

# Sound Archives, Metadata and New Web Technologies

JOSEPHINE SIMONNOT

*Centre de Recherche en Ethnomusicologie (CREM)  
Université Paris 10, France*

## INTRODUCTION

**I**n this paper, I present my experience to preserve, manage and disseminate a cultural heritage: Les Archives Sonores du CNRS-Musée de l'Homme, France, because, in order to keep alive this common world heritage, it must not disappear, and it must be accessible by the specialists and also shared with a large community of users to be enriched for the future generations.

Those archives are supported by the French National Research Center (CNRS), the National Museum of Natural History, the Paris 10 University and the Culture Ministry. Its history follows many steps: it began with the Trocadéro Ethnographic Museum and goes on with the Musée de l'Homme in 1937, under the direction of André Schaeffner. The Ethnomusicology Department, with his successor Gilbert Rouget, established a research team in 1968, in order to improve the research on music and sound. The aim was

also to go on the dissemination with the discographic publication.<sup>1</sup> The other main objective was to preserve this traditional heritage for the next generations. Since the beginning, Gilbert Rouget was convinced that it's important to disseminate this musical diversity and share the knowledge of this world. Publishing books was not enough to understand the musical phenomenon.

To illustrate the history and diversity of the collections deposited during one century, we listened the oldest recording of our sound archives from Mexico wa cylinder around 1932, a woman song: [http://archives.crem-cnrs.fr/archives/items/CNRSMH\\_I\\_1939\\_001\\_002\\_01/](http://archives.crem-cnrs.fr/archives/items/CNRSMH_I_1939_001_002_01/). And a young Mexican Lacandon boy song (“levent”) recorded on the first magnetic tape of our Mexican collections before 1953: [http://archives.crem-cnrs.fr/archives/items/CNRSMH\\_I\\_1953\\_005\\_001\\_03/](http://archives.crem-cnrs.fr/archives/items/CNRSMH_I_1953_005_001_03/).

The collections content about 5000 published records of traditional music and fields recordings (grabaciones de campo). Today, the database contents more than 60 000 items to inform the various recordings (42 000 from fields recordings).

For many years, we started a long process of digitization: first on CD-R, later on hard disc and now on the server of the National Infrastructure Huma-num, the service for humanities research and cultural data in France. <http://www.huma-num.fr>.

The Bibliothèque Nationale of France (BnF) continues the digitisation and we hope to finish it in 4 or 5 years. It consumes a lot of time because it is necessary to segment the sound files to have an easy access to the content. The treatment of the metadata is also a long process. For example, we can see on this paper document the description done in 1975 in the old catalogue with ethnographic and musical information:

---

1 All the phonographic publications are available online on the CREM database ([http://archives.crem-cnrs.fr/archives/fonds/CNRSMH\\_Editions/](http://archives.crem-cnrs.fr/archives/fonds/CNRSMH_Editions/)).

**Figure 1.** Aurore Monod Becquelin archives item, Bachajon, Mexico, 1973. Source: CREM-LESC, CNRS.<sup>2</sup>

Beçqin Monod (Aurore) 15 Bachajón I Mars 1973	COPIE sélective	Uher cassette	LD	0 13	B M	1975.15.1	1
Beçqin Monod A	M. H.	1/2 P			I A XI		
I 19 I 5'	Méxiqne, état de Chiapas, village de Bachajón.	Tzeltal groupe maya-			Carnaval de Bachajón : musique des kabalinal, hommes de la forêt.		
(Original : piste 1, de 0'00 à 5').	<i>Folklore mex = données pour l'original</i>				Parlé, conversation, puis deux filices amix, tambours <i>tayoh</i> à deux peaux tenu horizontalement et battus avec une baguette, deux coques <i>pay</i> , écailles de tortue <i>hax</i> battues avec deux gols de maïs, hochements et discours rituel <i>pa'to'tan</i> palmoindé par deux <i>ha</i> .		
	<i>Tel 1250'</i>						
	* N.B. cf. Beçqin Monod Aurore "Le Carnaval de Bachajón" J.E.A. 1975 (à paraître). Cf. documentation photo, Mission archéologique française au Mexique : réf. 73 E.						
PIÈCE	Vlt.	Pis.	Durée	ÉTAT	ETHNIE	CONTENU	

Sometimes, we don't have any information, only some notes on the original box tapes. But the priority is to share the audio document to facilitate the identification of the content.

So, how to give access to this documents, both audio and text? How to play audio on line and navigate inside the sound file? How to preserve data for long term? How to manage the database daily changing and to manage the access according to the intellectual property?

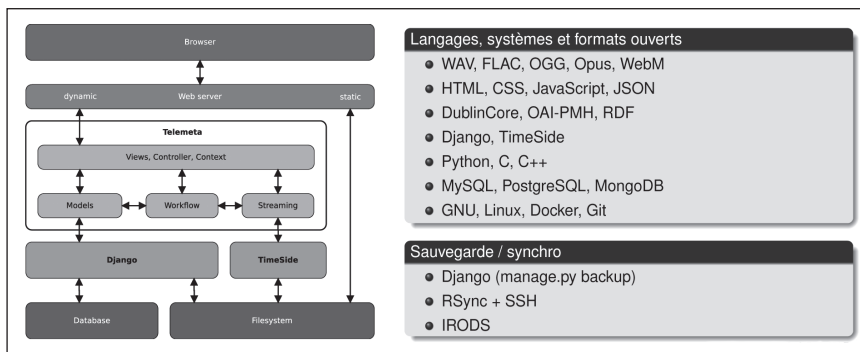
For many years, one of CNRS's objectives has been to improve access and to facilitate data sharing to the entire academic community. We do this through a toolbox of open source software's (HumaNum service). In anthropology, ethnomusicology and linguistics, researchers work on multiple kinds of documents such as sound recordings. The need to preserve and to easily access, visualize and annotate these materials is problematic because of their diverse formats, contents and the increasing quantity of data. With new audio technologies, questions linked to the preservation, the archiving and the availability of these audio materials have arisen. Since 2007, French ethnomusicologists and engineers have joined

2 Aurore Monod is research director at the CNRS Maya Civilization. More information on her work on: [http://archives.crem-cnrs.fr/archives/fonds/CNRS\\_MH\\_Monod-Becquelin/](http://archives.crem-cnrs.fr/archives/fonds/CNRS_MH_Monod-Becquelin/).

their efforts to develop a collaborative web platform for managing and improve access to digitized sound archives. This platform, an open-source web audio CMS dedicated to digital sound archives (Telemeta<sup>3</sup>), is developed through the expertise of the Parisson/Start up. Its architecture is associated with Time Side,<sup>4</sup> an open-source audio processing framework written in Python and JavaScript languages, which provides decoding, encoding and streaming capabilities together with an embeddable HTML audio player. Consequently, this CMS is able to produce visualization, annotation, segmentation, etcetera.

This tool allows to work on archives and to disseminate the sound collaboratively. A prototype started with the metadata integration in a MySQL database. It started with a prototype on line and the integration of a MySQL database to manage only metadata. It was yet a great improvement because we could browse the catalogue and create new metadata on line! The Telemeta platform launches in 2011 with 7000 sound files.

Figure 2. Telemeta architecture, formats. Source: Parisson.



Telemeta is a web content system to manage a database and the audio files (streaming player, graphic view, compression).

3 <http://telemeta.org>.

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

- Manage metadata and the audio/video files.
- Collaborative editing.
- Standardization: data and metadata. For example: the ISO standard for the languages vocabulary, Thesaurus of Geographic Names.
- Audio analysis, visualisation, encoding, embedding (i-frame link).
- Open source software and open formats.
- Access management with User count system.
- Catalogues accessible in open access.
- Right policy: sounds older than 50 years are public domain, they have free access but no re-use (the collector permission is required).
- Last change for the follow the evolution and the contributors work.
- Textual research and advanced research.

Each database has specific needs. An adapted research engine is required to have an easy access to the data: location, population, musical instrument, physical support, etc.

The first step was to describe the content with a structured model and create a specific form to organize the metadata. The collection form contains many fields: numbers, text, tables for ontologies. The next step in 2016-2017 will be to include authorities for personal names. For the interoperability, we must do the relation of the main metadata with the Dublin Core model so the database can be harvested by other web portals.

The catalogue is organized with four levels and you can group the collections as you wish. Items are grouped in collections, collections are grouped in corpus, and corpus are grouped in Funds. These Funds level describe mainly the works of a collector and propose a list of the different corpus of this person.

In the corpus level, you can see the collections, digitized or not digitized. The archivist can put markers on the sound files to show where the sound engineer should do the segmentation. The specificity of this tool is the possibility to add markers on the timeline

as well as notes on the sound file. The annotation is possible if you have a user account and they are available to everybody with the author of the notes. The spectral view is processed by the Time Side engine and offers another visualization of the content. You can copy the i-frame link to imbed the player in your website. It is not necessary to download the sound file and in fact, it is not allowed without permission. Interoperability is possible thanks to the Dublin Core and OAI-PMH protocol. For example: <http://www.rechercheisidore.fr>. A best practice guide is on line: <http://www.hu-ma-um.fr/ressources/guides>

Since the platform is launched, the database increase regularly and we have today more than 40 000 sound files on line, which 24 000 are on free access. And 2000 more visitors than before the platform!

The community of Telemeta users count 3 other teams: <http://telemeta.lam.jussieu.fr>, <http://sabiiod.telemeta.org/> and <http://pho-notheque.cmam.tn/>.

This project brings on other programs, international partnership and funding: Europeana Sounds for example, or analysis research program (Music Information Retrieval). So maybe we will all be connected together soon, because it will possible to connect Telemeta platforms with each other! This a rich experience to build common vocabulary, data models, and a task force about Intellectual property and ethical rules for the intangible world heritage.

The Wiki Telemeta project is available at: <http://telemeta.org> and <https://github.com/Parisson/Telemeta>. The benefits of this collaborative platform apply to numerous aspects of the field of ethnomusicology, ranging from musical analysis to comparative history and anthropology of music, as well as to the fields of anthropology, linguistics and acoustics. You can join us and download the tool, contribute on the database and you are welcome to visit the Sound Archives of Musée de l'Homme.<sup>5</sup>

---

5 <http://archives.crem-cnrs.fr>.

## BIBLIOGRAPHY

- Archivos de Aurore Monod Becquelin (1973). Bachajón.  
Disponible en: [http://www.mae.u-paris10.fr/lesc/IMG/pdf/bibweb\\_monod.pdf](http://www.mae.u-paris10.fr/lesc/IMG/pdf/bibweb_monod.pdf).
- Monod Becquelin, A. y Alain Breton (2002). *La guerre rouge, ou: une politique maya du sacré*. París: CNRS éditions.
- Monod Becquelin, A. (2002). Parlons tzeltal. *Une langue maya du Chiapas*. París: L'Harmattan.