



HAL
open science

Applications and implications of digital audio databases for the field of ethnomusicology, a discussion on the CNRS -Musée de l'Homme sound archives

Josephine Simonnot, Stephanie Khoury

► To cite this version:

Josephine Simonnot, Stephanie Khoury. Applications and implications of digital audio databases for the field of ethnomusicology, a discussion on the CNRS -Musée de l'Homme sound archives. First Monday, 2014. halshs-01280831

HAL Id: halshs-01280831

<https://shs.hal.science/halshs-01280831>

Submitted on 1 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Applications and implications of digital audio databases for the field of ethnomusicology,
a discussion on the *CNRS - Musée de l'Homme* sound archives**

by Josephine Simonnot

Research Center for Ethnomusicology, CNRS (UMR 7186)

josephine.simonnot@mae.u-paris10.fr

and Stephanie Khoury

Research Center for Ethnomusicology, CNRS (UMR 7186)

stephanie.khoury@mae.u-paris10.fr

Abstract

The online Web-based platform for the French *CNRS – Musée de l'Homme* audio archives offers access to about 26,000 published and unpublished recordings of traditional music from all over the world. Implemented as an archive database, it represents a collaborative scientific tool for the dissemination and production of knowledge. This article introduces these digital audio archives while dealing with issues of preservation of audio recordings, the intellectual property rights and ethics. It also deals with the new technological developments implemented to embed music archive into contemporary documentation and research dynamics.

Contents

Introduction

I. From wax cylinders to digital audio sounds

I.1 Historical constitution of the audio archives

I.2 The digital leap into online display of music

I.3 The Telemeta platform and its use for music documents

I.4 The collaborative documentation of the archives

II. Online audio database: ethics and intellectual property rights (IPR)

II.1 Intellectual property and public domain

II.2 Ethics and patrimony

II.3 Access and restrictions

III. Digital audio sounds as vehicules of knowledge.

III.1 Management

III.2 Usages and users

III.3 New perspectives: audio database as support and object of research

Conclusion

Introduction

In this contemporary technological era, the amount of available data online is countless and constantly increasing while the range of information displayed grows wider. Online music distribution, both as audio and video files, has generated a new way of relating to music, easing its massive consumption while providing virtual stores and libraries for music-producing companies and organizations of all kinds. If this is particularly striking in the entertainment music industry, a professional use of the music in the fields of humanities and social sciences has led some institutions, such as the French Ministry of Culture and Communications, to rethink the management and display of scientific audio data. For 15 years, French archivists working with audio materials have embraced the benefits of digitization for the conservation of documents recorded on obsolete formats and kept as such. The impalpable nature of digital data, and the quality and easiness of duplication, procures a feeling of permanence, as it is virtually protected from disappearance.

Music archive databases are exponentially present on the Web and display the recordings' collections of numerous institutions. These databases can be focused on one specific geographical area, on a population, on the data recorded during events linked to this institution, and so on. In the field of ethnomusicology, the audio recordings are generally of traditional music from all parts of the world and all kinds of contexts. Their online availability has raised numerous issues linked to the specific nature of the documents. What is the role and purpose of scientific music databases and what is their position in the spectrum of online distribution of music? As ethnomusicological archives gather musical artifacts of various populations and groups' cultural heritages, how do institutions approach the question of the ethical management and display of music recordings? Furthermore, how does the online availability of music

archives change the way we relate to audio documents in both scientific research and amateur interest?

In this paper, we aim to discuss the applications and implications of ethnomusicological audio databases through the example of the *CNRS - Musée de l'Homme* sound archives. Since 2011, this institution's archives have been available through a Web-based platform and represent a cutting-edge model for online collaborative databases dealing with audio recordings. It brings archivists, researchers and computer engineers together to work on this scientific sound database and reflect on its epistemology in order to manage its growing use and to develop tools that facilitate the emerging needs of users.

To address these points, we first introduce the nature of the music and audio documents that constitute the *CNRS - Musée de l'Homme*'s archives and the new method that has been implemented to process sound and contextual metadata with a single and collaborative computer platform. Then, we address questions related to intellectual property rights and the implicit ethics of the online diffusion of music archives. Finally, we discuss the way this database enhances archive-based scientific research.

I. From wax cylinders to digital audio sound

The *CNRS - Musée de l'Homme* music archives represent one of the largest archives of ethnomusicology in Europe and stand among the very few that provide an online access to the audio documents as well as detailed metadata. The establishment of the archives' collections followed an almost century-long process of thinking audio document diffusion and progressive adaptation to new technologies.

I.1 Historical constitution of the audio archives

The online sound archives database represents the achievement of a long process induced in the 1930's to collect, organize and archive published and unpublished audio recordings ranging from 1900's till today. Since the late nineteenth century and the invention of the first recorders, music materials, their classification and their preservation are central to the field of ethnomusicology, then called "comparative musicology", and shape our knowledge of the musical Man¹.

A *Phonothèque* (sound library) was created by the French historian and musicologist André Schaeffner in 1932, at the *Département d'Ethnologie Musicale* (Musical Ethnology Department) of the *Musée d'ethnographie du Trocadero* (Ethnography Museum of the Trocadero) in Paris. He was then returning from de "Dakar-Djibouti" expedition (1931), a large ethnographic fieldwork sponsored by Marcel Griaule that crossed the African continent on the sub-Saharan belt. Aware that collecting musical instruments was not enough to gain some understanding of musical practices, he made audio recordings of numerous musical performances from Senegal, Mali, Cameroun and Ethiopia engraved on wax cylinders. The *Phonothèque* was thus sheltering these precious audio materials and was associated to the organology section of the department, which was containing a significant musical instruments collection (these instruments are now located at the *Quai Branly*² Museum, Paris).

Many audio recordings joined the collection following Schaeffner first deposit, including numerous published records from domestic and foreign labels such as *Victor*, *Zonophone*, *His*

¹ Formulation in reference to Blacking, 1973.

² The instruments catalogue can be reached at: <http://collections.quaibrantly.fr/pod16/#f498e01e-2cf1-4f95-917d-5fc69726e770>

Master's Voice, Brunswick, Columbia, Parlophon, Odéon, Polydor, Pathé or *Gramofono*, as well as records produced at the Colonial Exhibit held in Paris in 1931.

If the Ethnography museum turned into the newly build *Musée de l'Homme*, in 1937, it's only after World War II that the next big ethnographic expedition crossed the central part of the African continent, and brought musical archives recorded by Gilbert Rouget (Gérard, 2012). In the meantime, the publication of edited 78rpm records³ started under the label "*Musée de l'Homme*" (Figure 1).



Figure 0. 78rpm records of Romanian music recorded by Brailoiu. (photo by Jean-Marc Fontaine, 2009)

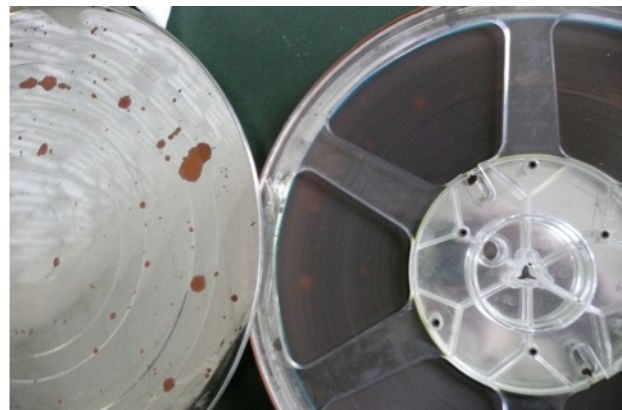


Figure 2. Analog audiotape from the *Musée de l'Homme* showing traces of rust. (photo by J.-M. Fontaine, 2009)

In the following years, an epistemology of French ethnomusicology emerged within the walls of the *Musée de l'Homme*. From 1954 onwards, about a hundred of published data were released as both 33 and 45 rpm records by editors like *Vogue Contrepoint, Boîte A Musique, Le Chant du Monde* and different scientific institutions such as the Peabody Museum of Archeology and Ethnology (Massachusetts, USA), the *Institut Fondamental d'Afrique Noire* (Senegal) and so on⁴. The department turned into the *Laboratoire d'ethnomusicologie* (Laboratory for ethnomusicology) of CNRS in 1968. It hosted researchers and students, including decisive

³ http://archives.crem-cnrs.fr/archives/corpus/CNRSMH_Editions_001/

⁴ http://archives.crem-cnrs.fr/archives/corpus/CNRSMH_Editions_002/

figures of the discipline such as Claudie Marcel-Dubois, Constantin Brailoiu, Alan Lomax, Curt Sachs or John Wright.

The collections, still located in the museum, were also renamed and became the *CNRS - Musée de l'Homme* archives. Between 1988 and 2001, 37 CDs were published under the label *Chant du Monde*, some of which became well-known references, such as the compilations “Voices of the world” and “Musical instruments of the world”⁵.

In 2008, the Laboratory for ethnomusicology⁶ and its associated audio archives moved to the Université de Paris X - Nanterre (now Université de Paris Ouest Nanterre), where it was renamed Research Center for Ethnomusicology (CREM). From the creation of the *Phonothèque* onwards, collections of unpublished audio materials have been in constant growth, today reaching a rate of about 30 collections deposited annually. Facing the thousands of raw recordings deposited since the 1930's and the hundreds of published records collected, the CREM had to implement a unified strategy to efficiently preserve, organize and classify all these documents.

I.2 The digital leap into online display of music

In Ethnomusicology, the nature of audio-visual materials used in the field raises special issues of preservation and accessibility. While digital formats are readable and duplicable to infinity, obsolete analog formats are not more permanent than a fragile tape ribbon and the equipments required to read them are disappearing (Figure 2). Since 2000, to preserve the audio archives, the *Laboratoire d'ethnomusicologie* (now CREM) developed and applied an on-going program of their systematic digitization. Priority was given to old recordings (cylinders and

⁵ http://archives.crem-cnrs.fr/archives/corpus/CNRSMH_Editions_003/

⁶ See Rouget, 2004.

direct cut discs) from iconic missions dating from the first half of the twentieth century. Ensuring the open-access to audio archives first answered a need from the archivists to optimize the digitization and the associated documentation with the use of one sole tool to both listen and fill-in the metadata in collaboration with collectors, when alive. Moreover, it fulfills a request from French state-sponsored academic institutions such as the Ministry of Culture and the National Library. They support and fund the digitization of these audio documents and of research projects attached to them at the condition of their open-access availability. Through the dynamism of the CREM, new audio materials constantly enrich the database and, from 2006 onwards, its researchers and engineers engage into a reflective process on the use and aims of these archives. Since 2011, the CREM uses the Telemeta platform architecture to organize, catalogue and display these archives on a Web server: <http://archives.crem-cnrs.fr>.

The CREM's initiative to implement its own database fits into the context of contemporary technologies applied to online music distribution. The availability of audio archives databases is a growing tool on the Web and majors institutions acknowledge the necessity of informing the scientific community as well as the broader public of what they have in their collections. This effort can be rendered as a catalogue through which the references and information about audio documents can be easily accessed. Useful, it yet does not convey the dynamic nature of audiovisual materials and still requires the action to go to the physical location of these archives. In such cases, consulting the audio archives is restricted to those who can go where the sound is, which can easily turn into an expensive and time-consuming enterprise. To facilitate the access to the audio document and embrace a philosophy of sharing knowledge, other interactive online

databases are being implemented, such as the French National Library⁷, the British Library Sounds⁸, the Digitization of the Ethnomusicological Sound Archive of the Royal Museum for Central Africa Project (DEKKMMA)⁹ and the recently released *Europeana Sound Library*¹⁰, a vast European project for which some documents come from the *CNRS-Musée de l'Homme* archives.

I.3 The Telemeta platform architecture and applications on music archives

Engineers from the CREM, from the Laboratory for Musical Acoustic (LAM)¹¹ and Web-developers from the Parisson Company¹² worked together for 7 years on a content management system (CMS), leading to the implementation of the Telemeta¹³ architecture to support the audio archives database. This platform, written in Python and Javascript languages, is an open-source software allowing to manage large audio databases and to easily index sound files (Figure 3).

⁷ <http://gallica.bnf.fr/html/enregistrements-sonores/fonds-sonores>

⁸ This database includes language and music recordings, classical Western music as well as world and traditional audio files from series attached to indentified collectors. It is available on <http://sounds.bl.uk>

⁹ The sound archive of the department of Ethnomusicology of the Royal Museum for Central Africa at Tervuren (Belgium) lists recording references from numerous countries in Africa. It is available on:

<http://music.africamuseum.be>

¹⁰ <http://www.europeanasounds.eu>

¹¹ <http://www.lam.jussieu.fr>

¹² <http://www.pariison.com/>

¹³ <http://telemeta.org>

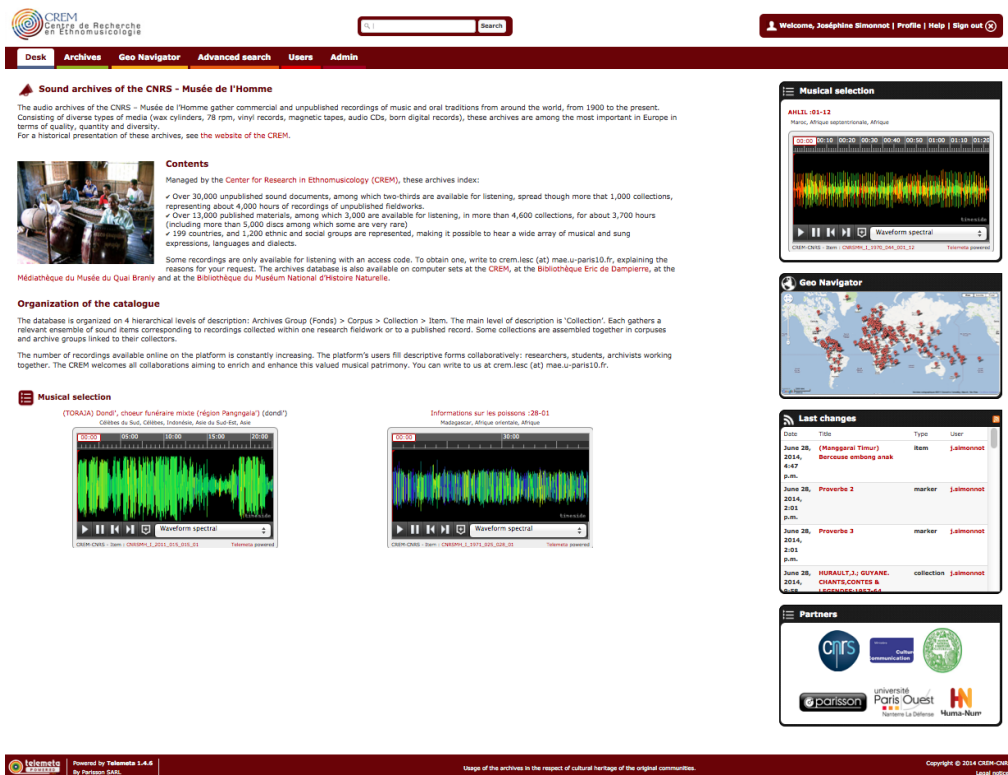


Figure 3. Screenshot of the Musée de l'Homme audio archives' database home page.

The sound archives metadata are structured on a MySQL CMS, which organizes the catalog in 4 levels: Series, Corpus, Collection, and Item. Collection is the main level of entry into the database. Each collection embeds a series of sound items that share a common contextual feature. It can appear as a published series of music recordings from a same area or on a shared theme, such as *Les Voix du Monde* (“Voices of the World”), which provides with examples of vocal techniques. But, the richness of the database also comes from the important amount of unpublished collections of raw recordings from one fieldwork in one specific area.

Collections are organized into corpuses, each one gathering either published versus unpublished collections of recordings or collections grouped by geographical area. These corpuses are gathered into Series and both are in reference to a collector or an organization.

Detailed presentations are provided for these two higher levels of the catalogue, leading to introduce the work of the collector or institution with attached documents, embedded short audiovisual materials and numerous hypertext links to relevant web pages (Figure 4).

The screenshot shows the 'Fonds Mireille Helffer (1928-)' page. At the top, there are navigation tabs: 'Desk', 'Archives', 'Geo Navigator', 'Advanced search', 'Users', and 'Admins'. Below the title, there are buttons for 'Edit', 'Copy', and 'Add to playlist'. The 'Fonds' section includes fields for Title, Description, Code, Public access, and Last modification. The 'Related media' section has a 'Media' tab and a 'Preview' button. It contains a 'Biographie' section with a detailed description of Mireille Helffer's work, a 'Photo' section with a portrait and a description of an article, and a 'Bibliographie' section with a list of publications. The 'Corpus' section features a table with columns for Title, Description, Date, and Sound. The table lists several recordings, including 'Les musiques rituelles du bouddhisme tibétain (Tibet-Népal-Indes)', 'Musiques tibétaines: Appelé du Roi Gê-Sar', and 'Musiques populaires du Népal et d'Inde'. At the bottom, there is a 'Dublin Core' logo and a footer with the text 'Stage of the archives in the respect of cultural heritage of the original communities' and 'Copyright © 2014 Centre CNRS'.

Figure 4. The series' page of Mireille Helffer, a renowned ethnomusicologists working on Tibetan music http://archives.crem-cnrs.fr/archives/fonds/CNRSMH_Helffer

As an user-friendly catalog, this platform offers different paths to access a sound item, whether the user search for music from a particular country or ethnic group, a collector, a type of instrument played on the recording, a title, the recording catalog number, the year of recording or of publication, etc. Each sound recording gets its own page on the platform containing audio and written details. Numerous text panels are dedicated to metadata listing all the information available about the recording in fields such as the date and place of the recording, its context, the original material it was recorded on, the various instruments or voices heard as well as remarks containing a wide range of elements available, from contextual description or content translation

to musical analysis. Archivists work on using commonly accepted norms from the International Organisation for Standardization (ISO) for languages. Geographical locations are uniformed using the integration of GeoEthno¹⁴ and GeoNames¹⁵ thesauruses, allowing Telemeta to manage the historicity of the location's terminologies (linking Benin to its former appellation of Dahomey, for example).

The sound is available to listen but also to be visualized via a dynamic audio player using TimeSide audio analysis and visualization framework¹⁶. This provides a signal processing tool for the display and streaming of audio sound on the Web. The on-demand processing of the graphic appearance of the sound selected and its compressed listening format are some of the major advantages of using this online platform. To enhance the visual experience, the visualization can be resized to full screen. Various graphical representations can be chosen, such as the waveform, the spectral analysis (spectrograms logarithmic and analogical), and the pitch level (aubio). These are very helpful and commonly used to spot speech or music sections and to navigate inside the recording.

The home page keeps tracks of the amount of documents digitized and offers a random selection of recordings leading directly to the audio document's page and brings the user into the collections.

I.4 A collaborative documentation of the archives

Online availability has changed the way traditional music is studied and shared in an academic context. The originality of the *CNRS-Musée de l'Homme* sound archives is to be

¹⁴ <http://www.mae.u-paris10.fr/dbtw-wpd/bed/index-lesc.html>

¹⁵ <http://www.geonames.org/>

¹⁶ <http://github.com/yomguy/TimeSide>. For more details, see the forthcoming Fillon *et al*, 2014.

embedded into the interactive platform system, Telemeta that allows the implementation of a collaborative concept of the database. This means that authorized people, i.e. researchers, archivists and people carrying knowledge, can add historical, contextual or analytical comments to the available metadata. They can also act on the audio recording through the possibility of placing time-embedded markers and associated comments, thus contributing to the knowledge displayed according to each one's own expertise (Figure 5). The long-term preservation of the database on a server is a strong argument to encourage depositors to work on the documentation related to their recordings. These annotations are available from the sound archive item Web page and are indexed through the database. A RSS (Rich Site Summary) flux automatically sends out modifications and additions to users who subscribed to it. As the digitization of all sound documents is an on-going process, only about 60% of unpublished materials are digitized yet. There is a checked symbol indicating the presence of the audio file on the database for each particular collection and items, thus stating its availability either for free listening or with a user account (one third of the digitized files is open access).

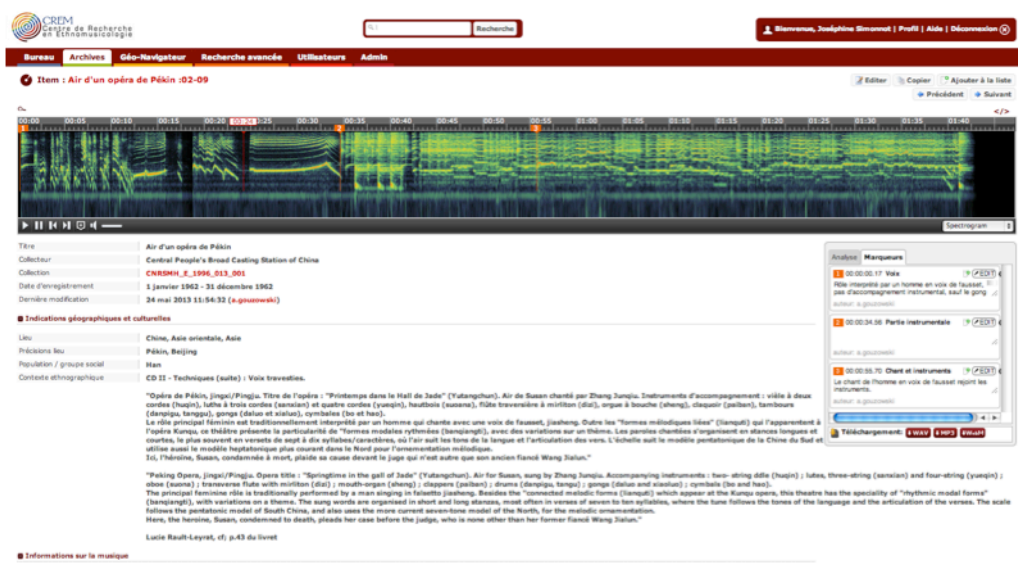


Figure 5. Screenshot of the spectral view of a song from an Opera of Beijing's performance with markers and the corresponding comments (on the right).

<http://archives.crem-cnrs.fr/archives/items/CNRSMH E 1996 013 001 002 009/>

II. Online audio database: ethics and intellectual property rights (IPR)

Numerous online digital sound libraries, including the *CNRS-Musée de l'Homme* one, allows for the online streaming of a part of their sound recordings. The access to the catalogue and to many audio recordings is free. Nonetheless, all is not open access. Restrictive options allow for control over the distribution of the audio documents and login-controlled access profiles are set. Through such decisions, parts of the world's intangible cultural heritage are virtually available to Web users. Yet, once the decision to put the archives online was made, numerous questions related to both the regulations related to intellectual property and the sensitivity of some recordings emerged.

II.1 Intellectual property and public domain

Recordings of traditional music, storytelling and mythological narratives that are made within the frame of a scientific work are different from music recordings produced in a commercial music industry. Most of the recordings in the archives, particularly among the oldest ones, do not make mention of the performers' names, as the protocol to collect music did not request them, and are rather associated with the collector. In such case and according to the legal system on this matter, as applied in France, the person producing the recording is not the sole owner of the sound document. It officially belongs to the performers, even if unknown, and to the institution that financed the fieldwork during which these recordings were made.

The intellectual property's related rights apply to the performers as well as collectors who produce the recordings and the institutions for which they did so. This status gives the producer

rights over the recording as an edited object while performers remain the owners of the music contained in that recording for a period of fifty years following the recording date. After such period of time, the recording enters the public domain. Therefore, the rule adopted is that of a sliding date and all the collections recorded before 1964 are publicly available in 2014, unless specific ethical issues prevent the recording to be displayed (i.e. a series of recordings containing secret ritual ceremonies from Australian native Aboriginals are not available to listen online in respect to the nature of their content). The authorized audio documents are to be listened online and there is no tool to download the audio file made available to visitors on the platform.

Since 2014, the CREM, which manages these audio archives, has adopted the *Europeana* model about right statements related to sounds and database. *Europeana* is a project of the European Commission and a consortium of digital libraries that aims to provide access to Europe's sound and music heritage¹⁷. Licenses and permissions are drawn from models available on the online organization Creative Commons website¹⁸. The main guidelines of the statements are that all metadata will be licensed under public domain (CC0) and the institutions aim to use the rights statements of public domain CC-BY and CC-BY-SA as much as possible. The tagline of the CREM is aligned with the *Europeana* project, which is that, as much as possible: public domain is the rule, copyright is the exception.

The open-access doesn't apply exclusively to recordings that are old enough and the audio documents recently deposited can be made available to be listened when the collector agrees. Thus, more recent collections can also be fully accessed. This represents an opportunity for researchers to save and secure the archiving of their own collections. In addition, today's

¹⁷ <http://www.europeanounds.eu/about>

¹⁸ <http://creativecommons.org>

collectors and depositors are more and more confident and aware about the benefits of Web sharing. On one hand, intellectual property is perpetual and inalienable. Attached to the artists and the collector so no one else is allowed to claim ownership over a recording or to have any commercial use of it, even if the data is published online. On the other hand, online accessibility allows sharing and interacting with other specialists as well as with performers of the recordings. On the matter of knowledge circulation, there is currently a switch from retention to display.

For example, Dana Rappoport, ethnomusicologist at the CNRS, has provided access to all her recordings of vocal music made in East Indonesia (1992-2012), which represent about 1,100 audio items¹⁹, with the agreement of people she worked with locally. In doing so, she expresses her and the local communities' wish to preserve and share this vanishing heritage, in order to keep their intangible heritage alive (Figure 6).

Moreover, the CREM committee has recently decided to give full access to all the records published by the Musée de l'Homme²⁰, and most of them are already online. While the editor (Chant du Monde/Harmonia Mundi) stopped the distribution of these CDs some fourteen years ago, and as most of the records are out-of-sale, the archives platform is for many the only way to access these recordings.

Thus, each collection has a status of access, which is specified on each sound item page, and that states whether it access is full or restricted to the available metadata without the sound, according to the intellectual property rights and the depositor wishes.

¹⁹ http://archives.crem-cnrs.fr/archives/fonds/CNRSMH_Rappoport/

²⁰ http://archives.crem-cnrs.fr/archives/fonds/CNRSMH_Editions/

We are well aware that online display of music may lead to illegal audio captures, but the importance of providing access to these archives is determinant to contribute to scientific research and to allow people to access to their own patrimony.



Figure 6. Singer in Flores, Indonesia (Photo by Luc Rivière, 2010)

II.2 Ethics and patrimony

The recording of traditional music from various places in the world early twentieth century was often linked to colonialism. Emancipated from it, the motive of pioneer ethnomusicologists was, on one hand, to promote a better understanding and respect of other cultures while bringing to light musical practices and complex musical systems that were unknown until then. On the other hand, another objective was to preserve objects and sounds that were threatened of

disappearance for future generations to see and learn about. The main motive was to share and reconstitute the patrimony collected. In the beginning of the twentieth century, as African art was in fashion in Europe, artists such as Picasso and later Giacometti would be regular visitors of the Musée de l'Homme's collections. Major artists of the time would regularly attend audiovisual events showing musical performances.

As the French school of ethnomusicology promotes the ethical position of returning the recordings to the communities who performed in them, the fragility of materials used for audio recordings made it difficult. Online streaming of these recordings in a digitized format allows to safely achieving this goal.

Today, many researchers wish to deposit the archives of the music and audiovisual material they have collected throughout their career in order to first keep a trace of these often unpublished materials, but also to make it available to the people it belong to within the population where these researchers have conducted fieldwork. This is not without raising many challenges, particularly related to the unique nature of ethnomusicological materials. What is the right of the researcher to display the audio duplicate of secret religious ceremonies or highly emotional moment? Are the content of the recording in any way threatening the integrity of the people recorded? Would the free audition of some recordings create trouble of any kind? In many cases, the access has to be restricted in order to protect ethical rights of the event or the person recorded. This decision is up to the collector who acts in accordance to the authorization he or she received from the recorded people and to his or her knowledge of the culture represented.

Online availability through the scientifically managed non-lucrative database of the *CNRS-Musée de l'Homme* sound archives is a form of giving back to communities part of their own

cultural heritage. Specific online access can be provided to local institutions, at their demand, to ease an access to the audio materials. France has a long tradition of opening culture to a wide audience and to make available sources of knowledge and an environment appropriate to understand it. Music is an intangible patrimony and as such it needs to be both protected and made available. But the transferring of this patrimony requests the contributions of different forms of knowledge as well as a strong technical support for storage, software and operating system server updates.

II.3 Access and restrictions predetermined by user profiles

Different profiles were implemented to moderate the access to contents, in respect to the intellectual property rights and related rights, in agreement with special restrictions or authorizations of collectors and according to the archivist aims and scientific purposes of the sound archives database. Such management leads to different approaches of the archives content, each determined in accordance to the uses and inputs expected. Therefore, through the Telemeta platform, the *CNRS-Musée de l'Homme* sound archives can be apprehended as a visitor, a researcher or as an archivist. People with an Administrator profile can access every element on the database, download files and assign a selection of authorized actions to each user profile.

Occasional and unauthorized visitors have access to all the metadata but only to the part of the recordings that are of free acces. For the audio files that are available, visitors have the possibility to share the recording in exporting and embedding the audio player into external web pages and blogs through an i-frame html link.

The researcher profile has additional prerogatives. Not only can people under this profile access to numerous music archives with restricted access, but they also have the ability to edit

the metadata and annotates the audio files. Doing so, this profile fits the collaborative dimension of the database. Developed to have its own space on the platform, a researcher profile can choose the language in which the platform is displayed (French, English or German). Researcher's profile also includes the possibility to create personal lists in which the user can save its own selections of sound items or of collections from the database. Such option is particularly helpful to organize a playlist for conferences or courses, to arrange a template for the publication of one's own recordings, to gather different music the user wish to use in its own research, etc.

Archivists are attributed the eponymous profile, which allows them to reorganize the catalog, in gathering collections into corpuses and series, to expand it, as well as to integrate new written and audio documentation. They are also in charge of the database thesauruses and their modifications, while having an on-going reflection on questions of terminology and of classifications in connection with other national archives, embedding the work made on the Telemeta platform for sound archives into a larger consortium of documentation databases.

With specific authorization, people involved in a particular research project can share data and notes online, thus allowing them to collaborate and to optimize the enrichment of the metadata. To facilitate such aims, the platform permits the export of metadata of series of items or of collections. It also allows to upload and download recordings, compressed (MP3 & OGG) or not compressed files (WAV, FLAC).

The usage options of the different profiles are regularly reassessed in accordance to users' experience, and the information from individual users who get directly in contact with the people in charge of the database. Complementarily, it is also evaluated through a qualitative survey and the detailed report of the statistics of users, including the URL address of websites setting a link to the platform. The combination of these modes of evaluation provide with information

regarding the way users connect and interact with the database, as well as bring to light individual approaches. This points to relevant issues and needs that developers, engineer and archivists address with adapted tools. Due to such interaction, the database is thus in constant evolution.

III. Digital audio sounds as vehicles of knowledge

Thanks to this tool, which uses common standards, researchers' work is made easier and the overall accessibility of the database is widely extended. Today, 47,700 items from 5,800 collections of the CREM are catalogued online on the *CNRS-Musée de l'Homme* archives' database. In May 2014, more than 26,300 sound files have been uploaded, among which about 12 000 are on public access. During this same month, 2,700 different visitors have consulted the platform, which represent a 145% increase compare to the consultation of the database on the same month of 2012. This leads to the diffusion of the methods and process to catalogue and manage audio archives that is adopted by other research programs in need of a platform to access and work on sound documents.

The CNRS-Musée de l'Homme platform, through the CREM, has brought researchers to use audio files within the frame of their own research while specialized blogs refer to some specific archives to illustrate related topics. As the uses end up being much broader and specialized than expected, improvements are regularly integrated to the platform and a project for the development of music analysis and indexation tools has been launched in 2013 in order to expand the database's archival and research possibilities.

III.1 Management

As digital humanities are at the forefront of new development in research, the implementation of the Telemeta platform for the music archives of the Musée de l'Homme appeared as a cutting-edge way to relate to sound archives. Soon after its release, two other platforms using the Telemeta framework were implemented to shelter research projects linked to the CNRS and for which people needed to manage their own collections of digital audio documents. One is the Laboratory for Musical Acoustics (LAM, UPMC/CNRS)²¹, using the platform to organize sounds from musical instruments of all kind considered separately and out of context in order to study their acoustic properties and characteristics. With the same objective of constituting an audio database of digitized sound, a consortium of research departments involved in the interdisciplinary project called Scaled Acoustic Biodiversity (Sabiod) use the program framework to gather audio signals of marine animals on their own platform²².

If the overall architecture remains the same, numerous elements were adapted to the specificities of each set of archives and the specific needs of the archivists in charge of it, such as the nature of the metadata or the preset representation of the audio sound. In these cases, the aim is to support the collective and collaborative work of research teams as well as individual researchers. As for the *Musée de l'Homme* sound archives, the interactivity with the platform is login protected and some audio recordings are not set in open access.

These two examples illustrate a new relation of researchers to sound archives and how such tools make them more accessible and user-friendly. Beyond uses within research departments or centers, such online representation brings individual platforms to contribute to broader archival

²¹ <http://telemeta.lam.jussieu.fr>

²² <http://sabiod.telemeta.org>

projects centralizing them into large audio databases. The *Europeana Sound* project²³ is an illustration of this. Launched in 2014 and sponsored by the European Commission, this project will give online access to a critical mass of audiovisual digital-objects. In the upcoming three years, over a million high quality sound recordings will be available via *Europeana*, from classical and folk music to environmental sounds of the natural world, as well as oral memories. Together, these collections reflect the diverse cultures, histories, languages and creativity of the peoples of Europe over the past 130 years. The project, coordinated by the British Library in London, bring together twenty-four national libraries, sound institutions, research centers and universities from twelve European countries. The *CNRS-Musée de l'Homme* sound archives platform participates in this European project in providing data to its online portal.

III.2 Usages and users

Through the few years since the Telemeta platform is operative, academic and public uses of the knowledge available on the database have arose. Beyond the scope of ethnomusicology, researchers such as anthropologists, linguists and acousticians find elements to integrate in they own work.

For archivists, the adoption of the Dublin Core format for the metadata allows to fit the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Information stored on the database can thus be harvested and referenced on web search engines and platforms dedicated to the digital humanities. This helps archivists to organize the metadata and to diffuse them.

Emerging researchers and students have conduct research primarily based on the music archives, considering the collections of languages and musical practices in diachronic

²³ <http://www.europeanasounds.eu>

perspectives. Ethnomusicologists have been used to compare former and current expressions of a same musical genre, repertoire, piece, or category of instruments (Khoury, 2004 and Lacombe, 2013). Linguists have integrated recordings from the database into their analysis of regional accents, cultural contacts among neighboring populations, and historical study of languages, including some that are not spoken anymore nowadays. Publications in both soft and hard formats, whether books or articles, an even museum exhibits can include audio illustrations in reference to specific recordings or collections archived on the database through the embedment of a URL link or a QR code (see Gérard, 2012).

Unexpected uses of the platform appeared since its online launching. They relate to the diffusion of knowledge, whether through direct teaching in a scholarly setting or through online specialized blogging. University professors from North America and Europe, as well as school teachers, use online streaming of music from the *CNRS-Musée de l'Homme* database to illustrate points raised in class and to get students to practices music analysis and transcriptions. The published records from the Musée de l'Homme's collection are particularly solicited as they offer didactical series gathering music recordings from diverse origins and provide a fully detailed documentation for each other them. Scientific blogs such as the *Jabal al-Lught: Climbing the mountain of languages*, set by Dr. Lameen Souag, a CNRS researcher in North African languages presents his works, including some in-depth analysis of recorded narrative from the *Musée de l'Homme*'s database²⁴.

As the open-access policy applies on a large amount of recordings and as the metadata are integrally accessible, the archive platform makes musical knowledge accessible outside of the academic range, leading people highly specialized and passionate about one domain to expand

²⁴ <http://lughat.blogspot.fr/2014/03/korandje-tale-conte-en-korandje.html>

their knowledge and share it (see for example the blog dedicated to 78-rpms, *Ceints de Bakelite*²⁵). Figure 7 summarizes the interactions of users with the platform according to their use of its content.

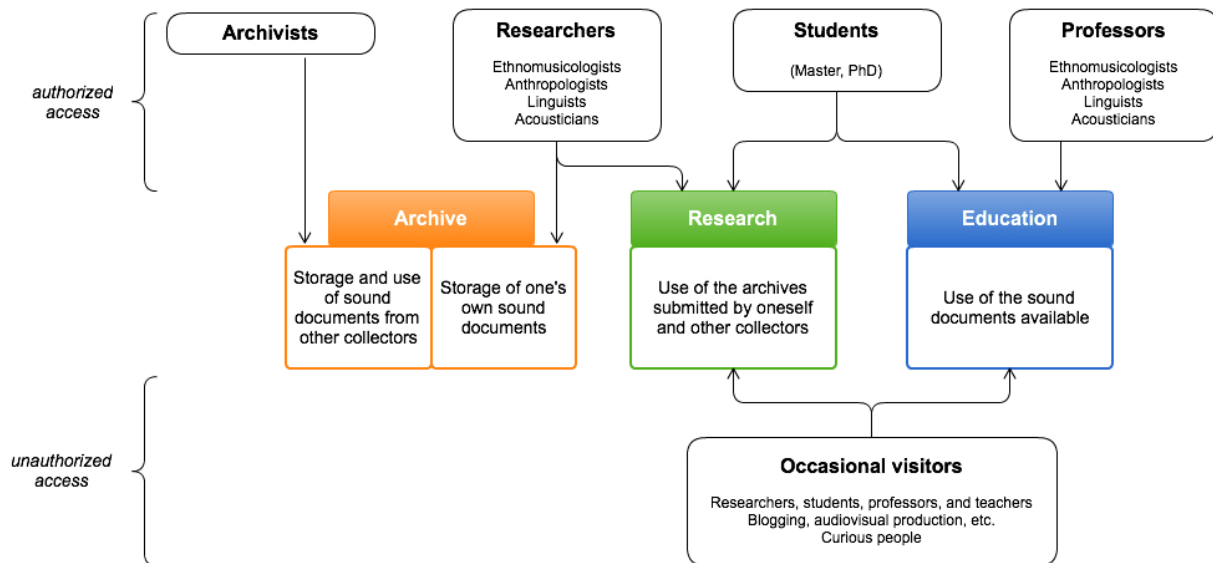


Figure 7. Diagram of users and usages summary

III.3 New perspectives: audio database as support and object of research

This new way of dissemination generates also innovative and ambitious research programs at the forefront of new technologies in the digital audio field. Since 2013, new analytical tools are being developed through a French national project in which different departments of the CNRS specialized on computer engineering, speech and Music Information Retrieval (MIR), acoustics and anthropology work together. The objective is to answer and to predict the platform users' needs and consequently to implement modifications and new options on the Telemeta framework.

²⁵ <http://ceintsdebakelite.com>

To be effective and in phase with the new audio technologies, new tools are expected to broaden the scope of their research activities linked to it. The reflection collectively engaged by engineers and researchers on the use of the sound archives database led to propose a set of tools for automatic indexing. Fields recordings contain speech, singing voice, instrumental music, technical noises, natural sounds, and all forms of concomitance of these different sound events. The automatic indexation of audio recordings directly from the audio signal itself aims to improve the access to anthropological archives.

Through the DIADEMS project (Description, Indexation, Access to Ethnomusicological and Sound Documents)²⁶, tools are implemented to develop advanced classification, segmentation and similarity analysis methods, thus helping in the management of large amount of digital audio documents. This allows for the automatic detection of audio events such as speech-music segmentation, speech recognition, as well as detection of tone, rhythm, and melodic patterns, as well as musical instruments families.

Conclusion

The digitization and online archiving of sound is embedded into a recent trend that aims to gather and preserve the broad audio heritage of a community, a region, or the world. The challenges faced by archivists are eased through the adaptability of the platform, allowing them to think further enumerations and thesauruses. Beyond the archival purpose of the database, scientific research and publications, educational programs and social medias have integrated contents of the online music archives into their mode of knowledge production and

²⁶ <http://www.irit.fr/recherches/SAMOVA/DIADEMS/en/welcome/>

dissemination. As it is based on public funds and requires perpetual maintenance, the main challenge today is the economic viability of such archive database, while maintaining the ethical choice to offer a free access to the musical heritage displayed through the *CNRS-Musée de l'Homme* archives online platform.

About the authors

Joséphine Simonnot is a research engineer at the CNRS, head of audiovisual section of the CREM, head of project for the Telemeta platform and CNRS representative in the European Sounds Project.

josephine.simonnot[at]mae[dot]u-paris10[dot]fr

Stéphanie Khoury has a PhD in cultural anthropology specialized in continental Southeast Asia performing arts and music and a research engineer at the CNRS on the DIADEMS project.

stephanie.khoury[at]mae[dot]u-paris10[dot]fr

Bibliographie

Blacking, John, 1973. *How musical is Man*, Seattle: University of Washington Press

Thomas Fillon, Joséphine Simonnot, Marie-France Mifune, Stéphanie Khoury, Guillaume Pellerin, Maxime Lecoq, Estelle Amy de La Bretèque, David Doukhan, Dominique Fourer, Jean-

Luc Rouas, Julien Pinquier, Julie Mauclair and Claude Barras, 2014. “Telemeta: An open-source web framework for ethnomusicological audio archives management and automatic analysis”, *working paper for the Digital Library for Musicology workshop*.

Gérard, Brice, 2012. “Gilbert Rouget et la mission Ogooué-Congo (1946). Institution et épistémologie dans l’histoire de l’ethnomusicologie en France”, *Gradhiva*, 16, 2012, pp. 192-215. [*Gilbert Rouget and the Ogooué-Congo Expedition (1946). Institution and Epistemology in the History of French Ethnomusicology*]

Khoury, Stéphanie, 2014, “D’un Trott, à l’autre... L’émergence d’un folklore au Cambodge, entre influence et individualité”, MPhil thesis in ethnology/ethnomusicology, Université de Paris Ouest-Nanterre.

Lacombe, Claire, 2013, “Pour une valorization des enregistrements d’archives de xylophones fang de 1908 à nos jours”, Master thesis in ethnomusicology, Université de Paris Ouest-Nanterre

Rouget, Gilbert, “Le Département d’ethnomusicologie du Musée de l’Homme. Maison mère de la discipline en France et dispositif en peril”, *L’Homme* 2004/3-4, N° 171-172, p. 513-523. [*The ethnomusicology department of the Musée de l’Homme. French headquarter of the discipline and organization in peril*]