DIGITIZATION OF HISTORICAL MUSIC ARCHIVES: PRESERVING THE PAST, EMBRACING THE FUTURE

Goffredo Haus, Luca A. Ludovico
Laboratorio di Informatica Musicale
Dipartimento di Informatica “Giovanni Degli Antoni”
Università degli Studi di Milano
Via Comelico, 39 - 20135 Milano - Italy
{goffredo.haus,luca.ludovico}@unimi.it

ABSTRACT
Cultural institutions dealing with music (opera houses, conservatories, public and private collections, etc.) often hold huge archives made of music-related heterogeneous materials. These subjects can greatly benefit from digitization campaigns and the consequent adoption of ICT techniques as it regards not only the preservation, but also the exploitation and revivification of their content. This paper, that summarizes the keynote speech held at the 3rd International Conference on Technologies for Music Notation and Representation (TENOR 2017), starts from the experiences of the Teatro alla Scala and the Ricordi Historical Archive in order to show the new possibilities emerging from the adoption of computer-based technologies and approaches.

1. INTRODUCTION
Opera houses, conservatories, public and private collections and other music institutions often hold in their archives an invaluable heritage made of heterogeneous materials, including scores, audio recordings, iconographic material, books and letters, etc. In these contexts, a digital vision – mainly based on digitization campaigns and the consequent adoption of technological approaches and methodologies – is becoming increasingly important, not only for the preservation, but also for the exploitation and reviving of their music and music-related content.

During more than 30 years of activity, the Laboratorio di Informatica Musicale (LIM) 1 of the University of Milan has been carrying out international projects and establishing collaborations with important music institutions, such as: Bach Archiv Leipzig, Bolshoi Theatre of Moscow, Orchestra Verdi di Milano, RAI Radiotelevisione Italiana, Ricordi Historical Archive, RSI Radiotelevisione Svizzera, and Teatro alla Scala of Milan.

The goal of this work is twofold: on the one side, summarizing the main lessons learned during the past experiences involving both technologies and music; on the other side, showing the new possibilities and the practical implications of advanced computer-based approaches applied to music archives.

2. THE HISTORICAL ARCHIVE OF THE TEATRO ALLA SCALA
As a relevant example of cultural institution holding a rich and heterogeneous archive, it is worth citing the case of the Teatro alla Scala of Milan, often briefly referred to as La Scala. Inaugurated on 3 August 1778, during the last two centuries it hosted the greatest opera singers, dancers, soloists and conductors, and it premiered renowned operas such as Nabucco, Otello, Falstaff, Madama Butterfly and Turandot.

In addition to pursuing artistic activities and cultural dissemination, the theater has also the key mission of preserving and exploiting the immense amount of music-related materials collected during its history. Such a cultural treasure embraces not only scores, but also recordings, photos, sketches, technical drawings, fashion plates, craft-made objects, and an intangible heritage of human skills and competences.

In recent times, thanks to the ArchivioLaScala web site, 2 the theater has provided enthusiasts and scholars with the opportunity to easily access its history and artistic heritage. This initiative is the public and tangible result of a 20-years-long effort involving a number of aspects: an extensive digitization campaign for thousands of materials physically preserved in the archives, a change in the operating procedures adopted by the technical staff in order to produce digital content, and the use of ad-hoc computer technologies to interact with the new platform, as explained below.

2.1 La Scala DAM
In order to preserve its cultural heritage and enhance its fruition by artists, technicians and scholars, in 1998 the superintendence of the theater decided to build an integrated asset management system, called La Scala DAM. 3

Copyright: ©2017 Goffredo Haus et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

URL: http://www.archiviolascala.org

3 DAM was conceived as a multi-lingual acronym standing for Digital Asset Management in English, and for Depositi, Archivi e Magazzini (i.e. repositories, archives and warehouses) in Italian.
The idea was to create the digital archive of all available content stored in different locations – i.e. to integrate several uncorrelated archives of the theater – in order to cover a timespan from the second decade of the 20th century to the present day.

The digitization campaign began in 1996 with the preservation of the phonic archive, a project carried out in collaboration with the LIM. As a result, about 5000 open-reel magnetic tapes since 1950 were preserved and restored, for a total amount of about 10,000 hours of audio. Analog media were thermally pretreated to recover the original audio information [1] thanks to a laboratory oven with forced convection (see Figure 1), and then digitized through ad hoc equipment (see Figure 2). In 2000, a similar project was carried out by the LIM in order to rescue about 200 magnetic tapes of high historical interest coming from the phonic archive of the Bolshoi Theatre of Moscow [2].

Since 1998, the digitization project was extended with a modular approach to all the departments of the theater, including Scores, Costumes and accessories, Sketches and fashion plates, Photos and playbills, Properties, and Editorial archives [3].

In this context of archive integration, the LIM research team proposed an advanced multimedia object-relational architecture to query and retrieve musical information in a multimodal way. The idea was to provide access to audio recordings and to the corresponding scores within an integrated environment, supporting different tools (e.g., queries by humming, symbolic inputs, etc.) to search in the digital repository. This platform, called Musical Archive Information System (MAIS), was profitably experimented at La Scala and documented in several scientific works. The MAIS introduced some innovative features with respect to the state of the art, such as different abstraction levels for the description of music information, multimedia content integration, and both symbolic and audio queries to retrieve data from the database. For further details, please refer to [4].

La Scala DAM was initially designed as a restricted-access application running on La Scala’s local network for theater workers and artists, with a very limited data exchange from and to the official web site. The project was internally released in 2006, making a huge amount of digital content available to technicians, artists and employees for the creation, production and documentation of each show.

Conversely, nowadays most materials are publicly available online. In this way, the theater not only responds to the mission of cultural dissemination, but also tries to involve and gain a new audience through multimedia content sharing.

The management of the project was entrusted to a group of institutional, technological and scientific partners, including Fondazione Milano per la Scala, University of Milan, Accenture, Fastweb, Hewlett-Packard, Oracle, and TDK.

At the moment of writing, the available digital contents cover approximately:

- 17,000 posters and playbills documenting all the artistic activities of the theater, including operas, ballets, concerts, recitals, and other cultural events. These materials are scanned at a high resolution, and their content (i.e. the schedule, the playlist, and the cast of the show) is manually transcribed;

- more than 1,000,000 on-stage and back-stage photos, portraying renowned singers (Carlo Bergonzi, Renato Bruson, Maria Callas, Mario Del Monaco, Giuseppe Di Stefano, Placido Domingo, Mirella Freni, Tito Gobbi, Luciano Pavarotti, Giulietta Simionato, Joan Sutherland, Renata Tebaldi, etc.), conductors (Claudio Abbado, Daniel Barenboim, Pierre Boulez, Guido Cantelli, Victor De Sabata, Gianandrea Gavazzeni, Carlo Maria Giulini, Carlos Kleiber, Riccardo Muti, Arturo Toscanini, Herbert von Karajan, etc.), dancers and choreographers (Roberto Bolle, Alessandra Ferri, Carla Fracci, Rudolf Nureyev, Luciana Savignano, etc.);

- 24,000 sketches, drawings and fashion plates signed by great artists, including Alberto Burri, Marc Chagall, Jean Cocteau, Salvatore Fiume, Piero
Fornasetti, Renato Guttuso, Marino Marini, Pablo Picasso, and Mario Sironi;

- 45,000 costumes prepared by designers such as Caramba, Emanuele Luzzati, Vera Marzot, Odette Nicoletti, and Franca Squarciapino;
- 60,000 costume accessories, including jewelry, lingerie, shoes, wigs and hats;
- 80,000 scene tools.

Since the theater is still active and the project running, the mentioned numeric data are growing day by day.

2.2 Technical Challenges and Remarks

The technological core of La Scala DAM project is a multimedia relational database, with a number of applications attached to feed the database and to retrieve information. Two goals guided the project: i) information entities had to be described in the most detailed way according to the models already in use in the theater, and ii) a network of relationships as rich as possible had to be established among those entities.

A first result is to provide a comprehensive description of a given object from multiple perspectives. This concept can be declined at different levels of aggregation and abstraction. For instance, if we focus on a single physical good, say a costume, the concept of "comprehensive description" may embrace its data sheet, the list of available accessories, the original fashion plate, a number of on-stage photos, and so on. Conversely, if we take into consideration an opera, this approach implies the possibility to easily obtain a synoptic view of all information somehow related to this work.

When the network of relationships is properly exposed through an ad-hoc application, it can provide the user with multiple navigation paths towards the requested information. An example is the following list of operations: search for a costume, retrieve the opera it was designed for, move to the list of all on-stage representations for that opera, find the name of the artist who played the role of the protagonist in a given date, watch the list of all the operas he/she has performed in the last season, choose one of those operas and open all the related photos, select a photo, and finally open its detailed data sheet.

These considerations guided the design of the web interface to browse content from all archives in an integrated way. Figure 3 shows some screenshots from the original web application. Much of the interface – not only icons and images but also most text parts – could be clicked, allowing to jump to new lists of object or to other meaningful aggregations of information.

During the project, not only technical issues but also relational problems emerged, mainly related to the paradigm shift required to the staff of "traditional" archives, as well as to the perceived feeling of autonomy loss in sectors managed independently so far.

Being the theater in activity, the design, implementation and testing of the solution did not have to affect the operating procedures of the theater staff, nor it had to hinder the production environment. The risk was the realization of a very powerful platform either unusable due to its complexity or poorly used because of the extra workload required to archive employees. For this reason, all the phases of the project were conducted in tight cooperation with the theater governance, archive managers, domain experts and workers, in a continuous exchange of critical observations and improvements.

During the digitization phase, a key problem was how to catch and represent physical objects. Needless to say that – in a multimedia database – only digital representations can be entered. Some archive contents were already in digital format, for instance recent photos and recordings as well as computer processed texts; but other materials had to undergo an analog-to-digital conversion, and in this case the main concern was to prevent information loss, sometimes in absence of international guidelines to follow. The real challenge concerned the digitization of 3D physi-

Figure 3. Screenshots from La Scala DAM web application. From top to bottom: the synoptic view of music works, the list of playbills, the list of photos.
cal objects, whose appearance and features could not be digitized with the available technologies. In this case, the digital representation could catch only some aspects considered relevant by experts. Let us cite the example of a stage tool or a costume, where multiple photos can capture shapes, colors and design details from different angles, but they do not allow a 3D reconstruction nor they provide information about materials.

Finally, the main question was how to structure information inside the database. An opera house like La Scala runs different kinds of performances, ranging from operas, ballets, and symphonic concerts to interviews, conferences, and public presentations; and each activity presents peculiar features concerning its structure, staging, and relationship with other information entities. Since the database had to cover all possible cases, the problem of information structuring was not a trivial one.

This issue was solved adopting a 3-tier hierarchy starting from the concept of base version, namely a work as it was conceived by its author(s). This definition is meaningful for operas, ballets, and concerts, but – clearly – it makes no sense for activities such as conferences or presentations. A base version has a number of metadata attached, including title, author names and roles, premiere date and place, and ensemble. The second key concept is the one of staging, that clusters a set of performances characterized by the same music program, belonging to the same season and having common features about production, staging, and cast. For example, “Le nozze di Figaro” by W.A. Mozart and L. da Ponte is a base version, whereas the 2005/06 production of that opera is a staging. In general terms, the same base version can have 1 to \( n \) stagings attached. Finally, the 3-tier hierarchy is closed by the concept of performance, namely the instance of a base version according to a particular staging, in a given date and place and with a specific cast. At La Scala, performances present a bivocal relation with the corresponding playbills. Moving from base version to staging and performance, information becomes more and more detailed.

Such a structure is suitable to represent also non-standard (yet frequent) cases. First, a single performance can include many independent music works, like in singing recitals, typically made of arias from different operas. This situation is managed by linking each subpart of the staging to a different base version, which requires an \( n \) to \( m \) relationship between the two entities in the database schema. Another issue is how to represent non-musical events, a case that can be easily solved by creating dummy base versions.

After defining the 3-tier spine, other entities are put in relationship with the mentioned concepts in the most proper way. For example, librettos and synopses are naturally linked to base versions, fashion plates and costumes to stagings, and audio recordings and on-stage photos to performances.

A simplified entity-relationship diagram for the resulting database is shown in Figure 4. For a more detailed discussion, please refer to [5].

### 3. THE RICORDI HISTORICAL ARCHIVE

Ricordi is an Italian music publisher that has promoted famous composers and musicians over more than two centuries of activity. Today, its Historical Archive is probably the most important private music collection in the world, including invaluable treasures such as the autograph scores of many operas by G. Verdi and G. Puccini. The collection embraces handwritten scores, printed scores, librettos, photographs, drawings, posters, letters, periodicals, and administrative documents.

Compared to the case of La Scala, the “core business” of the Ricordi Historical Archive is different: artistic production for the former, cultural heritage preservation for the latter. For this reason, the Archive produces no audio content (even if a collection of historical vinyl recordings is currently being bought), and its heritage is not going to grow significantly over time.

In the first decade of the new millennium, the Italian ministry for cultural heritage and activities (Ministero per i Beni e le Attività Culturali, MiBAC) promoted and funded an extensive digitization project involving the most important Italian musical institutions. The aim was to create the Italian music network (Rete della Musica Italiana, ReMI), a new management and diffusion data architecture focusing on Italian musical heritage. This goal was achieved by connecting peripheral repositories of music-related institutions each other and providing web users with centralized query services. On the one side, local repositories maintained their independence concerning the organization and management of their own contents, receiving public funding to adhere to the project; on the other side, centralized web services integrated and homogenized data access, guaranteed service continuity, optimized answering time, and offered unified user-friendly interfaces.

Launched on 20th June 2008, the new music-search interface was published within a thematic area of the Ministry’s portal known as InternetCultura.it, and offered a synoptic view and an integrated navigation of music documents preserved across Italy [6].

In order to support interoperability among heterogeneous systems, the digitization and cataloging of documents had to adhere to international standards, and specifically:

- Universal Machine Readable Catalogue (UNIMARC) [7], Dublin Core and Management Administrative Metadata (MAG) for metadata encoding;
- Open Archival Information System (OAIS) [8] and Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) [9] for preservation, interoperability and sharing models.

The LIM was in charge of the scientific coordination of the project, including the study of ad-hoc standards for cataloging music-related materials, the design and implemen-

---

4 Also analog scores, photos and playbills can be considered as physical objects, but – in general – their information content can be fully captured by scanning a single 2D side.

5 URL: http://www.internetculturale.it
In this context, MiBAC and Ricordi signed an agreement to make part of the Archive’s content publicly available. The digitization project involved a total amount of 6,586 documents (scores, texts, photos, drawings, sketches, fashion plates, scenic maps, etc.) and resulted in 12,660 digital scans related to about 100 music works. Digital scans were produced with high quality settings and saved in formats adequate for preservation and publishing purposes, and subsequently down-sampled and saved for web browsing. All the materials were cataloged by experts in compliance with international standards.

After the conclusion of the publicly funded project, the digitization and cataloging campaign was carried on at a local level, thus creating the current Ricordi digital archive.

The structure of both the peripheral (Ricordi) and the central (ReMI) database, as well as the design of the corresponding web applications to browse multimedia content, were clearly inspired by the experience made at La Scala. Similitudes between the two approaches are evident in the screenshots in Figure 5, extracted from InternetCulturale. The main difference concerns the kind of materials locally available, including for example handwritten scores instead of audio content. For the sake of clarity, it is worth underlining that – in the wider framework of ReMI / InternetCulturale – the idea was to integrate heterogeneous information and digital objects coming from different archives, so the lack of audio resources from Ricordi had to be filled by other institutions.

While maintaining the same approach, the design phase of the Ricordi platform was influenced by the different mission that a historical archive has with respect to an opera house in activity. In the former case, the goal is to highlight the richness and variety of the preserved heritage, to give new value to archive materials, to disseminate musical culture by providing easy access to sources, also through the interoperability with other platforms; in the latter case, the focus is on gaining new audience and supporting the local artistic production. These different approaches influence many aspects of digitization, indexing and browsing, from the compliance with international standards to the design of meaningful relationships oriented to specific navigation paths.

An interesting evolution of Ricordi’s digitization project – mainly conducted on graphic materials (see Figure 6) – is the reliving of its tangible and intangible cultural heritage through international exhibitions. These initiatives propose a mix among historical materials presented to the public in their original form and innovative technological
applications to add multimedia interactivity (see Figure 7). The availability of a digital catalog simplifies the organization of cultural events focusing on a specific theme (e.g., an author, a music work, a place, etc.), and a comprehensive library of digital objects makes it particularly easy to produce promotional and editorial materials and to implement computer-based applications.

The list of exhibitions organized by Ricordi in partnership with the LIM includes *Celeste Aida* (November 2006), *That’s Opera* (November 2008), *That’s Butterfly* (September 2009), *The Enterprise of Opera* (August 2013), and *Madama Butterfly - L’oriente ritrovato* (November 2016).

4. LEARNING FROM THE PAST

If we look back on the experiences conducted at important music institutions and archives, the outcomes are mainly satisfactory. There is a number of activities that have greatly benefited from digitization, computer-based approaches and technological innovations. Just to name a few examples:

- The adoption of correct procedures to ensure the preservation of cultural heritage;
- The application (or even the pioneering research) of standards for cataloging and sharing, with the side effect of fostering the interoperability with other systems;
- The experimentation of new paradigms to query and browse metadata and multimedia objects in an integrated way, thus improving the access to information and – consequently – facilitating organization, production, and communication processes;
- The release of innovative applications based on multimodal representations of music and music-related information.

Unfortunately, the introduction of computer-based approaches into “traditional” environments often implies additional work and requires the investment of extra resources, and – in these cases – long-term advantages need to be seen over short-term disadvantages. We can mention a number of real cases where innovation was not retained as an opportunity but as an obstacle:

- The database management system originally conceived by the LIM for *La Scala DAM* was object-oriented (OODBMS), but it was considered too advanced and somehow experimental by the other technological partners, so the idea was abandoned in favor of a more standard relational model;
- At La Scala, the new procedures oriented to a digital management of information, enthusiastically welcomed by those archives used to work with digital objects (e.g., the *Photos and playbills* and the *Phonic* archives), were initially rejected by some sectors of the theater (e.g., the *Sketches and fashion plates* archive);
- Sometimes, the experimentation of cutting-edge technologies (e.g., the *MAIS*) was not adequately appreciated, as people focused on the additional burden (e.g., the encoding of scores into a symbolic format) without seeing their potential (e.g., the possibility to query a huge archive of music works by humming or playing a digital instrument).

Experience teaches us that one of the most demanding tasks for computer experts is to clearly and convincingly illustrate the undeniable long-term benefits of ICT approaches and techniques.

Another cultural challenge is to explain that digitization does not imply preserving forever. A continuous investment of resources is required to ensure that: i) the media remain intact (the solution is to perform periodical tests and new copies, when needed), ii) the encoding formats continue to be known, documented and readable (open and commonly accepted standards are the best option, re-coding is an alternative), and iii) the devices to read them can be kept running or – in case – replaced.

5. SETTING OUT THE FUTURE

A cultural institution such as La Scala is economically supported by the income from its multiple activities (ticket sales, space and equipment rental, broadcasting rights and contracts, merchandising, etc.), not to mention public and private funding. Conversely, the economic sustainability of an archive is often based on the exploitation of its assets. From this point of view, the business model adopted by the
Ricordi Historical Archive is paradigmatic: the organization of exhibitions and events, the participation in national and international funded projects, editorial initiatives and sponsorships are the means to support the preservation of cultural heritage and to fulfill the mission of dissemination. This approach requires not only to showcase the available cultural assets, but also to find new ways to relive them. In this sense, information and communication technologies can play a key role.

A first way to benefit from digitization and new technologies is to integrate heterogeneous information according to a multilayer model, thus adding new value to the original uncorrelated objects. For the mentioned archives, aggregations led to a better access to information and to a more comprehensive view on assets. Thanks to ad-hoc file formats and computer applications, it is possible to extend such an approach from databases to other contexts. For example, music-related cultural heritage can be experienced adding synchronization features and interaction with multimedia content. These concepts have been applied in a number of Ricordi’s exhibitions, where computer-based interfaces let the audience browse autograph scores, follow music notation on different score editions, and compare historical audio performances through interactive stations, designed also for musically-untrained people. To that end, the cooperation between the Ricordi Historical Archive and the LIM suggested the adoption of the IEEE 1599 technological framework [10]. By integrating different and heterogeneous aspects of a music piece within a unique document, the IEEE 1599 standard creates a sort of local semantic network realized through a multilayer structure [11], which in turn can be integrated into a global network such as the Semantic Web. This approach fosters a more advanced music experience and allows the implementation of innovative services, such as interactive play-bills or the augmented fruition of a music show [12].

The integration – at a local and global level – of related content may provide new value to the original uncorrelated entities, potentially introducing new intellectual-property rights to be protected and exploited [13].

An interesting question for cultural heritage holders is whether technologically advanced approaches are pure scientific research and experimentation, or they can lead to the release of marketable products. An answer is provided by the mentioned experiences: both La Scala and Ricordi have improved their ability to go on the market, for example by attracting a new audience through their web archives, offering their goods and services (staging rentals, exhibitions, etc.) in a more effective way, and publishing editorial products based on digitized materials.

Concerning the revivification of archive content, many other successful examples could be mentioned: from interactive and adaptive products for music education [14], like a textbook recently published by Pearson, to new models to interact with live music performances [15].

By combining the efforts of cultural heritage holders and computer scientists, it is possible not only to create new value for music archives but also to perform a relevant cultural operation based on innovative models to experience music.

6. REFERENCES


