Building a home for Italian audio archives

Silvia Bianchi
DSFUCI, Siena University
Arezzo, Italy
bianchi.silvia83@gmail.com

Silvia Calamai
DSFUCI, Siena University
Arezzo, Italy
silvia.calamai@unisi.it

Monica Monachini
Institute for Computational
Linguistics, "A. Zampolli"
CNR, Pisa, Italy
monica.monachini@ilc.cnr.it

Niccolò Pretto
Institute for Computational
Linguistics, "A. Zampolli"
CNR, Pisa, Italy
niccolo.pretto@ilc.cnr.it

Maria Francesca Stamuli
Ministry of Culture and
Tourism, Florence, Italy
mariafrancesca.stamuli
@beniculturali.it

Pierangelo Bonazzoli
Unione dei Comuni Montani
del Cosentino, Italy
pierangelobonazzoli
@casentino.toscana.it

Abstract

Audio and audiovisual archives are at the crossroads of different fields of knowledge, yet they require common solutions for both their long-term preservation and their description, availability, use and reuse. Archivio Vi.Vo. is an Italian regional project financed by the Tuscany Region, aiming at (i.) exploring methods for long-term preservation and secure access to oral sources and (ii.) developing an infrastructure under the CLARIN-IT umbrella offering several services for different scholars in different domains interested in oral sources.

1 Introduction

Audio and audiovisual archives are scattered all over the Italian peninsula, from researchers' private houses, to universities and research centres, from cultural institutions (e.g., Istituti per la Resistenza), to State institutions, such as State Archives and Libraries. The pilot survey made by Galatà and Calamai in 2018 has emphasised precariousness, instability, and insecurity among different communities of Italian researchers dealing with audio and audiovisual archives. Almost half of the resources listed in survey (49.6%) was barely accessible. Only 9.2% of the resources was accessible and available, 4.6% was partially accessible, 35.1% was available upon request, 1.5% is available upon request and only for selected parts. As for the resources which were declared to be accessible, the access policies were as follows: only 9.2% of these resources was freely accessible online (with no authentication); 7.6% was accessible online via authentication; 29% was accessible onsite (i.e. where the resources are physically stored). As for the long-term maintenance and preservation the answer receiving the highest number of responses was nobody (43%), followed by reference Institutes, such as Associations, Foundations, libraries and their archives (17%), reference Universities (16%), the owners/individuals themselves (15%)[1]. Several research projects in recent years aim to disseminate audio and audiovisual archives, which are collected over the years by both researches and amateur fieldworkers: some examples are, among others, *Grammo-foni. Le soffitte della voce*, also referred as Gra.fo (grafo.sns.it), *Voci, parole e testi della Campania* (www.archivicampani.unina.it/archivi_campani_dev), *I granai della memoria* (www.granaidellamemoria.it), *Circolo Gianni Bosio Audio Archives* (www.circologiannibosio.it/archivio). Nevertheless, fragmentation and lack of common and shared standards are often the common features of certain initiatives of that sort, whose duration over time crucially appears to be dependent to the duration of external funding, if any. Moreover, the limited perspective (which often coincides with one and the only one field of knowledge of the researchers involved in the project) makes it tough to encompass audio and audiovisual archives which originated in different areas and for different purposes. Moreover, a linguist working with audio archives is not necessarily competent also in long-term preservation of audio data and data management. Eventually, not

[1]Further details available in (Galatà and Calamai, 2019)

all the research projects dealing with audio archives receive financing for all the different professional profiles involved in their preservation, managing and valorisation.

Given this picture, it appears urgent to provide an infrastructure offering: 1) a long term preservation service for audio archives, 2) a shared set of metadata compliant with the main international standards and FAIR principles, and 3) an access interface which takes into account the peculiarities of the audio modality and which is able to support researchers in different disciplines.

2   The Archivio Vi.Vo. project

In 2019, Regione Toscana decided to support the project Archivio Vi.Vo., which aims to catalogue and disseminate oral archives. The following partners were involved: Siena University (Silvia Calamai), Soprintendenza Archivistica e Bibliografica della Toscana (Maria Francesca Stamuli), CLARIN-IT (Monica Monachini), and Unione dei Comuni del Casentino (Pierangelo Bonazzoli). In order to reach the three ambitious objectives detailed in the previous section, Archivio Vi.Vo concentrate most of the efforts on the design and development of an IT infrastructure in the CLARIN-IT domain could be used by several other project concerning audio archives. An important part of the work concerns metadata definition, which is compliant with international archival standard, such as with ISAD (G) and ISAAR as well as several others derived from different disciplines will be detailed in Section 3.2.

2.1   The case study

The architecture described in Section 3 is in the process of being validated on a specific audio archive, namely Caterina Bueno's audio archive, which appears to be rather challenging, for the following reasons: i) the complex archival history, ii) the conservation condition, iii) the heterogeneity of the audio material.

Caterina Bueno (San Domenico di Fiesole, IT, 2nd April 1943 – Florence, IT, 16th July 2007) was an Italian ethnomusicologist and singer. Her work as a researcher has been highly appreciated for its cultural value, as it allowed the collection of many Tuscan and central Italy's folk songs that had been orally passed down from one generation to the next until the 20th century (when this century old tradition started to vanish). Her work as a singer was always oriented towards research. The Caterina parents were the Spanish painter Xavier Bueno and the Swiss writer Julia Chamorel. As a child, she was fascinated by Tuscan dialects, folk songs and peasant culture. At the age of twenty, she started travelling through the Tuscan countryside and villages recording Tuscan peasants, artisans, common men and women singing any kind of folk songs: lullabies, *ottave* (rhyming stanzas sung during improvised contrasts between poets), *stornelli* (monostrophic songs), narrative songs, social and political songs, and much more. These were the same songs that she sang in her performances, making them well-known and appreciated both in Italy and abroad in the second half of the 20th century, when she was at the pinnacle of her career. Caterina Bueno's sound archive is composed of about 476 analogue carriers (audio open-reels tapes and compact cassettes), corresponding to more than 700 hours of recording, and it was digitised during the PAR-FAS project Gra.fo. The analogue audio documents was located at two different owners': part of it was stored at Caterina's heirs' house, while the rest was kept by the former culture counsellor of the Italian Municipality of San Marcello Pistoiese, in the Montagna Pistoiese, where a multimedia library was supposed to be set up. Unfortunately, disagreements and misunderstandings between the two parties have so far made the archive fragmented and inaccessible to the community. Both owners, independently, have turned to Silvia Calamai for the reassembly of the whole archive in the digital domain, in respect of the artist's wishes. After digitising, the carriers were returned to their owners.

The archive is divided into several archival series (concerts, rehearsals, field research, interviews, conferences, music and theatre) which offer a vivid insight into the activity of the singer and ethnomusicologist. In several cases the original carriers were devoid of all the contextual information (place and date of recordings, speakers involved in the recordings). In other cases, the audio documents were recorded at different speeds and using different track head configurations, thus making rather complex the digitisation process and the creation of access copies. From this respect, Caterina Bueno's audio archive represents an extreme case where different levels of complexity call into question different professional

profiles and skills.

3 Building the home for audio archives

3.1 The infrastructure

As for the infrastructure, the ILC4CLARIN is improved through experimental approach to preservation, management and access to audio data and metadata. The experimental activity will aim to adopt the model and high-performance computing and archiving services of the new GARR network infrastructure, built along the Cloud paradigm. The project will also exploit the federated identity service of such infrastructure, in order to manage users' access. A robust system for managing authentication is essential for audio and audiovisual archives because of the frequent privacy, ownership, and copyright issues concerning their content. Several classes of users are considered, each of them with different access grants. The system could be divided in two different parts. The first concerns the data and metadata entry by the archivists, archive owners or in general researchers who wants to preserve the legacy. Considering the complexity of such system, it must be able to manage several international standards and several kinds of specific functionalities, the project would difficultly start from scratch. The starting point is the open-source software xDams (`www.xdams.org`). The three main characteristics that influenced the adoption of the software were the completeness of its standards coverage, its extensible no-sql database as well as the open-source license. The second part of the infrastructure is an access interface able to support researchers of different disciplines in discover and study audio or audiovisual documents.

3.2 The metadata

The relationship between the document and the carrier (e.g. the compact cassette or the open reels) appears to be rather controversial. The documents often need to be re-organised during the cataloguing process with the content privileged at the expense of the carrier. Given the absence of a one-to-one relationship between the physical carrier (i.e., compact-cassettes, open-reels) and the archival unit, the preservation copies are kept separately from the archival units (Mulè, 2003; Silvia et al., 2014; Stamuli, 2020), which are the outcome of a meticulous process of listening, analysis and comparison. In audio and audiovisual archives, the archival unit is therefore a set of data and documents pertaining to the very same communicative event, per unit of time and place. This approach leads to a very complex set of metadata, articulated in three different sub-groups: i) metadata for the description of the digital copies of the physical carrier (preservation copy); ii) metadata for the description and managing of oral sources as items of an (audio) archive; iii) metadata expressing the relationship between the preservation copy metadata and the digital archive metadata. Given this framework, the main set of metadata used for describing audio and audiovisual documents will be used, both for the archives description (such as EAC and EAD) and for the VLO catalogue implementation (CMDI).

4 Conclusion

Archivio Vi.vo will thus constitute a pilot study within CLARIN-IT to experiment methods and offer targeted services for both specialists and general public interested in oral sources. Archivio Vi.Vo. will develop a model which can be replicated on other audio archives, even outside the context of Tuscany.

References

Vincenzo Galatà and Silvia Calamai. 2019. Looking for hidden speech archives in italian institutions. In *Selected papers from the CLARIN Annual Conference 2018, Pisa, 8-10 October 2018*, number 159, pages 46–55. Linköping University Electronic Press.

Antonella Mulè. 2003. Le fonti orali in archivio. un approccio archivistico alle fonti orali. *Archivi per la storia*, 16(1):111–129.

Calamai Silvia, Biliotti Francesca, and Bertinetto Pier Marco. 2014. Fuzzy archives. what kind of an object is the documental unit of oral archives? In Marinos Ioannides, Nadia Magnenat-Thalmann, Eleanor Fink, Roko

Žarnić, Alex-Yianing Yen, and Ewald Quak, editors, *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, pages 777–785, Cham. Springer International Publishing.

Maria Francesca Stamuli. 2020. Fonti orali, documenti e archivi: riflessioni e proposte per la nascita di un 'archivio vivo'. In Duccio Piccardi, Fabio Ardolino, and Silvia Calamai, editors, *Studi AISV 6*.