchronisation frame | POSS | |
| Recommended buffer size | RBUF | |
| Relative volume adjustment | RVAD | RVA2 |
| Reverb | RVRB | |
| Seek frame | N/A | SEEK |
| Signature frame | N/A | SIGN |
| Synchronized lyric/text | SYLT | |
| Synchronized tempo codes | SYTC | |
| Album/Movie/Show title | TALB | |
| Beats per minute (BPM) | TBPM | |

| Description | Frame | |
|-------------------------------------|-------|-------|
| Subtitle/Description refinement | TIT3 | |
| Initial key | TKEY | |
| Language(s) | TLAN | |
| Length | TLEN | |
| Musiciancreditslist | N/A | TMCLa |
| Media type | TMED | |
| Mood | N/A | TMOO |
| Original album/movie/show title | TOAL | |
| Original filename | TOFN | |
| Original lyricist(s)/text writer(s) | TOLYc | |
| Original artist(s)/performer(s) | TOPEc | |
| Original release year | TORY | TDOR |
| File owner/licensee | TOWN | |
| Lead performer(s)/Soloist(s) | TPE1c | |
| Band/orchestra/accompaniment | TPE2 | |

| Description | Frame | |
|--|-------|-------|
| Set subtitle | N/A | TSST |
| Year | TYER | TDRCb |
| User defined text information frame | TXXX | |
| Unique file identifier | UFID | |
| Terms of use | USER | |
| Unsynchronized lyric/text transcription | USLT | |
| Commercial information | WCOM | |
| Copyright/Legal information | WCOP | |
| Official audio file webpage | WOAF | |
| Official artist/performer webpage | WOAR | |
| Official audio source webpage | WOAS | |
| Official internet radio station homepage | WORS | |
| Payment | WPAY | |
| Publishers official webpage | WPUB | |
| User defined URL link frame | WXXX | |

[Retour au paragraphe appelant](#)

Annexe 3: APE, métadonnées disponibles

Source : http://wiki.hydrogenaud.io/index.php?title=APE_key

| Nom du champ | Signification | Format Recommandé | Nom du champ | Signification | Format Recommandé | Nom du champ | Signification | Format Recommandés |
|--------------|--|-----------------------|-------------------|----------------------------------|--|--------------|--|-----------------------|
| Title | Music Piece Title | UTF-8 string | Copyright | Copyright holder | UTF-8 string | Media | Source | Media |
| | Music Work | | | | List of UTF-8 strings | | Source Media Number/Total Media Number | Media Integer/Integer |
| Subtitle | Title when TITLE contains the work or additional sub title | UTF-8 string | Publication right | Publication right holder | UTF-8 string | | Source Time | Media Time |
| Artist | Performing artist | UTF-8 string | File | File location | List of UTF-8 strings | Index | Indexes for quick access | Index Time |
| | List of Performing Artists | List of UTF-8 strings | | | Link | | | List of Index Times |
| Album | Album name | UTF-8 string | EAN/UPC | EAN-13/UPC-A bar code identifier | List of Links | Related | Location related information of | Link |
| Debut album | Debut album name | UTF-8 string | | | 12 or 13 digitInteger | | | List of Links |
| Publisher | Record label or publisher | UTF-8 string | ISBN | ISBN number with check digit | 9 digit number with a check digit (0...9 or X) | ISRC | International Standard RecordingNumber | ASCII String |

| Nom du champ | Signification | Format Recommandé | Nom du champ | Signification | Format Recommandé | Nom du champ | Signification | Format Recommandés |
|--------------|----------------------------------|-----------------------|-----------------|--------------------|---|--------------|--|-----------------------|
| Conductor | Conductor | UTF-8 string | Catalog | Catalognumber | sometimes in the EAN/UPC, often some letters followed by some numbers | Abstract | Abstract | Link |
| Track | Track Number | Integer | LC | Label Code | LC + 4 or 5 digit Integer | Language | Used Language(s) for music/spoken words | UTF-8 string |
| | Track Number/Total Tracks Number | Integer/Integer | Year * | Release date | Date | | | List of UTF-8 strings |
| Composer | Name of the original composer | UTF-8 string | Record Date | Record date | Date | Bibliography | Bibliography/Discography | Link |
| | Name of the original arranger | List of UTF-8 strings | Record Location | Record location(s) | UTF-8 string | Introplay | Characteristic part of piece for intro playing | Index Range Time |
| Comment | User comment(s) | UTF-8 string | Genre ** | Genre(s) | List of UTF-8 strings | Dummy | Place holder | Zero data |
| | | List of UTF-8 strings | | | UTF-8 string | | | |
| | | Link | | | List of UTF-8 strings | | | |
| | | List of Links | | | | | | |

[Retour au paragraphe appelant](#)

Annexe 4 : Vorbis Comment : métadonnées disponibles

| Nom du champ | Description du champ |
|--------------|--|
| TITLE | Titre |
| VERSION | Le champs version peut être utilisé lorsque plusieurs versions d'un même titre cohabitent au sein d'un album |
| ALBUM | Album ou nom de la collection à laquelle appartient le morceau |
| TRACKNUMBER | Numéro de piste |
| ARTIST | Responsable du travail. Dans la musique contemporaine, il s'agira de l'interprète ou du groupe, dans la musique classique, ce sera le compositeur. Pour un livre, il s'agira de l'auteur |
| PERFORMER | Artiste qui interprète l'œuvre. Dans la musique contemporaine, ce sera généralement le même que l'artiste. Pour un livre-audio, il s'agira de la personne lisant l'œuvre |
| COPYRIGHT | Attribution de copyright |
| LICENSE | Information de licence, par exemple " <i>All Rights Reserved</i> " ou un lien vers la licence <i>Creative Commons</i> utilisée |
| ORGANIZATION | Nom de la société productrice de l'œuvre |
| DESCRIPTION | Description du contenu |
| GENRE | Genre de l'œuvre |
| DATE | Date à laquelle la piste a été enregistrée |
| LOCATION | Endroit auquel la piste a été enregistrée |
| CONTACT | Contact pour les créateurs ou les distributeurs de l'œuvre (adresse email, url, adresse physique) |
| ISRC | Code ISRC |

Annexe 5: Exemples de schémas PBCore

Dans les exemples ci-dessous, les balises sont en couleurs, les valeurs des balises sont en italique.

i. Exemple 1 : un entretien référencé MCUa0567

```
<pbcoreDescriptionDocumentxmlns="http://www.pbcore.org/PBCore/PBCoreNamespace.html"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace.html
https://raw.githubusercontent.com/WGBH/PBCore_2.1/master/pbcore-2.1.xsd" >
<pbcoreIdentifieur source="MCU">MCU_a0567</pbcoreIdentifieur>
<pbcoreTitretitleType="Main" titleTypeSource="pbcoreTitle/titleType Controlled Vocabulary"
titleTypeRef="http://metadataregistry.org/concept/show/id/1773.html">Death Is A Poor Man's Doctor</pbcoreTitre>
<pbcoreDescription>Interviews from Detroit musicians</pbcoreDescription>
</pbcoreDescriptionDocument>
```

Dans cet exemple, l'entretien MCU_a0567 a pour titre *Death Is A Poor Man's Doctor*, et pour description la mention *Interviews from Detroit musicians*.

ii. Autre exemple de schéma

Pour cette interview de James Stallmeyer par Jack Bighton, on voit que de nombreuses métadonnées peuvent être utilisées : la durée, le nom du programme, le type de média....

```

<pbcoreCollectionxmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace.html https://raw.githubusercontent.com/WGBH/PBCore_2.1/master/pbcore-2.1.xsd"
xmlns="http://www.pbcore.org/PBCore/PBCoreNamespace.html"
collectionTitle="WILL World War II Oral History Project on WILL from Illinois Public Media"
collectionDescription="A collection of records from WILL World War II Oral History Project on WILL"
collectionSource="Illinois Public Media, University of Illinois at Urbana-Champaign"
collectionRef="http://will.illinois.edu/wwii/pbcorecollection"
collectionDate="2014-10-16T19:18:20-05:00">
  <pbcoreDescriptionDocument>
    <pbcoreAssetType> Item </pbcoreAssetType>
    <pbcoreAssetDatedateType="broadcast"> 2008-07-01T12:02:00-05:00 </pbcoreAssetDate>
    <pbcoreIdentifier source="Illinois Public Media"> james-stallmeyer-2008-07-01 </pbcoreIdentifier>
    <pbcoreTitletitleType="Program"> World War II Central Illinois Stories </pbcoreTitle>
    <pbcoreTitletitleType="Episode"> Oral History Interview with James Stallmeyer </pbcoreTitle>
    <pbcoreSubjectssubjectType="topic" source="Illinois Public Media"> History </pbcoreSubject>
    <pbcoreSubjectssubjectType="topic" source="Illinois Public Media"> Military </pbcoreSubject>
  <pbcoreDescriptiondescriptionType="Abstract"> </pbcoreDescription>
  <pbcoreGenre source="PBCore Genre list"> History </pbcoreGenre>

```

```

<pbcoreCreator>
  <creator> Brighton, Jack </creator>
  <creatorRole> web producer </creatorRole>
</pbcoreCreator>
<pbcoreContributor>
  <contributor> Stallmeyer, James </contributor>
  <contributorRole> interviewee </contributorRole>
</pbcoreContributor>
  <pbcoreInstantiation>
    <instantiationIdentifier source="Illinois Public Media">james_stallmeyer_20084501.mp3</instantiationIdentifier>
    <instantiationDate dateType="published">2014-10-14T16:01:45-05:00</instantiationDate>
    <instantiationDigital> audio/mpeg3 </instantiationDigital>
    <instantiationLocation> http://will.illinois.edu/nfs/james_stallmeyer_20084501.mp3 </instantiationLocation>
    <instantiationMediaType> Sound </instantiationMediaType>
    <instantiationGenerations> Copy </instantiationGenerations>
    <instantiationDuration> 48:46 </instantiationDuration>
  </pbcoreInstantiation>
</pbcoreDescriptionDocument>
</pbcoreCollection>

```


Annexe 6 : Technique de chapitrage d'un fichier audio

Le chapitrage d'un fichier audio, de type conférence, échange, émission, s'articule sur deux aspects :

- un **aspect technique** : comment découper le fichier en plusieurs fichiers à raison d'un sous-fichier par chapitre,
- un **aspect éditorial** : sur quels critères découper le fichier en chapitre.

Dans ce guide, seul l'aspect technique sera abordé. On suppose que l'aspect éditorial dépendra de la nature du contenu. En outre, quelque soit ce dernier, il n'influera pas sur l'aspect technique.

Le découpage d'un flux nécessite deux logiciels :

- un logiciel de lecture audio basique comme Itunes, Winamp, Windows Media Player
- un logiciel de découpe de fichier Audio comme *Cue Splitter* (PC) ou [Bigasoft Audio Converter](#) (Mac)

N-B : Tous les logiciels indiqués ici sont gratuits.

a) Etape 1 : détection des chapitre et création du CUESHEET

Pour ce faire, il faut simplement écouter le fichier audio et noter dans un tableur (Microsoft Excel par exemple) pour chaque chapitre :

- le time code du début de chapitre au format MM:SS
- l'intervenant (qui servira à remplir la donnée « Artiste » du fichier audio correspondant au chapitre)
- le titre du chapitre (qui servira à remplir la donnée « Artiste » du fichier audio correspondant au chapitre)

Remarque :

Un logiciel comme Audacity peut faciliter le travail de détermination des différents time code. En effet, ce logiciel de traitement audio affiche la forme d'onde du fichier tout en permettant la lecture.

De fait, il est possible de naviguer plus rapidement entre les chapitres qui correspondent généralement à de creux très prononcés dans la forme d'onde (creux qui correspondent à l'absence de son).

Une fois le tableur Excel rempli, cliquer sur le bouton *Générer le CUESHEET*.

Un fichier CUESHEET

- est un fichier regroupant les informations sur les différents chapitres d'un fichier audio,
- peut être utilisé par des logiciels pour extraire ces chapitres, ou naviguer entre ces chapitres,
- utilise une syntaxe lisible par l'œil humain qui peut être vérifiée en ouvrant le fichier dans un éditeur de texte,
- permet une gestion des métadonnées des chapitres/fichiers audio (il faut donc veiller à nommer correctement celles-ci).

Exemple de fichier CUESHEET :

```
TITLE "Ma Conférence"
PERFORMER "Divers"
FILE "Mon_Fichier1_A_Découper.mp3" MP3
TRACK 01 AUDIO
    TITLE "Chapitre 1"
    PERFORMER "Auteur 1"
    INDEX 01 0:00
TRACK 02 AUDIO
    TITLE "Chapitre 2"
    PERFORMER "Auteur 2"
    INDEX 01 07:44:00
```

Ici, le fichier *Mon_Fichier1_A_Découper.mp3*, qui correspond à la conférence intitulé *Ma Conférence*, contient deux titres :

- *Chapitre 1* dont l'auteur est *Auteur 1* et qui commence à *0 minute*
- *Chapitre 2* dont l'auteur est *Auteur 2* et qui commence à *07 minute*

44

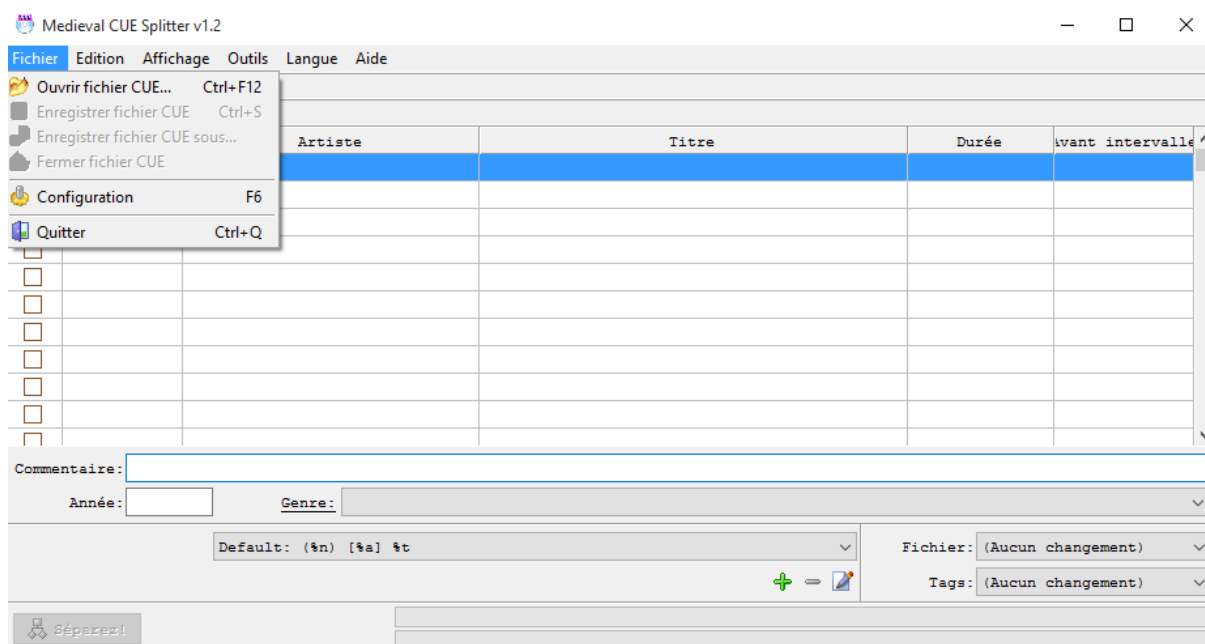
b) Etape 2 : découpage du fichier selon le chapitrage

Pour ce faire, il faut impérativement que le fichier audio et le fichier CUE correspondant soit dans le même répertoire.

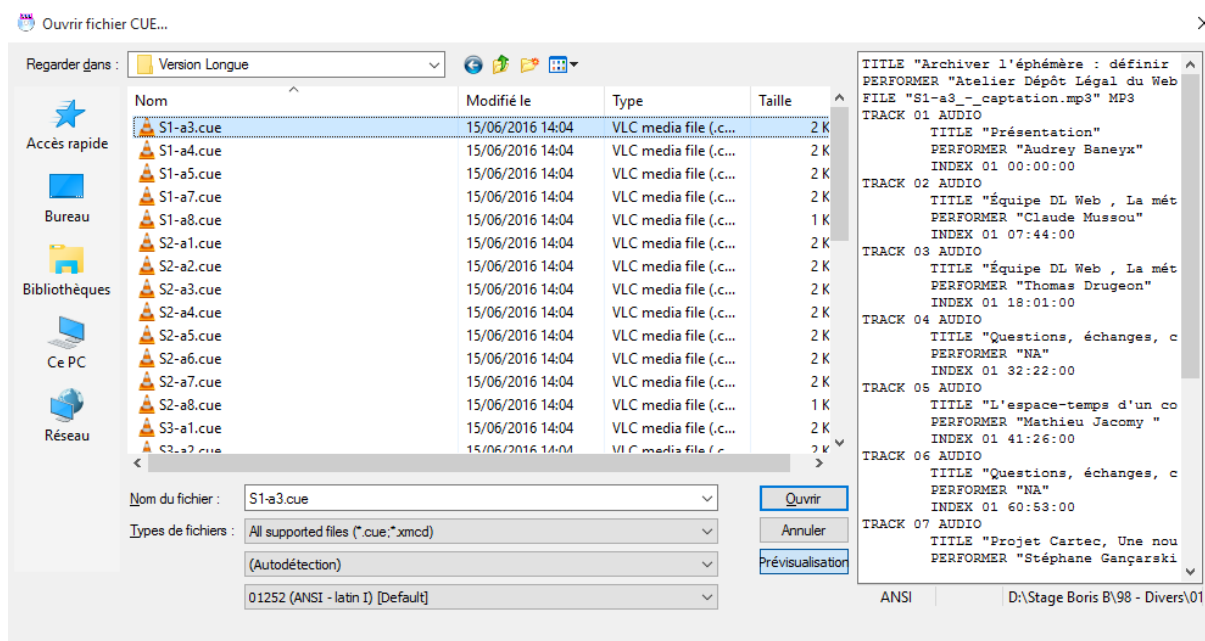
Procédure sur PC

→ Lancer le logiciel Cue Splitter.

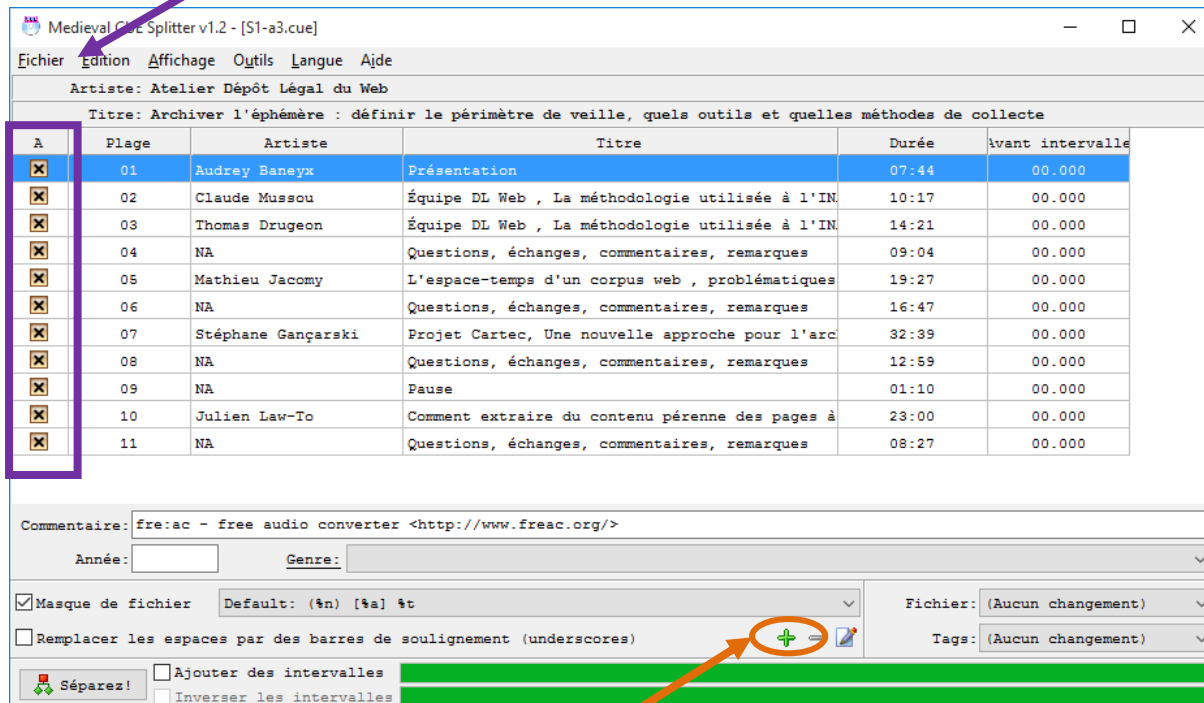
→ Cliquer sur *Fichier/Ouvrir fichier CUE*



→ Choisir le fichier CUE souhaité. Lors de cette étape, il est possible de vérifier le contenu du fichier CUE dans la droite de la fenêtre.



→ La liste des chapitres s'affiche. Il est possible dans la colonne A (cadre violet ci-après) de cocher ou décocher les titres que l'on souhaite extraire.



Medieval CUE Splitter v1.2 - [S1-a3.cue]

Echier Edition Affichage Outils Langue Aide

Artiste: Atelier Dépôt Légal du Web

Titre: Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte

| A | Plage | Artiste | Titre | Durée | Avant intervalle |
|-------------------------------------|-------|--------------------|---|-------|------------------|
| <input checked="" type="checkbox"/> | 01 | Audrey Baneux | Présentation | 07:44 | 00.000 |
| <input checked="" type="checkbox"/> | 02 | Claude Mussou | Équipe DL Web , La méthodologie utilisée à l'IN | 10:17 | 00.000 |
| <input checked="" type="checkbox"/> | 03 | Thomas Drugeon | Équipe DL Web , La méthodologie utilisée à l'IN | 14:21 | 00.000 |
| <input checked="" type="checkbox"/> | 04 | NA | Questions, échanges, commentaires, remarques | 09:04 | 00.000 |
| <input checked="" type="checkbox"/> | 05 | Mathieu Jacomy | L'espace-temps d'un corpus web , problématiques | 19:27 | 00.000 |
| <input checked="" type="checkbox"/> | 06 | NA | Questions, échanges, commentaires, remarques | 16:47 | 00.000 |
| <input checked="" type="checkbox"/> | 07 | Stéphane Gançarski | Projet Cartec, Une nouvelle approche pour l'arc | 32:39 | 00.000 |
| <input checked="" type="checkbox"/> | 08 | NA | Questions, échanges, commentaires, remarques | 12:59 | 00.000 |
| <input checked="" type="checkbox"/> | 09 | NA | Pause | 01:10 | 00.000 |
| <input checked="" type="checkbox"/> | 10 | Julien Law-To | Comment extraire du contenu pérenne des pages à | 23:00 | 00.000 |
| <input checked="" type="checkbox"/> | 11 | NA | Questions, échanges, commentaires, remarques | 08:27 | 00.000 |

Commentaire: fre:ac - free audio converter <http://www.freac.org/>

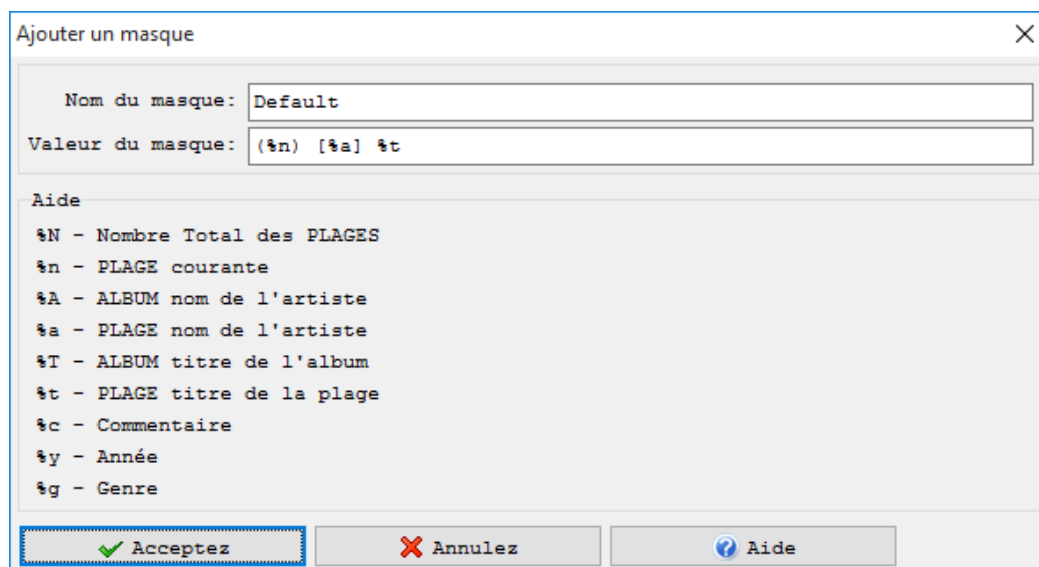
Année: Genre:

Masque de fichier Default: (%n) [%a] %t Fichier: (Aucun changement)

Remplacer les espaces par des barres de soulignement (underscores) Tags: (Aucun changement)

Séparez! Ajouter des intervalles Inverser les intervalles

→ Pour éditer le nom du fichier, cliquer sur la croix verte, la fenêtre ci-dessous permet d'éditer le masque qui sera utilisé pour nommer les chapitres résultant du découpage.



Ajouter un masque

Nom du masque: Default

Valeur du masque: (%n) [%a] %t

Aide

- %N - Nombre Total des PLAGES
- %n - PLAGE courante
- %A - ALBUM nom de l'artiste
- %a - PLAGE nom de l'artiste
- %T - ALBUM titre de l'album
- %t - PLAGE titre de la plage
- %c - Commentaire
- %y - Année
- %g - Genre

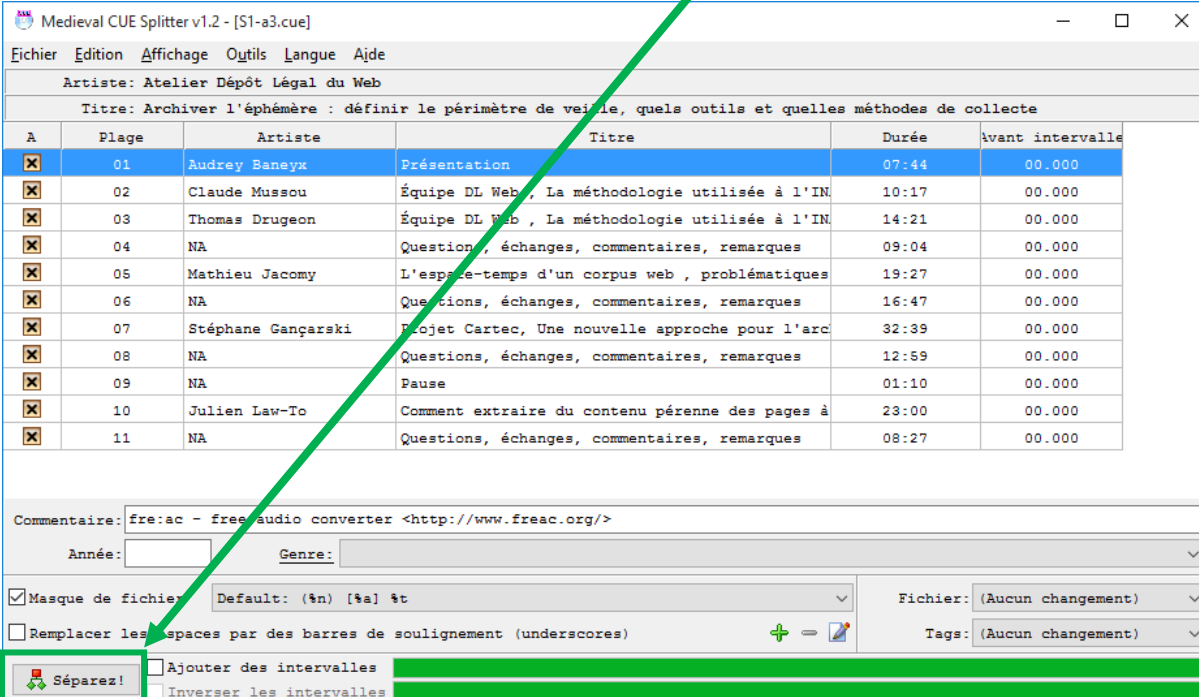
Acceptez Annulez Aide

Exemple pour le 1^{er} chapitre du fichier d'exemple :

| Artiste: Atelier Dépôt Légal du Web | | | | | |
|---|-------|---------------|--------------|-------|------------------|
| Titre: Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte | | | | | |
| A | Plage | Artiste | Titre | Durée | Avant intervalle |
| <input checked="" type="checkbox"/> | 01 | Audrey Baneux | Présentation | 07:44 | 00.000 |

| Masque | Résultats |
|-------------|--|
| %a - %t | Audrey Baneux - Présentation |
| %n_%a_%A_%t | 01_Audrey Baneux_Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte_Présentation |

→ Une fois le masque correct, cliquer sur le bouton *Séparer*



Medieval CUE Splitter v1.2 - [S1-a3.cue]

Fichier Édition Affichage Outils Langue Aide

Artiste: Atelier Dépôt Légal du Web

Titre: Archiver l'éphémère : définir le périmètre de veille, quels outils et quelles méthodes de collecte

| A | Plage | Artiste | Titre | Durée | Avant intervalle |
|-------------------------------------|-------|--------------------|---|-------|------------------|
| <input checked="" type="checkbox"/> | 01 | Audrey Baneux | Présentation | 07:44 | 00.000 |
| <input checked="" type="checkbox"/> | 02 | Claude Mussou | Équipe DL Web , La méthodologie utilisée à l'IN | 10:17 | 00.000 |
| <input checked="" type="checkbox"/> | 03 | Thomas Drugeon | Équipe DL Web , La méthodologie utilisée à l'IN | 14:21 | 00.000 |
| <input checked="" type="checkbox"/> | 04 | NA | Questions, échanges, commentaires, remarques | 09:04 | 00.000 |
| <input checked="" type="checkbox"/> | 05 | Mathieu Jacomy | L'espace-temps d'un corpus web , problématiques | 19:27 | 00.000 |
| <input checked="" type="checkbox"/> | 06 | NA | Questions, échanges, commentaires, remarques | 16:47 | 00.000 |
| <input checked="" type="checkbox"/> | 07 | Stéphane Gançarski | Projet Cartec, Une nouvelle approche pour l'arc | 32:39 | 00.000 |
| <input checked="" type="checkbox"/> | 08 | NA | Questions, échanges, commentaires, remarques | 12:59 | 00.000 |
| <input checked="" type="checkbox"/> | 09 | NA | Pause | 01:10 | 00.000 |
| <input checked="" type="checkbox"/> | 10 | Julien Law-To | Comment extraire du contenu pérenne des pages à | 23:00 | 00.000 |
| <input checked="" type="checkbox"/> | 11 | NA | Questions, échanges, commentaires, remarques | 08:27 | 00.000 |

Commentaire: fre:ac - free audio converter <http://www.freac.org/>

Année: Genre:

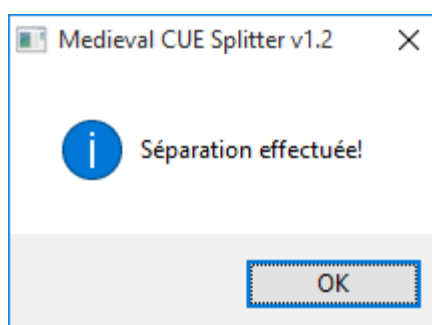
Masque de fichier: Default: (%n) [%a] %t Fichier: (Aucun changement)

Remplacer les espaces par des barres de soulignement (underscores) Tags: (Aucun changement)

Séparez! Ajouter des intervalles

Inverser les intervalles

→ Une fenêtre indique que la séparation s'est correctement déroulée :

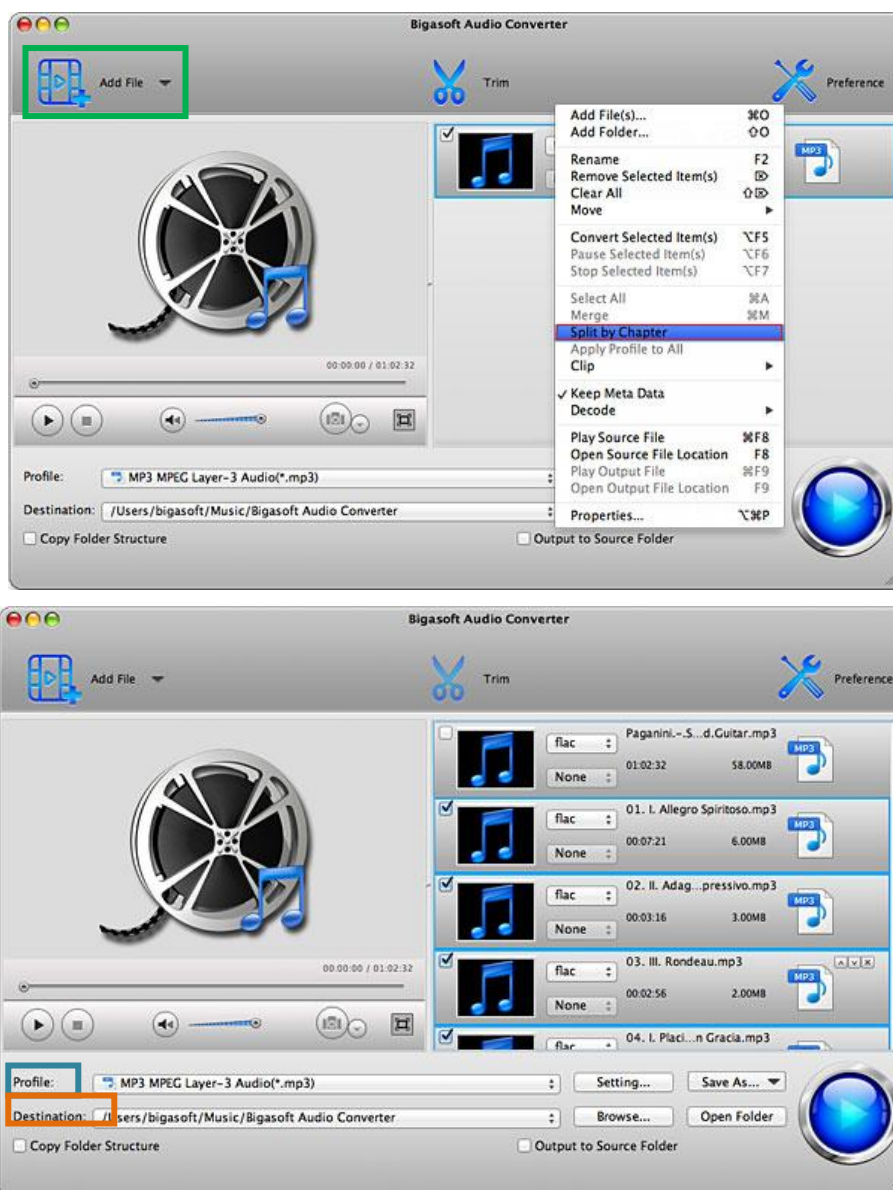


Procédure sur Mac

La procédure qui suit est une traduction de la procédure disponible en ligne⁶ et utilise le logiciel [Bigasoft Audio Converter for Mac](#).

→ Cliquer sur **"Add files"** (encadré vert ci-après) pour parcourir et ouvrir le fichier audio à découper. Il n'est pas nécessaire de lui préciser où se trouve le fichier CUE correspondant mais ce dernier doit se trouver dans le même répertoire que le fichier audio.

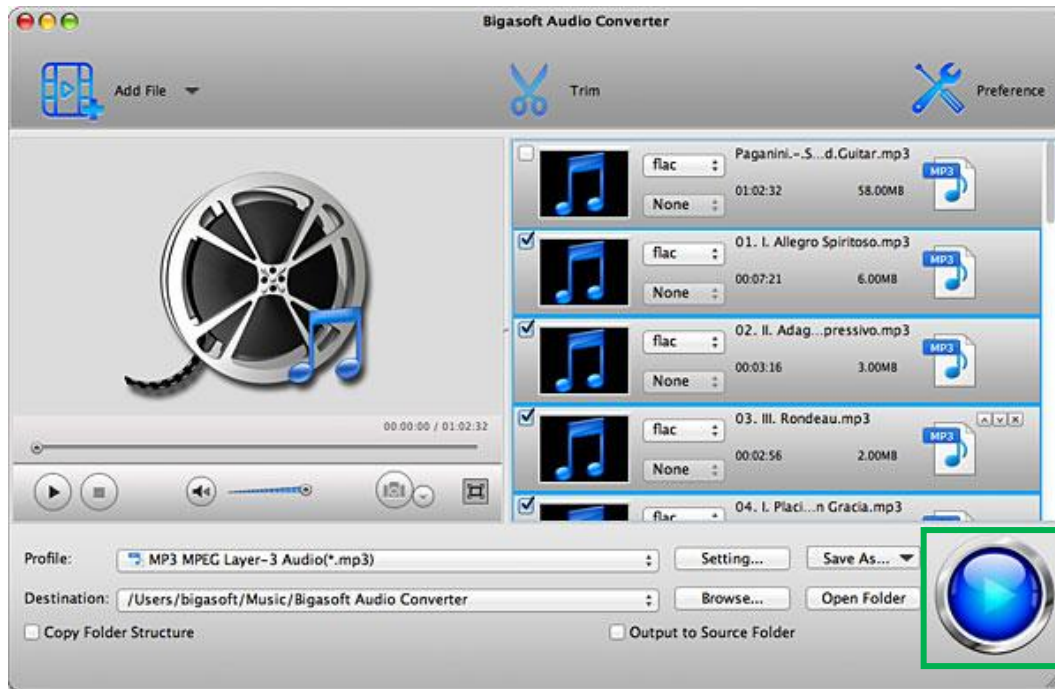
→ Dans la partie droite de la fenêtre, utiliser la souris ou CMD+A pour sélectionner tous les fichiers. Ouvrir le menu contextuel et choisir *Split by Chapter*.



⁶ <http://www.bigasoft.com/articles/cue-splitter-mac.html>

- ♦ Le champ *Profile* permet de choisir le format du fichier de sortie (sauf mention contraire, choisir mp3)
- ♦ Le champ *Destination* permet de choisir le répertoire dans lequel les fichiers issus du découpage seront enregistrés.

→ Cliquer sur le bouton bleu en bas à droite (encadré vert) pour lancer le découpage.



BIBLIOGRAPHIE

INTRODUCTION

[1] Perreault C, Mathew S. Dating the Origin of Language Using Phonemic Diversity. PLoS ONE, 2012 [en ligne]. [Consulté le 8 août 2016] <<http://dx.doi.org/10.1371/journal.pone.0035289>>

Cet ouvrage revient sur les origines du langage, son histoire et son inscription dans les cultures humaines. Bien que le sujet soit controversé, il constituera une lecture alternative aux productions de Jack Goody.

PARTIE 1 : DEFINITION D'UN FONDS DOCUMENTAIRE AUDIO

[2] ISO 01.14020 (12/1988) et AFNOR Z44-066 (03/1994), d'après citation dans [31]

Bien que ces normes n'aient pas été consultées, de nombreux textes y font référence et il paraissait difficile de faire l'économie de leur citation dans ce mémoire pour une consultation ultérieure à la discrétion du lecteur.

[3] Schneider Claire. Numérisation et traitements numériques des phonogrammes musicaux. Mémoire, DESS en Sciences de l'information et de documentation spécialisées, CNAM-INTD, 2005. 121 p

Ce mémoire donne les définitions de base pour décrire et appréhender un corpus de phonogramme d'un point de vue théorique mais aussi via un cas pratique.

[4] Dewalt Bryan. Points tournants : Si l'histoire du phonogramme et du tourne-disque m'était contée [en ligne]. Bibliothèque et Archives Canada. [Consulté le 8 août 2016]

<www.collectionscanada.gc.ca/gramophone/028011-3021-f.html>

Ce site présente de manière concise et précise l'histoire du son, d'un point de vue système et format de diffusion.

[5] Ganz Jacob. The MP3: A History Of Innovation And Betrayal [en ligne]. [Consulté le 8 août 2016] <www.npr.org/sections/therecord/2011/03/23/134622940/the-mp3-a-history-of-innovation-and-betrayal>

Excellent site sur les différentes dates clés et les détails de l'histoire de la création du format mp3.

[6] Radio France. Le Son en 3D : DÉFINITIONS [en ligne] Paris. [Consulté le 8 août 2016] <<http://nouvoson.radiofrance.fr/le-son-en-3d-definitions>>

[7] Cohen Louis, Mauborgne Joseph O, Squier George O. sur le Muzak. Electrical signaling. Brevet US1641608, Etats-Unis. 1922.
Ce brevet contient la description du système électronique mis en œuvre dans le MUZAK.

[8] Savetz Kevin, Randall Neil, Lepage Yves. MBONE. Multicasting Tomorrow's Internet. John Wiley, 1996. 234 p. ISBN 1-56884-723-8.
Cet ouvrage décrit la technologie MBONE et plus généralement la technologie de Multi-Casting.

[9] Kozamernik Franc, Mullane Michael. An introduction to Internet Radio [en ligne]. Mis à jour le 26 Octobre 2005. [Consulté le 8 août 2016] <https://tech.ebu.ch/docs/techreview/trev_304-webcasting.pdf>
Cet article donne un historique très complet de l'adoption de la technologie de streaming par les différentes radios afin de diffuser leurs programmes. Sont abordées les implications techniques, pratiques, les problématiques rencontrées.

[10] Oswald Nadia. L'histoire du streaming musical en 15 dates.. [en ligne]. Publié le 18 juin 2015. [Consulté le 8 août 2016] <<http://mybandnews.com/2015/06/histoire-du-streaming-musical-en-15-dates-spotify-deezer-youtube-apple/>>

Cet article, relativement récent, expose les dates jalons de l'histoire du streaming musical ainsi que les acteurs clés du marché.

[11] SNEP. Bilan positif du marché de la musique enregistrée au 1er semestre 2016 [en ligne]. Publié le 21 juillet 2016. [Consulté le 8 août 2016] <<http://www.snepmusique.com/actualites-du-snep/bilan-positif-du-marche-de-la-musique-enregistree-au-1er-semester-2016/>>

Cet article présente un bilan récent (à la date d'écriture de ce mémoire) rédigé par un acteur majeur de la gestion des phonogrammes, le Syndicat National de l'édition Phonographique.

[12] Oury Antoine. Le Syndicat National de l'édition dresse un panorama du livre audio [en ligne]. Publié le 9 juillet 2015. [Consulté le 8 août 2016] <<https://www.actualitte.com/article/monde-edition/le-syndicat-national-de-l-edition-dresse-un-panorama-du-livre-audio/59393>>

Cet article présente un bilan récent (à la date d'écriture de ce mémoire) du marché du livre audio.

[13] Gatineau Julie. Le livre audio : quel destin pour un objet hybride en bibliothèque. Mémoire - Diplôme de Conservateur de Bibliothèque. ENSSIB - Lyon, janvier 2015. 87 p.

En complément de la source [12], ce mémoire propose un historique, une étude complète du livre audio et des problématiques inhérentes à ce format disruptif.

[14][15] Fabien Muller. L'utilisation des phonogrammes du commerce à la télévision : la licence légale. Mémoire, Diplôme d'Etudes Spécialisées en Droit de l'audiovisuel. Faculté de Droit et de Science Politique d'Aix-Marseille, 2004. 155p.

Ce mémoire aborde de manière détaillée mais accessible au néophytes, les problématiques, les contraintes liées à la gestion des droits d'un phonogramme.

**PARTIE 2 : EDITORIALISATION D'UN FONDS DOCUMENTAIRE
AUDIO**

[16] Guyot Brigitte. Processus éditorial : faire passer un document d'un monde à l'autre *dans* Où va le travail à l'ère du numérique, dirigé par Kogan Anne-France et Metzger Jean-Luc. Editions des Mines de Paris. 2007. ISBN 978-2-91-176278-9.

Cette publication rédigée à la suite d'un travail collectif du RTP-DOC sur le document numérique contient la première référence au terme éditorialisation tel qu'il est appréhendé actuellement et en amène par là-même les concepts fondateurs.

[17] Marcello Vitali-Rosati. Éditorialisation : état de la recherche [en ligne]. Publié le 21 Février 2015. [Consulté le 8 août 2016]. <<http://blog.sens-public.org/marcellovitalirosati/editorialisation-etat-de-la-recherche/>>

Ce article issu du blog de Marcello Vitali-Rosati intitulé "Culture numérique, Pour une philosophie du numérique" contient de nombreuses réflexions sur l'éditorialisation et son articulations dans le paysage web actuel. En outre, la bibliographie de l'article fournit de très nombreuses bonnes références en la matière.

[18] Société Civile des Producteurs Phonographique. Le Code ISRC [en ligne]. [Consulté le 8 août 2016]. <www.scpp.fr/scpp/home/lascpp/guidepratiqueisrc/tabid/107/default.aspx>

Cet article francophone donne des informations de bases pour comprendre la notion d'ISRC. Le fait qu'il soit rédigé par la SCPP, société qui s'appuie fortement sur cette notion, lui confère une valeur de référence.

[19] Matteo Louis. Two Sides Of The Same Coin: ISRC And ISWC [en ligne]. Publié le 09 octobre 2015. [Consulté le 8 août 2016]. <www.hypebot.com/hypebot/2015/08/two-sides-of-the-same-coin-isrc-and-iswc-draft-1.html>

Cet article anglophone explicite clairement les différences entre l'ISRC et l'ISWC, notamment par le biais d'exemples tout à fait compréhensibles et pertinents.

[20] Valérie Dhiver. Enjeux et apports de l'éditorialisation dans les sites de VOD français. Mémoire, Titre professionnel "Chef de projet en ingénierie documentaire". 2008, CNAM-INTD. 103 p

Dans ce mémoire, Valérie Dhiver a effectué un travail de recherches particulièrement complet sur l'éditorialisation d'un corpus de vidéos : étude des méthodes mais aussi comparaison de stratégies de portails de VOD. Le travail de ce présent mémoire a été très inspiré par ses réflexions et sa manière de présenter les problématiques et leur résultats.

[21] Patrick Gros. L'indexation multimédia : Description et recherche automatiques. 23 mai 2007. Hermes Science Publications, Collection : IC2 signal et image. ISBN-13: 978-2746214927.

[22] Beauchef Hélène. "Concevoir un projet éditorial pour le web". E. Sinatra Michael, Vitali-Rosati Marcello (édité par), Pratiques de l'édition numérique, collection « Parcours Numériques », Les Presses de l'Université de Montréal, Montréal, p. 205-219, ISBN: 978-2-7606-3202-8

Cet ouvrage contient de très nombreuses informations sur le principe de l'éditorialisation numérique, ses tenants et ses aboutissements.

[23] Pauline Airy. Pratiques et enjeux de l'éditorialisation des contenus dans l'édition numérique. Mémoire, Master professionnel Communication rédactionnelle dédiée au multimédia. 2013. 28 p.

Ce mémoire contient des informations sur le processus de l'éditorialisation d'un corpus documentaire numériques.

[24] Brylawski Sam, Lerman Maya, Pike Robin, Smith Kathlin. ARSC Guide to Audio Preservation. ISBN 978-1-932326-50-5. 240 p.

Dans ce guide anglophone, les auteurs reviennent de façon très complète et très précise sur les métadonnées des fichiers audio, que ce soit sur la façon de le déterminer, de les présenter, de les utiliser et de les partager.

[25] Crozat Stéphane. Scenari. La chaîne éditoriale libre: structurer et publier textes, images & son. 2007. Paris : Eyrolles. ISBN 978-2-212-12150-6. 200 p.

Cet ouvrage présente une autre définition de l'éditorialisation, complémentaire à celles des chercheurs en documentation et/ou sciences humaines et sociales, plus orientée vers l'aspect technique du processus.

[26] Gaillard Laurent, Crozat Stéphane. Perspectives pour une (re)éditorialisation collaborative de contenus audiovisuels. 2011. In Documentaliste - sciences de l'information, dossier "Vidéo en ligne", 47, 4.

A nouveau, cet ouvrage de Stéphane Crozat revient sur les aspects pratiques de l'éditorialisation. A noter que le concept de ré-éditorialisation y est abordé.

[27] Vincent Puig and Yves-Marie L'Hour, « Vers de nouveaux outils pour les *Digital Studies* », *Anthrovision* [en ligne]. version 2.1, publié le 15 septembre 2014 [Consulté le 8 août 2016]. <<http://anthrovision.revues.org/628>>

Cet article traite des techniques d'annotations des vidéos. Bien que cela dépasse le cadre de ce mémoire, certaines idées ont été inspirantes.

[28] BON Hugo. Encadrement des pratiques et des métadonnées: les nouveaux enjeux de la musique en ligne. Mémoire, Master 2 Droit du Numérique, Administration & Entreprises. Université Paris 1 - Panthéon Sorbonne. 2012. 119 p.

Ce mémoire revient sur la description et l'importance des métadonnées, en en détaillant les aspects juridiques et leurs implications.

[29] What Is ID3v2 [en ligne]. Publié le 8 octobre 2010. [Consulté le 8 août 2016]. <<http://id3.org/ID3v2Easy>>

Ce site est le site officiel de l'id3 et contient toutes les informations nécessaires pour en faire une utilisation pertinentes et maximales.

[30] OggVorbis, format specification: comment field and header specification [en ligne]. [Consulté le 8 août 2016]. <www.xiph.org/vorbis/doc/v-comment.html>

De même que la source précédente, ce site est le site du développeur du codec audio Ogg et du VobisComment. Il contient donc toutes les informations nécessaires à son propos.

[31] Ouvrage collectif de la Commission Documentation de la FAMDT sous la coordination de Claire MARCADÉ. Patrimoine culturel immatériel - Traitement documentaire des archives sonores inédites - Guide des bonnes pratiques. 2014. Éditions FAMDT.

Ce guide donne à la fois des informations sur la FAMDT mais aussi des prescriptions sur la gestion d'un fonds documentaire de phonogrammes.

[32] Monnin Alexandre. Qu'est-ce qu'un tag ? Entre accès et libellés, l'esquisse d'une caractérisation. Présentation effectuée dans le cadre de des 20es Journées Francophones d'Ingénierie des Connaissances, « Connaissance et communautés en ligne ». Mai 2009, Hammamet, Tunisie. 13p.

Cet article définit et décrit l'objet "tag", son importance, ses contextes, ses règles de détermination, de lecture et d'interprétation.

[33] Simac-Lejeune Alain. Recherche de documents similaires sur le web par segmentations hiérarchiques et extraction de mots-clés. 13e Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances. Janvier 2013, Toulouse, France. 7p.

Cet article met en exergue l'importance des mots-clés et leur intérêt pour la navigation au sein d'un corpus.

[34] Maurel Fabrice, Ferrari Stéphane, Beudin Alexandre, Lecarpentier Jean-Marc. Tag Cloud to Tag Thunder: Towards New Oral and Nonvisual Reading Strategies. 27ème conférence francophone sur l'Interaction Homme-Machine., Octobre 2015, Toulouse, France. ACM, Actes de la 27ème conférence francophone sur l'Interaction Homme-Machine.

Cette présentation revient sur les différentes méthodes pour afficher, valoriser des tags.

PARTIE 3 : COMPARAISON DES STRATEGIES D'EDITORIALISATION DE SITES WEB

[35] André Nicolas. Etat des lieux de l'offre de musique numérique au premier semestre de l'année 2013. Philharmonie de Paris. Paris, 2013. 44p.

<http://philharmoniedeparis.fr/sites/default/files/mnum_s1_2013.pdf>

Cet ouvrage étudie fournit de bons indicateurs sur le marché, ses acteurs, ses problématiques.

[36] Spotify. Discover Weekly reaches nearly 5 billion tracks streamed since launch [en ligne]. Publié le 25 mai 2016. [Consulté le 8 août 2016].

<<https://press.spotify.com/fr/2016/05/25/discover-weekly-reaches-nearly-5-billion-tracks-streamed-since-launch>>

Dans cet article du blog de Spotify, des chiffres récents et leur analyse est disponible.

[37] Deezer. About Deezer [en ligne]. [Consulté le 8 août 2016].

<www.deezer.com/fr/company>

Page officielle de Deezer fournissant des informations sur la société (historique, chiffres officiels, implantations...).

[38] Le Saux Laurence. "Le site de France Culture sera une bibliothèque infinie d'émissions", selon sa directrice Sandrine Treiner [en ligne]. Publié le 22 janvier 2016 [Consulté le 8 août 2016].

<www.telerama.fr/radio/le-site-de-france-culture-sera-une-bibliotheque-infinie-d-emissions-sandrine-treiner-directrice,137176.php>

Cet article récent permet de bien saisir l'importance qu'est donnée par une radio aux phonogramme, à leur mise à disposition auprès des auditeurs.

[39] Enquête Médiamétrie 126 000 radios, L'audience de la Radio en France en Avril - Juin 2016 [en ligne]. Publié le 13 juillet 2016. [Consulté le 8 août 2016].

<www.mediametrie.fr/radio/communiqués/telecharger.php?f=ced556cd9f9c0c8315cfbe0744a3baf0>

Ce résultat d'enquête, conduite par l'un des leaders sur le marché, fournit les chiffres d'audience de radio à la période d'écriture de ce mémoire.

[40] Ginouvès Véronique. "Quand le renard raconte ses histoires au monde. La naissance du portail du patrimoine oral, catalogue collectif d'archives sonores et audiovisuelles", in Internationale de l'imaginaire no 25. Le Patrimoine culturel immatériel, premières expériences en France, Paris, Actes Sud, 2011. ISBN 978-2-7427-8977-1

Dans cet écrit, Véronique Ginouvès revient sur l'importance des phonogrammes dans un patrimoine culturel et sur les processus éditoriaux utilisés pour les valoriser.

[41] INA, l'entreprise en chiffres [en ligne]. Mis à jour le 31 décembre 2015. [Consulté le 8 août 2016]. <www.institut-national-audiovisuel.fr/nous-connaitre/entreprise/chiffres-cles.html>

Cette page officielle de l'Institut National de l'Audiovisuel contient toutes les chiffres clés qui permettent de se rendre du volume pharaonique de ses archives audiovisuels.

[42] Bibliothèque Sonore de Paris, notre histoire [en ligne]. [Consulté le 8 août 2016]. <<http://www.bibliotheque-sonore-paris.com/notre-histoire.html>>

Cette page officielle de la Bibliothèque Sonore de Paris décrit l'histoire de l'association.

SITOGRAFIE

<http://pbcore.org/> : site anglophone qui décrit en des termes simples le schéma PBCore, en donne des exemples d'utilisations.

<http://editorialisation.org/>: journées d'études - sous la direction de Marcello Vitali-Rosati, Michael E. Sinatra et Benoît Melançon.

GLOSSAIRE

Back-office

Appliqué aux sites web, le terme back office désigne l'interface, l'ensemble logiciel et/ou le service permettant de gérer l'environnement technique du site (serveurs, aspect graphique du site, gestion des bases de données contenant les médias...). Dans le cas d'un service, celui-ci ne communique pas avec les clients ou les utilisateurs dont les demandes et retours sont gérés par le front-office.

CMS

Content Management System ou Système de Gestion de contenu) est un site web disposant de fonctionnalités de publication et offrant en particulier une interface d'administration (back-office) permettant à un administrateur de site de créer ou organiser les différentes rubriques.

(Source : <http://www.commentcamarche.net/contents/1316-cms-content-management-system>)

Creative Commons/ Creative Commons Plus

Contrats-type ou licences pour la mise à disposition d'œuvres en ligne. Inspirés par les licences libres, les mouvements open source et open access, ces licences facilitent l'utilisation d'œuvres (textes, photos, musique, sites web...).

Ces licences s'adressent aux auteurs souhaitant :

- partager et faciliter l'utilisation de leur création par d'autres,
- autoriser gratuitement la reproduction et la diffusion (sous certaines conditions),
- accorder plus de droits aux utilisateurs en complétant le droit d'auteur qui s'applique par défaut,
- faire évoluer une oeuvre et enrichir le patrimoine commun (les biens communs ou Commons),

- économiser les coûts de transaction,
- légaliser le peer to peer de leurs œuvres.

Les *Creative Commons Plus* sont une extension des *Creatives Commons* présentant des options de partage supplémentaires

(Source : <http://creativecommons.fr>)

Cuesheet

Fichier texte ayant l'extension .cue et contenant les métadonnées d'une image CD ou DVD, notamment les timecodes des chapitres ou des pistes.

Exemple : cuesheet d'un fichier mp3 correspondant à un concert de l'artiste Faithless à Berlin en 1998.

```
TITLE "Live in Berlin, 1998"
PERFORMER "Faithless"
FILE "faithless - live in berlin.mp3" MP3
TRACK 01 AUDIO
  TITLE "Reverence"
  PERFORMER "Faithless"
  INDEX 01 00:00:00
TRACK 02 AUDIO
  TITLE "She's My Baby"
  PERFORMER "Faithless"
  INDEX 01 06:42:00
TRACK 03 AUDIO
  TITLE "Take The Long Way Home"
  PERFORMER "Faithless"
  INDEX 01 10:54:00
TRACK 04 AUDIO
  TITLE "Insomnia"
  PERFORMER "Faithless"
  INDEX 01 17:04:00
```

DAM

Digital Asset Management ou *Gestion des Actifs Numériques*. Ensemble des tâches et des logiciels qui permettent de collecter, de classier, de rechercher, de visualiser et plus généralement gérer différents contenus multimédia (photo, vidéos, enregistrements audio).

Dublin-Core

Format descriptif générique d'œuvre comprenant 15 éléments différents, qui a été créé en 1995 à Dublin (Ohio) par OCLC et le NCSA (National Center for Supercomputing Applications) dont l'objectif est de fournir un

socle commun d'éléments descriptifs pour améliorer le signalement et la recherche de ressources au-delà des diverses communautés et des nombreux formats descriptifs propres à chaque spécialité, tout en restant suffisamment structuré.

Source: http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html

FAQ

Frequently Asked Questions ou *Foire Aux Questions* : aide destinée à l'utilisateur et proposée sous forme de listes de questions types habituellement posées par les utilisateurs. Elle permet d'éviter aux services utilisateurs d'éviter de répondre aux mêmes questions.

Freemium

Mot valise composé de Free (Gratuit) et Premium. Il s'agit d'une offre commerciale proposant un accès gratuit mais réduit dont certaines fonctionnalités peuvent être débloquentées par l'utilisateur sous réserve d'un abonnement payant.

Front-office

Par opposition au back-office, le front-office est le nom du service qui gère la relation avec les clients et/ou utilisateurs.

ISO 8859-1

Norme de codage informatique de caractères prenant en charge 191 caractères, évolution de l'ASCII (American Standard Code for Information Interchange) qui n'en prend en compte que 128.

Liker

Du verbe anglais "like" qui signifie "aimer", "liker" est utilisé sur les réseaux sociaux par un utilisateur pour dire qu'il apprécie un contenu.

OAI-PMH

Open Archives Initiative - Protocol for Metadata Harvesting ou *protocole pour la collecte de métadonnées de l'Initiative pour les Archives ouvertes*

Ce protocole libre d'utilisation permet d'échanger sur Internet des métadonnées entre plusieurs institutions, afin de multiplier les accès aux documents numériques, d'en accroître la visibilité, de constituer des corpus à partir de ressources présentes en différents lieux

Source: www.bnf.fr/fr/professionnels/protocoles_echange_donnees/a.proto_oai.html

Pair-à-Pair

Aussi appelé *Peer To Peer* ou *P2P*. Protocole de partage de fichiers permettant de partager des fichiers entre plusieurs ordinateurs connectés entre eux via Internet. Chaque utilisateur peut être serveur et receveur d'un autre, formant ainsi des pairs.

Plugin

Aussi appelé *extension* ou *plugiciel*. Logiciel pouvant s'ajouter à un autre pour ajouter ou modifier des fonctionnalités natives.

Podcast

Mot valise issu de *Pod* (en référence à l'*Ipod* d'Apple) et *Cast* (pour *Broadcast* ou *diffusion*), aussi appelé *Baladodiffusion*. Emission (de radio ou de télévision par exemple) que l'on peut télécharger depuis internet vers un baladeur pour une écoute ultérieure.

QQOPQC

Méthode d'analyse systématique d'une problématique se basant sur les questions *Quoi? Qui? Où? Pourquoi? Quand? Comment?* afin de n'oublier aucune information connue.

Timecode

Référence temporelle utilisée dans les domaines du son et de l'image, pour la synchronisation et le marquage de matériaux enregistrés. Il s'exprime en heures, minutes, secondes ou en images (*frames*).

UNICODE/UTF-8

Norme de codage informatique dont le but est de supporter un maximum de caractères et de symboles afin d'éviter les problèmes de compatibilité. Il contient 128 172 caractères couvrant une centaine d'écritures.

XML

Extensible Markup Language ou *langage de balisage extensible* en français. Langage informatique de balisage interprétable par un humain ou par une machine permettant, entre autres, de structurer des données sous forme d'arbre hiérarchique.