

Telemeta : un système libre de gestion d'archives musicales orienté web

article pour la Revue Francophone d'Informatique Musicale
version 0.8 - 19 janvier 2012

Auteurs :

Guillaume Pellerin : directeur R&D, Parisson SARL
Joséphine Simonnot : ingénieur de recherche CNRS, CREM



Table des matières

1	Introduction	2
1.1	Historique	2
1.2	Objectifs	2
1.3	Partenariats	3
1.4	Fonctionnalités	4
1.5	Licence	4
2	Technologies et philosophie	4
2.1	Les fondamentaux du logiciel libre	4
2.2	Standards et normes	4
2.2.1	Web, navigateurs	4
2.2.2	Audio	5
2.2.3	Métadonnées	5
2.3	Briques principales	5
3	Architecture serveur	5
3.1	Modèle de données	6
3.2	Droits d'accès	6
3.3	Traitement audio	6
3.4	Moissonnage et sauvegarde	7
3.5	Installation	7
4	Interface web	7
4.1	Espace utilisateur	8
4.2	Edition	8
4.3	Moteur de recherche	8
4.4	Lecture et analyse audio	8
4.5	Media associés	8
4.6	Administration	8
5	Développement	9
5.1	Contexte	9
5.2	Modélisation	9
5.3	Controlleurs	9
5.4	Vues	10
5.5	Feuille de route	10
6	Conclusion et perspectives	10

Résumé

1 Introduction

Telemeta est une solution libre d'archivage et d'indexation audio basée sur une interface de type web. Elle offre des fonctions avancées d'édition, d'indexation, de transcodage, de publication et de sauvegarde d'archives audio et de leurs méta-données. Il donne accès aux ressources distantes publiées ou privatisées issues de collections numérisées (fichiers, CD audio, vinyl ou bandes magnétiques) par l'intermédiaire d'un simple navigateur ou d'une application dédiée en accord avec les standards du web.

1.1 Historique

Initié en 2007 lorsque le besoin en système d'archivage pérenne du CREM et les développements en cours de Parisson se sont rencontrés, le projet Telemeta a proposé plusieurs objectifs :

- valoriser le fonds d'archives en faisant converger les archives sonores numérisées et leurs méta-données vers une plateforme de type web,
- concevoir et produire une application orientée recherche pour la gestion collaborative des archives,
- définir le processus de migration de la base de données du CREM mais aussi de l'importation des fichiers numériques existants.

L'assistance à la maîtrise ouvrage fournie par Parisson pour la conception puis le développement de l'application a permis de livrer en 2009 un prototype logiciel qui a trouvé un écho favorable dans la communauté de chercheurs et des développeurs. En juin 2011, après une phase d'optimisation, Parisson a opéré la migration et l'importation finale de la base de données structurée du CREM (méta-données, ontologies géographiques ou thésaurus d'instruments) et des fichiers numérisés. Depuis, les chercheurs et les documentalistes participent de manière collaborative à l'amélioration de l'indexation de plus 35000 archives.

Le code du logiciel "Telemeta" dans sa version 1.0 est désormais publié et les archives du CREM disponibles (<http://archives.crem-cnrs.fr>).

Chronologie :

- 2007 : rencontre du CREM avec Parisson et le LAM, contour des besoins mutuels, soutiens du projet Anthroponet
- 2008 : définition du cahier des charges, choix du framework, modélisation, migrateur 4D > MySQL
- 2009 : prototypage (v0.5), recherche de partenaires
- 2010 : corrections, développement agile (v0.9), soutien du TGE Adonis
- 2011 : migration définitive, v1.0 déployée sur serveur, mise en production

1.2 Objectifs

D'un point de vue fonctionnel :

- Pérenniser les archives audio numériques (logiciels et formats)
- Valoriser le patrimoine culturel par la consultation légale

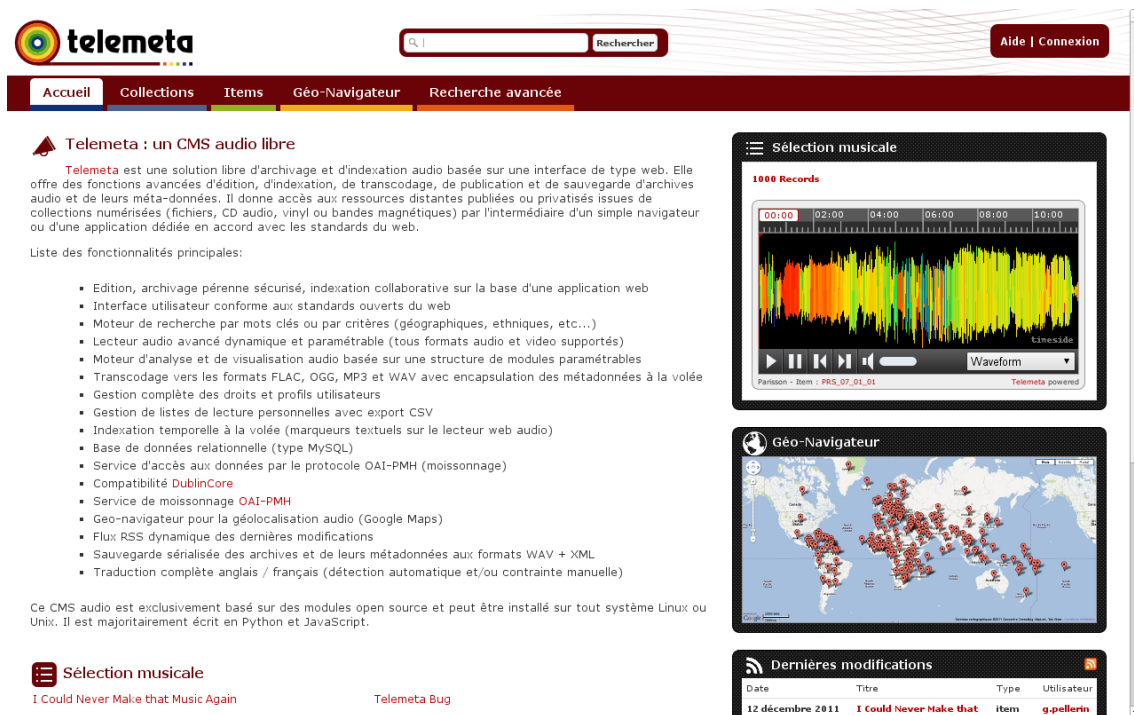


FIGURE 1: Page d'accueil de Telemeta

- Faciliter et optimiser la transmission des méta-données par l'intermédiaire de protocoles et de standards ouverts
- Augmenter les capacités de recherche et de travail collaboratif (saisie en ligne, géo-localisation, web sémantique, interopérabilité, statistiques)
- Définir une ergonomie d'indexation et de publication collaborative, i.e. aussi un principe de flux de travail dans un contexte de sauvegarde pérenne d'oeuvres audio-visuelles.

D'un point de vue technique :

- architecture "fluide" : streaming, SOD
- langage objet, déploiement
- standards (W3C, open source, ressources JavaScript, HTML5)

1.3 Partenariats

Les partenaires participants au projet :

- CREM : Centre de Recherche en Ethnomusicologie du Laboratoire d'Ethnologie et de Sociologie Comparée (LESC), UMR 7186
 - LAM : équipe Lutheries, Acoustique et Musique de l'Institut Jean le Rond d'Alembert (IJLRA), UMR 7190
 - MAE : Médiathèque Eric-de-Dampierre de la MAE, Nanterre
 - MNHN : Museum National d'Histoire Naturelle
 - IRI : Institut de Recherche et d'Innovation
 - MuCEM : Musée des Civilisations de l'Europe de la Méditerranée
 - MMSH : Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme
 - TGE Adonis : Très Grand Equipement pour les sciences humaines et sociales du CNRS
- exemple : CREM

1.4 Fonctionnalités

Liste des fonctionnalités principales :

- Edition, archivage pérenne sécurisé, indexation collaborative sur la base d'une application web
- Interface utilisateur conforme aux standards ouverts du web
- Moteur de recherche par mots clés ou par critères (géographiques, ethniques, etc...)
- Lecteur audio avancé dynamique et paramétrable. Tous formats audio et video supportés.
- Moteur d'analyse et de visualisation audio basée sur une structure de modules paramétrables
- Transcodage vers les formats FLAC, OGG, MP3 et WAV avec encapsulation des métadonnées à la volée
- Gestion complète des droits et profils utilisateurs
- Gestion de listes de lecture personnelles avec export CSV
- Indexation temporelle à la volée (marqueurs textuels sur le lecteur web audio)
- Base de données relationnelle (type MySQL)
- Service d'accès aux données par le protocole OAI-PMH (moissonnage)
- Compatibilité DublinCore
- Geo-navigateur pour la géolocalisation audio (Google Maps)
- Flux RSS dynamique des dernières modifications
- Sauvegarde sérialisée des archives et de leurs métadonnées aux formats WAV + XML
- Flux RSS dynamique des dernières modifications
- Traduction complète anglais / français (détection automatique et/ou contrainte manuelle)

Telemeta incorpore ainsi un ensemble de fonctions spécifiques au travail des archives temporelles comme la musique ou la voix parlée / chantée, l'interface ayant été pensée par et pour les chercheurs. Les services de transposition des méta-données au format DublinCore à travers le protocole OAI-PMH permet de proposer l'ensemble des méta-données aux services de moissonnage institutionnels tels que Isidore créé par le TGE Adonis pour les sciences humaines (<http://www.rechercheisidore.fr>).

1.5 Licence

Telemeta est publié selon les termes de la licence libre CeCILL¹ conforme au droit français et publiée par le CNRS, l'INRIA et le CEA.

2 Technologies et philosophie

2.1 Les fondamentaux du logiciel libre

Pérenniser les ressources informatiques, copies à grandes échelles, réutilisation des innovations
Dynamiser le développement (partage, communautés internationales)
Limiter les coûts de déploiement à grande échelle
sécurité

2.2 Standards et normes

Contraintes

2.2.1 Web, navigateurs

- HTML5 : langage hypertextuel avec balises <audio> <video>
- CSS : feuilles de styles
- JavaScript : langage côté navigateur (interfaces et lecteur dynamique)

1. <http://www.cecill.info>

2.2.2 Audio

Alors que l'application est capable de décoder tous les formats audio-visuels connus grâce à la librairie GStreamer², nous avons fait le choix de certains formats d'export pour l'encodage audio. Alors que, par défaut, la lecture en ligne s'effectue en streaming au format MP3 ou OGG Vorbis selon la disponibilité du navigateur, certains voudront télécharger un fichier FLAC compressé sans pertes encapsulant les méta-données DublinCore ou encore au format d'origine comme le WAV.

2.2.3 Métadonnées

- DublinCore (OAI-PMH)
- SQL : base de données
- OWL : Web Ontology Language

2.3 Briques principales

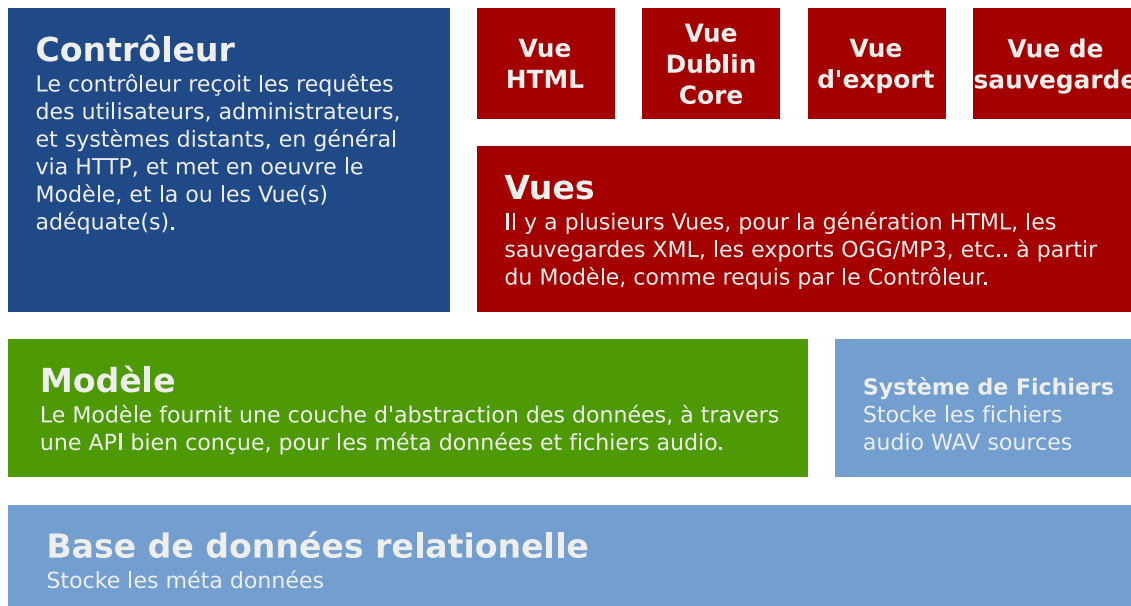
Telemeta utilise de nombreuses briques logicielles libres, éprouvées et conformes aux standards de la programmation scientifique :

- Linux : noyau serveur conseillé
- Apache : serveur web frontal (+ module WSGI)
- Python : langage interprété orienté objet
- Django : framework web pour Python
- MySQL : base de données relationnelle
- Scipy / Numpy : modules d'analyse mathématique en Python
- TimeSide³ : traitement des signaux audio, lecteur audio web dynamique

3 Architecture serveur

côté serveur

L'architecture de Telemeta

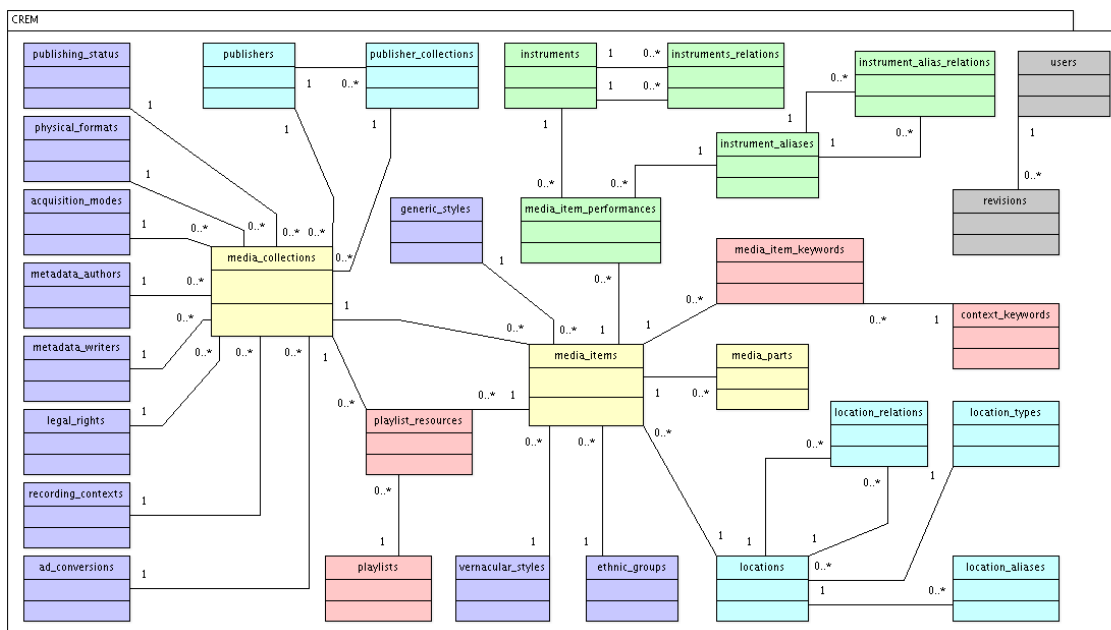


2. <http://gstreamer.freedesktop.org>

3. <http://code.google.com/p/timeside>

3.1 Modèle de données

Modèle de données du CREM



3.2 Droits d'accès

Droits d'accès

- **Utilisateurs** : profils, droits, gestion des mots de passe
- **Groupe**s : administrateur, documentaliste, chercheur, membre, anonyme
- **Droits** : ajouter, supprimer ou modifier les objets selon les règles
- Pour tous les utilisateurs : **listes de lecture** personnelle, ajout de **marqueurs temporels**
- **Droits particuliers** pour la lecture audio (date glissante de 50 ans, paramètre “public access”)

Détail des droits des groupes

- **Anonyme** : parcours du site, lecture selon les autorisations de chaque objet
- **Membre** : ajout de listes de lectures personnelles, ajout de marqueurs
- **Chercheur** : ajout et edition des fiches documentaires
- **Documentaliste** : suppression d'objets, tous droits sur la base de données (hors utilisateurs), lecture audio de tous les items
- **Administrateur** : ajouter des utilisateurs

- + droits spéciaux au cas par cas :

- Lecture audio de tous les items
- Téléchargement de tous les items

3.3 Traitement audio

TimeSide : web audio components

Pour l'analyse, le transcodage et la lecture des signaux, le projet connexe TimeSide a été créé par Parisson et d'autres développeurs externes. Cette librairie est dotée d'une architecture de processeurs de signaux temporels donnant accès à un ensemble de modules de calculs. Ces modules - écrits entièrement en Python grâce aux fonctions avancées du projet de calcul scientifique Numpy - font référence à une API simple et publiée. L'accès aux flux audio, aux flux de méta-données et aux ressources de calcul est définie par une API simple et publiée. Une couche graphique dynamique

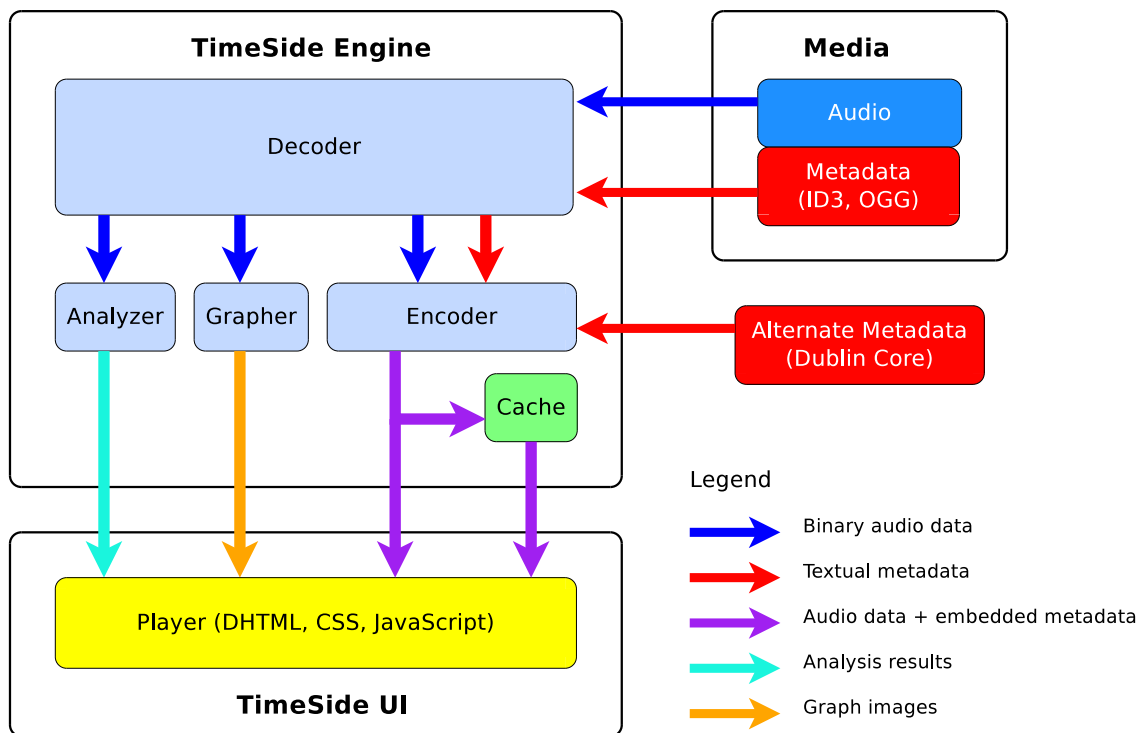


FIGURE 2: Schéma fonctionnel de la librairie TimeSide

- écrite en AJAX / CSS / Javascript y est également disponible pour la lecture et l'interfaçage du navigateur, côté client, pour l'affichage des résultats d'analyse et d'autres fonctions avancées comme l'indexation temporelle. La liste des modules disponibles peut être agrémentée aisément de sorte que les plateformes externes, comme Telemeta, disposent automatiquement des nouveaux outils publiés. A ce jour, TimeSide est utilisé par Telemeta pour calculer les images (formes d'onde, spectrogramme, etc), les propriétés acoustiques et informatiques globales et moyennées des médias sonores, le décodage et l'encodage des flux audio. Pour plus d'informations : <http://code.google.com/p/timeside/>

3.4 Moissonnage et sauvegarde

Moissonnage et sauvegarde

- mapping Dublin Core
- serveur OAI-PMH intégré
- flux RSS (revisions)
- sauvegarde sérialisée : XML + WAV
- upload accumulatif

3.5 Installation

voir INSTALL.rst

4 Interface web

côté client

4.1 Espace utilisateur

The screenshot displays the CREM (Centre de Recherche en Ethnomusicologie) user interface. At the top, there's a header with the CREM logo, a search bar, and user information: 'Bienvenue, Guillaume Pellerin | Profil | Aide | Déconnexion'. Below the header is a navigation bar with tabs: Accueil, Collections, Items, Géo-Navigateur, Recherche avancée, Utilisateurs, and Admin. The main content area is titled 'Listes de lecture' (Reading Lists) and includes an 'Ajouter' (Add) button. It is divided into two sections: 'Sandbox' and 'Tests'. The 'Sandbox' section shows a table with columns: Titre, Type, Cote, Collecteur, Période d'enregistrement, Sonore, and Action. It contains one item: 'MUSIQUE RITUELLE :04' (item, 1195, 1959). The 'Tests' section shows a table with the same columns, containing multiple items including 'CNRSMH_I_1961_001_01-01', 'ROUMANIE : POLYPHONIE VOCALE DES AROUMAINS', 'SALUTATIONS AUX DIEUX :A01', 'GABON ; SALLEE, P. 1970', 'SOLLISTE, CHOEUR, NKOOKO, TAMBOURS, HOCHET, SIFFLET (3/14)', 'MUSIQUE DE GYMNASE ZUR-XANEN :01-01', 'GABON ; SALLEE, P. 1965', 'GABON ; SALLEE, P. 1978', 'CHANT ET SAAMBI', 'SAAMBI ET CHANT (1/3)', 'INDONESIE : CD TORAJA. FUNERAILLES ET FETES DE FECONDITE', 'MVET :59-01', 'We ready now _12', and 'J'ouvert Barrio _01'. On the right side, there is a 'Dernières modifications' (Recent Modifications) section showing a list of changes with columns: Date, Titre, Type, and Utilisateur. It lists several modifications by users like j.simonnot and a.julien.

FIGURE 3: Espace de travail personnel : listes de lectures, suivi des dernières modifications

4.2 Edition

formulaires générés automatiquement selon le modèle de données

4.3 Moteur de recherche

par mots clés
avancée

4.4 Lecture et analyse audio

formats standards orientés web
TimeSide UI : lecteur audio dynamique

- Lecture audio
- Affichage audio
- Indexation temporelle
- Portabilité
- Principes du modèle de développement
- code.google.com/p/timeside/wiki/UiGuide

4.5 Media associés

4.6 Administration

- Collections, Items
- Enumérations ou thésaurus

The screenshot shows the Telemeta search interface. At the top, there is a logo for CREM (Centre de Recherche en Ethnomusicologie) and a search bar with the text "Recherche". To the right of the search bar, there is a user profile section with the text "Bienvenue, Guillaume Pellerin | Profil | Aide | Déconnexion". Below the search bar, there is a navigation menu with the following items: Accueil, Collections, Items, Géo-Navigateur, Recherche avancée (highlighted), Utilisateurs, and Admin. The main content area is titled "Recherche avancée" and contains a form with several fields: LIEU, POPULATION / GROUPE SOCIAL, TITRE, DÉPOSANT / CONTRIBUTEUR, COLLECTEUR, ANNÉE D'ENREGISTREMENT, ANNÉE DE PUBLICATION, and SONORE. The LIEU field is currently selected, and a dropdown menu is open, showing a list of geographical locations with the number of items associated with each: Gabon (1612 items), Gambela (1 item), Gambie (10 items), Gandaki (6 items), Gangtok (6 items), Garoua (8 items), Gascogne (1 item), Gaule (106 items), and Gaziantep (10 items). The dropdown menu is currently showing "Gabon (1612 items)".

FIGURE 4: Exemple de requête de recherche avancée avec complétion dynamique des termes géographiques

- Ontologie ethnographique
- Ontologie instrumentale
- Compositions
- Mots clés
- Marqueurs temporels

5 Développement

5.1 Contexte

Plateforme de développement : telemeta.org

Présentation, installation, documentation, Blog, tickets, bugs, Mailing list, Wiki utilisation de [Git](https://git-scm.com/) (versionnement décentralisé)

dépôts : vcs.parisson.com, github, vous !

```
git clone http://vcs.parisson.com/git/telemeta.git
```

5.2 Modélisation

Voir <http://github.com/yomguy/Telemeta.git/telemeta/models/>

5.3 Contrôleurs

Voir web/ et url.py

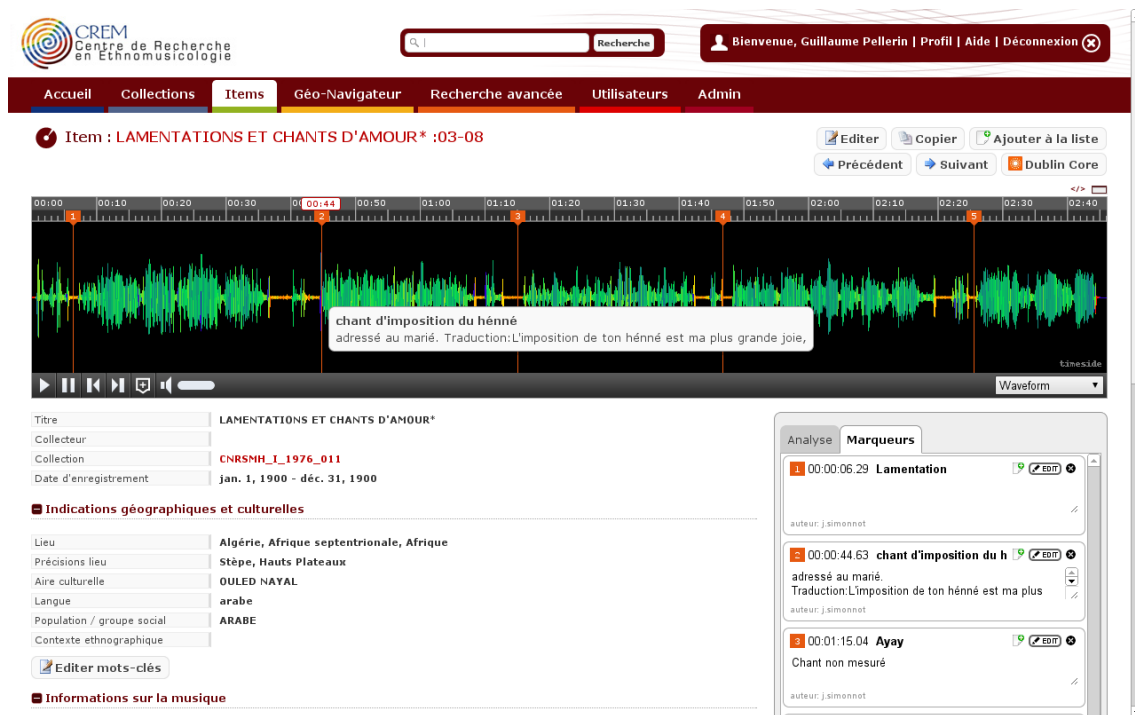


FIGURE 5: Exemple d'un item en cours de lecture avec affichage dynamique des contenus des marqueurs

5.4 Vues

Voir templates/telemeta_default/

5.5 Feuille de route

Objectifs pour Telemeta 1.x :


- Compatibilité HTML5 totale du lecteur audio
- Accès spéciaux par adresses IP
- Intégration et MAJ des **ontologies** (langues, géographie)

Objectifs pour Telemeta 2.x :


- **Modèle générique** de données
- **Fonctions d'analyse** augmentée (reconnaissance, recoupement statistique) : ANR CONTINT 2011 **DIADEMS** et ANR CORPUS 2011 **DicA2Ref**
- **Déploiement mutualisé** (CNRS IN2P3)
- Définition d'une **API**

6 Conclusion et perspectives


- Technologie et ergonomie prometteuse pour la **sauvegarde et la valorisation** du patrimoine musical
- **Déploiement** et **pérennité** optimisés avec les briques open source
- **Intégration souple** de données métiers hétérogènes (sciences humaines et sciences informatiques)
- **Plateforme collaborative** à un niveau international
- Exemple au CREM du 18/05 au 25/06 : plus de **450** fiches en moyenne éditées par semaines !
- Plateforme de **développement** ouverte



CREM
 Centre de Recherche
 en Ethnomusicologie


 Bienvenue, Guillaume Pellerin | [Profil](#) | [Aide](#) | [Déconnexion](#)

[Accueil](#)
[Collections](#)
[Items](#)
[Géo-Navigateur](#)
[Recherche avancée](#)
[Utilisateurs](#)
[Admin](#)


Administration

- Administration générale
- Utilisateurs
- Énumérations
- Instruments

Administration de Django
Bienvenue, Guillaume. [Modifier votre mot de passe](#) / [Déconnexion](#)

Administration du site

Auth	
Groupe	Ajouter Modifier
Utilisateurs	Ajouter Modifier
Sites	
Sites	Ajouter Modifier
Telemeta	
Instrument alias relations	Ajouter Modifier
Instrument aliases	Ajouter Modifier
Instrument relations	Ajouter Modifier
Instruments	Ajouter Modifier
Langues	Ajouter Modifier
Lieux	Ajouter Modifier
Location aliases	Ajouter Modifier
Location relations	Ajouter Modifier
Media collections	Ajouter Modifier
Media item relateds	Ajouter Modifier
Media items	Ajouter Modifier
Media parts	Ajouter Modifier
Revisions	Ajouter Modifier

Actions récentes
Mes actions

- [D. Rappoport](#) Utilisateur
- [Documentaliste](#) Groupe
- [test](#) Utilisateur
- [test](#) Utilisateur
- [Membre](#) Groupe
- [Documentaliste](#) Groupe
- [Chercheur](#) Groupe
- [Administrateur](#) Groupe
- [r.zaccarelli](#) Utilisateur
- [r.zaccarelli](#) Utilisateur

FIGURE 6: Page d'accueil de l'administration générale du site



<http://telemeta.org>

Telemeta 1.3 “Bell”

<http://pypi.python.org/packages/source/T/Telemeta/Telemeta-1.3.tar.gz>

Ce document est mise à disposition selon un [contrat Creative Commons](#)