The CNRS - Musée de l'Homme Sound Archives from 1900 to the present: a long way between heritage, knowledge and technologies

Josephine Simonnot

The *CNRS - Musée de l'Homme* Sound Archives from 1900 to the present: a long way between heritage, knowledge and technologies.

Josephine Simonnot, Research Centre for Ethnomusicology, Laboratory of Ethnology and Comparative Sociology, CNRS, France.

The online web-based platform for the French *CNRS – Musée de l'Homme* sound archives offers access to published and unpublished recordings of music from all over the world, collected from the 1900's through the present day. We first introduce these sound archives managed by the Research Centre for Ethnomusicology[1]. We then explain their entry in the new web eco-system as well as their availability to a broad audience through a suite of tools for analysis and visualization. Finally, we study how a sound archive database can become a collaborative tool for the production and dissemination of knowledge.

1. Historical constitution of the archives

During the last century, audio documents have become increasingly important in ethnomusicology. Recordings are not only a way to fix and preserve music for future generations, they also support musical transcriptions for researchers and shape our knowledge of musical phenomena. In this context, the *CNRS - Musée de l'Homme* music archives became one of the largest archives of ethnomusicology in Europe, providing online access to audio documents as well as detailed metadata such as the collector, the country or the year of each recording. The establishment of these archives represents a long process begun by the musicologist André Schaeffner in the 1930's to collect, organize and archive published and unpublished audio recordings, such as the Paris Universal Exhibition cylinders (1900). André Schaeffner was returning from the "Dakar-Djibouti" expedition (1931), an ethnographic field trip headed by Marcel Griaule on the Sub-Saharan belt. Aware that collecting musical instruments was not enough to understand musical practices, he made audio recordings engraved on cylinders[2] of numerous performances from Senegal, Mali, Cameroon and Ethiopia. These audio materials were preserved in the newly established *Phonothèque* in the *Département d'Ethnologie Musicale* of the *Musée d'ethnographie du Trocadero* in Paris, associated with the organology section of the department, which contained a significant collection of musical instruments. These instruments are now located at the *Musée du quai Branly*[3] in Paris.

In the following years, hundreds of audio recordings were added to the collection, including those from other explorative missions in North Africa, involving Germaine Tillon

[1] http://lesc-cnrs.fr/crem
[2] http://archives.crem-cnrs.fr/archives/fonds/CNRSMH_Cylindres/
[3] The instruments catalogue : http://collections.quaibranly.fr/pod16/#f498e01e-2cf1-4f95-917d-5fc69726e770

and Thérèse Rivière (1936). Many published records were also collected from domestic and foreign labels, such as *Victor, Zonophone, His 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. In 1937, the *Musée d'ethnographie* became the new *Musée de l'Homme*.

Fig. 1. A 78rpm recorded in Romania by **Brăiloiu**, 1937. Photo by Jean-Marc Fontaine, 2009

A new ethnographic expedition crossed the central part of the African continent in 1946 and brought back musical recordings by Gilbert Rouget (Ogooue Congo Mission[4]), that were added to the collection of the *Musée de l'Homme*. Around the same time, the museum began releasing their collections of fieldwork recordings to the public on 78 rpm records[5] under the label "*Musée de l'Homme*" (Figure 1). In 1952, Rouget recorded the first tapes on the field, and since 1954, around one hundred recordings were published as 33 and 45 rpm records[6] by editors like *Vogue Contrepoint* and *Boîte à Musique (BAM),* as well as different scientific institutions such as the Peabody Museum of Archeology and Ethnology (Massachusetts, USA) and the *Institut Français d'Afrique Noire* (IFAN). The *Département d'Ethnologie Musicale* became the *Laboratoire d'ethnomusicologie* (Laboratory for Ethnomusicology) of the CNRS in 1968. Simultaneously, a sound laboratory was established for frequency analysis, with a Stroboconn in the 1950s and a Sonagraph in the 1970s.

[4] See Gérard, 2012
[5] http://archives.crem-cnrs.fr/archives/corpus/CNRSMH_Editions_001/
[6] http://archives.crem-cnrs.fr/archives/corpus/CNRSMH_Editions_002/

The museum's collections later became the *CNRS – Musée de l'Homme* Archives. Between 1988 and 2001, 37 CDs were published under the label *Chant du Monde*[7], such as "Anthology of Voices of the world" (1996) and "Musical instruments of the world" (1990). At 1999, hired in the staff and awarded about the insubstantiality of the tapes, I started the digitization of some emblematic collections and the indexation of the native digital recordings. Later, the French Ministry of Culture supported a digitization program to continue this huge work.

In 2009, the *Musée de l'Homme* was closed for renovation. The audiotapes were transferred to the French National Library, where they were preserved in an air-conditioned environment and where the digitising could be finished. The *Laboratoire d'ethnomusicologie*[8], renamed as the Research Center for Ethnomusicology (CREM), moved with the records collections to the *Université Paris Nanterre*. Over the years, collections of unpublished audio materials have been in constant growth, today reaching a rate of about 30 collections deposited annually, which represents between 1,000 and 5,000 audio items referenced in the database. Thousands of field recordings are deposited by collectors (or by their heirs): wax cylinders, celluloid records, magnetic tapes, DAT, together with descriptive documentation, more or less detailed, and dissemination agreements. However, the temporal nature of audio-visual materials raises specific issues of preservation and accessibility. Although digital formats can be duplicated without loss of information, obsolete analogue formats are fragile and technology for accessing content on these formats is disappearing quickly.

To handle issues with changing technologies, since 2007, the CREM has been implementing a strategy to classify and preserve these documents with an online sound database. This is the latest step taken in an almost century-long process of audio document diffusion and progressive adaptation to new technologies. Moreover, organizing the digitalized archives into a standardized system fulfils a mandate from the French Ministry of Culture and the CNRS to provide access to digitized content on the Internet. Since 2011, these archives are available through a web-based platform and represent a model for online collaborative tools (http://archives.crem-cnrs.fr). Today, 57,900 items from 6,100 collections are catalogued on this database. As of October 2016, more than 37,200 sound files have been uploaded[9], among which about 22,100 are available in streaming to the public without user count. With an average of 2,500 single visitors each month who consult the platform, this database serves as a case study for the online diffusion of music archives.

2. Telemeta: a new open and collaborative web audio platform

2.1 Specific features of the web platform

Telemeta is an open-source software platform designed by the Parisson Company[10], which facilitates the management and indexing of sound files in the audio database (Figure 2). As

[7] http://archives.crem-cnrs.fr/archives/corpus/CNRSMH_Editions_003/
[8] See Rouget, 2004
[9] In 2020, 53,800 sounds and videos are uploaded.
[10] http://www.parisson.com/

such, this digital assets management system (DAMS) is free and can be downloaded online[11], embracing a philosophy of knowledge sharing. Sounds are available for listening in a compressed format delivered from a streaming server.

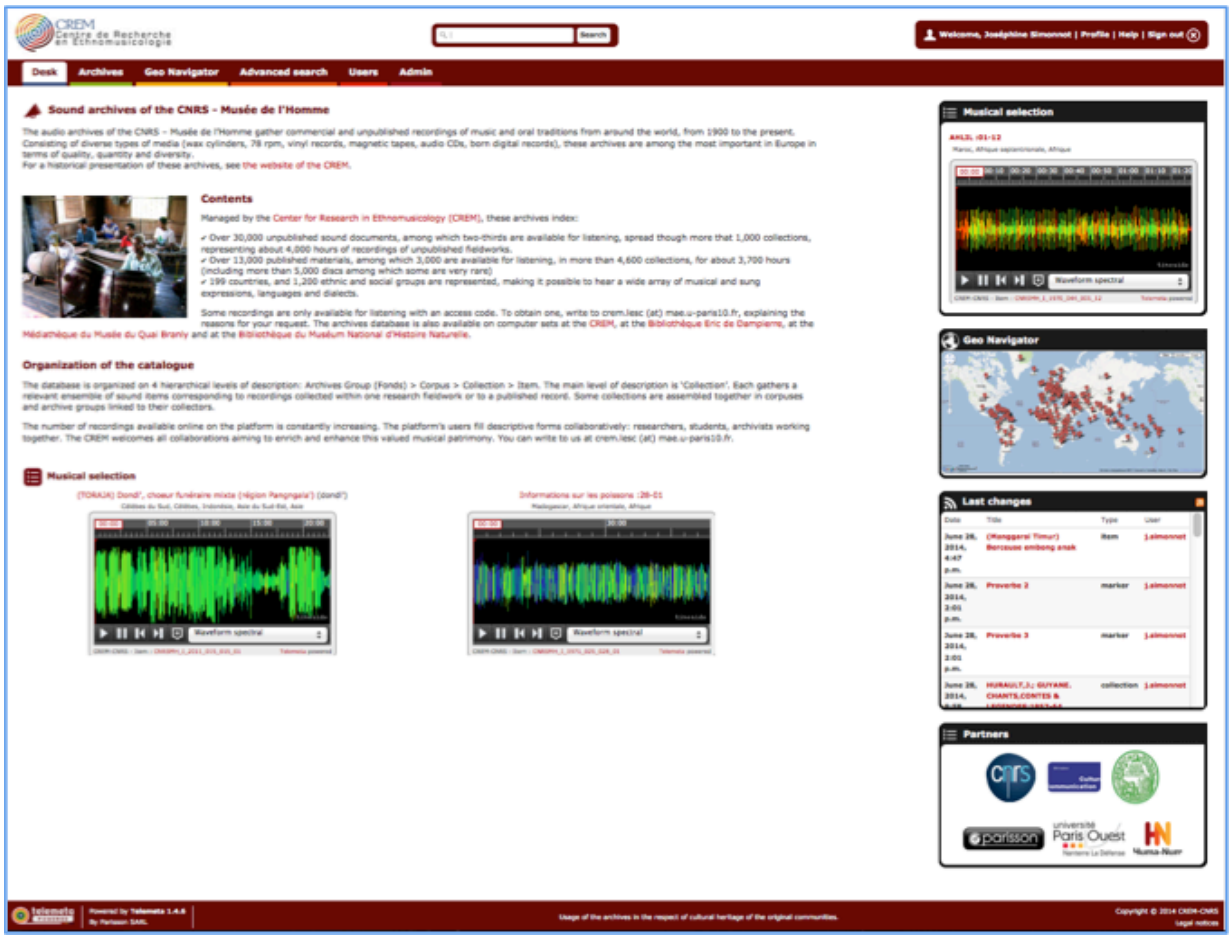

Fig. 2. CNRS – Musée de l'Homme audio archives' home page

Furthermore, sound can be visualised with a dynamic audio player using *TimeSide*[12], an open web audio processing framework. Various graphical representations can be chosen, such as the waveform or the spectral view. These are useful for spotting speech or music sections and for navigating inside a recording (Figure 3).

The *CNRS-Musée de l'Homme* platform allows the enrichment of the sound database by researchers, students, archivists and people with special knowledge who can edit the metadata with a user count, according some usual rules, given by the CREM staff. In particular, they can add historical, contextual or analytical comments to the available metadata. They can also add to the audio recording time-embedded markers and associated comments, thus contributing to the displayed information according to their own expertise. Contributors are also encouraged to provide pictures and references (publications, links). The monitoring of provided

[11] http://telemeta.org
[12] https://github.com/Parisson/TimeSide

information is based on trusting that authorised contributors are driven by an honest willingness to share their knowledge.

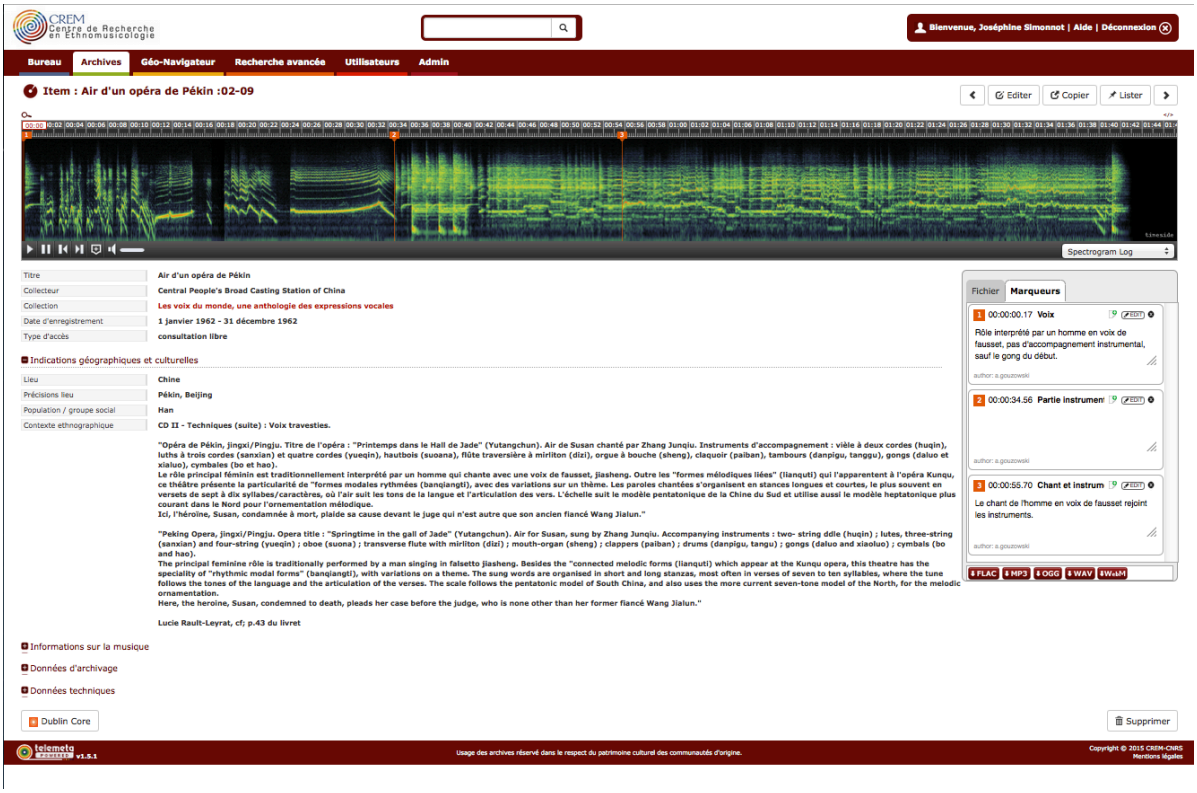

Fig. 3. A song from Opera of Beijing; spectral view with markers and comments

Beyond the collaborative aspect of the database, the long-term preservation of the archives on a server is also a strong argument to encourage contributors to work on the documentation related to their recordings. High definition copies of the sound content from the database are kept and backed up daily on the server of the CNRS Infrastructure for Digital Humanities, *Huma-Num*[13], along with other scientific content such as nuclear or astronomical data. But only compressed sounds are accessible by users thanks to the platform. Online availability for publication is another motivation for depositing material (i-frame, QR code or url link embedded). An RSS (Rich Site Summary) flux automatically sends out updates about modifications and additions to users who subscribe to it. Thus, it is easy for anyone to follow the changes that have been made to the database. Today, many researchers from all over the world wish to deposit the personal archives they have collected throughout their careers in order to keep a record of these unpublished materials.

Written in Python and Javascript languages, the sound archives metadata are based on an open-source relational database management system, called MySQL (Structured Query Language). In order to contextualize the collections, CREM has chosen to structure its catalogue in 4 levels: Archive Series, Corpus, Collection, and Item. The *Item* is the minimal unit, the individual sound file with its own set of information. Items are gathered in a

[13] http://www.huma-num.fr/

Collection, which is the main level of entry into the database. Each *Collection* can appear as a published or unpublished set of music recordings from a field study. Collections deposited by an individual collector or institution are organized into separate *Corpuses*, each one in reference to a theme, such as an ethnic group, a geographic area, or a set of fieldwork trips. These corpuses are gathered into *Series* (*Fonds* in French) that are named in reference to the collector or the depositor of the audio archive. Detailed profiles are provided for the latter two levels of the catalogue, introducing the work of the collector or institution with attached documents when available, embedded short audiovisual materials and hypertext links to relevant web pages. For instance, see the page of Mireille Helffer who works on Tibetan music (http://archives.crem-cnrs.fr/archives/fonds/CNRSMH_Helffer).

Of course, the design of this platform is a financial investment, but as an open source software, other government (or non-government) organisations can use it without incurring the first costs. This is a "sustainable" investment because all the IT data are available and the digital humanity community can use and improve it. The software is under Affero GPL Licence (Free Software Foundation). The entire database could also be exported by its owner to be reuse and, in this way, we send a selection of our metadata each semester to the Europeana Library aggregation.

2.2 Ways to access the *CNRS - Musée de l'Homme* sound archives

The Telemeta platform offers different paths to access a sound or video item. The home page offers access by geographical region, a random selection of recordings leading directly to each audio document's page and a direct access to all free access items (Figure 2). The user may browse with a simple search function or using the "advanced research" with which a user can select a particular country or ethnic group, a collector, a type of instrument, a title, the year of recording or of publication, and so on. Each sound recording has its own page containing audio and written metadata. Numerous text panels are dedicated to the listing of all information available about each recording in fields such as the date and place of the recording, its context, and the various instruments or voices heard, as well as information about the format and musical analysis. Archivists can use commonly accepted norms, such as the International Standardization Organization for languages (ISO 639). The names of geographical locations are standardized using the integration of *GeoEthno*[14] and *GeoNames*[15] thesauruses, allowing Telemeta to manage the different historical names of a location (Dahomey is the former name of Benin, for example).

The level of access to the collections depends on the status of a user. Users with **administrator** profiles can access every element on the database, download audio files and assign a selection of authorized actions to each user profile. Archivists have a specific profile, which allows them to organize the catalogue and ontologies, and to integrate new documentation. It also allows for the uploading (audio and video) and downloading of recordings as compressed (MP3 & OGG) or uncompressed formats (WAV, FLAC).

[14] http://www.mae.u-paris10.fr/dbtw-wpd/bibliotheque/g/MON.html
[15] http://www.geonames.org/

The **researcher** profile (with a login) can access all audio archives and has the ability to edit metadata and to annotate the audio files, in keeping with the collaborative nature of the database. A user's browser sets the language in which the platform is displayed (French, English, Chinese, Arabic or German). This profile also includes the possibility to create personal lists in which the user can save selections of items. Such an option is particularly helpful for organizing playlists for conferences or courses, arranging a template for the publication of recordings, and gathering different musical recordings the user wishes to use for research. The aim is to fit to the researcher's, needs, allowing them to work on their archives.

Visitors (users without a login) have access to all the metadata of the catalog. In addition, they can listen to all free access recordings: records published by the *Musée de l'Homme*, and all recordings that the collector and musicians agree to disseminate online. Any visitor can listen and visualise these recordings or export the audio player into external web pages through an i-frame html link. Bloggers can share recordings from the database in that way. However, the visitor profile does not allow downloading sounds (only streaming) or editing metadata. Statistics about which recordings are shared on which blogs or websites are recorded, allowing the administrators to keep track of the audio files circulation. This form of online access can also be provided to organisations through IP addresses to allow broader dissemination of all audio materials without password, as in *Musée du Quai Branly* or *Musée de l'Homme*.

The usage options of the different profiles are regularly reassessed, based on surveys about user experiences. Thus, the database is in constant evolution. The next step is to create a working group for sharing annotation during a work in progress.

2.3 Access and intellectual property rights

The *CNRS – Musée de l'Homme* archive allows for the online streaming of a part of their sound recordings. Access to the metadata catalogue is free. Nonetheless, restrictive options allow for control over the distribution of the audio documents with login-controlled access profiles. Through such decisions, numerous questions related to both intellectual property and ethical issues of some recordings have emerged.

Oral culture that is transmitted from one generation to another often has no identifiable author or composer. Moreover, most archives, particularly the oldest ones, do not mention the names of performers. In such cases, under the legal system in France, the person producing the recording is not the sole owner of the sound document. The content officially belongs to the performers, even if unknown, and to the institution that financed the fieldwork. This status gives the producer rights over the recording as an edited object while performers remain the owners of the music. Until 2015, these performer and producer rights lasted for a period of 50 years following the recording date. After this period, the recording enters the public domain. But the European law change recently to become 70 years after the recording year. Only collections recorded before 1963 stay available, unless there are specific ethical issues (e.g. recordings containing a secret ritual). Therefore, recordings done since 1963 will be available

in 2033, according the present law. Maybe it is a good deal for music producers or lawyers. However, for academic archivists, that just means much more work: how to contact every collector or performer from all over the world? How to convince public institutions to pay for digitising this restricted data? Furthermore, most of the time, the communities demand free access to their own heritage. They have no restrictions, especially for non-commercial use, as is the case of our institutions.

Thus, each collection has a level of access specifying whether it can be accessed with or without the sound, according to the current intellectual property rights and the depositor's wishes. Moreover, the CREM committee has recently decided to give full access to all the records published by the *Musée de l'Homme*[16] that are already online, although the editor (*Chant du Monde/Harmonia Mundi*) stopped the distribution of CDs 15 years ago and most of the records are out-of-commerce.

Since 2014, the CREM has adopted the *Europeana* model concerning rights related to sounds. *Europeana Sounds*[17], a project of the European Commission, is a consortium of digital libraries that aims to provide access to Europe's sound and musical heritage. The main guidelines are that all metadata will be licensed under the public domain and that institutions are encouraged to use the rights statements of the Creative Commons (CC0, CC-BY and CC-BY-SA)[18]. The tagline of the CREM, aligned with *Europeana*, is: "Public domain is the rule, copyright is the exception." Thus, metadata provided to Europeana Sounds could be reused and sounds under the public domain would be available with free access, but with a policy of "no re-use".

Open-access is also applied to some recent collections: researchers and depositors are increasingly aware of the benefits of web sharing. On the one hand, intellectual property is considered to be perpetual and inalienable in France. No one is allowed to claim ownership over a recording produced by another individual or to have any commercial use of it, even if the data is published online. On the other hand, online accessibility has changed the way traditional music is shared in an academic context. This allows interactions with the recording's performers in the local communities: it is an opportunity for ethnomusicologists to provide original performers access to content. The benefit of the tool is to introduce a collaborative exchange between researchers and musicians around their practices. For example, Dana Rappoport, an ethnomusicologist at the CNRS, has provided access to all her recordings[19] of vocal music made in East Indonesia (1992-2012), which represent about 2,210 audio items, with the agreement of the local people. In doing so, she expresses the local community's wish to preserve and share this vanishing heritage. Recently, after the online discovery of their forgotten ritual, the Toraja performers have requested that all data should be available and propose their contribution as transcribing song texts. Furthermore, they are very grateful for

[16] http://archives.crem-cnrs.fr/archives/fonds/CNRSMH_Editions/
[17] http://www.europeanasounds.eu/about
[18] http://creativecommons.org
[19] http://archives.crem-cnrs.fr/archives/fonds/CNRSMH_Rappoport/

the preservation of their almost lost heritage. With the recordings online, they could process a re-appropriation of their ritual tradition and teach the young people.

3. Contribution of the new technologies to sound databases and research

Soon after the release of this platform, two other sound databases began using the Telemeta framework to host their recordings. One is the Laboratory for Musical Acoustics (LAM, UPMC/CNRS) [20], which is using this system to organise sound archives from musical instruments in order to study their acoustic properties. Another database, *Scaled Acoustic Biodiversity*[21] (Sabiod), is a consortium of research departments involved in studying the audio signals of marine animals. Since 2015, the *Centre des Musiques Arabes et Mediterranéennes*[22], hosted in Sidi Bou Said, Tunisia, had implemented an Arabic interface of Telemeta to manage the Tunisia National Phonotheque.

Although the overall architecture remains the same, numerous elements were adapted, such as the organisation of the metadata. For archivists, the compatibility with the Dublin Core metadata format allows the content of their databases to fit the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Therefore, information stored on the database can be referenced on web search engines dedicated to the digital humanities, which ensures the distribution of data.

Online platforms, such as *Isidore*[23] (the French meta-portal for digital humanities) or *Europeana Sound,* harvest the free-access archives provided by the *CNRS-Musée de l'Homme*. Launched in 2014 and sponsored by the European Commission, this project will give online access to a critical mass of audio-visual digital objects. In 2017, about a million items will be available via *Europeana*, from classical and folk music, to environmental sounds and oral histories. The project, coordinated by the British Library, brings together 24 partners (national libraries, research centers and universities) from 12 European countries that will work together on a new structure for audio data: the European Data Model for Sounds (EDMS[24]).

Since 2011, when the Telemeta platform became operational, academic and public use of the database has increased. Projects concerning sound databases also increased, providing wider financial support. Beyond the scope of ethnomusicology, researchers such as anthropologists, linguists, musicians and acousticians find elements for their own work. Students have conducted research by using collections, including research from a diachronic perspective (Khoury 2014, Lacombe 2013). Publications in many formats, whether books or

[20] http://telemeta.lam.jussieu.fr
[21] http://sabiod.telemeta.org/
[22] http://phonotheque.cmam.tn/
[23] https://isidore.science/
[24] Europeana Data Model Profile for Sound : http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/EDMSound//TF_Report_EDM_Profile_Sound_301214.pdf

articles, and even museum exhibits, can include audio illustrations in reference to specific recordings archived on the database by embedding a URL link or a QR code, as in Gérard (2012).

Numerous unanticipated uses of the platform have emerged, including teaching and specialised 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* to illustrate points raised in class and to help students practice musical analysis and transcription. Scientific blogs with embedded sounds from the database are published such as the *Jabal al-Lught*[25]*, Climbing the mountain of languages*. Dr. Lameen Souag, a CNRS researcher in North African languages, present an analysis of a recording in the early 1950s in Algeria (Korandjé tale (قصة ب), with a transcription in Korandjé, Arabic, English and French. He writes:

> Champault's recordings have recently been made available online by the Centre de Recherche en Ethnomusicologie, through Cécile Funke's archival work. Many are in Arabic or French, but the Korandjé ones are an irreplaceable resource for the study of the language; in her time, the language was under rather less pressure and verbal arts were in much better health (Jabal al-Lught, 2014).

Because the open-access policy applies to a large number of recordings, the archive platform makes musical knowledge accessible outside of the academic sphere, allowing individuals to expand their knowledge (see, for example, the blog dedicated to 78-rpms, *Ceints de Bakelite*[26]).

Another community of users appeared on Telemeta last year: researchers in Musical Information Retrieval (MIR[27]). These researchers require many musical recordings to improve algorithms for the analysis of audio content. This use of Telemeta generates innovative research programs at the forefront of new technologies in the digital audio field. Since 2013, new analytical tools have been developed through the *DIADEMS* project[28] (Description, Indexation, Access to Ethnomusicological and Sound Documents), a French national project involving several CNRS departments specialising in computer sciences, speech and music detection. The objective of *DIADEMS* is to implement new options on the Telemeta framework and improve access to the musical data by users. In his thesis, « Spectre de rythme et sources multiples: au cœur des contenus ethnomusicologiques et sonores », Maxime Le Coz (2014) present results of some analysis tools implemented in this project. Since 2015, new tools are tested by the CREM and the Museum staff on a sandbox[29]. Now, to index and to segment the audio content, we are being helped in this long and careful process with on line tools, like start recorder detection, speech/singing detection, monophonic/polyphonic parts detection, etc. With these analysis tools and annotation markers, the CREM experiments with new collaborative work

[25] http://lughat.blogspot.fr/2014/03/korandje-tale-conte-en-korandje.html
[26] http://ceintsdebakelite.com
[27] http://www.ismir.net/index.html
[28] https://www.irit.fr/SAMOVA/site/pagediadems.html
[29] http://diadems.telemeta.org/

between the sound engineer and the archivist to identify the content and to segment the sound files in an efficient way. The next step is the proposal of specific ontologies or tagging to describe the categories and to manage the public access to annotations and enrichments of some work in progress from researchers. Exploring all the functionality of such analysis tools will be a long road. But, without doubt, the music information retrieval community is deeply interested in our anthropological material, looking for a "real life" sound database to build the audio research engine of the future. They have come to us with providential financial support to save our endangered old tapes.

Conclusion

Fifteen years ago, when the production and dissemination of CDs were economically difficult, the main concern was providing access to all individuals interested in the music archive of the *Musée de l'Homme* and facilitating work and study with these documents. Today, the digitisation of music and its publication online has facilitated the exchange of music on a massive scale never before imagined. Thanks to the Internet, new uses and new opportunities are available to scholars, and to the performers and their communities. The contemporary mechanisms for music sharing also create innovative research approaches and present new opportunities for connections between oral culture disciplines.

New challenges also appear, such as questions of intellectual property in an increasingly commercial world, even when culture and science are considered to be common property. Governments have always sponsored academic institutions that collected our musical heritage. Now, new tools exist to preserve, disseminate and study this knowledge if stakeholders agree to sustainably support the necessary technological infrastructure. An exciting fact is that the world's musical heritage is no longer limited to a few specialists. Everyone can discover these resources, even local musicians and their descendants.

Acknowledgements

This work is partly supported by a grant from the ANR (French National Research Agency) with reference ANR-12-CORD-0022.

References cited

Le Coz, Maxime, 2014. « *Spectre de rythme et sources multiples : au cœur des contenus ethnomusicologiques et sonores* ». École doctorale Mathématiques, Informatique, et Télécommunications de Toulouse. http://publications.ut-capitole.fr/16314/

Gérard, Brice, 2012. "Gilbert Rouget et la mission Ogooué-Congo (1946). Institution et épistémologie dans l'histoire de l'ethnomusicologie en France". *Gradhiva*, 16: 192-215.

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 valorisation des enregistrements d’archives de xylophones fang de 1908 à nos jours”. Master thesis in ethnomusicology, Université de Paris Ouest-Nanterre

Rouget, Gilbert, 2004. “Le Département d’ethnomusicologie du Musée de l’Homme. Maison mère de la discipline en France et dispositif en péril”, *L’Homme*, 3-4, N° 171-172: 513-523.