

**Guide pratique
pour la plateforme Telemeta
des Archives sonores du CNRS-Musée de l'Homme**

<https://archives.crem-cnrs.fr/>

Documentation métier de la plateforme Telemeta, version CREM
V.6

Joséphine Simonnot, Ingénieur de recherche, LESC CNRS, 12 mai 2020

Sommaire

1	Présentation de l'interface (utilisateur et administrateur)	7
1.1	Page d'accueil.....	7
1.2	Page Bureau.....	8
1.3	Menu Archives	9
1.4	Page Géo-navigateur	10
1.5	Page Recherche avancée	10
1.6	Menu Terminologie	10
1.7	Page d'aide	11
1.8	Page Admin	11
2	Organisation de la base de données.....	11
2.1	Le modèle de données : historique.....	11
2.2	La base de données	11
2.3	Niveaux de catalogue	15
2.4	Plan de nommage des fichiers (système de cote).....	16
2.5	Description des champs documentaires	16
2.5.1	Fiche Fonds	16
2.5.2	Fiche Corpus	17
2.5.3	Fiche Collection.....	17
2.5.4	Fiche Item	20
2.6	Description des listes de vocabulaires.....	22
2.7	Fonctionnalités de modification et d'ajout d'éléments dans la base de données.....	23
2.7.1	Import de média.....	23
2.7.2	Ajout et modification des métadonnées	24
2.7.3	Convention de saisie	24
2.7.4	Structuration des collections	25
2.7.5	Modification des listes de vocabulaire.....	25
2.7.6	Modification en masse (automatique).....	26
2.8	Interopérabilité.....	26
2.8.1	Moissonnage via le <i>Dublin Core</i>	26
2.8.2	Mapping par le format CSV	29

3	Fonctionnalités de recherche et de sélection	29
3.1	Recherche simple.....	29
3.2	Recherche avancée.....	29
3.3	Affichage des résultats (recherche simple et avancée).....	30
3.3.1	Affichage en liste.....	30
3.3.2	Filtrage des résultats	31
3.4	Recherche dans les niveaux “Collections” et “Items” dans le menu “Archives”	31
3.5	Recherche par pays	31
3.6	Recherche par listes/énumérations.....	32
4	Fonctionnalités de navigation et d'affichage	33
4.1	Dans un <i>Fonds</i> ou un <i>Corpus</i>	33
4.2	Dans une <i>Collection</i>	33
4.3	Dans un <i>Item</i>	34
5	Fonctionnalités de gestion des accès aux données	35
5.1	Gestion de la connexion	35
5.1.1	Création de compte utilisateur.....	35
5.1.2	Connexion à la plateforme	36
5.2	Les types d'accès aux collections et aux items (embargo).....	36
5.3	Les profils utilisateurs.....	38
5.4	Fonctions d'accès aux médias et métadonnées.....	39
6	Fonctionnalités de gestion des média (son, vidéo, images)	40
6.1	Données produites par TimeSide, transcodages (analyse audio).....	40
6.2	Lecteur de streaming.....	42
6.3	Représentation graphique des médias	44
6.3.1	Avec un compte utilisateur :.....	44
6.3.2	Sans compte utilisateur :	45
6.3.3	Les algorithmes de l'ANR DIADEMS:.....	45
6.4	Annotation temporelle.....	46
6.5	Illustration des fiches avec une image ou un lien externe.....	47
6.6	Réutilisation d'un média	49
6.6.1	Lien « i-frame »	49
6.6.2	Livre électronique (export Epub).....	50
7	Divers	50
8	Conclusion	52

Bibliographie

Introduction

Telemeta est un système de gestion de contenus audiovisuels en ligne, conçu et développé à partir de 2007. Ce document expose son fonctionnement qui est le reflet des objectifs fixés par notre ancienne directrice Rosalia Martinez. L'urgence était de sauver les métadonnées de l'ancienne base 4D¹ menacée d'obsolescence. Il était également demandé de faciliter son accès et le travail des chercheurs. Ce patrimoine culturel audiovisuel mondial fragile, totalement inaccessible au public et voué à rester confiné pendant des décennies, devait être partagé plus largement sur le web.

Contexte²

Comme il n'existe que des plateformes de gestion documentaire propriétaires sur le marché³, l'objectif était de proposer une plateforme *open source* adaptée aux besoins “métiers” de l'archivage du son afin de gérer, annoter et partager en ligne les archives sonores et vidéo. Actuellement, le CNRS propose des solutions d'hébergement et d'exposition de données numériques mais pas de système de gestion de bases de données audiovisuelles en ligne.

L'interface de Telemeta est collaborative et orientée tout public, sans compétences particulières en informatique, afin d'impliquer directement les producteurs de données et de savoirs (chercheurs, déposants divers, étudiants...) dans l'enrichissement des métadonnées et d'apporter une plus-value scientifique propre à une équipe de recherche.

Le développement du logiciel Telemeta a débuté en 2007 suite à la rencontre de Guillaume Pellerin au LAM⁴, avec deux contraintes majeures :

- Migrer la base de données 4D existante contenant des milliers de fiches documentaires déjà saisies (intervention de la société Samalyse) vers un système ouvert, libre et compatible avec les standards du Web,
- Fournir des outils d'analyse et de transcodage à la volée qui permettent de valoriser les données selon les usages métiers.

En effet, l'analyse de la base de données faite par la Société Samalyse à cette époque était la suivante : « *Il apparaît que le fonds documentaire que constitue cette base n'est actuellement pas mis à profit comme il pourrait l'être. Cette base est largement sous-utilisée, les chercheurs, pour certains, ne l'utilisent pas du tout. La raison principale de cette situation est le fait que l'environnement technique et la structure de la base ont été conçus pour s'inscrire dans un environnement documentaire traditionnel, qui implique généralement de prendre contact avec un documentaliste pour effectuer une recherche. Le documentaliste possède les connaissances techniques nécessaires à sa consultation. Mais à l'heure d'Internet, ce processus de recherche est trop lourd et inadapté, en un mot : obsolète* » (Cf. O. Guiyadi, Docref-1.1-stable, sur Mycore)

Les fonctionnalités développées dans cette plateforme ont permis de rendre les Archives sonores du CNRS-Musée de l'Homme accessibles, interopérables et réutilisables depuis 2011, date de la mise en ligne de la version de production de la plateforme, bien avant la promotion actuelle des principes “FAIR⁵”.

¹ La base de données 4D ne fonctionnait que sous Mac OS 9 et aucune mise à jour n'a pu être faite.

² Cf. Article du Bulletin de l'AFAS sur le contexte historique : <https://journals.openedition.org/afas/4056>

³ Portail du type Alexandrie

⁴ Équipe Lutherie, Acoustique, Musique de l'Institut d'Alembert (CNRS, Sorbonne Université) : <http://www.lam.jussieu.fr/>

⁵ FAIR : Facile à trouver, Accessible, Interopérable et Réutilisable

Langages et formats

L'application Web s'appuie sur des briques et des dépendances 100% *open source*. Outre son modèle de données spécifique et dédié à l'ethnomusicologie, sa particularité réside dans la gestion automatique des transcodages et des analyses audios opérées sur les enregistrements au fil de l'eau. Une base de données relationnelle MySQL gère les métadonnées (ou documentation descriptive).

La totalité de la documentation informatique ainsi que le programme complet sont accessibles sur GitHub.

<https://github.com/Parisson/Telemeta>

<https://github.com/Parisson/TimeSide/>

Features :

<https://github.com/Parisson/Telemeta#overview>

Changelog :

<https://github.com/Parisson/Telemeta#news>

Documentation technique d'installation et de configuration (dépendances logicielles) :

<https://github.com/Parisson/Telemeta-doc>

Remarque : grâce à cette documentation, le CMAM⁶ en Tunisie ou l'Université de Mexico ont installé Telemeta sur leur serveur pour leur usage propre. Les archives sonores d'acoustique musicale utilisent également Telemeta, ainsi que l'Université de Poitiers qui s'appuie sur la technologie TimeSide.

L'application est basée sur le *framework*⁷ Django de type MVC (Modèle-Vue-Contrôleur) écrit en langage Python. Ce choix permet de faire évoluer le modèle de données et d'y inclure des méthodes d'analyse issues des laboratoires de recherche français et étrangers. Le **frontend** (interface web) est classiquement basé sur HTML5, CSS et JavaScript. Elle peut être exécutée sur n'importe quel serveur grâce à Docker qui permet d'empaqueter l'application et ses dépendances dans un conteneur.

Composition Docker (module d'installation multi-système) :

<https://github.com/Parisson/Telemeta/blob/master/docker-compose.yml>

Liste des dépendances **backend** :

<https://github.com/Parisson/Telemeta/blob/master/setup.py#L51>

Liste des dépendances **frontend** :

<https://github.com/Parisson/Telemeta/blob/master/app/settings.py#L349>

Environnement technique

Le projet dispose d'une machine virtuelle (VM) de production basée sur OpenStack déployée sur l'infrastructure du CNRS **Huma-Num**⁸.

Système d'exploitation (OS) : Debian Stretch (V.10)

Serveur web / load balancing : Nginx

Gestion des conteneurs : Docker

Mémoire RAM : 80 Go

Stockage de données du CREM (hors calcul pour la NYU) : 5 To

Mémoire allouée à la VM de développement début 2020 : 160 G

⁶ Telemeta du CMAM, Tunisie : <http://phonotheque.cmam.tn>

⁷ Framework : infrastructure logicielle, conçue pour modeler l'architecture des logiciels applicatifs

⁸ Très Grande Infrastructure de Recherche de l'INSHS du CNRS : www.huma-num.fr

Formats des données audio/vidéo

Les formats utilisés sont les standards les plus répandus et/ou acceptés par l'archivage à long terme (sauf le MP3) :

- Fichiers non compressés audio pour la conservation : WAV (48kHz/24bit ; 44.1kHz/16bit), FLAC
- Fichiers compressés audio : MP3 et OGG Vorbis
- Fichiers compressés Vidéo : MP4 et WebM

Le choix des formats est détaillé dans le chapitre 6.1. La compression des fichiers sources audio (wav) et le calcul de l'affichage graphique (forme d'onde spectrale par défaut) est automatique.

L'interopérabilité avec d'autres plateformes (comme Isidore⁹) est assurée par la correspondance avec la norme Dublin Core et le protocole OAI-PMH. La correspondance de la base de données documentaire avec la norme de description « Dublin Core » est faite pour les champs principaux dans la version du CREM (Cf. chapitre Interopérabilité).

Remarque : les approches “métiers” étant nombreuses et diverses, le mode d’emploi contient des répétitions de description pour que chaque rubrique soit complète.

Remerciements

L'idée de cette plateforme est née grâce à Hugues Genevois (LAM¹⁰) et sa conception est le fruit d'une collaboration entre le CREM et plusieurs partenaires : la start-up Parisson puis l'IRCAM (avec G. Pellerin et T. Fillon), la MMSH (structuration des métadonnées) et des laboratoires de recherche en traitement informatique du signal (IRIT, LIMSI, LaBRI, Université Queen Mary de Londres, Université de New York)¹¹.

Nous remercions également notre ancienne directrice, Rosalia Martinez, d'avoir soutenu ce projet malgré les défis nombreux et l'absence de financements dédiés, ainsi que les participants actifs comme Aude Da Cruz Lima et Sara Tandar et les chercheurs qui nous ont fait confiance. Il faut également souligner le soutien du TGE Adonis (Stéphane Pouyllau et Richard Walter), et de Caroline Bodolec pour le Cristal collectif du CNRS. Les contributions de Nedra Mellouli-Nauwynck¹² (Paris 8) et des stagiaires de l'IUT Montreuil ont permis d'améliorer le moteur de recherche (modules Haystack et ElasticSearch) : Killian Mary, Angy Fils-Aimé, Raouf Benmansour, Mathieu Boyé, Nicolas Neto, Thomas Balastegui. De nombreux contributeurs bénévoles ont participé à la traduction de l'interface dans plusieurs langues (Anas Ghrab pour l'Arabe, Susanne Fürniss pour l'allemand, Carlos Sandroni pour le Portugais, Zaichao Xia pour le Chinois).

Mais rien n'aurait pu se faire sans l'enthousiasme et les compétences de Guillaume Pellerin, le principal bâtisseur de ce projet qui met les nouvelles technologies au service d'un patrimoine culturel immatériel unique au monde. Nous rendons également hommage aux fondateurs de ces archives, André Schaeffner et Gilbert Rouget, dont la volonté était de transmettre ce patrimoine aux générations futures. Espérons que ce vœu soit ainsi réalisé.

⁹ Isidore : <https://isidore.science/>

¹⁰ Équipe Lutherie, Acoustique, Musique de l'Institut d'Alembert (CNRS, Sorbonne Université).

<http://www.lam.jussieu.fr/> ; Telemeta LAM : <http://telemeta.lam.jussieu.fr>

¹¹ Cf. ANR et Labex chapitre 6

¹² Suivi des développements et encadrement bénévole des stagiaires informatiques

1 Présentation de l'interface (utilisateur et administrateur)

Ce chapitre décrit la page d'accueil à la connexion de manière générale. Chaque élément est ensuite détaillé dans les autres chapitres.

1.1 Page d'accueil

La page d'accueil du site est construite avec les éléments suivants, de haut en bas, de gauche à droite.

The screenshot shows the main interface of the CREM-CNRS website. At the top, there's a header with the logo 'Centre de Recherche en Ethnomusicologie' (CREM), a search bar, and a user menu ('Bienvenue, Joséphine Simonot | Aide | Déconnexion'). Below the header, a navigation bar includes 'Bureau', 'Archives', 'Géo-navigateur', 'Recherche avancée', 'Terminology', and 'Admin'. A language selector 'Français' is also present. The main content area is divided into several sections:

- Archives sonores du CNRS - Musée de l'Homme**: A section about the audio archive, mentioning it contains recordings from 1900 to the present, including cylinders, 78 records, vinyls, magnetic tapes, cassettes, CDs, DVDs, and digital supports. It's described as one of the largest in Europe in terms of quality, quantity, and diversity. A link to the CREM website for historical presentation is provided.
- Actualités**: A news section mentioning participation in the European Sounds project (2014-2017).
- Content**: A section detailing the content of the archive, including over 42,000 unedited documents, 32,000 digitized ones, and nearly 4,000 hours of field recordings. It also lists 199 countries represented by over 1,200 ethnic groups or social contexts, covering a wide range of musical practices (instrumental and vocal) and languages/dialects.
- Géo-navigateur**: A map showing the locations of recordings around the world.
- Partenaires**: A section listing partners: CIRFS, Université Paris Nanterre, Parisson, Huma-Num, and europeana sounds.
- Sélection musicale**: Two waveform spectral displays for 'Recit héroïque (bharat) de dâna Gollâ 67-04' (Utar Pradesh, Inde du Nord, Inde, Asie méridionale, Asie) and 'Appel ..03 Guyane française, Amérique du Sud, Amérique' (Guyane française, Amérique du Sud, Amérique). Each display includes a play button and a 'Waveform spectral' button.

- Logo de l'équipe
- Moteur de recherche simple (bouton « loupe »)
- Bouton de connexion/aide : fenêtre de login, affichage du nom de l'utilisateur
- Barre horizontale comprenant 6 menus : Page **Bureau**, Menu **Archives**, Page **Géo-navigateur**, Page **Recherche avancée**, Menu **Terminologie**, Page **Admin**
- Menu **Langue** de l'interface (choix entre 7 langues : français, anglais, allemand, chinois, arabe, espagnol, portugais). Sélection automatique de la langue selon la langue du navigateur de l'utilisateur.
- Un texte de présentation avec une photo (modifiable dans GitHub¹³). Le texte est traduit en anglais, en allemand et en portugais. Sélection automatique de la langue selon la langue du navigateur de l'utilisateur.
- Un bouton “Accéder aux items en libre accès”
- Une fenêtre “Géo-navigateur” cliquable avec la géolocalisation par pays

¹³ <https://github.com/Parisson/Telemeta-pages/blob/crem/fr/home.rst>

- Une fenêtre avec les logos des partenaires
- Une **Sélection musicale** avec 2 players incrustés avec le titre de l'item, le lieu et le lien hypertexte vers la ressource (sélection aléatoire pour inviter l'utilisateur à consulter les archives)
- Bandeau avec le lien vers le wiki **Telemeta.org**, une mention pour les communautés d'origine, les mentions légales.

1.2 Page Bureau

La page **Bureau** est accessible avec un compte utilisateur. C'est un espace personnel (privé).

The screenshot shows the Telemeta Bureau interface with four main sections:

- Mes listes de lecture**: A list of user-created playlists, including "Collections Inédites", "Collections Éditées", and various MP3 export sets like "SET1. Arabic area", "SET2. Arabic area", "SET2. East-Africa", "SET2. India", and "Moussali, Published records".
- Mes recherches**: A search history table showing queries like "loncke", "Exposition Coloniale de Paris, 1931", and "sardaigne 1979".
- Mes dernières modifications**: A table of recent changes made by the user, showing items added or modified on specific dates.
- Toutes les dernières modifications**: A comprehensive log of all recent changes across the system, including user "jduring" performing various actions on different dates.

• Listes de lecture

L'utilisateur peut créer une ou plusieurs listes personnelles de lecture avec une sélection d'éléments de la base grâce au bouton “**Ajouter**” et lui donner un titre et une description. Les listes de lecture affichent des listes des collections, d'items ou de marqueurs temporels sélectionnés par l'utilisateur dans la base de données avec le bouton “**Lister**” ou en cochant dans une liste de résultats de recherche avec “**Ajouter à la liste**”. Chaque liste peut être “dépliée” ou “repliée” pour alléger l'affichage.

Chaque liste contient les boutons suivants :

- “**Éditer**” : pour modifier un titre et une description d'une liste
- “**CSV Collections**” : export des métadonnées de la fiche *collection* sous format csv (tableur)
- “**CSV Items**” : export des métadonnées de la fiche *item* sous format csv (tableur)
- “**Supprimer**” : suppression de la liste

Chaque liste affiche ces informations en colonnes :

- Le titre
- Le type (collection/item)
- Le collecteur
- La date (pour les items)

- Numérisé (pour les items)
- Suppression (corbeille) : supprime la sélection uniquement dans la liste

Il est possible de trier chaque colonne par ordre alphabétique, croissant/décroissant.

Le bureau affiche également les fonctions suivantes :

- Bouton “Ajouter” : ajouter une liste de lecture avec un titre et une description
- “Mes recherches” : affiche les 10 dernières recherches de l’utilisateur avec la date
- “Mes dernières modifications” : affiche les 25 dernières modifications de l’utilisateur avec le titre, le type (collection/item/marqueur), la date et le flux RSS.
- “Toutes les dernières modifications” : affiche les 100 dernières modifications faites par tous les utilisateurs avec le titre, le type (collection/item/marqueur), la date et le flux RSS.

1.3 Menu Archives

Le menu “Archives” permet de sélectionner un niveau de catalogue. Le catalogue est organisé en 4 niveaux : **Fonds**, **Corpus**, **Collection** et **Item**. Ces niveaux sont décrits dans le chapitre 2. L’item correspond à un document audiovisuel et à sa documentation descriptive (métadonnées) et doit faire partie d’une **Collection**. Une collection regroupe des enregistrements et peut éventuellement faire partie d’un ou plusieurs **Corpus** et un Corpus peut éventuellement faire partie d’un ou plusieurs **Fonds**.

A chaque niveau de catalogue, la page indique une liste d’éléments avec :

- L’affichage du nombre de pages de résultats, flèches de saut de page
- Le titre
- La présence du média (numérisé)
- Le collecteur
- Année d’enregistrement
- La cote
- Le bouton “Ajouter” permet d’ajouter une référence dans le niveau sélectionné
- Le bouton “Dublin Core” permet de voir la vue des métadonnées dans ce format
- Le bouton “Supprimer” permet de supprimer la fiche

Remarque :

Le bouton “Supprimer” est accessible **uniquement aux comptes Admin/Documentalistes**. Dans une fiche “Collection », ce bouton supprime la Collection et ses items. **Ne pas confondre avec les cases de sélection des items qui servent pour les listes de lecture et non à sélectionner des items à supprimer.**

Pour la navigation dans les niveaux **Collection** et **Item** :

- Un bouton de **filtrer** en haut à droite pour sélectionner tout, les inédits, les édités ou les références avec des média (numérisés) parmi les collections ou les items

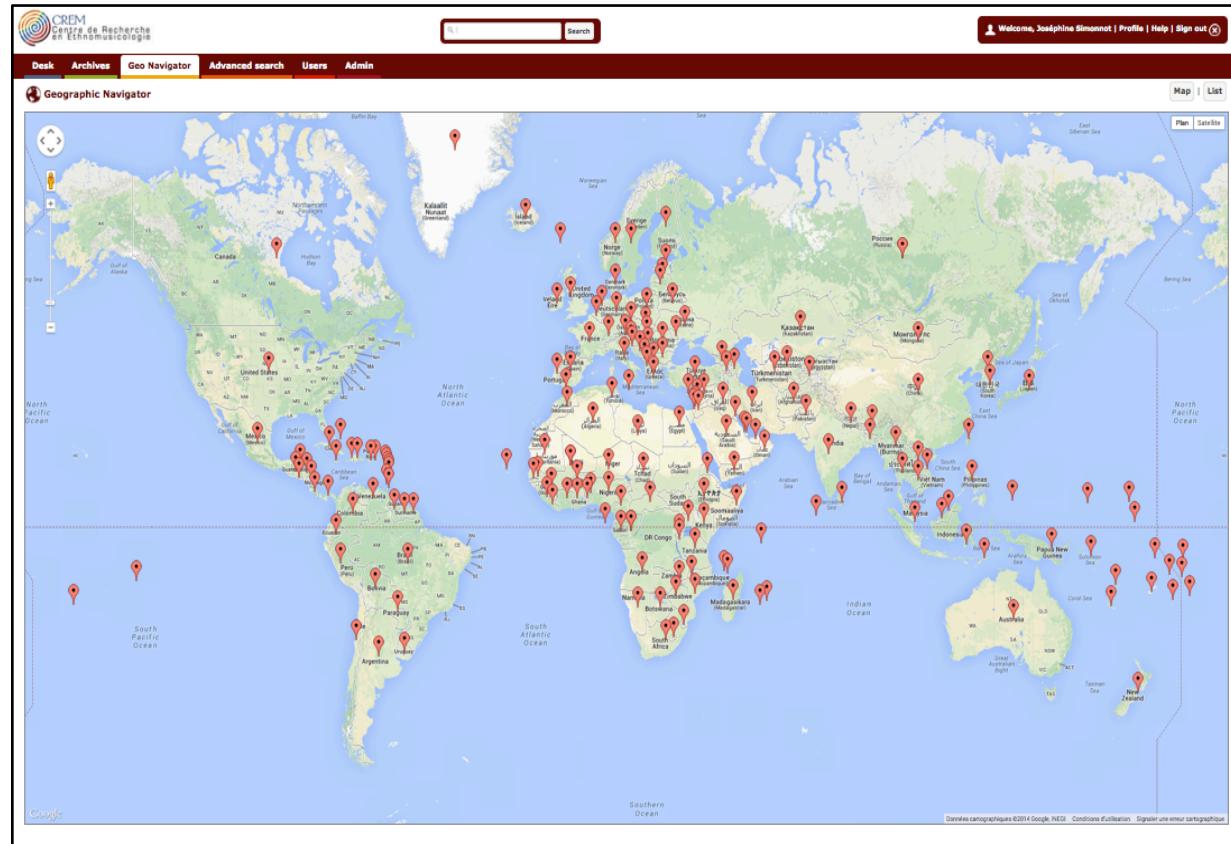
Pour le niveau **Item** uniquement :

- Un menu de sélection du nombre de **Résultats par page** : 20, 50, 100, 250, 500, 1000…

- Une colonne de sélection : toute la page ou item par item (à cocher) pour ajouter à une **Liste de lecture** personnelle déjà créée (Bureau) avec le bouton du bas “Ajouter à la liste” qui ouvre un menu déroulant pour choisir parmi ses listes.

1.4 Page Géo-navigateur

Le Géo-navigateur permet de visualiser les collections par pays, soit sur une carte du monde (par défaut), soit dans des listes de pays par continent avec le menu “Lister” en haut à droite (Cf. chapitre 3.5).



1.5 Page Recherche avancée

Pour affiner les recherches, il existe un moteur de recherche avancée à partir de 13 critères. Il a été développé en 2015-2016 selon les besoins prioritaires définis par les archivistes du CREM (Cf. Chapitre 3.2).

1.6 Menu Terminologie

Ce menu propose la liste des *instruments* (termes génériques et termes vernaculaires), accessibles par tous les utilisateurs.

Les autres *Énumérations* (listes de vocabulaire) sont accessibles par les comptes “super utilisateurs” et permet un accès aux items indexés par les termes d'une liste. (Cf. Chapitre 3.6).

1.7 Page d'aide

Cette page indique les personnes à contacter en cas de problème ou question. La page modifiable via GitHub, version de test (Cf. chapitre 7)

1.8 Page Admin

Cette page est réservée aux administrateurs de la base de données. Elle sert principalement à gérer et à créer les comptes utilisateurs et attribuer les profils et les permissions « à la carte » (Cf. chapitre 5). Elle permet également aux administrateurs d'accéder à toute la base de données. Un moteur de recherche simple permet de sélectionner l'item, la collection, le corpus ou le fonds, de le modifier et d'enregistrer la modification. On peut également rechercher un utilisateur et consulter les informations de son compte.

2 Organisation de la base de données

2.1 Le modèle de données : historique

Le modèle de données est le fruit des réflexions d'ethnomusicologues depuis les années 1930, visant à décrire de manière adaptée des archives de musiques traditionnelles du monde entier. Les données actuelles contenues dans la base de données documentaire sont conditionnées en grande partie par ce modèle. Les descripteurs (ou champs) sont détaillés dans le chapitre « Description des champs documentaires ». Le modèle contient de nombreuses listes (ou tables) de vocabulaires spécifiques à la discipline décrites plus loin. Avant la migration, les métadonnées étaient structurées selon 3 niveaux hérités du système de gestion des collections conservées au Musée de l'Homme : Collection, support physique et Item. Le modèle s'est enrichi en 2011 de champs supplémentaires décrits plus loin.

Remarque : La migration de ce modèle pour un modèle plus simple faciliterait peut-être la gestion de l'ensemble des métadonnées mais implique la perte d'informations et la baisse du niveau de granularité de la description scientifique ou de la contextualisation. Le choix de supprimer certains champs ne peut être décidé qu'en interne, selon l'utilité effective des informations contenues dans les métadonnées.

2.2 La base de données

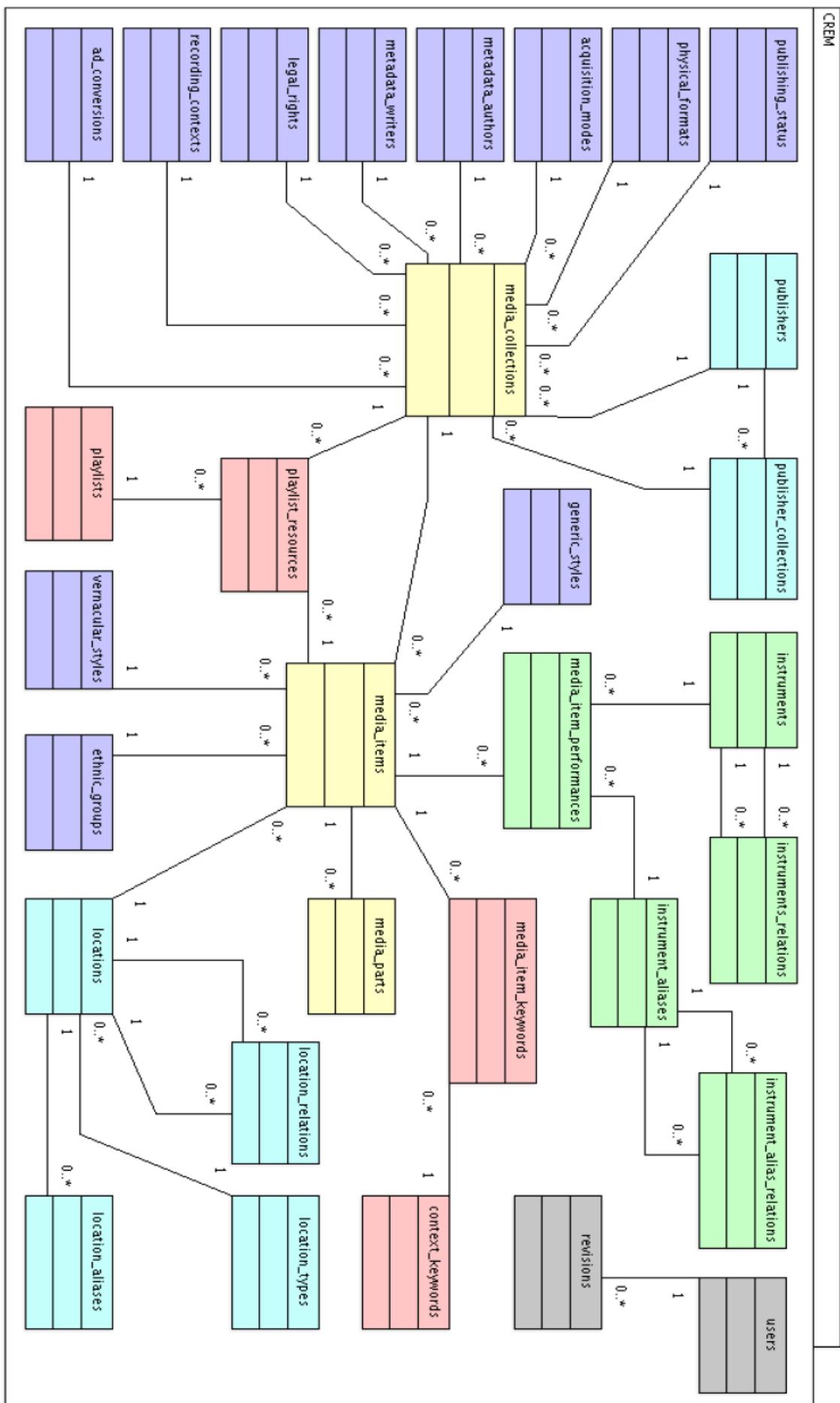
La base actuelle est héritée d'une base de données conçue par Pribislav Pitoëff¹⁴ sous 4D, un format propriétaire. Le schéma UML simplifié de l'ancienne base montre l'ensemble des relations internes recommandées pour la base de données actuelle.

Pour plus d'informations, consulter le document de référence de 2009 qui indique les recommandations pour la réorganisation de la base de données du CREM produit par Olivier Guilyardi, Société Samalyse.

Ce document ainsi que les schémas techniques sont consultables sur :

<https://mycore.core-cloud.net/index.php/s/48dedcecb574f9cf6c38f66f365b26f2>

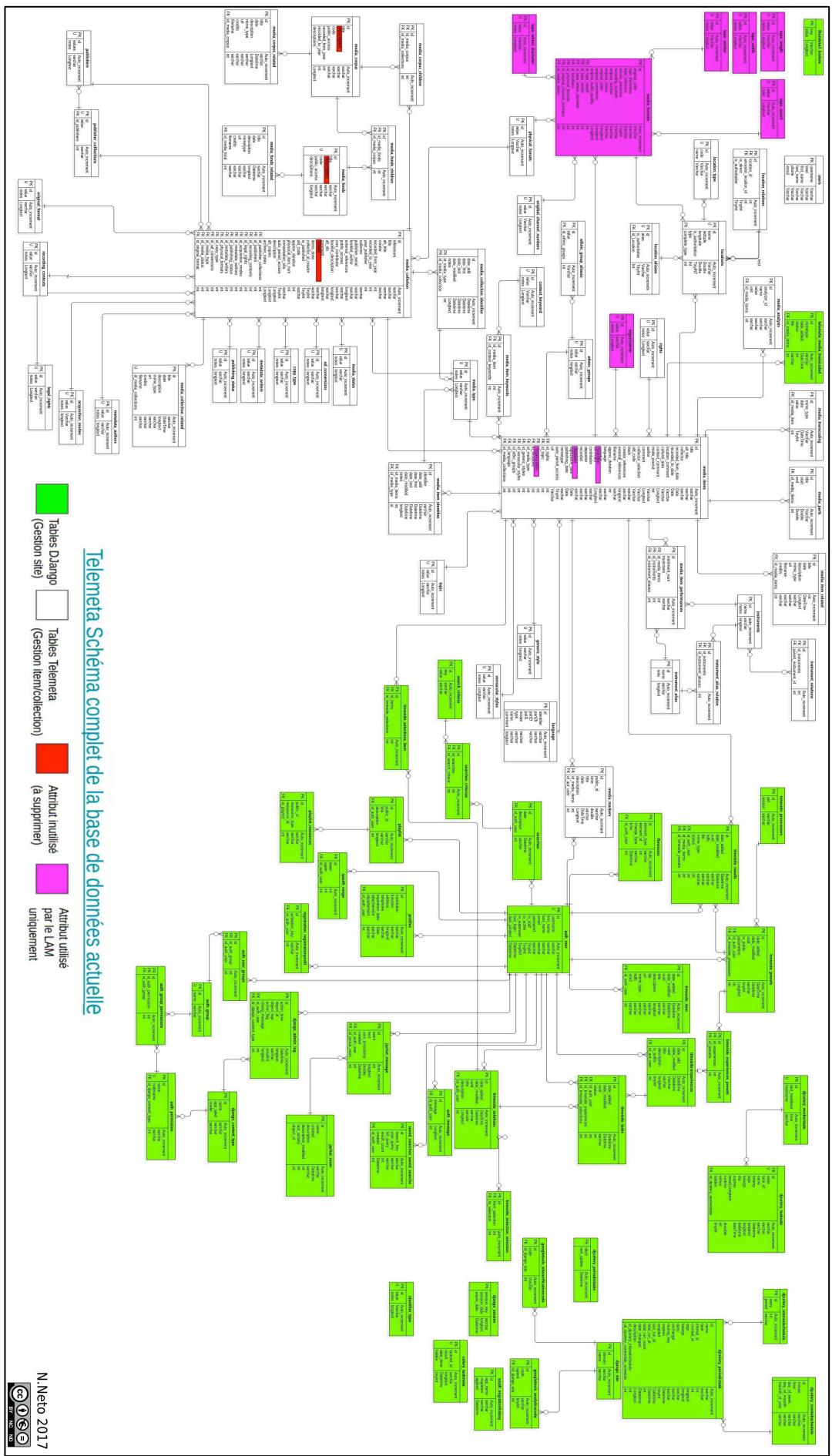
¹⁴ Ingénieur CNRS responsable des archives sonores du Laboratoire d'ethnomusicologie jusqu'en 2008



Cette migration a été l'occasion d'opérer de profonds remaniements structurels et sémantiques. Cette opération particulièrement complexe a fait l'objet d'une spécification séparée (préconisations et « moteur » d'exportation de la base de données développé entre 2009 et 2011 pour récupérer fiches **items** (phono) et **collections** saisies sous 4D).

Les principales métadonnées antérieures à 2011 sont issues d'une migration effectuée depuis l'ancienne base vers une structure SQL. Cette migration a été l'occasion d'opérer de profonds remaniements structurels et sémantiques de la base de données pour éviter les redondances d'informations. Le “moteur” d'exportation, développé entre 2009 et 2011, a fait l'objet d'une spécification séparée. Entre 2011 et 2015, d'autres fonctionnalités ont été ajoutées.

Le Schéma UML actuel (consultable sur MyCore <https://mycore.core-cloud.net/index.php/s/48dedcecb574f9cf6c38f66f365b26f2>), complété par Nicolas Neto en 2017 lors d'un stage informatique au CREM, est le suivant :



La plateforme respecte les standards de sauvegarde et de transmission de données :

- Langage des pages web
 - **HTML** : langage hypertexte pour représenter les pages web,
 - **CSS** : styles des pages web,
 - **XML** : sauvegardes sérialisées de la base de données et échange de métadonnées,
- Base de données
 - **SQL** : langage de la base de données relationnelle,
- Description des métadonnées et des ontologies
 - **Dublin Core** : format de données permettant le moissonnage de l'ensemble de la base de données par des entités externes (Isidore),
 - **OAI-PMH** : protocole de transfert des métadonnées encapsulant DublinCore (moissonnage).
 - **OWL** (Web Ontology Language) : structures sémantiques des données (théâtre, ontologies).

2.3 Niveaux de catalogue

Depuis 2011, le catalogue est structuré en 5 niveaux afin de mieux contextualiser les données dans des ensembles thématiques.

- **Fonds** : le 1er niveau hiérarchique de métadonnées comprenant des données textuelles et un ensemble cohérent de **Corpus**.
- **Corpus** : le 2e niveau hiérarchique de métadonnées comprenant des données textuelles et un ensemble cohérent de **Collections**,
- **Collection** : le 3e niveau hiérarchique de métadonnées comprenant des données textuelles et un ensemble cohérent d'**Items**, souvent regroupés par supports physiques originaux,
- **Item** : le 4e niveau hiérarchique de données comprenant des métadonnées textuelles et un **média**. Ce média peut contenir des parties (ou Marqueurs),
- **Partie** (marqueur) : le 5e niveau hiérarchique de données comprenant les mêmes métadonnées de l'Item parent, plus une plage temporelle (ou séquence).

Les 4 niveaux principaux du catalogue sont visibles dans le menu “Archives” de la page d'accueil : **Fonds**, **Corpus**, **Collection** et **Item**. L'organisation du catalogue est décrite en page d'accueil pour les utilisateurs.

La **collection** est l'ensemble de base hérité du Musée de l'Homme. Elle regroupe un ensemble cohérent de fichiers audio et/ou vidéo (items) d'une même mission, d'un même disque ou d'une même thématique, collectés par une même entité morale ou physique. Certaines collections sont elles-mêmes regroupées en corpus selon certaines thématiques ethniques ou géographiques ou éditoriales. Les corpus peuvent être regroupés en fonds, également selon certaines thématiques ou collecteurs emblématiques. Les niveaux **Fonds** et **Corpus** sont facultatifs, alors que le niveau **Collection** est nécessaire pour ajouter des données dans la plateforme.

L'**item** correspond à un document audiovisuel et à sa documentation descriptive (métadonnées) et doit faire partie d'une collection. Une collection peut éventuellement faire partie d'un ou plusieurs **Corpus** et un **Corpus** peut éventuellement faire partie d'un ou plusieurs **Fonds**. Un item doit être obligatoirement dans une collection et sa cote doit être issue de la cote de sa collection et doit être unique. L'**Item** peut être dupliqué avec une nouvelle cote dans sa collection ou dans une autre collection (en choisissant une autre collection dans le champ « Cote collection »).

2.4 Plan de nommage des fichiers (système de cote)

Lors de la migration de la base 4D en 2011, les cotes ont été modifiées selon un nouveau système de cote avec des *Under score* (« _ ») pour remplacer les caractères interdits (« . » ; « / » « : »). Le fonds d'archives étant la copropriété du CNRS et du Musée de l'Homme (MNHN), les caractères suivants ont été choisi comme en-tête à toutes les références du catalogue : **CNRSMH**.

Les anciennes cotes ont été conservées lors de la migration de 2011 dans le champ “Ancienne cote” (Old Cote). L'ancien système de cote était trop hétérogène pour faire une conversion automatique. Peu à peu, ces anciennes cotes ont été remplacées par les nouvelles car chaque modification d'une ancienne fiche impose l'attribution d'une nouvelle cote pour être enregistrée. A ce jour, il reste environ 10 000 fiches item avec une ancienne cote.

Les documents de terrain inédits et les disques publiés commercialement se distinguent au niveau de la cote :

CNRSMH_I_ : archives de terrain ou inédites, anciennement **BM**.

CNRSMH_E_ : archives publiées commercialement (éditées), anciennement **DI**.

Les cotes des items doivent respecter la cote de leur collection.

Exemple de cote **Item** : CNRSMH_I_9999_999_999_99

- La première série de chiffres : année de dépôt de la collection
- La deuxième série de chiffres : n° de collection
- La troisième série de chiffres : n° de support
- La quatrième série de chiffres : n° item

Les disques édités étant parfois regroupés en coffret ou séries, la **cote collection** comporte 3 chiffres de plus selon l'ancien système hérité du musée de l'Homme.

Exemple de cote collection d'un disque : CNRSMH_E_1963_007_001

Le moteur de recherche avancé permet de sélectionner le type de document recherché (Inédit/Édité), selon le statut de la collection qui est indiqué dans la cote (CNRSMH_I_ ou CNRSMH_E_).

2.5 Description des champs documentaires

Certains champs sont constitués de chaîne de caractères, de zone de texte extensible ou de listes fermées. Chaque niveau du catalogue possède un formulaire de saisie accessible par le bouton “**Éditer**” quand l'utilisateur a un compte utilisateur. À chaque niveau du catalogue, un bouton « **Dublin Core** » en bas à gauche de la page permet de visualiser dans une nouvelle fenêtre les correspondances de la fiche dans le format de métadonnées Dublin Core simple et étendu.

Les niveaux **Fonds** et **Corpus** disposent d'un nombre de champs limité :

2.5.1 Fiche Fonds

- Titre
- Description
- Cote
- Durée totale disponible (calculée) : alphanumérique HH:MM:SS
- Dernière modification de la fiche : Date, heure et utilisateur (renseignement automatique)
- **Média associés** (repliable)
- Liste des **Corpus** du Fonds

The screenshot displays a detailed record page for the 'Fonds Dana Rappoport' collection. At the top, there's a header with the GREM logo, a search bar, and links for 'Aide | Connexion'. Below the header, the navigation menu includes 'Accueil', 'Archives', 'Géo-Navigateur', and 'Recherche avancée'. The main content area is titled 'Fonds : Fonds Dana Rappoport'. It contains several sections:

- Média associés**: A section for associated media, featuring a thumbnail of a woman (Dana Rappoport) and a preview window.
- Biographie de D. Rappoport**: A section containing a brief biography of Dana Rappoport, mentioning her work at the Centre national de la Recherche Scientifique and her research on various Indonesian archipelago musics.
- Fiche de présentation des travaux de Dana Rappoport**: A section detailing her publications from 1991 to 2011, including her thesis and articles.
- La métamorphose en chantant: plurivocalités d'Indonésie**: A section about a radio broadcast on France Musique.
- Corpus**: A table listing various recordings collected by Dana Rappoport across different Indonesian archipelago regions, each with a unique Cote (e.g., CNRSMH_Rappoport_001 to 005).

At the bottom of the page, there are links for 'Dublin Core', 'telemeta', and 'Usage des archives réservé dans le respect du patrimoine culturel des communautés d'origine'. A copyright notice at the very bottom right reads 'Copyright © 2015 CREM-CNRS Mentions légales'.

2.5.2 Fiche Corpus

- Titre
- Description
- Cote
- Année d'enregistrement (depuis) : format : "AAAA"
- Année d'enregistrement (jusqu'à) : format : "AAAA"
- Durée totale disponible (calculée) : alphanumérique HH:MM:SS
- Dernière modification de la fiche : Date, heure et utilisateur (renseignement automatique)
- **Média associés** (repliable)
- Fonds parents
- Liste des **Collections** du Corpus

2.5.3 Fiche Collection

La fiche collection contient toutes les informations valables pour tous les items de cette collection. Certains champs compilent automatiquement les informations issues des métadonnées des items, comme les lieux ou les noms de populations. Ces champs, calculés ou affichés automatiquement par l'application, sont les suivants :

- États / nations : concaténation des pays des items
- Populations / groupes sociaux : concaténation des populations des items
- Dernière modification de la fiche
- Durée calculée
- Volume de la collection
- Nombre d'items

La fiche Collection est organisée par **blocs thématiques** pour le confort de la lecture.

Bloc généraliste

Champs	Descriptions	Remarques
(Nombre) items : Liste	Nombre d'items dans la collection. Lien dynamique vers la liste Items	Calcul automatique
Titre		Chaîne de caractères
Titre original/traduction	Titre original si titre traduit ou traduction du titre si besoin.	Chaîne de caractères
Déposant/contributeur	Nom du déposant de la collection dans les archives	Chaîne de caractères. Format de saisie "Nom, Prénom ; Nom, Prénom"
Statut du document	Renseignement automatique en fonction de la cote collection (CNRSMH_I = Inédit ; CNRSMH_E = Édité)	Calcul automatique
Description	Texte de description	Zone de texte
Contexte d'enregistrement	Contexte technique	Énumération
Période d'enregistrement	Année de début et année de fin d'enregistrement	Format de saisie "AAAA" dans les 2 champs "Année enregistrement" concaténés
Année de parution	Date de parution pour un document publié	Format de saisie "AAAA"
Type d'accès	Condition de communication : libre, restreinte, réservée, mixte	Énumération. Modifiable par Admin/Doc
Corpus	Liens vers les Corpus "parents"	Affichage automatique

Bloc *Indications géographiques et culturelles*

États / nations	Compilation des pays des items	Affichage automatique
Populations / groupes sociaux	Compilation des noms de populations, ethnies.	Affichage automatique

Bloc *Mentions légales*

Collecteur	Collecteur, compilateur (personne morale ou physique)	Chaîne de caractères. Format de saisie "Nom, Prénom ; Nom, Prénom"
Éditeur	Éditeur de la publication commerciale	Énumération. Ne s'affiche pas si nul
Collection éditeur	Collection de l'éditeur de disque	Énumération
Numéro dans la série	N° dans la collection de l'éditeur	Chaîne de caractères
Auteur de la notice éditée	Auteur du livret	Chaîne de caractères. Format de saisie "Nom, Prénom ; Nom, Prénom"
Référence de l'éditeur	N° de référence de l'éditeur de disque	Chaîne de caractères
Références bibliographiques	Bibliographie concernant la collection archivée	Zone de texte
Accès automatique après la date glissante	Case à décocher si l'accès public reste restreint après 50 ans	
Droit d'utilisation	Instruction de communicabilité et d'utilisation	Énumération

Bloc **Données d'archivage** (replié par défaut pour alléger l'affichage)

Cote	Référence unique	Cf. plan de nommage
Ancienne cote	Cote utilisée dans l'ancienne base de données (BM ou DI).	
Mode d'acquisition	Condition de dépôt, don, d'achat ou d'échange	Énumération
Déposant CNRS	Appartenance au CNRS du déposant	
Type d'exemplaire	Exemplaire principal ou secondaire (version du document)	Énumération
Rédacteur fiche	Auteur de la description	Énumération
Documentation associée	Concernant la collection archivée	Zone de texte
Réédition	Type de réédition de publication	Énumération
Statut de la collection	Original ou copie	Énumération
Autres exemplaires	Indications sur les autres exemplaires de la collection dans les archives	Zone de texte
Commentaires	Commentaires divers	Zone de texte
Saisie fiche	Nom de la personne ayant fait la saisie dans la base de données	Énumération
Dernière modification de la fiche	Nom et date de la dernière modification	Affichage automatique
Note de l'archiviste	Notes sur l'archivage	
Fiche items faites	État de la documentation	chaîne de caractères
Collecteur identique au déposant	Case à cocher	
Édité	Case à cocher si document publié commercialement	
Lieu de conservation	Lieu de conservation de l'archive	chaîne de caractères

Bloc **Données techniques** (replié par défaut pour alléger l'affichage)

Type de média	Audio/vidéo	Énumération
Durée estimée	Durée indiquée dans la documentation. Ne s'affiche pas si le champ "Durée calculée" est renseigné	Format de saisie "HH:MM:SS"
Durée calculée	Durée totale des items	Calcul automatique
Volume de la collection	Volume total de la collection en Mo	Calcul automatique
Nb de composant (support/pièce)	Nombre de supports ou de pièces dans la collection	Alphanumérique
Nombre d'items	Nombre d'items dans la collection.	Calcul automatique
Support original	Nature du support audiovisuel original	Énumération. Ne s'affiche pas si nul
Support archivé	Nature du support audiovisuel archivé	Énumération. Ne s'affiche pas si nul
Numérisation	Opérateur de numérisation	Énumération. Ne s'affiche pas si nul

Bloc **Média associés**

Titre	Titre du document associé	chaîne de caractères
Description	Description du document associé	Zone de texte
Url	Lien URL vers une ressource externe	chaîne de caractères
Crédits	Responsabilité juridique	chaîne de caractères
Fichier	Lien vers le document/ vignette de visualisation (JPEG)	Téléchargement

Bloc **Identifiant** (saisie dans un onglet spécial). Ce bloc a été ajouté pour accueillir un **identifiant pérenne** généré par l'institution.

Identifiant	Référence	Chaîne de caractères
Type d'identifiant	Type ARK ou Handle ou DOI	Liste 3 entrées
Date de première attribution	Date d'attribution de l'identifiant	Alphanumérique. "AAAA-MM-JJ"
Date de dernière attribution	Date d'attribution de l'identifiant	Alphanumérique. "AAAA-MM-JJ"
Notes	Remarques	Zone de texte

2.5.4 Fiche Item

Certains champs non renseignés ne s'affichent pas pour alléger la consultation. La fiche **Item** est organisée par **blocs thématiques** pour le confort de la lecture.

Bloc généraliste

Champs	Description	Remarques
Titre	Titre de la pièce	Chaîne de caractères
Titre original/traduction	Titre original si titre traduit ou traduction du titre si besoin.	Chaîne de caractères
Collecteur	Nom du collecteur de l'item	Chaîne de caractères. Format de saisie "Nom, Prénom ; Nom, Prénom"
Collection	Liens vers la Collection "mère"	Affichage automatique
Date d'enregistrement (depuis)	Date	Format de saisie "AAAA-MM-JJ"
Date d'enregistrement (jusqu'à)	Date	Format de saisie "AAAA-MM-JJ"
Type d'accès	Condition de communication : libre, restreinte, réservée, mixte	Liste. Modifiable par Admin/Doc

Bloc *Indications géographiques et culturelles*

Lieu	Lieu d'enregistrement	Énumération
Précisions lieu	Lieu d'enregistrement plus précis	Chaîne de caractères
Aire culturelle	Zone d'influence géo-culturelle	Chaîne de caractères
Langue	Langue du média	Si absent de la liste ISO
Langue (norme ISO)	Langue du média	Énumération
Populations/groupes sociaux	Populations/groupes sociaux des interprètes	Énumération
Contexte ethnographique	Contexte ethnographique de l'enregistrement	Chaîne de caractères
Mots-clés	Thématique	Énumération (onglet de saisie spécial)

Bloc *Informations sur la musique*

Modalité d'exécution	Mode d'exécution musicale	Chaîne de caractères
Style vernaculaire	Style musical vernaculaire	Énumération
Style générique	Style musical générique	Énumération
Auteur/compositeur	Auteur de la musique ou du texte	Chaîne de caractères
<i>Saisie dans un onglet spécial</i>		
Nombre	Nombre d'élément	Alphanumérique
Voix/instruments	Nom générique (voix ou instrument)	Liste
Nom vernaculaire	Terme vernaculaire de l'instrument	Liste
Interprètes	Nom du musicien	Chaîne de caractères "Nom, Prénom ; Nom, Prénom"

Bloc **Données d'archivage**

Cote	Nouvelle cote	Chaîne de caractères
Cote originale	Ancienne cote	Chaîne de caractères
N° de l'item	N° du support et/ou de la plage	Chaîne de caractères
Sélection collecteur	Indication qualitative du collecteur	Chaîne de caractères
Collecteur identique à la collection	Collecteur identique à celui de la collection	Case à cocher
Référence du collecteur	Numérotation du collecteur	chaîne de caractères
Référence éditée	Références discographiques ou bibliographiques	Zone de texte
Accès automatique après la date glissante	Case à décocher si l'accès public reste restreint après 50 ans	Case cochée par défaut
Remarques	Remarques diverses	Zone de texte
Dernière modification de la fiche	Nom et date de la dernière modification	Affichage automatique

Bloc **Données techniques** (replié par défaut)

Type de média	Audio ou vidéo	Énumération
Type MIME	Format numérique du média (norme web)	Affichage automatique
Durée estimée	Durée indicative fournie par l'archive	Chaîne de caractères
Durée calculée (affiché "durée")	Durée calculée par Timeside	Calcul automatique
Titre	Titre du document associé	Chaîne de caractères
Description	Description du document associé	Zone de texte
Url	Lien URL vers une ressource externe	Chaîne de caractères
Crédits	Responsabilité juridique	Chaîne de caractères
Fichier	Lien vers le document/ vignette de visualisation (JPEG)	Téléchargement

Bloc **Média associés** (saisie dans un onglet spécial)

Titre	Titre du document associé	Chaîne de caractères
Description	Description du document associé	Zone de texte
Url	Lien URL vers une ressource externe	Chaîne de caractères
Crédits	Responsabilité juridique	Chaîne de caractères
Fichier	Lien vers le document/ vignette de visualisation (JPEG)	Téléchargement

Bloc **Marqueurs** (saisie dans un onglet spécial)

N° du marqueur	N° dans l'ordre chronologique du média	Alphanumérique
Time code	Position dans le temps du marqueur	HH:MM:SS:FF
Titre	Titre de la séquence	Chaîne de caractère
Commentaire	Commentaire sur la séquence	Zone de texte
Auteur du marqueur	Nom du compte utilisateur	Affichage automatique

Bloc **Identifiant** (saisie dans un onglet spécial). Ce bloc a été ajouté pour accueillir un identifiant pérenne généré par l'institution.

Identifiant	Référence	Chaîne de caractères
Type d'identifiant	Type ARK ou Handle ou DOI	Liste à 3 entrées
Date de première attribution	Date d'attribution de l'identifiant	Alphanumérique. "AAAA-MM-JJ"
Date de dernière attribution	Date d'attribution de l'identifiant	Alphanumérique. "AAAA-MM-JJ"
Notes	Remarques	Zone de texte

2.6 Description des listes de vocabulaires

Les Collections et les Items contiennent des tables externes, dont la majorité sont des “**Énumérations**” simples et d’autres plus complexes avec des relations que l’on nomme “**Listes**”.

Les énumérations simples sont les suivantes :

- **Contexte d'enregistrement** : terrain, spectacle, radio, ...
- **Droit d'utilisation** : condition de communication (libre, restreinte, interdite...)
- **Éditeur** : Maison d'édition de disque
- **Mode d'acquisition** : don, dépôt, échange, achat...
- **Mot-clé** : Mots de vocabulaire de la discipline
- **Numérisation**: opérateur de la numérisation
- **Population / groupe social**: Noms des populations/ethnies
- **Rédacteur fiche** : Personne ayant rédigé la notice documentaire
- **Réédition** : Type de réédition de la publication (partielle, augmentée...)
- **Saisie fiche** : Personne ayant saisie la notice documentaire dans la base de données
- **Style générique** : Terme général décrivant le genre de l'archive (populaire, religieux...)
- **Style vernaculaire** : Terme vernaculaire décrivant l'archive
- **Support archivé** : Format du support archivé physiquement par l'équipe
- **Support original** : Format original du support archivé par l'équipe

Les listes suivantes sont incluses dans une table (Media_item_performance/Formation) qui contient 4 éléments :

- **Instruments** : noms d'instruments de musique (termes génériques) et des types de production vocale. 534 termes en 2020
- **Instruments (vernaculaire)** : noms d'instruments de musique (termes vernaculaires). 2023 termes en 2020
- **Nombre** : nombre d'instruments dans chaque catégorie ou nombre d'interprètes
- **Interprètes** : Nom, Prénom de(s) l'interprète(s)

Dans le menu **Admin**, il existe une table nommée “Instrument alias relation” indiquant les relations entre ces 2 listes. Par exemple : **Ampongavilany > Timbale**

Les listes suivantes ont été ajoutées dans la nouvelle base de données en 2011 :

- **Type d'exemplaire**: Version de l'exemplaire, principal ou secondaire
- **Statut de la collection** : Copie ou original
- **Langue ISO** : Liste norme « ISO 639-3 » (version 2010). Quand la langue n'existe pas dans la liste « Langue ISO », on peut l'indiquer dans le champ texte « Langue »
- **Type d'identifiant** : ARK ou Handle ou DOI (en prévision de l'attribution d'un identifiant pérenne). Le champ « type d'identifiant » n'est pas utilisé actuellement.
- **Type de média**: audio ou vidéo

Les métadonnées ne sont pas toujours renseignées avec ces nouvelles listes dans les collections antérieures à 2011.

Remarque : Certaines énumérations sont utilisées par le LAM¹⁵ et peuvent être masquées par les utilisateurs. Elles sont vides dans la version du CREM :

- Diamètre de bobine

¹⁵ Telemeta du LAM : <http://telemeta.lam.jussieu.fr>

- Discipline
- Droit
- Nombre de pistes
- Organisation
- Vitesse de bande

Liste alphabétique des énumérations/listes

Nom tables	Origine	Type	Nom interne	Ajout	Niveau catalogue
Collection Éditeur	Interne	Liste	publisher_collection	4D	Collection
Contexte d'enregistrement	interne	Énumération	recording_context	4D	Collection
Droit d'utilisation	interne	Énumération	legal_rights	4D	Collection
Éditeur	interne	Énumération	publisher	4D	Collection
Instruments	interne	Liste	instruments	4D	Item
Instruments (vernaculaire)	interne	Liste	Instruments Alias	4D	Item
Instruments Alias Relation	interne	Liste	Instruments Alias Relation	4D	
Instruments Relation	interne	Liste	Instruments Relation	4D	
Langues ISO	Norme ISO 639-3	Liste	langage_iso	2011	Item
Lieux (géographique)	GéoEthno	Liste	location	2011	Item
Mode d'acquisition	interne	Énumération	acquisition_mode	4D	Collection
Mot-clé	interne	Énumération	keywords	4D	Item
Numérisation	interne	Énumération	ad_conversion	4D	Collection
Population / groupe social	interne	Énumération	ethnic_group	4D	Item
Rédacteur fiche	interne	Énumération	metadata_author	4D	Collection
Réédition	interne	Énumération	publishing status	4D	Collection
Saisie fiche	interne	Énumération	metada_writer	4D	Collection
Statut de la collection	interne	Énumération	status_id	2014	Collection
Statut du document	interne	Énumération	doc_status	2014	Collection
Style générique	interne	Énumération	generic_style	4D	Item
Style vernaculaire	interne	Énumération	vernacular_style	4D	Item
Support archivé	interne	Énumération	physical_format	4D	Collection
Support original	interne + LAM	Énumération	original_format	2014	Collection
Type de média	interne	Énumération	media_type	2014	Collection
Type d'accès	interne	Énumération	public_access	2011	Collection/Item
Type d'exemplaire	interne	Énumération	copy_type	2014	Collection
Type d'identifiant	interne	Énumération	archive_id	2014	Collection
Type MIME	norme IANA	Variable	mime_type	2011	Item

2.7 Fonctionnalités de modification et d'ajout d'éléments dans la base de données

2.7.1 Import de média

L'import manuel de média se fait dans les fiches ***Item*** en téléchargement soi-même sur le serveur le fichier numérique (non compressé pour le son) via le bouton “**Parcourir**” en bas de l'onglet principal et en cliquant sur “**Enregistrer** ».

L'import d'une image (Jpeg ou PDF) se fait dans l'onglet “**Média associés**” de l'***Item, Collection, Corpus*** ou ***Fonds*** de la même manière. Le nom du fichier média doit respecter si possible la cote de l'item. Le cas échéant, le fichier téléchargé par un administrateur aura le nom de l'item (cote de l'item).

2.7.2 Ajout et modification des métadonnées

A chaque niveau du catalogue, il est possible d'ajouter un élément avec le bouton “**Ajouter**” ou de dupliquer un élément sélectionné avec le bouton “**Dupliquer**”. Lors d'une duplication, les informations présentes dans la partie “Voix/Instruments” sont conservées par défaut. Pour les modifier, il faut enregistrer et refaire “**Éditer**” pour modifier les éléments. L'ajout d'un nom de population ou d'un lieu dans un ***item*** est automatiquement ajouté dans la fiche ***Collection***.

La base de données peut donc être modifiée simplement et progressivement, sans compétence informatique particulière. Il suffit de respecter les conventions de saisies et le système de cote définis par les archivistes (voir plus loin) pour garder un système homogène d'identification des fichiers numériques. Il est important de noter que les métadonnées existantes ou fournies par le déposant sont très rarement complètes. La disponibilité simultanée du média et de sa documentation, même partielle, permet un enrichissement « au fil de l'eau », au gré des recherches de l'archiviste ou du déposant, de ses disponibilités d'écoute et des contributions externes.

Certains champs sont duplifiables en cas de besoin : ***Mots clés, Voix/Instruments, Média associés***. Pour avoir des champs supplémentaires de ce type, il faut enregistrer la fiche puis « **Éditer** » à nouveau pour avoir des nouveaux champs disponibles à la saisie.

Le menu “**Révisions**” dans la page « Admin » permet de voir l'historique des modifications et l'ID de la référence depuis l'ouverture de la plateforme (le 18 mai 2011). Il y avait eu 280867 modifications en mars 2020.

Les utilisateurs peuvent être tenus au courant de l'historique des modifications. Ces informations sont disponibles dans les flux de type RSS.

La modification des métadonnées est également possible via PhpMyAdmin avec un mot de passe :
https://archives.crem-cnrs.fr/phpmyadmin/db_structure.php?server=1&db=telemeta&token=f448e42ec974fd94167bd6c54876ce6f

Dans cet outil, les données s'affichent en liste et dans certains cas la modification en série est plus rapide ou plus pratique. Des modifications en masse sont également possibles. Cf. chapitre 2.7.6

2.7.3 Convention de saisie

- **Les cotes :**

Chaque cote devant être unique, il faut une nouvelle cote pour chaque nouvel élément qui respecte le modèle existant.

- La cote Fonds est du type : CNRSMH_Berthe
- La cote Corpus est du type : CNRSMH_Berthe_001
- La cote collection est du type : CNRSMH_I_9999_999 (inédit/terrain) ou du type CNRSMH_E_9999_999_999 (Édité/publication)

En particulier, la cote item doit être conforme à la cote de la collection :

Exemple : Les items de la collection CNRSMH_I_9999_999 seront de la forme CNRSMH_I_9999_999_999 ou CNRSMH_I_9999_999_999_999

- **Les champs Dates, Nom, N° Item :**

Il faut respecter les conventions de saisie pour les dates et les noms de personnes :

- Date : AAAA-MM-JJ
- Nom de personne : Nom, Prénom
- Noms de plusieurs personnes : Nom, Prénom ; Nom, Prénom
- Pour les disques : la face et le numéro de la plage (: Face A - Plage 1)
- Pour les bandes : le numéro de la bande et le numéro de la plage (:02-15)

Les formulaires de saisie sont organisés par **onglet**. Certains champs ont une liste de vocabulaires ou une chaîne de caractères avec l'auto-complétion pour faciliter la saisie d'une fiche ou lors d'une duplication de fiche. A chaque modification, il faut “**enregistrer**” et passer à la fiche suivante.

2.7.4 Structuration des collections

La structuration d'un fonds lors d'un dépôt se fait en concertation avec le déposant, le chercheur et les archivistes afin d'organiser et de contextualiser au mieux les collections en fonction du parcours scientifique du collecteur (ou de l'institution) et de rendre compte de la cohérence des données collectées. Les **Collections** peuvent être regroupées en **Corpus** et les **Corpus** peuvent être répartis en **Fonds**.

L'ajout d'une **Collection** dans un **Corpus** ou un **Corpus** dans un **Fonds** est une opération simple et accessible sans compétences particulière (avec compte Admin/documentaliste).

Le formulaire d'édition (bouton **Edit**) permet de sélectionner les éléments à ajouter dans un Corpus ou dans un Fonds. La liste des éléments du niveau inférieur s'affiche et il faut sélectionner un ou plusieurs éléments, soit avec l'ascenseur de la liste, soit avec le filtre/moteur de recherche. Une flèche permet de transférer sa sélection dans la liste “Choix des collections” ou “Choix des Corpus”. Cette structuration peut être modifiée à tout moment par les administrateurs et les documentalistes, ce qui permet de faire des regroupements et de contextualiser les collections et les corpus (Champs Titre, Description, dates).

2.7.5 Modification des listes de vocabulaire

La grande majorité des listes sont modifiables par les Administrateurs et les Documentalistes dans le menu « **Terminologie** » de la page d'accueil. L'ajout d'un mot de vocabulaire se fait avec le bouton « **Ajouter** ». La modification d'un mot de vocabulaire se fait en sélectionnant un terme dans une liste et en enregistrant la modification.

La fusion de 2 mots de vocabulaire redondants se fait en sélectionnant un terme dans “**Remplacer par**” et en enregistrant la modification. Si on coche la case “**effacer la valeur après remplacement**”, le terme remplacé par un autre disparaît. Toutes les fiches items sont ainsi indexées avec le même mot de vocabulaire sans à avoir besoin de changer chaque fiche. Une note ou un commentaire peuvent être ajoutés pour chaque terme.

2.7.6 Modification en masse (automatique)

Via l’outil PhpMyAdmin et un mot de passe, il est possible de faire des modifications en masse en prenant les **précautions d’usage** et de faire un essai sur l’instance de test systématiquement. Ce type de modification comporte des risques car les fonctionnalités de protection et de sécurité présentes dans Telemeta ne fonctionnent pas dans cet outil.

Instance de test :

<http://sandbox.crem.telemeta.org/phpmyadmin/>

Instance de production :

https://archives.crem-cnrs.fr/phpmyadmin/db_structure.php?server=1&db=telemeta&token=f448e42ec974fd94167bd6c54876ce6f

Voici un exemple de requête SQL pour modifier le nom du collecteur dans les items d’une collection :

`UPDATE media_items SET collector = 'Nouveau collecteur' WHERE code LIKE 'Cote collection%';`

ou

`UPDATE media_items SET collector = 'Beaudet, Jean-Michel' WHERE code LIKE 'CNRSMH_I_2011_005%'`

Une liste de requêtes types est partagée sur ce document :

https://docs.google.com/document/d/1saMv6_Ff3WUR3c67onI6-F_bq5VMS72XgOWhxgB3Ur8/edit

2.8 Interopérabilité

2.8.1 Moissonnage via le *Dublin Core*

Pour faciliter le moissonnage, des fonctions de correspondance sont implémentées entre les données de la base et le format cible, par exemple le Dublin Core. Ces fonctions proposent des métadonnées “pivots” et les proposent sous forme de tableaux au format CSV.

La correspondance des métadonnées sont visibles dans le format Dublin Core pour chaque niveau du catalogue, **Fonds**, **Corpus**, **Collection** et **Item** grâce au bouton « **Dublin Core** ». Le format Dublin Core est compatible avec le protocole de l’Open Archives Initiative (OAI-PMH¹⁶). Cette fonctionnalité a été développée avec le soutien financier du MNHN en 2009 (projet Anthroponet¹⁷). Elle permet le moissonnage opéré par Isidore, le moteur de recherche de l’INSHS du CNRS :

<https://isidore.science/organization/http://archives.crem-cnrs.fr/>

Les équivalences Dublin Core sont présentées dans les tableaux suivants (collection et item)

- Vue *Dublin Core* de la Collection

¹⁶ <http://www.openarchives.org/OAI/1.1/openarchivesprotocol.htm>

¹⁷ <https://www.iri.centre Pompidou.fr/projets/anthroponet/>

Element	Refinement	Valeur	Remarque
identifier		media_collections.code	Nouvelle cote
type		"Collection"	
title		media_collections.title	Titre
title		media_collections.alt_title	Traduction du titre
creator		media_collections.creator	Auteur / Cédant
contributor		media_collections.collector	Collecteur
contributor		metadata_authors.value référencé par media_collections.metadata_author_id	Rédacteur fiche
subject		« Ethnologie »	Valeur constante
subject		« Ethnomusicologie »	Valeur constante
publisher		publishers.value référencé par media_collections.publisher_id	Editeur
publisher		« CNRS – Musée de l'homme »	Valeur constante
date	created	1. media_collections.recorded_from_year au format AAAA 2. OU media_collections.recorded_from_year et media_collections.recorded_to_year au format : start=AAAA; end=AAAA;	Année ou période d'enregistrement (conforme à W3C-DTF et DCMI Period)
date	issued	media_collections.year_published au format AAAA	Année de publication (conforme à W3C-DTF)
rightsHolder		media_collections.creator	Détenteur des droits
rightsHolder		media_collections.collector si différent de media_collections.creator	Détenteur des droits
rightsHolder		publishers.value référencé par media_collections.publisher_id Si édité	Détenteur des droits
rightsHolder		media_item_performances.musicians des items liés si renseignés	Détenteur des droits (interprètes)
rights	license	legal_rights.value référencé par media_collections.legal_rights_id	Statut juridique
rights	accessRights	selon la valeur de media_collections.public_access : 1. full : public 2. metadata : restricted 3. none : private	Droits d'accès
format	extent	media_collections.approx_duration OU durée calculée des items, au format HH:MM:SS	Durée de la collection
format	medium	physical_formats.value référencé par media_collection.physical_format_id	Format physique
format*		si des items sont numérisés, types MIME disponibles (audio/x-wav, application/ogg, etc...)	Formats de fichier
relation*	hasPart	pour les items liés à cette collection: media_items.old_code OU media_items.code si présent	Anciennes ou nouvelles cotes des items liés

- Vue *Dublin Core* de l'Item

Element	Refinement	Valeur	Remarque
identifier		media_items.old_code OU media_items.code Si présent	Ancienne ou nouvelle cote
type		"Sound"	
title		media_items.title Si présent, ou media_collections.title de la collection lié, éventuellement suivi d'un suffixe (n° d'item, etc..)	Titre de la collection ou de l'item si disponible
title		media_items.alt_title	Traduction du titre
creator		media_collections.creator de la collection liée	Auteur / Cédant de la collection
contributor		media_items.collector si présent ou media_collections.collector de la collection liée	Collecteur de l'item ou de la collection
contributor		metadata_authors.value référencé par media_collections.metadata_author_id, de la collection liée	Rédacteur fiche
subject		« Ethnologie »	Valeur constante
subject		« Ethnomusicologie »	Valeur constante
subject*		context_keywords.value liés à cet item via media_item_keywords	Mots clefs du contexte ethnographique
description	abstract	media_item.context_comment	Commentaire contexte ethnographique
publisher		publishers.value référencé par media_collections.publisher_id, de la collection liée	Editeur
publisher		« CNRS – Musée de l'homme »	Valeur constante
date	created	Par priorité, selon disponibilité : 1. media_items.recorded_from_date et media_items.recorded_to_date au format: start=AAAA-MM-JJ; end=AAAA-MM-JJ; 2. media_items.recorded_from_date au format AAAA-MM-JJ 3. (collection liée) media_collections.recorded_from_year et media_collections.recorded_to_year au format : start=AAAA; end=AAAA; 4. (collection liée) media_collections.recorded_from_year au format AAAA	Année ou période d'enregistrement de l'item ou de la collection (conforme à W3C-DTF et DCMI Period)
date	issued	media_collections.year_published au format AAAA, de la collection liée	Année de publication de la collection (conforme à W3C-DTF)
coverage	spatial	Pays, région (si disponible), ville/village (si disponible) du lieu référencé par media_items.location_name (exemple: Sarawak, Malaisie)	Nom complet du lieu
rightsHolder		media_collections.creator de la collection liée	Détenteur des droits
rightsHolder		media_collections.collector de la collection liée si différent de media_collections.creator	Détenteur des droits
rightsHolder		publishers.value référencé par media_collections.publisher_id de la collection liée si éditée	Détenteur des droits
rightsHolder		media_item_performances.musicians si renseignés	Détenteur des droits (interprètes)
rights	license	legal_rights.value référencé par media_collections.legal_rights_id de la collection liée	Statut juridique
rights	accessRights	selon la valeur de media_items.public_access : 4. full : public 5. metadata : restricted 6. none : private	Droits d'accès
format	extent	media_items.approx_duration OU durée calculée de l'item, au format HH:MM:SS	Durée

Références :

- DDCMI Metadata Terms : <https://www.dublincore.org/specifications/>
- W3C-DTF : <http://www.w3.org/TR/NOTE-datetime>
- DCMI Period : <https://www.dublincore.org/specifications/dublin-core/dcmi-period/2006-04-10/>

2.8.2 Mapping par le format CSV

Les métadonnées sont aussi exportables dans un format CSV, ce qui permet de faire le « data mapping » des métadonnées. Dans le projet Europeana Sounds, les données ont été moissonnées de cette façon pour le portail Europeana : www.europeana.eu/

Les métadonnées sont donc réutilisables de cette façon : une ou plusieurs collections sont sélectionnées dans une liste de lecture (bureau, avec compte utilisateur). L'export des fiches **Collection** ou des fiches **Items** sous le format ouvert CSV se fait selon le schéma décrit sur le document partagé suivant :

<https://docs.google.com/spreadsheets/d/1Hpvk-5ZvKTr7AOTdrtIt0FkFjrw4IFIMKTYCi0neyOo/edit#gid=359255607>

3 Fonctionnalités de recherche et de sélection

La majorité des utilisateurs qui consultent les Archives du CNRS-Musée de l'Homme arrivent par une recherche directement dans Google. Une fois sur le site, le visiteur peut faire une recherche simple ou avancée grâce à un moteur de recherche interne spécifique. Toutes les recherches faites par un utilisateur ayant un compte sont mémorisées et peuvent être rechargées (visibles page “**Bureau**” dans “**Mes recherches**”).

3.1 Recherche simple

Il s'agit d'une recherche plein texte dans les champs « texte » des niveaux **Item**, **Collection**, **Corpus** et **Fonds**. Les résultats sont classés selon les 6 champs listés dans les 2 blocs de filtres indiqués plus loin (Cf. Chapitre 3.3 Affichage des résultats).

3.2 Recherche avancée

Lors de la mise en ligne en 2011, la recherche avancée proposait 7 critères. Actuellement, il est possible de croiser 13 critères.

The screenshot shows the CREM (Centre de Recherche en Ethnomusicologie) search interface. At the top, there's a navigation bar with links for Bureau, Archives, Géo-Navigateur, Recherche avancée (which is highlighted), Utilisateurs, and Admin. On the right, there's a user profile with "Bienvenue, Joséphine Simonnot" and links for Aide and Déconnexion. Below the navigation, there's a search bar with a magnifying glass icon. The main area is titled "Rechercher" and contains several search fields and dropdown menus:

- title :** (text input field)
- collecteur :** (text input field)
- lieu :** (text input field with "Indonésie" placeholder)
- population / groupe social :** (dropdown menu with "no preference" selected)
- instruments :** (text input field)
- type de média :** (dropdown menu with "no preference" selected, showing options "audio" and "video")
- Available for consultation media :** (dropdown menu with "no preference" selected, showing options "online and public" and "online (account required)")
- Date d'enregistrement:** (dropdown menu with "1920" selected)
- Statut du document :** (dropdown menu with "no preference" selected, showing options "Édités" and "Inédits")
- année de parution :** (dropdown menu with "----" selected)
- contexte d'enregistrement :** (dropdown menu with "Terrain" selected)
- nature du support original :** (dropdown menu with "Bande(s) magnétique(s)" selected)
- cote :** (text input field)

At the bottom center is a "Rechercher" button with a magnifying glass icon. At the very bottom of the page, there are footer links for "telemeta", "Usage des archives réservé dans le respect du patrimoine culturel des communautés d'origine.", "Copyright © 2015 CREM-CNRS", and "Mentions légales".

Ces critères ont été choisis par le CREM en 2015 selon les descripteurs jugés prioritaires et grâce aux moyens de développement offert par plusieurs stagiaires informatiques encadrés par Nédra Mellouli-Nauwynck (IUT d’Informatique de Montreuil). Ce moteur est devenu ainsi un module indépendant (Haystack et ElasticSearch). Certains critères font appel à des listes très lourdes (populations, Instruments) et la mise en place de la complétion est complexe.

Les critères de recherche sont les suivants :

- **Le titre** (Collection ou item)
- **Collecteur** (avec auto-complétion) : Nom de la personne morale ou physique ayant collecté l’archive
- **Lieu** (avec auto-complétion) : lieu d’enregistrement
- **Population/groupe social** : liste et auto-complétion
- **Instruments** : Instruments de musique, nom générique et vernaculaire (avec auto-complétion)
- **Type de media** : Audio / Vidéo / Sans préférence
- **Disponibilité du média** : En ligne (numérisé et en accès libre) / accessible avec un compte (numérisé mais pas en accès libre) / sans préférence
- **Période d’enregistrement** : Concerne la date d’enregistrement du média. Entrée d’une date de début et d’une date de fin, par saisie simple ou par menu déroulant avec les dates possibles de la base de données
- **Statut du document** : Édités /Inédits/Sans préférence. La sélection se fait selon si la cote contient **CNRSMH_I** (inédit/terrain) ou **CNRSMH_E** (édité/publié)
- **Année de parution** : pour un document édité (disque, cassette...). Entrée d’une date de début et d’une date de fin, par saisie simple ou par menu déroulant avec les dates possibles de la base de données
- **Contexte d’enregistrement** : saisie simple ou par un menu déroulant
- **Nature du document archivé** : Format du support physique ou numérique archivé, qui peut être différent du format original
- **Cote** : les cotes collection sont toutes de la forme **CNRSMH_I_AAAA_NNN** (Inédits) ou **CNRSMH_E_AAAA_NNN** (Édités). Ce champ est très utile pour les archivistes. (Cf. plan de nommage Chapitre 2.4)

3.3 Affichage des résultats (recherche simple et avancée)

3.3.1 Affichage en liste

- Les résultats de la recherche s’affichent en liste de **Collections** par défaut, avec le nombre de résultats par niveaux de catalogue. Il est possible également de choisir d’afficher les résultats de la requête en liste d’**Items**, de **Corpus** ou de **Fonds**.
- Il est possible de trier chaque colonne par ordre alphabétique, croissant/décroissant, et de faire une sélection en cochant dans la liste à gauche pour ajouter à une **Liste de lecture** (à choisir dans le menu “**Select your playlist**”). La liste des items sélectionnés s’affiche pour vérification avant de faire “**Ajouter à la liste**” à nouveau. L’utilisateur est prévenu si l’élément a déjà été sélectionné. On peut aussi revenir à la liste des résultats de la requête en choisissant “**Return to search results**”.

- Si on coche “**Select all**”, tous les items affichés par la requête sont sélectionnés et peuvent ajouter à une **Liste de lecture**.
- Pour chaque niveau de catalogue, on a un menu de sélection du nombre de résultats par page : 20, 50, 100, 250, 500, 1000...

3.3.2 Filtrage des résultats

Pour affiner les résultats de la recherche, un filtrage est proposé selon 6 critères répartis dans **2 blocs thématiques**. En sélectionnant plusieurs critères, les filtrages se cumulent et le résultat tient compte de tous les critères activés.

- **Bloc 1**, concerne l'**accessibilité** aux résultats de la recherche (déplié par défaut)
 - Statut d'accès : Les résultats comprenant les Inédits ou les édités
 - Statut d'accès : Les résultats avec le média numérisé (sound)
 - Type d'accès : Les résultats avec le média accessible librement sans compte (viewable)
- **Bloc 2**, concerne les **caractéristiques du média** (replié par défaut pour alléger la fenêtre)
 - Type de media : Les résultats avec de l'audio ou de la vidéo
 - Contexte d'enregistrement : Terrain ou Studio ou Radio
 - Nature du support archivé : CDR ou Disque ou Cylindre

Remarque : La vitesse d'accès est ralentie car le module ElasticSearch-Haystack nécessite une mise à jour pour que sa compatibilité soit maintenue avec la dernière version de Telemeta 1.7 qui est en Django 1.8 (en 2020).

3.4 Recherche dans les niveaux “Collections” et “Items” dans le menu “Archives”

Par le menu “**Archives**” de la page d'accueil, on peut faire un tri aux niveaux **Collections** et **Items** grâce au bouton “**Filtres**” en haut à droite de l'écran, parmi toutes les collections ou tous les items. Pour ces niveaux du catalogue (**Collections** et **Items**), ce bouton de tri permet de sélectionner soit les références inédites, soit les références éditées (disques), soit les références ayant des médias (numérisées). Ce bouton de filtres est surtout utile pour les archivistes qui ont besoin d'avoir des informations quantitatives : combien d'items numérisés ? Combien de collections inédites. Il existait avant l'amélioration du moteur de recherche avancée. Depuis 2014, il fait double emploi mais son accès est plus direct quand on est dans le menu “**Archives**” et il a été maintenu.

Pour le niveau **Item** du menu “**Archives**”, on peut sélectionner également le nombre de résultats affichés par page : 20, 50, 100, 250, 500, 1000... pour une lecture plus facile des résultats et pour cocher ses sélections pour une liste de lecture personnelle.

3.5 Recherche par pays

Le Géo-navigateur est accessible à tous les utilisateurs et permet d'afficher les collections par pays, soit sur une **carte** du monde, soit dans des listes de pays par continent dans le menu “**Lister**” en haut à droite de l'écran.

La localisation des collections est possible en combinant le thésaurus GéoEthno¹⁸ du LESC (qui prend en compte les termes historiques des noms de lieux) et Geonames¹⁹ (qui fournit les coordonnées géographiques). En sélectionnant un pointeur, on obtient le nombre de collections et d'items par pays, ainsi que le lien vers le résultat.

Remarques :

- L'utilisation de GéoEthno permet d'associer dans une recherche, par exemple, Burkina Faso et Haute-Volta comme un même lieu. De plus ce thésaurus est en langue française...
- La mise à jour de Geo-Ethno devrait être effectuée avec l'intégration éventuelle des nouveaux termes saisis dans le champ « Précision lieu » (Cf. consignes, Chapitre 7).
- La plateforme utilise actuellement l'application Google Map. L'utilisation d'Open Street Map, qui est gratuit et collaboratif, est envisagée pour les projets futurs, afin d'utiliser un outil plus ouvert et d'affiner la géolocalisation des sources.

3.6 Recherche par listes/énumérations

La sélection d'une liste dans le menu “Terminologie” de la page d'accueil permet d'accéder à tous les termes composant la liste et de rechercher tous les items indexés par chacun des termes d'une liste. En effet, dans chaque liste, le nombre d'items s'affiche en face de chaque terme correspondant. Par exemple, dans la liste “Mot-clé”, on trouve 201 berceuses. En cliquant sur le nombre « 201 », on obtient la liste des 201 items de berceuse, on peut les sélectionner et les ajouter dans une liste de lecture. Cette fonctionnalité de recherche est utile pour les documentalistes et la gestion des listes.

Les énumérations suivantes sont accessibles par les profils administrateurs et documentalistes ayant le statut « super-utilisateur » :

- Contexte d'enregistrement
- Droit d'utilisation
- Éditeur
- Mode d'acquisition
- Mot-clé
- Numérisation
- Population / groupe social
- Rédacteur fiche
- Réédition
- Saisie fiche
- Statut de la collection
- Style générique
- Style vernaculaire
- Support archivé
- Support original
- Type d'exemplaire
- Type d'identifiant
- Type de média

Les listes d'**instruments** (termes génériques et termes vernaculaires) sont consultables tout public dans le menu “Terminologie”, ce qui permet de faire une recherche par instrument de musique, de la même manière que pour les énumérations ci-dessus.

¹⁸ GéoEthno: <http://www.mae.u-paris10.fr/dbtw-wpd/bed/index-lesc.html>

¹⁹ <https://www.geonames.de/>

Remarque : la consultation des listes par chaque profil utilisateur devrait être ajoutée dans les permissions par profil afin que ces listes soient visibles (en lecture seule) par les profils autres que “Super-utilisateur”. Cette permission n'a pas été spécifiée lors du développement de l'accès aux listes en page d'accueil par les stagiaires en informatique.

4 Fonctionnalités de navigation et d'affichage

4.1 Dans un *Fonds* ou un *Corpus*

Lorsque l'accès se fait par le menu “Archives”, les possibilités de navigation sont les suivantes :

- Affichage en liste des titres de *Fonds* ou de *Corpus*,
- Titres avec liens hypertexte (en rouge) pour accéder au *Fonds*, au *Corpus*
- Affichage des numéros de pages des listes
- Flèches de saut de page
 - Navigation dans un *Fonds* :

Dans la fiche *Fonds* sélectionnée, les titres des *Corpus* “fils” s'affichent en liste. La fiche peut comporter des ressources associées visible dans la rubrique “**Média associés**” (images, liens, bibliographie...) et la liste des *Corpus* se trouve en bas de la fiche. Comme ailleurs, les liens vers les ressources externes et les *Corpus* “fils” sont en rouge. La liste des *Corpus* affiche les titres, la description, la présence de média numérisés et la cote.

- Navigation dans un *Corpus* :

Dans la fiche *Corpus* sélectionnée, les titres des *Collections* “filles” s'affichent en liste. La fiche peut comporter des ressources associées visible dans la rubrique “**Média associés**” (images, liens, bibliographie...) et la liste des *Collections* se trouve en bas de la fiche. Le ou les Fonds “père(s)” sont en rouge (lien hypertexte) et permettent de “remonter” dans le catalogue. La liste des *Collections* affiche les titres, la description, la présence de média numérisés et la cote.

Le bouton **Dublin Core** permet de visualiser dans une nouvelle fenêtre les correspondances de chaque fiche descriptive Fonds ou Corpus dans le format de métadonnées Dublin Core simple et étendu. Les valeurs moissonnées par le protocole OAI-PMH sont ainsi accessibles.

4.2 Dans une *Collection*

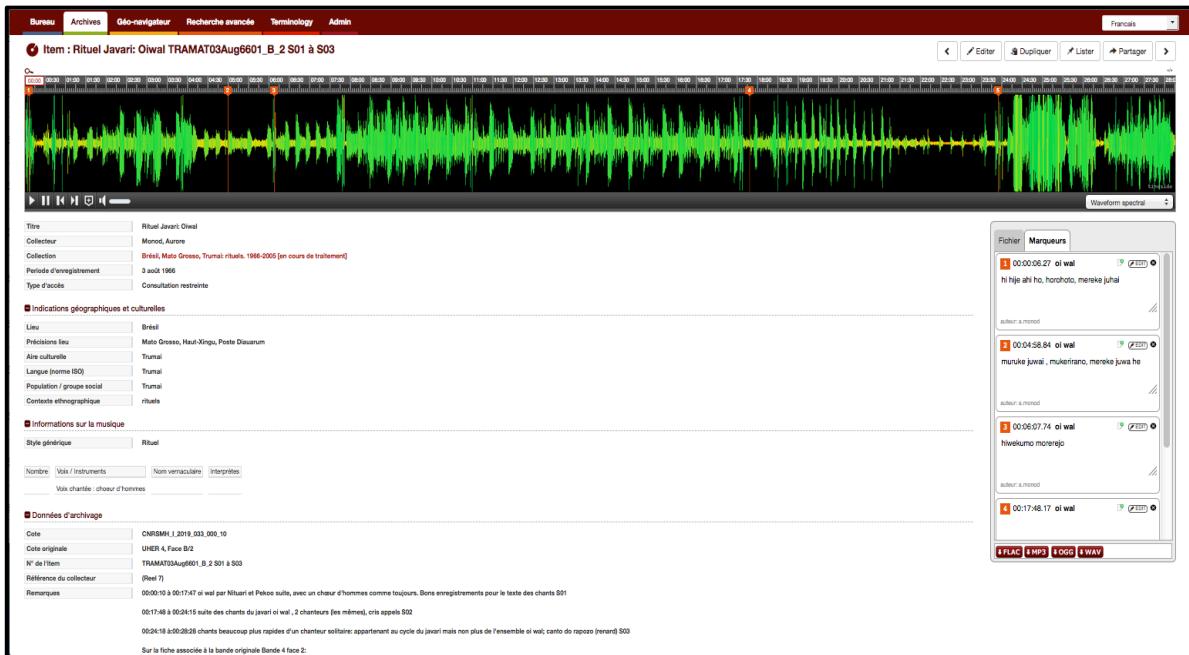
La collection est l'ensemble de base qui rassemble des items d'une même mission, d'un même disque ou d'une même thématique, collectés par une même entité morale ou physique.

Une fois la collection sélectionnée, la fiche « **Collection** » apparaît avec le nombre des items. Un lien dynamique vers la liste des items permet d'aller directement aux items qui sont en bas de la fiche **Collection**. Le ou les *Corpus* “père(s)” sont en rouge (lien hypertexte) et permettent de “remonter” dans le catalogue. On peut ainsi naviguer dans le catalogue par les corpus et fonds « parents ». Les éléments suivants s'affichent :

- La liste des *Items* affiche les titres, la description, la présence de média numérisés et la cote.
- La présence du média audio ou vidéo est indiquée par une marque verte.

- Certains champs non renseignés ne s'affichent pas pour alléger la consultation.
- Selon le statut d'accès de la collection et des items et selon son profil, l'utilisateur peut écouter directement un item dans la liste (bouton « lecture/pause » à gauche du titre) ou sélectionner la fiche item pour consulter la fiche documentaire et le son/vidéo en streaming (si le media est présent dans la base de données).
- Le bouton « **Dublin Core** » permet de visualiser dans une nouvelle fenêtre les correspondances dans le format de métadonnées Dublin Core simple et étendu.

4.3 Dans un Item



Une fois l'**Item** sélectionné, la fiche descriptive de l'Item apparaît. Le lien vers la Collection “mère” est en rouge (lien hypertexte) et permet de “remonter” dans le catalogue. Certains champs non renseignés ne s'affichent pas pour alléger la consultation. Le son ou la vidéo s'affiche avec la description documentaire (Cf. TimeSide chapitre 6).

Les fonctions de navigation dans la fiche item sont les suivantes :

- Passage à l'item suivant ou précédent (**flèches** en haut à droite)
- Affichage possible du player plein écran (loupe)
- Onglet « **Marqueurs** » : liste des marqueurs temporels éventuels
- Onglet “**Fichier**” : métadonnées techniques calculées
- Navigation dans le média : lecture, pause, retour au marqueur précédent/suivant, marqueur (drapeau), niveau d'écoute
- Curseur mobile de lecture avec affichage du Time code (MM:SS)
- Bouton “**Marqueur**” pour positionner une annotation (titre et commentaire) dans le temps.
- Affichage en rouge de la position du marqueur temporel avec son numéro et de l'annotation au passage de la souris ou lors de la lecture
- Menu option d'affichage graphique (Forme d'onde (waveform) par défaut)
- Lien hypertexte en rouge pour remonter au niveau **Collection**
- Lien hypertexte en rouge pour accéder au “**Média associé**”

Le bouton **Dublin Core** permet de visualiser dans une nouvelle fenêtre les correspondances dans le format de métadonnées Dublin Core simple et étendu.

5 Fonctionnalités de gestion des accès aux données

Telemeta permet de gérer les accès aux données, selon le statut des collections *ET* le profil de l'utilisateur.

Cette fonctionnalité a été améliorée en 2014. A cette période la législation sur les droits des artistes-interprètes préconisait un embargo de 50 ans. Le Code du Patrimoine²⁰ prévoit également un délai de communicabilité des archives publiques de 50 ans (Catégorie Vie privée). Cette règle peut être modulée selon les consignes du déposant et des communautés autochtones, au cas par cas selon les conditions éthiques. Les utilisateurs du monde académique sont invités en page d'accueil à demander un compte utilisateur pour accéder à l'ensemble des ressources audiovisuelles numérisées et consultables en ligne. Les données descriptives (métadonnées) sont en accès libre, exceptées les fiches **items** des collections consultables "sur demande". Cf. questions juridiques chapitres 7.

5.1 Gestion de la connexion

L'application permet de mettre un **embargo** de 50 ans (ou plus) sur des collections de la base de données. Le compte utilisateur permet d'accéder à toutes les métadonnées (sauf les items en consultation sur demande) et à tous les sons/vidéos sous embargo (sauf les items en consultation sur demande).

5.1.1 Crédation de compte utilisateur

Le compte est attribué par les administrateurs du site (dans la page "Admin"/Auth/Utilisateurs) et la création d'un compte se fait avec le nom de l'utilisateur et son adresse mail au minimum.

Un compte peut aussi être attribué à une plage d'adresses IP fournie par une institution qui souhaite donner accès dans ses locaux sans demander un compte pour chaque utilisateur (consultation seule).

A chaque compte, l'administrateur peut attribuer un profil parmi 5 groupes : Administrateur, Documentaliste, Chercheur, Membre, Visiteur.

Il existe **275 permissions** différentes assignées aux groupes Administrateur, Documentaliste et Chercheur. Les permissions (lecture, écriture, suppression) peuvent être modulées pour chaque compte utilisateur en plus du lot de permissions de son groupe. Ces permissions sont consultables sur PhpMyAdmin²¹ (mot de passe) :

https://archives.crem-cnrs.fr/phpmyadmin/sql.php?server=1&db=telemeta&table=auth_permission&pos=0&token=a53d4d8e81e82df4dd6e7ced8ea322ca

Tous les comptes utilisateurs y sont également consultables (mot de passe) :

https://archives.crem-cnrs.fr/phpmyadmin/sql.php?server=1&db=telemeta&table=auth_user&pos=0&token=a53d4d8e81e82df4dd6e7ced8ea322ca

Les administrateurs peuvent choisir une date d'expiration du compte et voir la date de la dernière connexion. Ils peuvent également attribuer un ou plusieurs statuts au compte de l'utilisateur :

- **Actif** : Précise si l'utilisateur doit être considéré comme actif. Décrocher la case plutôt que de supprimer le compte.
- **Statut équipe** : Précise si l'utilisateur peut se connecter au site d'administration.

²⁰ Code du Patrimoine, Livre II, Articles L.213-1 et L.213-2

²¹ https://archives.crem-cnrs.fr/phpmyadmin/db_structure.php?server=1&db=telemeta&token=f448e42ec974fd94167bd6c54876ce6f

- **Statut Super-utilisateur** : Précise que l'utilisateur possède toutes les permissions sans les assigner explicitement.

5.1.2 Connexion à la plateforme

La connexion à la plateforme se fait avec le bouton “**Connexion**” sur la page d'accueil. Le choix de la langue de l'interface se fait automatiquement selon la langue utilisée dans le navigateur. Le visiteur est invité à mettre son nom de compte (en général en minuscule, par exemple : p.dupont) et son mot de passe. En cas d'oubli, il peut cliquer sur “**Mot de passe oublié**”. Dans ce cas, le visiteur reçoit un nouveau mot de passe à l'adresse mail qu'il a donné pour la création du compte.

5.2 Les types d'accès aux collections et aux items (embargo)

Les administrateurs peuvent choisir une durée d'embargo de 50 ans pour l'accès public aux média. L'accès évolue chaque année grâce à l'utilisation d'une « **date glissante** ». Actuellement, la moitié des données sont concernées et donc ne sont pas en accès libre. Les collections sont configurables par item et par collection grâce au champ **Type d'accès** au niveau **Collection** et au niveau **Item** (avec un compte administrateur ou documentaliste). Le statut de la collection est prioritaire sur le statut de l'item. S'il faut changer le statut de tous les items, il suffit de changer celui de la collection et le type d'accès aux items sera le même que celui de la collection. Par défaut, une collection est en accès restreint et les items aussi. Si la collection devient accessible à tous (“**Consultation libre**”), il n'est pas nécessaire de modifier le statut de tous les items de cette collection, uniquement le statut de la collection. Le statut “**Accès restreint**” des items reste toujours affiché mais ce statut est désactivé et l'accès devient libre.

Les items d'une même collection peuvent avoir besoin de statuts d'accès différents (libre, restreint ou réservé). Dans ce cas, il faut choisir le statut « **Accès partiel aux items** » pour la collection.

- **Le Type d'accès d'une collection**

Il existe 4 types d'accès au choix, afin de répondre à tous les cas de figure de communicabilité des contenus :

- **Consultation libre** : l'accès aux métadonnées et au son est autorisé au public (full). Statut par défaut.
- **Consultation restreinte** : seul l'accès aux métadonnées est autorisé au public (metadata)
- **Consultation sur demande** : l'accès aux métadonnées des items et aux médias est réservé aux profils administrateur et documentaliste (none).
- **Accès partiel aux items** : Les items de la collection peuvent avoir des statuts différents (mixte) parmi les 3 autres statuts (Full/Metadata/None). Le statut de chaque item sera alors à configurer selon les besoins.

La case “**Accès automatique après la date glissante**” permet d'indiquer que la collection va passer en accès libre 50 ans après la date d'enregistrement. Si la collection contient plusieurs dates d'enregistrement, la date de l'item le plus récent est prise en compte automatiquement. Le statut “**Consultation restreinte**” reste affiché, même si la collection est devenue librement consultable automatiquement selon la date glissante. Chaque année, les collections ayant plus de 50 ans passent en accès libre. Si la case est décochée, la collection ne passe pas en accès libre après 50 ans et reste en consultation restreinte. Cette fonctionnalité encourage de nombreux chercheurs à déposer et à archiver leurs collectes de terrain, même si ces données sont « sensibles », ce qui est le cas le plus fréquent.

- Le Type d'accès d'un *item*

Il existe 3 types d'accès au choix :

- **Consultation libre** : l'accès aux métadonnées et au son est autorisé au public (**full**). Statut par défaut.
- **Consultation restreinte** : seul l'accès aux métadonnées est autorisé au public (**metadata**).
- **Consultation sur demande** : l'accès aux métadonnées des items et aux médias est réservé aux profils administrateur et documentaliste (**none**). La fiche **Collection** reste accessible (métadonnées).

Quand le média d'un item n'est pas consultable, un message invite l'utilisateur à contacter l'équipe pour obtenir l'autorisation avec un lien hypertexte vers la page des mentions légales.

Le tableau suivant indique tous les cas de figure pris en compte pour l'accès aux données tel qu'il a été implémenté en 2014.

Tableau récapitulatif d'accès aux métadonnées et aux médias :

Statut Collection	Statut Item	Statut prioritaire	Soumis à date glissante	Accès Admin Doc	Accès Chercheur Membre	Accès Public
full	Full, metadata ou none	Collection	oui	full	full	full
metadata only	Full, metadata ou none	Collection	oui	full	full	metadata (pendant 50 ans)
metadata only	Full, metadata ou none	Collection	non	full	metadata	metadata
none	Full, metadata ou none	Collection	oui	full	none (pendant 50 ans) Mais fiche coll. accessible	none (pendant 50 ans) Mais fiche coll. accessible
none	Full, metadata ou none	Collection	non	full	none	none
mixte	full	Item	oui	full	full	full
mixte	metadata		oui	full	full	Metadata (pendant 50 ans)
mixte	metadata		non	full	metadata	metadata
mixte	none		oui	full	none (pendant 50 ans) Mais fiche coll. accessible	none (pendant 50 ans) Mais fiche coll. accessible
mixte	none		non	full	none mais fiche coll. accessible	none mais fiche coll. accessible

Admin = administrateurs + documentalistes

Pendant 50 ans = statut temporaire (soumis à la date glissante de 50 ans)

5.3 Les profils utilisateurs

Les droits d'accès varient selon les profils utilisateur.

Les 5 profils sont consultables et modifiables dans le menu “Admin” :

- Admin,
- Documentaliste,
- Chercheur,
- Membre,
- Visiteur.

La liste des permissions est indiquée dans chaque profil (page Admin). Les principales caractéristiques sont les suivantes :

- **Profil Admin**

- Accès à toutes les données et à toutes les options de visualisation graphique audio
- Permission pour toutes les modifications des données (Création, Suppression, Duplication)
- Exportation de fiches en zip ou sous Epub
- Téléchargement des sons (Wav, Flac, Mp3, Ogg)
- Gestion des thésaurus : modifier, supprimer, fusionner et ajouter les nouvelles entrées
- Gestion des accès (permissions), création des comptes utilisateurs, attribution des codes d'accès (login)
- Accès au bureau (playlist, dernières modifications)
- Exportation des métadonnées sous format CSV (via une Playlist)
- Modification et suppression de tous marqueurs.
- Ajout d'un « media associé » : pdf, jpg, png, URL

- **Profil Documentaliste**

- Accès à toutes les données et à toutes les options de visualisation graphique audio
- Permission pour toutes les modifications des données (Création, Suppression, Duplication)
- Exportation de fiches en zip ou sous Epub
- Export des médias (wav, Flac, Mp3, Ogg)
- Gestion des thésaurus : modifier, supprimer, fusionner et ajouter les nouvelles entrées
- Accès au bureau (playlist, dernières modifications)
- Exportation des métadonnées sous format CSV (via une Playlist)
- Modification et suppression des marqueurs
- Ajout d'un « media associé » : pdf, jpg, png, URL

- **Profil Chercheur**

- Accès à toutes les données (sauf statut **None**/Consultation sur demande) et à toutes les options de visualisation graphique audio.
- Pas de téléchargement proposé
- Permission pour les modifications des données Item, Corpus et Fonds (Création, Duplication, pas de suppression)
- Pas de création de fiche collection

- Accès au bureau (playlist, dernières modifications)
- Exportation des métadonnées sous format CSV (via une Playlist)
- Pose de « marqueurs » et saisie commentaires. Pas de suppression et de modification des autres marqueurs.
- Ajout d'un « media associé » : pdf, jpg, png, URL
- Signature du marqueur (automatique).

- **Profil membre** (compte institutionnel avec adresse IP)

Accès aux métadonnées et média (sauf statut **None**/Consultation sur demande).

Pas de téléchargement proposé

Écoute des fichiers sons compressés mono

Accès au bureau (playlist, dernières modifications)

Pas de pose de « marqueurs »

Recherche simple et multicritère

Pas de modification, consultation simple

- **Profil visiteur**

Accès aux métadonnées (sauf none) et média en accès libre

Pas de téléchargement proposé

Recherche simple et multicritère

Lien avec le site web du Laboratoire et les sites partenaires.

5.4 Fonctions d'accès aux médias et métadonnées

Différentes fonctionnalités sont prévues pour accéder aux données et métadonnées. Sans compte utilisateur, tout le catalogue (métadonnées) est accessible à tous publics (exceptés les items des collections en “**accès sur demande**”). Les médias sont accessibles selon le statut des collections (et des items).

On peut écouter directement un item dans la liste des items (bouton « lecture » à gauche des titres) ou sélectionner la fiche item pour consulter la fiche documentaire en même temps que l'affichage du média en streaming (si le media est présent dans la base de données).

Les métadonnées sont également accessibles par **exportation** et donc **réutilisables** : soit dans la collection au format ePub (livre numérique avec les métadonnées et les médias), soit dans une liste de lecture (bureau) au format **CSV** (avec compte utilisateur). L'export des fiches **Collection** ou des fiches **Items** de la “**liste de lecture**” sous le format ouvert CSV se fait selon le schéma décrit sur le document partagé suivant :

<https://docs.google.com/spreadsheets/d/1Hpvk-5ZvKTr7AOTdrtIt0FkFjrw4IFIMKTYCi0neyOo/edit#gid=359255607>

Les médias audio et vidéo d'une collection sont accessibles par téléchargement avec le bouton “**Zip**” dans la fiche collection (Compte Admin et Doc). Les exports sous divers formats sont également possibles dans l'application web de gestion de base données **PhpMyAdmin** (Cf. chapitre 2.8).

Les “**Médias associés**” (images pdf, jpg, png) sont accessibles en vignette de prévisualisation ou en téléchargement.

6 Fonctionnalités de gestion des média (son, vidéo, images)

TimeSide²² est un **Framework**²³ qui gère l'analyse audio, la représentation graphique, le transcodage audio, le streaming audio et vidéo, ainsi que l'annotation temporelle avec des marqueurs éditorialisables. C'est une API en langage Python qui permet d'avoir une interface web avec un affichage dynamique du média dans la fiche descriptive. Elle est gratuite (développée avec l'ANR Wasabi depuis 2016), téléchargeable et peut être installée via Docker sur tous les systèmes d'exploitation.

L'avantage pour la gestion des archives sonores est le transcodage en temps réel et à la demande d'un fichier audio pour la consultation en ligne sous forme compressée (OGG). La consultation d'un son est donc quasi immédiate après son téléchargement sur le serveur car on n'a pas besoin d'attendre une campagne de calcul de compression ni de gérer la sauvegarde de 2 fichiers sons pour chaque item (compressé et non compressé).

La lecture utilise un player **HTML5** extensible, ce qui permet de zoomer pour avoir un affichage plein écran. L'analyse audio permet aussi d'extraire des caractéristiques sonores et techniques grâce à un jeu de "librairies" audio, telles que Aubio, Yaafe ou VAMP. Tous les « Plug-in » ou algorithmes ont été développés sous licence libre et sont réutilisables (Cf. <https://github.com/Parisson/TimeSide/>)

6.1 Données produites par TimeSide, transcodages (analyse audio)

Une des fonctionnalités les plus importantes de TimeSide pour l'archivage est la compression des sons. La compression des fichiers son est calculée automatiquement, ce qui permet de consulter un son compressé, plus adapté au streaming, dès son chargement sur la plateforme. L'opération de compression et celle de la double sauvegarde sur le serveur (le son non compressé Wav et le son compressé Ogg) ne sont pas faites par l'archiviste, ce qui fait gagner du temps. A chaque calcul, le fichier compressé automatiquement par TimeSide est stocké en mémoire sur le serveur. Le son haute définition (stéréo, WAV, 48kHz ou 44kHz au CREM) n'est pas accessible en consultation par les utilisateurs. Il peut être téléchargé uniquement par les administrateurs. Le format WAV est le plus répandu, bien documenté et non compressé. Il est accepté pour l'archivage à long terme (CINES, Archives Nationales).

TimeSide gère automatiquement les formats audiovisuels suivants :

- Fichiers non compressés pour la conservation : WAV (48kHz/24bit ; 44.1kHz/16bit), FLAC
- Fichiers compressés audio : MP3 et OGG Vorbis
- Fichiers compressés Vidéo : MP4 et WebM

Remarque :

Le format **OGG** est un format ouvert alors que le MP3, très utilisé, est un format propriétaire (sous licence). De plus, à débit égal, il est de meilleure qualité.

²² <https://github.com/Parisson/TimeSide/>

²³ Framework : infrastructure logicielle, conçue pour modeler l'architecture des logiciels applicatifs

Les caractéristiques techniques des données audios gérées par TimeSide sont les suivantes :

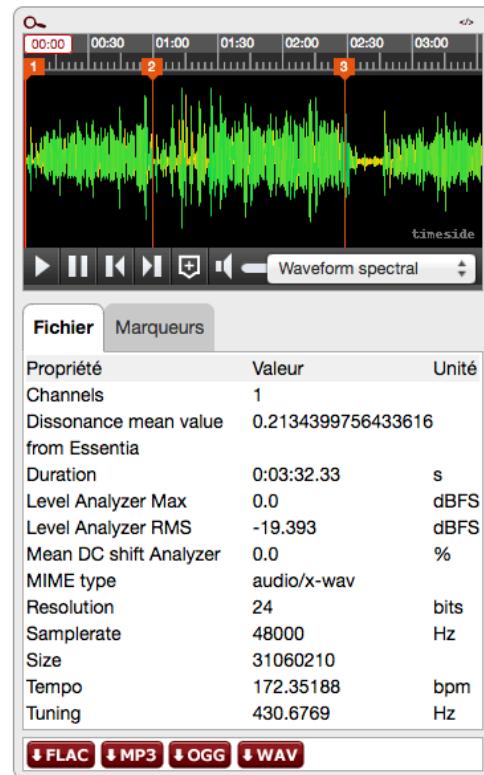
- **WAV** : format privilégié pour l'archivage et la sauvegarde audio, ce format sans perte a une qualité au moins égale à celle d'un CD audio. N = 8, 16, 24 ou 32 bits. Fe de 8000 à 192000 Hz.
- **MP3** : utilisé pour la transmission en streaming et la lecture par les lecteurs audio en ligne, il utilise une compression avec pertes pour diminuer le débit. Très largement utilisé de part portabilité, il peut encapsuler partiellement les métadonnées. Ce format n'est pas libre de droit (royalties pour l'encodeur, non pour le décodeur). N = 8 ou 16 bits, fe de 8000 à 48000 Hz ; k de 0,05 à 0,2.
- **OGG Vorbis** : qualité supérieure à débit égal que le MP3 et libre de droit (open source), encapsulation totale des métadonnées. N = 8 ou 16 bits, fe de 8000 à 96000 Hz, k de 0,01 à 0,3
- **FLAC** : compression sans pertes, qualité égale au WAV à l'écoute, multi-pistes, open source, encapsulation totale des métadonnées. N = 16 ou 24 bits, fe de 11050 à 96000 Hz, k = 0,6. Ce format n'est pas utilisé par le CREM.

Le lecteur comprend une fenêtre de visualisation et de navigation dans le média, ainsi qu'un onglet "Fichier" avec les données techniques calculées.

Cet onglet est affiché par défaut. Certains algorithmes (Dissonance, Tempo, Tuning) sont calculés lors du chargement du son sur le serveur depuis 2018 depuis leur intégration dans TimeSide lors du projet DaCaRyH (Cf. tableau page suivante).

Remarque :

Dans la version du CREM, le téléchargement des formats de fichiers sons visibles ici n'est possible que par les profils "Administrateurs » (Wav, Flac, Ogg, Mp3), mais cette permission peut être étendue dans le menu Admin (profils utilisateurs).



Trois des algorithmes du projet DaCaRyH sont donc opérationnels directement dans cette fenêtre, les autres sont disponibles dans le menu déroulant du player avec un compte (Cf.chapitre 6.3).

Financé conjointement par le LABEX « Les Passés dans le Présent » et par son homologue britannique l'AHRC *Care for the Future*, le projet DaCaRyH, *Data science for the study of Calypso Rhythm through History* (2016-2017), a rassemblé des ethnomusicologues et des ingénieurs du CREM-LESC, la Société Parisson et le Centre de Musique Numérique (C4DM) de l'Université Queen Mary de Londres.

Les données calculées qui s'affichent dans la fenêtre “**Fichier**” du player sont les suivantes :

Nom / valeur calculée	Signification	Unité	Remarque
Channels	Nombre de canaux		1 ou 2
Duration	Durée	S	
Level Analyzer Max	Niveau maxi	dBFS	
Level Analyzer RMS	Niveau moyen	dBFS	
Mean DC shift Analyzer	Différence entre les canaux	%	
MIME type	Type MIME		
Resolution	Résolution audio	bits	16 ou 24
Samplerate	Taux d'échantillonnage	Hz	
Size	Taille fichier numérique (Mo)	Octet	
<i>Valeurs musicales :</i>			
Tempo	Tempo moyen	bpm	Battement par mn. Algorithme (DACARYH)
Tuning	Fréquence du LA3	Hz	Algorithme (DACARYH)
Dissonance mean value from Essentia	Taux de dissonance	entre 0 et 1	Algorithme Essentia (DACARYH)

La taille du fichier (en Mo) s'affiche aussi dans la fiche item et la somme des fichiers items dans la fiche collection (durée et taille).

La somme des durées s'affiche dans les **Corpus** et les **Fonds**.

6.2 Lecteur de streaming

Dans chaque fiche **Item**, il est possible (mais non obligatoire) d'associer un média audio ou vidéo à la fiche descriptive, ce qui permet de vérifier la documentation pendant la description ou consultation de l'archive audiovisuelle simultanément à la lecture du média, dans la même page. En effet, la description d'un son musical est complexe (moins immédiate qu'une photo) et certains détails organologiques, par exemple, échappent souvent à la mémoire du collecteur. Ses notes de terrain sont généralement incomplètes et seule l'écoute permet de valider ou de corriger les métadonnées, qui varient d'une pièce à l'autre. On constate à chaque séance de travail d'archivage avec un chercheur que le son réveille la mémoire quasi systématiquement. Les données sont donc souvent enrichies progressivement.

Pour ce faire, un affichage dynamique du média dans la fiche descriptive est très utile. Il est possible grâce au langage informatique “orienté objet” utilisé par TimeSide (Python). Cette caractéristique apporte un confort de travail très apprécié par les utilisateurs. Il permet également de « naviguer » dans la temporalité du document en repérant visuellement les différentes séquences de la matière sonore.

The screenshot shows a digital library interface with a top navigation bar in French: Bureau, Archives, Géo-navigateur, Recherche avancée, Terminology, Admin, and a language switch to Francese. Below the navigation is a toolbar with icons for Edit, Duplicate, List, and Share.

The main area displays a media player with a green waveform visualization. The timeline at the bottom ranges from 00:12 to 13:30. To the right of the player is a thumbnail image of a group of people in traditional Kenyan attire performing a dance.

Item Details:

- Titre:** Chapitre 3 : Les danses-chants des hommes circoncis
- Collecteur:** Marmone, Giordano
- Collection:** Exemples pour la thèse de G. Marmone, 2017 : Vidéos d'illustration
- Période d'enregistrement:** 1 janvier 2012 - 31 décembre 2012
- Type d'accès:** consultation libre

Indications géographiques et culturelles:

- Lieu:** Kenya
- Précisions lieu:** Samburu County
- Langue:** Samburu
- Langue (norme ISO):** Samburu
- Population / groupe social:** Samburu
- Contexte ethnographique:** Cérémonie

Informations sur la musique:

- Modalités d'exécution:** Chœur et solistes

At the bottom left are buttons for Nombre, Voix / Instruments, Nom vernaculaire, and Interprètes. On the right, there are buttons for Fichier (selected) and Marqueurs, along with download links for FLAC, MP3, OGG, and WAV formats.

Le lecteur (audio et vidéo) propose les fonctionnalités suivantes :

- Affichage du player plein écran dans la fiche item pour l'audio (bouton **loupe** dans la barre supérieure)
- Affichage du player plein écran dans la fiche item pour la vidéo (bouton plein écran en surimpression du lecteur vidéo)
- Une *Time line* avec un compteur de temps pour l'audio et la vidéo (synchrone)
- **Curseur** mobile de lecture avec affichage du Time code (MM:SS)
- Navigation dans le média : **lecture** (>), **pause** (II), retour au marqueur précédent/suivant, volume d'écoute
- </> : Bouton “**i-frame**” en haut à droite, pour insérer le lien HTML du player dans un autre site web (très utilisé)
- Bouton “**Marqueur**” pour positionner une annotation (titre et commentaire) dans le temps. Affichage en rouge de la position du marqueur temporel avec son numéro et de l'annotation au passage de la souris ou lors de la lecture.
- Onglet « **Marqueurs** » : liste des marqueurs temporels éventuels, avec le time code,
- Onglet “**Fichier**” : métadonnées techniques calculées

6.3 Représentation graphique des médias

La visualisation du son est calculée par **TimeSide** directement sur le serveur (sans passer par un autre outil localement). Elle facilite la consultation du document dans sa durée et la pose des marqueurs temporels pour l'annotation. Elle permet de repérer les différentes séquences dans le document et ses motifs mélodico-rhythmiques.

L'affichage est calculé par TimeSide sur le serveur, ce qui peut prendre du temps si le fichier est long (supérieur à 10 mn).

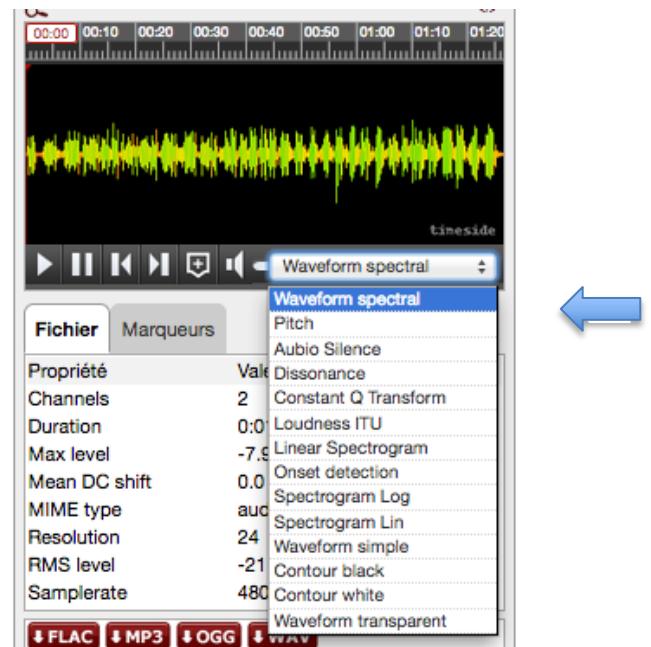
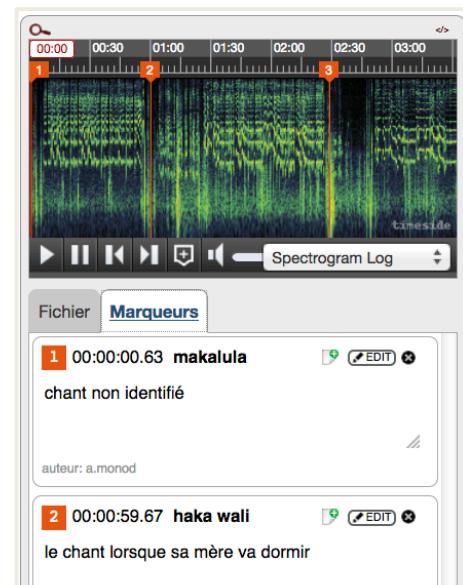
L'illustration ci-contre montre la visualisation avec “**Spectrogram Log**” qui permet un repérage des parties du chant.

Par défaut, l'affichage graphique du son est la **forme d'onde spectrale** (*waveform spectral*) tel qu'on le voit p. 43 : elle représente la variation de l'intensité en fonction du temps avec une variation des couleurs en fonction de la fréquence. Un menu déroulant dans la barre de navigation audio propose d'autres représentations graphiques du son. Pour limiter la saturation du serveur, certaines options sont en accès libre, d'autres sont disponibles avec un compte.

6.3.1 Avec un compte utilisateur :

Tous les outils de visualisation (ou option d'affichage graphique) et d'analyse audio disponibles dans TimeSide sont accessibles avec un compte utilisateur (version de 2020) :

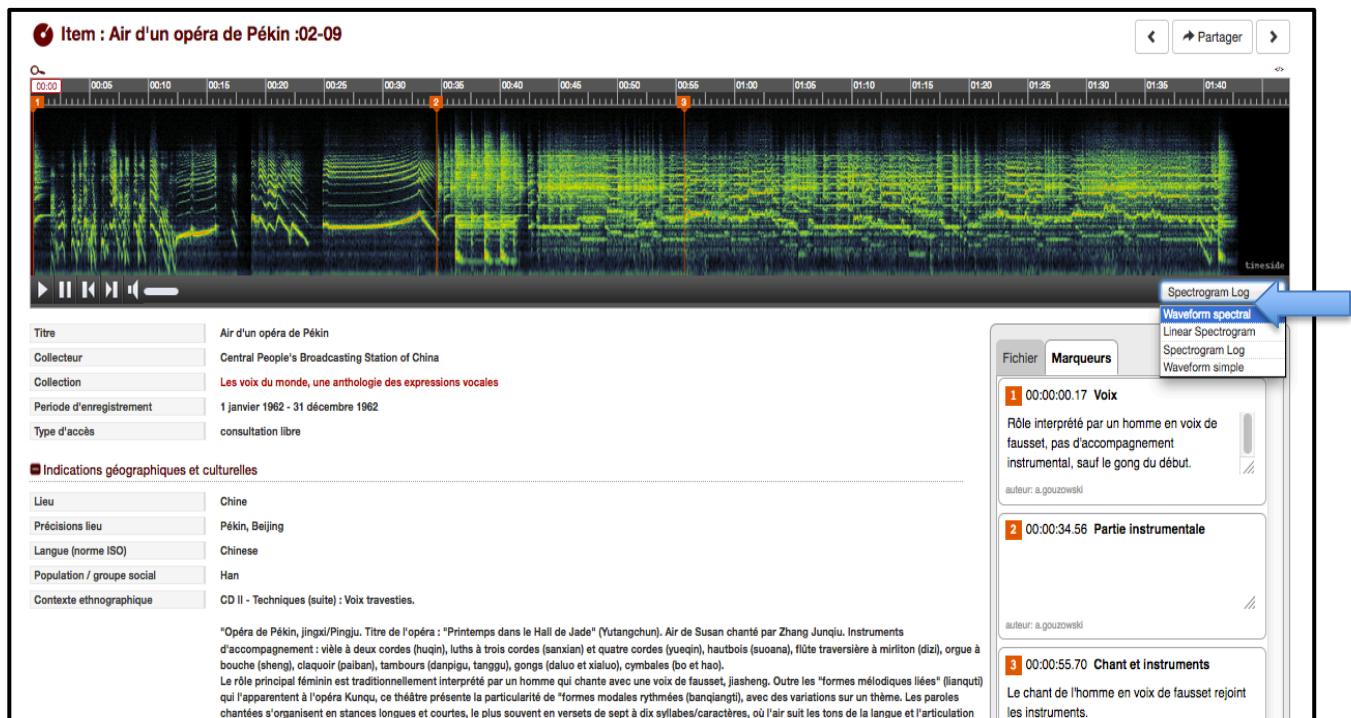
- **Waveform spectral** : forme d'onde spectrale
- **Pitch** : affichage des hauteurs relatives (utilisable sur des fichiers courts)
- **Aubio Silence** : détection des silences audio
- **Dissonance** : calcul de la dissonance harmonique en fonction du temps)
- **Constant Q Transform** : autre formule pour afficher le spectre
- **Loudness** : calcul de l'énergie en fonction du temps
- **Spectrogram Lin** : Spectrogramme (Temps/fréquences, échelle linéaire)
- **Spectrogram Log** : Spectrogramme (Temps/fréquences, échelle logarithmique)
- **Onset detection** : détection des attaques, utilisable sur des fichiers courts



- **Waveforme Simple** : forme d'onde, couleur noire
- **Contour black** : Contour de forme d'onde, couleur noire
- **Contour White** : Contour de forme d'onde, couleur blanche
- **Waveforme transparent** : forme d'onde

6.3.2 Sans compte utilisateur :

Les outils de visualisation accessibles sans compte utilisateur dans le menu du player sont les suivants : Waveform spectral, Spectrogram Log, Spectrogram Lin et Waveform Simple



6.3.3 Les algorithmes de l'ANR DIADEMS²⁴:

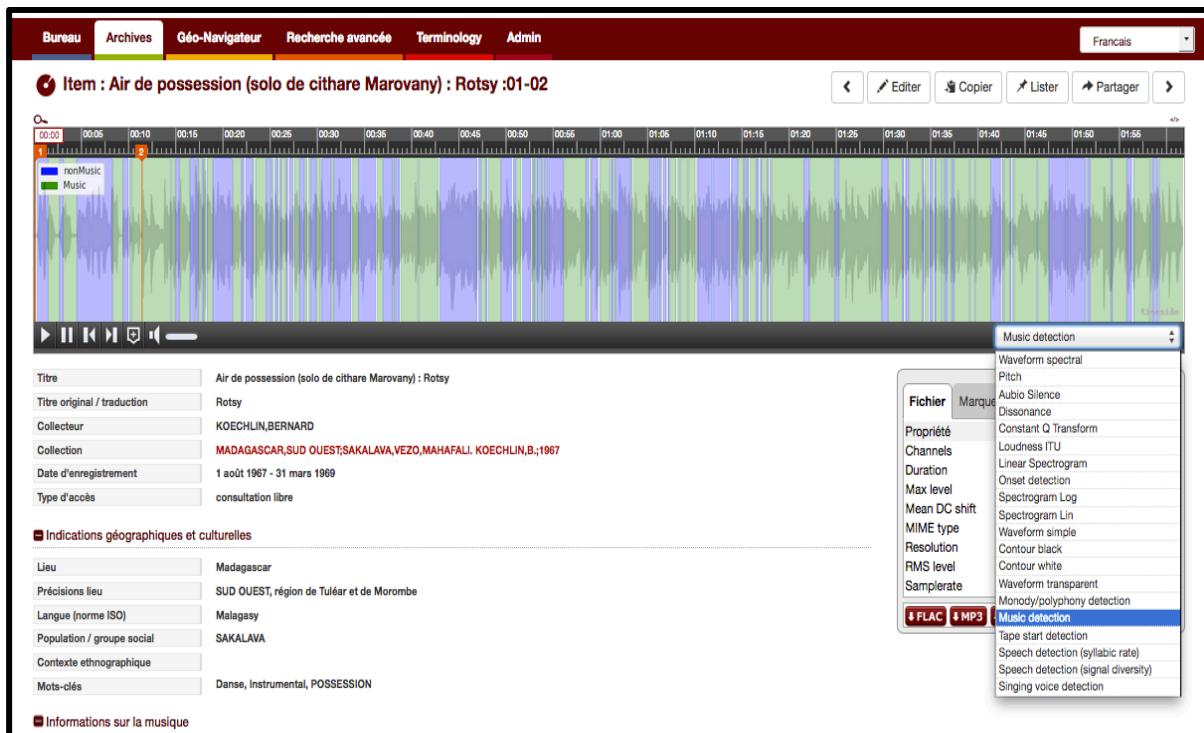
Les outils de détection (algorithmes d'analyse audio) implémentés lors de l'ANR Diadems (2013-2016) sont en attente de réintégration dans l'instance de production (avril 2020). Ils sont visibles en 2020 avec un compte utilisateur sur l'instance Telemeta suivante : <http://diadems.telemeta.org>. Cette instance contient 16 000 fichiers sons.

Le tableau suivant donne la définition des algorithmes développés :

Nom	Définition
Monody/Polyphony detection	mono/polyphonique
Music detection	détection des zones de musique
Tape start detection	démarrage de bande
Speech detection (syllabique rate)	détection de parole/non-parole
Speech detection (signal diversity)	détection de parole/non-parole (selon contexte, variation...)
Singing voice detection	détection de voix chantée

²⁴ <https://www.irit.fr/SAMOVA/site/pagediadems.html>

L'affichage graphique des analyses est explicité dans le player avec un jeu de couleur en surimpression sur la forme d'onde et des marqueurs d'annotation manuelle. Voici un exemple de visualisation du résultat de l'algorithme « Music detection » fourni par le LaBRI dans le cadre de l'ANR DIADEMS :



Détection musique/non musique, menu déroulant, instance de Diadems

6.4 Annotation temporelle

Les utilisateurs ayant un compte peuvent faire des annotations temporelles dans les médias de leur choix. Cette fonctionnalité proposée par TimeSide a été conçue pour résoudre les problèmes d'indexation des données de terrain qui ne peuvent être segmentées car il faut parfois garder la continuité d'un événement rituel.

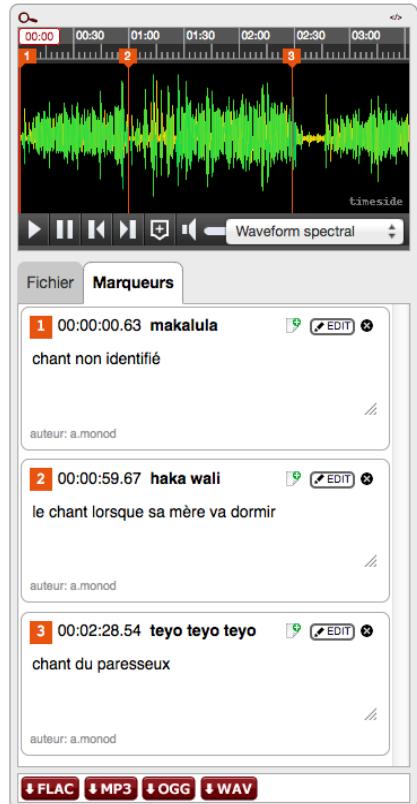
Par ailleurs, l'archiviste ne dispose pas toujours des informations culturelles et ethnographiques ou n'a pas les compétences linguistiques pour identifier et segmenter les séquences d'une archives de terrain. La plateforme permet alors de positionner des marqueurs sur la représentation graphique du son grâce au bouton **Marqueur** (drapeau avec une croix) placé à côté du haut-parleur, afin d'annoter le contenu sonore. L'utilisateur peut le poser “à la volée” pendant la lecture ou le placer (et le déplacer par un simple “cliquer-tirer”). L'annotation peut être faite au moment du chargement en ligne par l'archiviste ou plus tard par le chercheur qui retrouve progressivement les informations, ou par un utilisateur ayant les compétences culturelles ou musicales nécessaires.

Une fenêtre de couleur orange s'ouvre alors dans l'onglet “**Marqueurs**” comme ci-contre, avec la liste des marqueurs.

L'utilisateur peut saisir un titre, un commentaire, enregistrer ou supprimer son marqueur. Le Time Code s'affiche ainsi que le nom de son compte utilisateur afin de “signer” chaque marqueur.

A la lecture, un curseur rouge apparaît ainsi que le numéro du marqueur. Le titre et le commentaire apparaissent en surimpression pendant la lecture ou au passage de la souris.

- Le bouton “+” permet d'ajouter le marqueur à une « **Liste de lecture** ».
- Le bouton “**Edit**” permet de modifier son marqueur si on est l'auteur ou administrateur.
- Le bouton “x” permet de supprimer le marqueur et la numérotation des marqueurs est incrémentée automatiquement. Un marqueur peut être supprimé uniquement par son auteur ou par un administrateur.



Les marqueurs sont visibles tout public et plusieurs utilisateurs peuvent annoter le même document. Cependant il est rare que plusieurs utilisateurs annotent le même document.

6.5 Illustration des fiches avec une image ou un lien externe

Il est possible d'intégrer une ou plusieurs images pour illustrer une fiche à tout moment avec un compte utilisateur. Par ailleurs, des métadonnées sont parfois indiquées sur le support physique de l'archive et l'ajout d'une image consultable à la lecture du média est appréciable pour identifier son contenu.

Chaque niveau de catalogue peut donc être illustré par un document image ou par une autre ressource web (lien URL) grâce au menu “**Média associés**” du formulaire de saisie. Ce menu est accessible en mode “Edition” (Cf. bouton “**Editer**”).

Lors de la saisie, il est possible de décrire l'image ou la ressource externe avec **un titre, une description, les droits** et de télécharger un document image (pdf, jpg, png).

Les “**Médias associés**” sont accessibles en vignette de prévisualisation (images au format JPEG) ou en téléchargement (avec ou sans compte utilisateur).

- Exemple de fiche **Fonds** de Mireille Helffer :

Bureau Archives Géo-navigateur Recherche avancée Terminology Admin Français

Fonds : Fonds Mireille Helffer (1928-)

Média associés

Media

Biographie

Mireille Helffer, née le 6 janvier 1928, directeur de recherche honoraire au CNRS, chargée de mission au Musée national des Arts Asiatiques-Guimet, membre de l'UMR 9957 (CNRS/EPHE/Université Paris-Sorbonne) qui lui dirigeait jusqu'à sa retraite en 1996 la collection de musiques tibétaines et bouddhistes de l'Institut Guimet où elle a enseigné jusqu'en 1992. Spécialiste des musiques himalayennes, elle a effectué de nombreuses missions de terrain au Népal, au Ladakh, et parmi les communautés tibétaines réfugiées en Inde. Ses publications ont touché à des domaines aussi divers que les castes de musiciens au Népal, le chant épique tibétain (objet de sa thèse de doctorat, publiée en 1977), le répertoire liturgique des monastères du bouddhisme tibétain, les notations musicales tibétaines, les instruments de la musique tibétaine. Elle se consacre aujourd'hui à l'étude des traditions musicales du monastère mying-ma-pa de Zhe-chen au Népal.

Frequent visits to numerous Tibetan monasteries, both in India (Ladakh, Himachal Pradesh, Karnataka) and in Nepal, allow one to carry out a study of fieldwork practices covering two decades. After a period of "across the board" collection involving written material (musical scores and ritual writings) and sound (ten or so hours of recording) enabling a comparative study of the domain, the need was felt for a more in depth study of a particular aspect. Thus in this way, a closer involvement with a monastery of the mying-ma-pa tradition, Zhe-chen in Bodnath, Nepal came about. Here, the author stayed several weeks a year when the end of the year ritual celebration took place. Straightforward collection of musical material during short stays was replaced by attentive listening and a genuine involvement in complex rituals of which it became possible to grasp the structure and distinctive musical features; increased informality with the monks, the song master and the monastic authorities allowed for fruitful exchanges, some examples of which are cited here.

revue Les Cahiers d'Ethnomusicologie

Photo

Extrait de l'article : Laurent Aubert et Mireille Helffer, « « Piégée par les musiques d'Asie » : le besoin de comprendre », Cahiers d'ethnomusicologie [En ligne], 10 | 1997, mis en ligne le 06 janvier 2012, consulté le 16 avril 2012. URL : <http://ethnomusicologie.revues.org/904>

Crédits

Trân Quang Hai ; revue Les Cahiers d'Ethnomusicologie

Type mème

image/png

Télécharger

CNRSMH_Helffer.png

Bibliographie

Publications de Mireille Helffer disponibles à la bibliothèque Frédéric-Damourier_Nanterre



- Exemple de fiche **item** d'un disque Africa Vox (illustration avec la rondelle du disque) :

Bureau Archives Géo-navigateur Recherche avancée Terminology Admin Français

Item : Louange des chefs Songhay Disque05_FaceA_01

Informations sur la musique

Waveform spectral

Données d'archivage

Media

Marqueurs

Fichier	Marqueurs	
Propriété	Valeur	Unité
Channels	2	
Duration	0:03:15.12	s
Level Analyzer Max	0.0	dBFS
Level Analyzer RMS	-25.348	dBFS
Mean DC shift Analyzer	0.001	%
MIME type	audio/x-wav	
Resolution	24	bits
Samplerate	48000	Hz
Size	56197714	

Dublin Core

Supprimer

Telemeta

Usage des archives réservé dans le respect du patrimoine culturel des communautés d'origine.

Copyright © 2000 CRDM-CNRS
Mémoires Numériques



6.6 Réutilisation d'un média

La réutilisation d'un média est la partie la plus délicate des principes "FAIR²⁵" pour les données anthropologiques, qui sont très rarement en ligne. Les données du fonds en libre accès sont réutilisables de diverses manières.

6.6.1 Lien « i-frame »

Grâce au lien « **i-frame** » proposé dans le player de **TimeSide**, il est possible d'encapsuler le player et le média sélectionné dans un autre site web, sans avoir besoin de le télécharger. La plus-value scientifique et technique de ce type de réutilisation est triple :

- Le média est accessible en lecture et les marqueurs éventuels sont visibles,
- Le lecteur, ses fonctionnalités de navigation sont utilisables,
- L'accès à la documentation complète des archives est possible via le lien inclus dans le lecteur.

Ces avantages multiples pour l'utilisateur permettent d'éviter les téléchargements abusifs des sons. Il complète avantageusement le lien URL classique. Les administrateurs des archives gardent en outre la **traçabilité** de l'utilisation du média, grâce aux outils de statistiques, tel que Google Analytics ou Matomo d'Huma-Num).

Des exemples de réutilisation sont visibles sur le site "**A la naissance de l'ethnologie française**" :

The screenshot shows a detailed page from the website 'A la naissance de l'ethnologie française'. The top navigation bar includes links for ACCUEIL, MISSIONS, MÉTHODES, PERSONNES, CONTEXTE, ÉVÉNEMENTS, RESSOURCES, and A PROPOS. There are also buttons for Incription, Se connecter, and a search bar. The main content area features a map of West Africa with labels like Guinée, Mali, Burkina Faso, and Niger. A specific location in Mali is highlighted. The title of the page is 'Suite de chants de circoncision'. The left sidebar has sections for 'Contexte' and 'Contenu'. The 'Contenu' section contains a waveform visualization with numbered markers (1-10) and playback controls. Metadata includes: Auteur/Producteur : Schaeffner, André; Mission : Mission Dakar-Djibouti (1931-1933); Source : <http://archives.crem-cnrs.fr/archives/collections/1215/>; Institution de conservation : Centre de Recherche en Ethnomusicologie, LESC, Université Paris Nanterre. A red button on the right says 'Connectez-vous pour ajouter cette notice dans votre panier'. Below the content area is a large map of the Goundourou region in Mali. At the bottom of the page is a URL: <http://naissanceethnologie.fr/items/show/20351>.

Le site de l'ANR *Islamic Learning in Morocco*²⁶ utilise ainsi les enregistrements d'Anis Faraji déposés dans la plateforme Telemeta. Depuis le site de cette ANR ILM (2017-2020), 3000 utilisateurs en 2 ans ont accédé à ses enregistrements.

²⁵ FAIR : Facile à trouver, Accessible, Interopérable et Réutilisable

²⁶ ANR *Islamic Learning in Morocco* http://anrilm.cnrs.fr/presentation_anr_ilm/

Le plus grand nombre de connections par « *i-frame* » provient du site gabonais Dungandzi (qui a pour but de promouvoir la langue yipunu) avec 18 000 visites en 3 ans :
<http://bajag-mujabitsi.blogspot.com/>

Grâce à cette fonctionnalité, les sites web du monde de la recherche et le grand public via les réseaux sociaux utilisent largement ces archives audiovisuelles (30% des connexions), dans des proportions bien supérieures au référencement des plateformes de moissonnage (Isidore ou Europeana) qui représentent entre 1 à 2 % des connexions. Les connexions en provenance des réseaux sociaux, très majoritaires, se répartissent ainsi : 65% sont des blogs et 33% sont des pages Facebook.

6.6.2 Livre électronique (export Epub)

Cette fonctionnalité permet d'exporter toute une collection sous forme de livre électronique (e book) afin de consulter les contenus sur une tablette ou liseuse. Le bouton « Epub » est situé en haut à droite de la fiche **Collection**. Cette fonction a été développée pour le livre de Michèle Castellengo afin d'écouter les exemples sonores pendant la lecture : « Écoute musicale et acoustique, avec 420 sons et leurs sonagrammes décryptés²⁷ ». Les exemples sonores sont téléchargeables depuis la plateforme Telemeta du LAM. Cette fonctionnalité n'est pas encore exploitée à notre connaissance pour les données gérées par le CREM.

7 Divers

- Import automatique de média

- Utilisation d'un accès SFTP fourni par Guillaume Pellerin ou Huma-Num (mot de passe),
- Versement sur le serveur via l'application Cyberduck dans la branche Telemeta, dossier Simonnot/Import/new.
- Un dossier par collection nommé avec la cote collection, contenant les sons nommés avec exactement le même nom que les fiches items (cote items) est chargé sur le serveur.
- Si nombre de sons > nombre de fiches : création automatique des fiches items (avec le son mais sans métadonnées).
- Vérifier l'absence d'espace dans les cotes et la conformité des cotes avec les médias

- Consigne pour la gestion de la géolocalisation avec GeoEthno

Les recommandations de la Société Samalyse concernant les informations de lieux étaient les suivantes en 2009 :

« Dans la base 4D, les informations géographiques étaient contenues dans la table des items. Les anciens champs **Continent**, **État** et **Région_Village** indiquaient la zone de couverture de l'item, les deux premiers de ces champs étant des énumérations dans 4D. De façon à moderniser la gestion des informations géographiques, qui sont cruciales dans la base de données du CREM, et à permettre des évolutions intéressantes à plus ou moins long terme (géolocalisation, calcul de distance, représentation sur carte, etc.) il fut préconisé d'utiliser un thesaurus géographique. De cette façon, un seul champ "Lieu" est nécessaire dans la table des items, ce champ faisant référence à un élément du thésaurus, qui est une arborescence de lieux. Le thésaurus choisi pour ce faire est **GeoEthno**, maintenu par le **LESC** à la Maison de l'Archéologie et de l'Ethnologie de Nanterre, dont dépend le CREM. **GeoEthno** est disponible dans un format de données XML.

Ces tables incluent également des indicateurs booléens (*is_authoritative*) qui devraient permettre aux

²⁷ <https://www.editions-eyrolles.com/Livre/9782212138726/ecoute-musicale-et-acoustique>

utilisateurs, lors de la conception de la nouvelle application du CREM, d'entrer de nouveaux noms de lieux, s'ils n'existent pas dans le thésaurus GeoEthno. Puis, périodiquement, les responsables du CREM et de GeoEthno pourront se réunir pour décider lesquels de ces nouveaux lieux peuvent être intégrés officiellement à GeoEthno. Grâce aux dits indicateurs booléens, il sera possible de retrouver les lieux créés au CREM, et qui ne font pas encore partie de GeoEthno.

Il faut aussi noter que l'emploi d'un thesaurus préviendra les erreurs de saisie (comme placer un pays dans un mauvais continent, un village dans une mauvaise région, etc..) et permettra la normalisation des régions et villages. En outre, le thesaurus, qui devra être réimporté régulièrement dans la base du CREM, est censé suivre les éventuels changements de nom survenant pour des raisons politiques ou autres. Lors de la migration de la base, il sera simple de déterminer le nom de lieu (terme descripteur GeoEthno) des pays.

Pour ce qui est des régions et villages (ancien champ), il est difficile de déterminer le terme descripteur GeoEthno, et il est préconisé de migrer cette information de manière brute dans le champ "Précision lieu". Cependant, une tentative de mise en correspondance du champ Region_Village de la table Phono sera effectuée, et considérée avoir réussi si le terme exact est trouvé une seule fois dans GeoEthno. Par la suite, et au fur et à mesure du maintien de la nouvelle base, des termes descripteurs GeoEthno plus précis pourront être sélectionnés par les documentalistes, les informations du champ "Précisions lieu" devenant peu à peu inutiles. » (Cf. O. Guilyardi, Docref-1.1-stable, sur Mycore)

- **Modification des contenus textuels** (page d'accueil, aide, mentions légales...)

Les changements des contenus textuels (page d'accueil, aide, mentions légales...) se font dans Github, dans la branche CREM, version de test, menu "pages" :

<https://github.com/Parisson/Telemeta-pages/blob/crem/fr/home.rst>

La nouvelle version est contrôlée dans l'instance de test (contenu et mise en page) avant d'être intégrée dans la version de production.

- **Questions juridiques et éthiques**

Dans la grande majorité des cas, il n'y a pas d'auteur ou de compositeur déclarés ou connus en musique traditionnelle. Parfois les noms des interprètes sont connus, mais les personnes sont difficiles à joindre. Nous référant aux écrits de la juriste Anne-Laure Sterin²⁸ qui définissent le cadre éthique et juridique pour les archives sonores, les enregistrements antérieurs à 1963 sont librement consultables, sauf cas particuliers pour des questions d'éthiques. Après 1963, la majorité des déposants ont besoin d'un embargo sur leurs données de terrains avec une restriction d'accès pendant 70 ans après sa collecte. Mais si leur contenu n'a pas de caractère secret ou « sensible » pour les communautés d'origine, et que le collecteur souhaite diffuser ses enregistrements postérieurs à 1963, nous considérons que c'est la personne la mieux placée pour respecter la volonté des interprètes et de leurs descendants. Quand la mise en ligne est libre d'accès (sans mot de passe), les témoignages des communautés sont nombreux et enthousiastes, ce qui nous encourage à poursuivre dans cette voie. En cas d'avis contraire, la diffusion en ligne peut être arrêtée à tout moment, ce qui n'est pas possible dans le cas d'un disque. En l'absence d'exploitation commerciale et donc d'éventuel préjudice financier pour les ayants droits, les risques juridiques sont inexistant pour l'institution selon les juristes spécialisés. Par ailleurs, les fichiers haute définition ne sont jamais accessibles, contrairement à un CD, ce qui limite les exploitations commerciales.

²⁸ <http://ethiquedroit.hypotheses.org/1135>

8 Conclusion

En l'absence d'outil web adapté dans nos institutions et de moyens informatiques en interne, la collaboration avec Parisson puis l'IRCAM a été une opportunité inespérée pour valoriser un fonds d'archives sonores d'une grande richesse mais peu accessible²⁹. Il faut rendre hommage également aux nombreux collaborateurs et contributeurs qui ont participé à ce projet de manière progressive (ou “agile”), parfois bénévolement, motivés par sa dynamique innovante. Il faut souligner également l'impulsion donnée par le TGE Adonis³⁰ (devenu Huma-Num) et le soutien du Ministère de la Culture.

La plateforme collaborative Telemeta correspond aux besoins définis en 2009 par l'équipe du CREM, en adéquation avec les moyens financiers d'une petite structure de recherche, quant à la gestion, la diffusion et l'annotation de données audio et vidéo. Cette interface multi-utilisateur respecte le modèle de données spécifique à l'ethnomusicologie et les besoins particuliers en matière de gestion des accès. La gestion de l'embargo sur les données encourage de nombreux chercheurs à déposer leurs collectes de terrain, même si ces données ne peuvent être consultables tous publics, ce qui est souvent le cas.

Le modèle de données, défini et enrichi progressivement depuis la création de l'équipe du Musée de l'Homme, suit les besoins de la discipline. Il pourrait être simplifié (certains champs sont peu utilisés), les listes de vocabulaires peuvent être améliorées en interne et un petit nombre d'entre elles peuvent être externalisées, en examinant la balance bénéfices/constraintes de dépendre de thésaurus externes.

Les performances de l'outil sont satisfaisantes compte tenu du volume des données en ligne. La prise en main par les utilisateurs est effective, spécialistes ou public large (en moyenne 350 000 pages vues par an, 48 000 visites par an). La vitesse d'accès est limitée par la vitesse du serveur Huma-Num (le nouveau système de stockage Active Circle n'est pas conçu pour un accès ultra-rapide) mais cette lenteur relative est compensée par l'offre de service de l'infrastructure (hébergement, sauvegarde, maintenance machine...). La vitesse d'accès est également ralentie actuellement par le module ElasticSearch-Haystack qui nécessiterait une mise à jour.

La base de données n'est plus réservée à quelques archivistes initiés, l'enrichissement de la base est constant depuis 2011 (53 700 médias à ce jour) et de nombreux projets de recherche innovants ont vu le jour. Cette disponibilité et cette flexibilité des données de recherche dans une interface ergonomique ainsi que la réutilisation des médias représente une preuve de concept qui pourrait être utile en ce qui concerne l'équilibre entre l'ergonomie et le degré de complexité des fonctionnalités attendues pour une base de données audiovisuelles.

L'outil n'a pas de vocation éditoriale mais se veut surtout utile à l'archivage et donc utilisable par les collecteurs et les utilisateurs de ces archives. Des améliorations sont possibles (licence libre) et dépendent des moyens financiers et humains disponibles. Depuis 2017, des propositions de modularisation afin de mutualiser les fonctionnalités communes aux différentes instances, ainsi que l'autonomie du framework Timeside ont été proposées mais restent en attente. Aucune solution technique présentant l'ensemble des fonctionnalités de cette outil n'est proposé à ce jour par Huma-Num et les spécialistes du domaine ne sont pas favorables à un outil en langage PHP. En outre, la solution d'hébergement du CNRS (Nakala) ne propose pas la gestion d'un embargo sur les médias.

²⁹ Cf. Article du Bulletin de l'AFAS : <https://journals.openedition.org/afas/4056>

³⁰ Très Grand Équipement Adonis, INSHS, CNRS, 2007-2010

Le projet Telemeta/TimeSide a suscité d'autres plateformes en France et à l'étranger et a contribué à faire connaître la richesse de ce patrimoine scientifique et culturel du CNRS, ainsi que l'excellence des laboratoires de recherche dans le domaine des Data Sciences.

Bibliographie

- Simonnot, Joséphine, 2008, Telemeta : un projet pour les archives sonores de la recherche. Traitement, valorisation, diffusion, en ligne : <http://www.imageson.org/document1045.html>, mise en ligne le 31 octobre 2008. - Simonnot, Joséphine, 2008, Telemeta : un projet d'application collaboratif dédiée à l'archivage des documents sonores, La lettre d'information du programme ARSHS, http://mshdijon.u-bourgogne.fr/msh_cnrs/arshs/Lettre_arshs_2.html, mise en ligne en novembre 2008.
- Simonnot, Joséphine, 2011, Telemeta, un projet Web pour les archives sonores de la recherche, *Lettre de l'INSHS*, en ligne : <http://www.cnrs.fr/inshs/recherche/telemeta2.htm>, mise en ligne le 24 février 2011
- FillonThomas, Simonnot Joséphine , Mifune Marie-France , Khoury Stéphanie , Pellerin Guillaume, Lecoz Maxime , Amy de La Bretèque Estelle , Doukhan David , Fourer Dominique , Rouas Jean-Luc, Pinquier Julien , Mauclair Julie and Barras Claude , 2014. "Telemeta: An open-source web framework for ethnomusicological audio archives management and automatic analysis", *working paper for the Digital Library for Musicology workshop*, London, Septembre 2014
- Simonnot Joséphine, Khoury Stéphanie, "Applications and implications of digital audio databases for the field of ethnomusicology: A discussion of the CNRS — Musée de l'Homme sound archives", *First Monday*, Volume 19, Number 10 - 6 octobre 2014
- <http://firstmonday.org/ojs/index.php/fm/article/view/5554/4133>
- SIMONNOT, Joséphine, "The CNRS - Musée de l'Homme Sound Archives from 1900 to the present: a long way between heritage, knowledge and technologies", in ZIEGLER, Susanne, Ingrid AKESSON et LECHLEITNER, Gerda (éds), *Historical Sources of ethnomusicology in contemporary debate*, Cambridge Scholars Publishing (Cambridge, s.n.): 103-116,2017 en ligne : <http://www.cambridgesscholars.com/historical-sources-ofethnomusicology-in-contemporary-debate>
- SPIELMANN, Florabelle, HELMLINGER Aurélie, FILLON Thomas, PELLERIN Guillaume, STURM Bob L., BEN-TAL Oded, QUINTON Elio et SIMONNOT Joséphine 2017, Zoom arrière. L'ethnomusicologie à l'ère du Big Data, *Cahiers d'ethnomusicologie. Anciennement Cahiers de musiques traditionnelles*, 30 : 9-28.
- Simonnot Joséphine, « Partager les archives sonores du musée de l'Homme sur le web avec la plateforme Telemeta », *Bulletin de l'AFAS* [En ligne], 46 | 2020, mis en ligne le 14 mars 2020, consulté le 07 mai 2020. URL : <http://journals.openedition.org/afas/4056>
- Marie-France Mifune et Florabelle Spielmann, « Les archives sonores ethnomusicologiques du CNRS-musée de l'Homme. Quelles implications dans des projets d'indexation et d'analyse du contenu audio», *Bulletin de l'AFAS* [En ligne], 46 | 2020, mis en ligne le 14 mars 2020, consulté le 07 mai 2020. URL : <http://journals.openedition.org/afas/4070> ; DOI : <https://doi.org/10.4000/afas.4070>