AR 3209

Frédéric Lemmers

Digitizing sound archives at Royal Library of Belgium. Challenges and difficulties encountered during a major digitization project

Abstract: Music in general and recorded music in particular are rarely a priority for libraries' digitization policies, although wax cylinders and 78rpm discs might be digitized for preservation and accessibility reasons. The respect of the original recording technique during the digitization process will ensure the scientific and artistic credibility of the digitized sources. The Royal Library of Belgium started in 2016 the digitization of its whole collection of 78rpm. Realized by subcontracting, this project of about 4,000 hours will constitute a large corpus of sources for the digital musicology upcoming needs.

Keywords: Sound recordings; digitization; user expectations; audio-turn

Zusammenfassung: Musikalien im Allgemeinen und Tonträger im Besonderen erhalten in bibliothekarischen Digitalisierungsstrategien häufig nur wenig Beachtung, obwohl gerade für Wachszylinder und Schellackplatten sowohl aus Gründen der Bestandserhaltung als auch zur Verbesserung der Zugänglichkeit ihre Digitalisierung dringend geboten wäre. Um bei der Retrodigitalisierung von historischen Tonaufnahmen künstlerisch und wissenschaftlich zuverlässige Ergebnisse zu erreichen, ist den originalen Aufnahmetechniken große Aufmerksamkeit zu schenken. Die Königliche Bibliothek Belgiens lässt seit 2016 ihre vollständige Sammlung an Schellackplatten digitalisieren. Mithilfe eines Dienstleisters werden ca. 4 000 Stunden Tonaufnahmen produziert, die für aufkommende Forschungsfragen der digitalen Musikwissenschaft einen gewichtigen Quellenkorpus darstellen.

Schlüsselwörter: Tonaufnahmen; Digitalisierung; Nutzererwartungen; Audio-Turn

1 Introduction

Patrimonial institutions in general and libraries in particular are really active in the field of digitization since the beginning of the years 2000. With the internet revolution, patrimonial institutions founded new opportunities to ensure both conservation and accessibility missions of the rich heritage they have

to preserve, to manage and to transmit to the next generations. Henceforth, the systematic description of digitized items within well-structured online databases and their indexation by search engines offer to the world wide community of researchers and enlightened amateurs the possibility to discover relevant documentation and artistic sources from all over the world, directly from home.

Nobody ignores that digitization projects are very expensive and considerably time consuming activities, especially for the documentation sector (libraries and archives) that are counting their collection in neither current meters nor kilometers and in millions of pages. Logically, libraries gave the priority to their fragile and damaged collections of historical newspapers and printed books from the 19th century.[1] At the same time, many of them intended to highlight the wealth and the diversity of their printed and handwritten collection by realizing significant digitization projects of treasures. Within this context, everyone will understand that it was very difficult to elect musical collections in general and collection of sound recordings in particular within libraries digitization programs. It easily explains the reasons why digitization of sound has been postponed for many years by many libraries in the world.

When libraries, archives, documentation centers and museums started to digitize their collections of music, they decided to select as a priority their collections of music manuscripts, unique or rare printed music books, music iconography or music instruments. This is just the logical consequence of the occidental tradition of writing and its hierarchy of sources. On the European level, the first digitized corpus about music to be available in EUROPEANA[2] concerned musical instruments. By bringing together metadata and digital reproductions of the most important collections of music instruments conserved in Europe, the MIMO project aimed to create a "single access point to digital content and information on the collection of musical instruments held in European museums".[3] Ten years later, a similar project has been set up with focus on sound material, including recorded music. Led by the British Library with a consortium of many national libraries and sound archive institutes, the project EUROPEANA SOUND aimed to select, to describe, to digitize and to publish online in EUROPEANA a huge corpus of sound recordings, with the goal to "double the number of audio items accessible through EUROPEANA to over 1 million and improve geographical and thematic coverage by aggregating items with widespread

---

[1] See Jeanneney, Jean-Noël: Le programme de numérisation de la presse française. In: *Le Journal dans tous ses états – Het rijk van de krant*, ed. by Marc D'hoore, Dirk Luyten and Thierry Delplancq, Archives et Bibliothèques de Belgique, Brussels, 2007, vol. LXXVIII, (1-4) 165–73.

[2] At the very beginning, the Europeana initiative aimed to set up a European Digital Library with the collections of European national libraries; launched in 2008, it became immediately a more extensible project which would include digitized sources from the whole European cultural heritage sector including text, image and audiovisual material. See online www.europeana.eu.

[3] See EUROPEANA Pro website: https://pro.europeana.eu/project/mimo; the MIMO project webpage http://www.mimo-international.com/MIMO.

popular appeal such as contemporary and classical music, traditional and folk music, the natural world, oral memory, and languages and dialects".[4]

This project was also an opportunity for EUROPEANA to launch a dissemination portal dedicated to music.[5]

Next to these European initiatives, it has to be reminded that several big institutions realized significant digitization projects on sound, e.g. the Bibliothèque nationale de France and its amazing collection of 400.000 digitized sound recordings that are now available on GALLICA.[6]

The situation is not so different in Belgium, where the main collections of sound recordings are conserved by patrimonial libraries and broadcasting companies. By counting about 80.000 items, the collection of the Royal Library of Belgium is one of the most important ones.[7]

Music collections of the Royal Library of Belgium were not included in the list of priorities that should be digitized within the first part of the federal digitization program (2004-2012).[8] For this program, the Royal Library of Belgium has pointed out two absolute priorities: the retro-conversion of about 1.5 million typewritten descriptions of its collections with the goal to enrich its general catalogue of descriptive metadata; and the digitization of 30 titles of Belgian historical newspapers from the period 1830-1950, that were in danger of self-destruction due to the acidification of the paper.

Initiated administratively in 2005, these two projects operationally started in 2008 and remained running until 2012. Considering the long duration of these projects, the Belgian federal government decided in 2010 to set up a public private partnership for the five next years with the goal to increase the number of digitization projects realized within this new period. Scientific institutes had been asked to update their priorities on this occasion. Music in general and sound recordings in particular appeared on the list of the Royal Library's priorities. At the same time, the Royal Library created its own centralized digitization

---

[4] See EUROPEANA Pro website: https://pro.europeana.eu/project/europeana-sounds ; and the EUROPEANA SOUNDS Project website: http://www.eusounds.eu/about.html.

[5] https://www.europeana.eu/portal/fr/collections/music.

[6] This project is the result of successive digitization programs among them a huge partnership between the Bibliothèque nationale de France through its affiliated branch BNT-PARTENARIATS, MEMNON Archiving Systems and BELIEVE DIGITAL, which aims to digitize and to provide online access to a selection of about 200.000 discs published between 1900 and 1962 and conserved at the library (See "Synthèse des premiers accords conclus par BnF-Partenariats", Paris, 28 juillet 2014, en ligne: http://www.bnf.fr/fr/la_bnf/anx_actu_bib/a.partenariats_numerisation_bnf.html#SHDC__Attribute_BlocArticle0BnF).

[7] For a description of the collections, see the library's website: https://www.kbr.be/en/collections/music.

[8] See Mettens, Philippe: Het Belgisch digitaliseringsplan voor de Federale Wetenschappelijke Instellingen en het Koninklijke Belgisch Filmarchief. In: *La numérisation du patrimoine écrit: du virtuel à la réalité – De digitalisering van het geschreven erfgoed: van denkbeeld tot feit*, ed. By Marc D'Hoore, Frédéric Lemmers & Dirk Leyder. (Archives et Bibliothèques de Belgique – Archief-en Bibliotheekwezen in België, 2011, vol. LXXXII (1-4) 15–30.

department with the goal to coordinate the implementation of the digitization strategy of the institution through internal production and with subcontracting. Although this digitization department has been well equipped with several and diversified scanning stations, the library decided not to purchase specific digitization equipment for audio collections and not to commit any specialized staff members like sound engineers. Subcontracting would be the per default principle for all digitization projects about sound recordings. Although this ambitious program was never funded, Belgian musical heritage was henceforth written on the digitization plans of the institution. Implementation would become a reality within the third federal program, called DIGIT-03 (2013-2018). Indeed, within this actual running program, the Royal Library of Belgium is digitizing and publishing online its well-known FETIS collection and its collection of historical sound recordings (1900-1950).

2   Sound recordings, an heritage in its own

In contrast to printed and handwritten music material that can be read without any machine sound recordings cannot be read on their own. The musical execution that is fixed on the source is not intelligible on its own; a reproduction machine is needed to display it and its characteristics will depend on the quality and technical specificities of the reading installation. Sound recordings need to be read by appropriated material to deliver the information they are fixing. This particularity has great impact on the digitization. By constituting the ultimate analogue reading of the original recording, conditions of the digitization have great impact on the result and on the properties of the sound provided on digital format.

For that reason, the major characteristics of the sound recordings over time will be summarized, from the Edison age until today. The intention is certainly not to make the history of sound recording but well to highlight some major aspects of the diversity of recording techniques and recorded supports to illustrate the heterogeneity of this heritage and the related problematics for digitization projects. Indeed, a good understanding of this historical, technical and commercial context of the sources is a key success factor for audio digitization projects.[9]

The history of sound recording covers a period of about 140 years of technical and industrial development, since the invention of the phonograph by Edison in 1877. This evolution can be looked at

---

[9] See Oliver Read & Walter Leslie Welch, *From Tin Foil to Stereo: evolution of the phonograph*, New York, Bobbs-Merrill, 1976; Timothy Day, A century of recorded music: listening to musical history, New Haven, Yale University Press, 2000.

through different angles: the recording techniques and recorded supports, the conditions of listening to sound recordings or the recorded repertoire in function of the evolution of bandwidth.

Considering the recording techniques and recording supports, three successive periods have to be kept in mind: the acoustic era (before 1925), the electric era (1925-1982) and the digital era (since 1982). Considering the physical supports that have been commercialized at these three periods, there was a huge diversity of supports and formats. From about 1893 until 1912 (1929 for Edison's production), cylinders in wax or celluloid have been produced in parallel with shellac discs. If several formats of cylinders coexisted, the most frequent ones had a length of 9.5-11 cm and a diameter of 5.5 cm. They fixed an acoustic audio signal recorded by vertical etching. Most of these turned on 160rpm and previewed a recording duration of 2-3 minutes. This very short duration had consequences on the recorded repertoire: vocal music was resumed as short arias and songs with arrangements for the longest ones. Although they appeared at the same moment as cylinders, shellac black discs were produced during a longer period (1890-1950) on several formats (17-25-30-35 cm). Originally recorded acoustically by vertical etching (before 1925), shellac discs progressively benefited from the appearance of electricity within the recording process and became systematically recorded by lateral etching (after 1925). If, before 1925, their recording speed was very variable from one label to another, with a speed comprised between 60 rpm and 120 rpm, most of the labels fixed it to 78 rpm after 1925. This standardization of the speed is a very important factor to take into consideration within a digitization process as the speed will determine the duration of the recording and then, in music, the diapason of the execution of the recorded work. Practically, it means that the correct speed has to be identified before the ultimate analogue reading of an original sound recording or corrected after digitization by audio correction to avoid the production of any wrong or false document, historically or artistically. Concerning their duration, shellac discs offered recordings of 3 to 5 min by side. This augmentation of the duration had positive consequences on the recorded repertoire: most of opera arias became recordable on one side. Nevertheless, complete operas or symphonies remained complex to commercialize on recording as it requested an important number of sides.

Fig 1: Example [of 78rpm electric disc form the year 1928] Richard Wagner "Siegfried": "Réveil de Brunnhilde" [recorded in French translation] (Act III), Laure Bergé (Belgian soprano from La Monnaie, Brussels), Grand orchestre symphonique, dir. Bruno Seidler-Winkler, Berlin, TRI-ERGON TE10036 matr. 3343.

After the Second World War, shellac discs continued to be commercialized for several years but started to be progressively replaced by microgroove discs in vinyl. At the beginning of the 1950's, this new support was the only one remaining. This revolution was a starting point for the systematic recording of instrumental music and orchestral lyric works because of the upgrade of the quality of the recordings and the extension of time duration by side. Commercialized on several formats (17-28-30cm), LP's were only recorded with electricity by lateral etching. Their speed was immediately standardized to 33rpm or 45rpm and their duration reached 20 or 30 min by side in function of their format. Thanks to this technical evolution, the recorded repertoire increased significantly in wealth and diversity for including complete operas, complete instrumental chamber music works and symphonies, etc.

Up to the beginning of the years 1980's, LP's coexisted with magnetic tapes (1950-) and mini-tapes (1963-), that are also very interesting sources to fix music archives (unpublished).

The digital age started in 1982 with the appearance of the compact disc, followed by the mini-disc (1992-2011). This little disc of plastic was able to fix a digital signal which was readable by laser. As there was no more needle to read the recorded signal, friction noises completely disappeared from the listened sound. Thanks to this innovation, many discographic labels published remastered anthologies of previous recordings.

At the end, it becomes clear that sound recordings, commercial editions and unpublished sound archives have to be considered as heritage on their own. They constitute a special kind of patrimonial source which is reflecting many aspects of the musical life in the 20th century. First of all, they are the physical testimonies of the development of a new musical industry, which contributes to disseminate music out of the walls of concert halls or churches. This new segment of the musical life entered into competition with domestic musical practices. Thanks to recorded music, it suddenly became possible to access music at home without being able to play an instrument.

This heritage keeps for the upcoming generations a memory of the tastes of the last 100 years of musical life in terms of repertoire as well as of stylistic practices. For the first time in the history of music, musical performances could be documented and kept for the future.

Finally, it constitutes the fundamental source of certain musical practices especially for popular music or ethnomusicology, jazz, and for all unwritten music traditions.

For all those reasons, sound archives collections have physically to be preserved in good conditions; the performances they are fixing on their supports have to be saved digitally and given into access for the actual generation. In the digital world, to continue to give access inside reading rooms to physical analogue recordings is absurd and not responsible regarding the negative possible impact of these reading sessions on the long term preservation of these sources. Furthermore, this option is no longer an appropriate answer to the public expectations. Made in good conditions, digitization of historical recording sources will constitute the ultimate analogue reading of the original. This unique operation will ensure the preservation of the original support, the preservation of the recorded musical performance and make its access easier.

3   Listening historical sound recordings

Listening conditions for sound recordings have evolved over time differently than recording techniques. Indeed, acoustic recordings were produced to be played on acoustic phonographs with an ironic needle. They can also be played on an analogue pick up with an electronical cell. At the very beginning, electric shellac discs were played on acoustic phonographs despite their better sound quality and the huger specter of frequencies that they were able to fix. The quality of their listening was incredibly better after the arising of electric record players and the hi-fi stereo installations. Digital sound recordings fixed on physical supports (CD, DVD, MD, DAT) were played during two decades on dedicated digital players coupled to analogue hi-fi installations.

Although the first digital recordings arose in the years 1980's, recorded sound remained fixed on a physical support up to the years 2000's. Then, with the arising of the internet and the informatics networks, it became possible to dissociate the recording from its physical support and to save it as an electronic file on hard disks of a computer, a NAS, on storage or, more recently, in the "cloud".

This last technical evolution is probably the most important for the evolution of the conditions of commercialization of sound recordings and for the renewing of listening practices as it gives the same materiality to sounds, images, videos and texts. From one day to another, the commercialization of recorded music didn't need any physical support anymore. Also the display evolved a lot. The recording reader is no longer a physical tool but appropriated software which is generally available by default on electronic devices.

Thanks to the internet and the development of downloading platforms, it is now possible to purchase sound recordings online. Instead of getting a physical object, the public only downloads audio files directly from the internet or listens to it online without owning it, thanks to streaming possibilities (ex. iTunes, Amazon, etc.). People started to set up their own sound libraries on their own personal laptops. More recently, these databases can be shared with all mobile devices. Repertoires of references can be indexed by applications to disseminate music in a car or at home thanks to wireless sound installations (ex. SONOS).[10]

Today, everyone is expecting to be able to listen to his music catalogue everywhere and on all his electronic devices. To find a certain recording, the majority of the population is directly going online and believes that there is an access to everything. Although this perception is wrong, it perfectly illustrates the evolution of the expectations of the public of the libraries.

From a library perspective, this evolution of mentalities means that recordings that are not accessible online don't exist. But the titles recorded on old phonograms are generally not recognized by mobile applications like SHAZAM which is however able to recognize more than 15 billion titles of recorded music based on a short sample played.[11] The same ascertainment has been made with GRACENOTE,[12] a music database which iTunes especially is cooperating with to enrich music files with descriptive metadata. Both examples illustrate the gaps that are remaining existing in the world wide accessible recorded music catalogue.

---

[10] https://www.sonos.com/en/system
[11] https://en.wikipedia.org/wiki/Shazam (company)
[12] https://en.wikipedia.org/wiki/Gracenote

Libraries can contribute to fill-in these gaps by realizing digitization projects of recorded sources. This means that libraries have a responsibility in prioritizing digitization programs of their collections of recorded music because it is the only possible way to ensure a neutral and huge access to the memory of musical practices.

Libraries and patrimonial institutions have to understand this significant evolution if they want to reach their official mission. Sound extraction of their collection of physical sound recordings constitutes the best way to offer an access to the repertoire they fix. This reality constitutes one of the most valuable arguments to start digitization projects on sound recordings.

Until now, digitization projects concerning sound recordings were mainly developed to offer digital files on the internet via online catalogues and digital libraries but not via commercial downloading platforms. This marketing strategy assumes that clients of libraries are spontaneously going on libraries websites and their digital catalogues when they are searching for a certain title of recorded music. However, since a couple of years, it has been established that the majority of internet queries are made via GOOGLE.[13] On the basis of this observation, libraries that intend to catch a broad audience might change their digital strategy of valorization in general and for audio recordings in particular. Several options seem to be possible: they could negotiate with search engines a better ranking of their catalogues and webpages of, even better, the indexation of the content of their catalogue. They could also offer their digitized collections on downloading platforms. In addition to facilitate access to their collections of sound recordings, they would contribute at the same time to enrich the global online music catalogue.

4   The Royal Library of Belgium's sound recordings collections: a corpus of about 100.000 items

In Belgium, sound recordings are not covered by the legal deposit legislation (1965). Sound recordings and film material were even especially excluded from the obligation of deposit in the first legal deposit law.[14] If this formal exclusion disappeared in 2008 after the revision of this article 1, there is actually no executive disposition which organizes the obligation of deposit for sound recordings.[15]

---

[13] See Bruno Racine, *Google et le nouveau monde*, Paris, Plon, 2010.

[14] See art. 1 of the Belgian law on the Legal Deposit at the Royal Library of Belgium dd. 8 April 1965 (publication dd. 18 June 1965) (https://www.kbr.be/fr/loi-du-8-avril-1965-instituant-le-depot-legal).

[15] See art. 2 of the law dd. 19 December 2006 modifying the law of 1965 (publication dd. 23 March 2007) and its executive Royal Decree dd. 14 February 2008 (publication dd. 21 March 2008) (https://www.kbr.be/fr/loi-du-19-decembre-2006).

Nevertheless, the Royal Library of Belgium conserves the most important collection of sound recordings in Belgium. Largest parts of this corpus have been constituted by private collectors and have joined the library's collections within the last 15 years thanks to the King Baudouin Foundation. These private collections are reflecting a collecting life and their author's preferences and musical interests. These are covering a specific musical genre or a specific period of the sound recording history. Collection Yves Becko[16] is covering lyric singing in the first part of the 20th century in Belgium and France; Marc Danval's collection[17] is dedicated to the jazz in Belgium and the Clément Dailly's collection[18] aims to cover the huge diversity of the recorded repertoire of classical music since the arising of the LP's. If sound recordings constitute the most important significant part of these three collections, black shellac are also available within 20 several private music collections conserved at the Library.[19] The remaining 20.000 LP's and compact discs have been purchased by the library to compensate the lack of legal deposit.

5   A digitization project for which target audience?

If, in the past, digitization projects were essentially made and funded for preservation reasons, making content available online progressively became the first goal and the best argument to get funding for digitization programs. This general rule is also valuable for sound recordings. Even more, to facilitate access to this heritage is probably the strongest argument to justify its digitization.

Giving online access to sound recordings asks the question of the target audience. Which kind of public could be interested in listening to historic recorded sources that are by definition technically imperfect and sometimes unpleasant to listen to? Through which dissemination channels do libraries have a chance to reach the right target audience? Which criteria should libraries follow to define the technical expectations of the reproduction process?

A first main target group includes academics, researchers and students who could be interested in getting access to relevant data sets of digitized sound recordings for scientific research needs in the humanities, studies in ethnomusicology and in popular music as well as studies in musical practices and musical genres. Everyone who is interested in the study of the musical life since the end of the 19th

---

[16] https://www.kbr.be/fr/fonds-yves-becko.

[17] https://www.kbr.be/fr/fonds-marc-danval.

[18] https://www.kbr.be/fr/fonds-clement-dailly.

[19] For an exhaustive list of collections that are conserved at the music department of the Royal Library of Belgium, see the Music collection presentation webpage: https://www.kbr.be/fr/fonds-de-la-section-de-la-musique.

century can find a wide range of sources illustrating the diversity of musical genres and repertoires, artistic influences, evolution of performance practice.

From a scientific perspective, digitization projects of sound recordings need to prioritize the scientific credibility over the comfort of listening. Otherwise, their online offer would create a risk of falsification of the artistic and the historical variables of the recordings. Online published results should be neutral in its audio specifications. No restoration of the recorded signal would take place. Listening would be less comfortable as historical sources would present rubbing noises.

A second target audience group could be defined as people who are interested in music in general and who are spending their free time in surfing online in order to discover new pieces of music and rare interpretations. This second target group of amateurs could give more importance to the comfort of listening than to the access of the neutral extracted sound signal. The listening of an entire song could depend on the audio quality of the files. Accessibility through user friendly applications is a must. For this audience, the availability of the outcomes of digitization projects could be more appreciated on downloading platforms than on digital libraries of public institutions. This target audience is also interested in sharing their preferred recordings through social networks with their virtual community of friends.

Considering this second target group's expectations, libraries could focus their selection process on the recordings from the post-1950 period as they have a better sound quality; for the more ancient issues, libraries should foresee restoration of the audio signals by erasing original noise and clicks of the original recording.

To combine both target group expectations requires a delicate arbitrage. Priorities in the selection of the pieces are different in function of the interests of the target groups as well as sound quality levels expectations, what makes a very difficult balance to reach for a library.

Against this background at the Royal Library of Belgium, it has been decided to focus on the historical sources and the scientific credibility of the online proposed sources. In this approach, digitization only consists in capturing the recorded signal and saving it in digital format. No restoration is foreseen. Online published digitized recordings have to represent the material history of the works.

6 Criteria for selection and justification

At Royal Library, priority for digitization in general has been given to fragile and damaged supports, followed by patrimonial treasures, with the aim to highlight the large wealth and the diversity of its

collection.[20] Sound recordings are in danger if they are made available to the public on their original form. The rubbing of the needle in the furrow of the disks combined with multiple listening sessions constitutes a strong risk of degradation for historical sound recordings. For that reason, it has been decided that wax cylinders and shellac discs couldn't stay available for the public on their original form and that digitization was mandatory to make them accessible again.

Fig 2: Example [of wax cylinder] Jan Blockx, Princesse d'auberge, Maurice Decléry (baritone from La Monnaie), Edison 17010, published before 1914

Normally, there is no digitization project of sources that are not already described in the online catalogue of the library. This reason is very simple: digital files have to be linked to descriptive metadata sets to be findable by the public. Although sound recordings were not already described in the general catalogue, descriptive metadata already existed on an electronic form within some Excel file. Checking this metadata sets and converting them into the online catalogue are part of the preparation tasks for this project.

In general, priorities for digitization at Royal Library have to relate to the history of Belgium, its regions and its people. Danval and Becko's collections of shellac discs are following this rule as they are

---

[20] See the digitization criteria on the Library digitization webpage: https://www.kbr.be/en/digitisation

respectively related to the lyric musical life and to jazz in Belgium. Incidentally, both are also reflecting the importance of the contribution of Belgian musicians to the world musical life of the years 1880-1950.

Copyright and neighboring rights are also a very important element of digitization projects. As the library hasn't got any budget for right clearance, priority is given to out-of-copyrighted material. Copyright was a supplementary reason to prioritize the collection of cylinders and 78rpm discs to 33rpm and 45rpm LP's. Nevertheless, copyright remains a complex topic within this project as several works sound as unknown material or orphan works. The digitization department is checking rights for each item and registers search results. In some situations, diligent searches are made by following the legal procedure[21] and its results are registered into the European Orphan Database.[22] At the end, the Royal Library gives access online without any restriction to recordings that have been recognized as out-of-copyrighted or orphan works, while in-copyrighted material is made accessible only inside the library reading rooms.

7   Description of the operations: to scan, to control, to store and to disseminate

The digitization of the collection of 78rpm discs of the Royal Library takes place within the Belgian federal digitization program previously mentioned, whose . the third covers the years 2013-2018. The audio project of this program involves three scientific institutes: the Royal Library, the State Archives[23] and the Centre for Historical Research and Documentation on War and Contemporary Society CEGESOMA.[24] Articulated on three transversal platforms for digitization, long term preservation and valorization, this program includes commitment of extra staff members inside institutions and the use of subcontracting for some digitization projects that cannot be realized inside the institutions. Public tendering and financial administration are under the responsibility of the Belgian Federal Science Policy Administration while scientific monitoring of the project and technical operations are under the responsibility of the institutions.

The audio project requires the participation of a private company for the extraction of the analogue recording and its copy into digital files. The Belgian company MEMNON Archiving Services[25] has been chosen as the outcome of the public tendering process. Started in 2017, the outstanding project

---

[21] A Belgian law dd. 2016 has transposed the Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works into the Belgian legislation. Two Royal Decrees dd. 17 October 2016 (publication 29 November 2016) are respectively designating the Royal Library of Belgium as National Competent Authority for Orphan Works and defining the list of appropriate sources for diligent search.

[22] https://euipo.europa.eu/orphanworks.

[23] http://www.arch.be/index.php?l=en.

[24] http://www.cegesoma.be/cms/index_en.php.

[25] https://memnon.com/

represents 4.000 hours of recorded music, 30.000 shellac discs, 60.000 sides as much as cataloguing descriptions and has to be finished by 2020. With this project, the Royal Library intends to digitize its complete collection of 78rpm.

To meet its deadlines and reach the final project goals, the library monthly provides about 600 discs to MEMNON. With the help of the music department, a team of four staff members from digitization department is preparing the deliveries, by checking the state of the discs, the consistency of the metadata available on the Excel sheet before copying them into the general catalogue. After checking copyright issues, they digitize the central stickers of the discs by using a book scanner.[26] Scanned stickers are produced with the perspective to publish digitized audio recordings with a visual thumbnail. At the end of the preparation, discs are displayed into special boxes that will be transmitted to the subcontractor. A checksum list is accompanying each box and is sent by email to the project manager by MEMNON.

Three technical processes are possible for the digitization of 78rpm sound recordings: to take the sound of a phonogram played on a phonograph, to realize an analogue reading of the original into a professional studio, or to make an optical reading. The first option could be interesting to give an idea of the context of listening at the age of the phonograph but it is not recommended for the preservation of the sources as the technical process is significantly damaging the grooves of the discs.[27] The second option consists in coupling an analogue disc player to a digital recording studio. Operations are made by sound engineers. Third option is a variant of the second one and consists in replacing the disc player by an optical audio extraction process.[28]

The Royal Library of Belgium chose the second option as the most appropriate for the preservation of the original objects. Digitization process consists in the ultimate analogue reading of the original disc. To do it, MEMNON has to follow the technical specifications that have been set up by the library. The most important issues are to identify the correct reading speed for each disc and the appropriated needles. To select a wrong needle or to read the disc on a wrong speed would have strong implications on the length of the tracks, the timbre of the instruments and voices, on the color of the performance. Difficulties are

---

[26] About 125 days of scanning on a A2 scanner "Copybook COBALT" from i2s company will be necessary to scan the whole collection of discs centers.

[27] We can cite the example of a Belgian anthology published on LP, dedicated to Great Voices of the Province of Hainaut at the age of the 78rpm, published in 1985 by EMI-La Voix de Son Maître (ref. 1A 153 2906583).

[28] See the Swiss experience of the University of Fribourg: http://www.fonoteca.ch/gallery/visualAudio/home_en.htm

coming from the fact that there are few pieces of documentation available about this. To do that, MEMNON is checking several databases.

All discs cannot be digitized by MEMNON: about 7% of the items are rejected either because they have a bad quality of sorrow, either because they have a hole bad-centered on the disk center, or because the object is already too much damaged to resist on the analogue reading.

Each month, the digitization department of the library receives a production report from MEMNON with includes references that have been successfully digitized and those for which digitization has been aborted.

MEMNON has instructions to carry out a direct and neutral reading of the originals. If the Royal Library accepts to erase some impacting surface noises related to the ware of the grooves, the library asked not to use restoration technologies like CEDAR software.[29] This request is to ensure the preservation and the intelligibility of the original acoustic properties of the physical recordings. After digitization, acoustic and electric recordings have to be recognizable between them. It is important that digitized recordings keep their original characteristics despite less comfortable listening experience.

Two different formats of files are produced: a 24 bit 96kHz WAV file is produced as digital master of the recording while a 256kbps MP3 file is produced for dissemination. All files are named following the naming convention of the library and are accompanied by XML METS files. Created files and related metadata are copied by MEMNON on hard disks following a certain structure.

Every received hard disks are copied by the Royal Library on its storage infrastructure. After copying, the library is realizing a MD5 checksum verification to make sure that the delivered content does not include corrupted files.

For the quality control the digitization department of the library systematically verifies the adequacy between the file name, its corresponding metadata description and the audio content. This operation needs to open the file and to listen to it. This is a very time consuming task.

After validation, all files related to one bibliographic unit are encapsulated within one SIP[30] and stored on the infrastructure of the library. This operation triggers the dissemination of the content and the preservation in the digital archive.

---

[29] https://www.cedar-audio.com.

[30] Submission Information Package (SIP) is part of the Open Archive Information System OAIS: https://en.wikipedia.org/wiki/Open_Archival_Information_System.

The main online valorization of the digitized sound collections will be managed by the library through its usual online tools like its general catalogue and its digital gallery. Every reference will be indexed into the OPAC. Search possibilities include: composer names, librettist names, performer names, record label names, commercial references, matrix number, and date of publication.

For the whole collection, bibliographic references will be displayed with thumbnails. Access to audio files will depend on the copyright status of the recordings: in-copyrighted material will be only accessible inside the library. Afterwards, the library will valorize the digitized corpus as a new digital collection that will be findable into its digital gallery. Sub-collections could be created to highlight a certain artist, a certain performer or a certain discographic label. Finally, if there is no project of commercial offer at this moment, the library could set up some partnership with discographic labels that are active in publishing discographic anthologies dedicated to a certain artist.

8   Conclusion

Ten years after the beginning of its digitization program, the Royal Library of Belgium has started digitizing its collection of 78rpm shellac discs. Although digitization is here subcontracted, this project constantly requires dedicated team staff members in the digitization department of the library. With a time schedule of 54 months, this project has an exceptional duration. By the end of 2020, this will generate a corpus of 4.000 hours of digitized music, representative of both Belgian lyric and jazz lives in the first part of the 20th century.

This outstanding project for the library is a real mass digitization project. However, everything has been done to ensure the scientific and the artistic credibility of the digitized sources. Great attention has been put in the preparation of the project and in the selection of the criteria for the quality control of the produced files. All produced files will be saved on both library storage infrastructure and on the Belgian federal digital preservation infrastructure. Digitized corpora will be valorized online with respect of the copyright legislation as well as through the library dissemination channels and other digital portals like Europeana Music.

Beyond the technical and human challenge that such projects represent for heritage institutions like the Royal Library of Belgium, these initiatives are gradually constituting a rich, huge and scientifically credible corpus of digitized sources with great potential for further research topics in the field of the digital humanities.

If, today, expectations of the researchers in digital musicology seem to be less-identified than concrete and frequent requests of historians of linguists who need big data for their text and data mining investigations, it seems reasonable to think that such digitized sound recordings collections will constitute amazing opportunities for the researchers of tomorrow. Finally, are such digitization projects perhaps the expected latest step before an audio-turn in musicology research topics?

Frédéric LEMMERS

Numérisation des collections – Digitalisering van de collecties

Bibliothèque royale de Belgique – Koninklijke Bibliotheek van België

Boulevard de l'Empereur – Keizerslaan, 4

1000 Bruxelles

Belgien

frederic.lemmers@kbr.be