

## The Grey-side of Audio Archives

**Monica Monachini**, Institute of Computational Linguistics “A. Zampolli” - CNR, Pisa, Italy  
**Maria Francesca Stamuli**, Ministry of Culture and Tourism – Florence, Italy  
**Silvia Calamai**, Human Sciences and Intercultural Communication, Siena University  
**Niccolò Pretto**, Institute of Computational Linguistics “A. Zampolli” - CNR, Pisa, Italy  
**Silvia Bianchi**, Human Sciences and Intercultural Communication, Siena University, Italy

### **Abstract**

*Archives often include documents that can hardly be considered publications or grey literature as such, yet they maintain their documentary value and play a role of primary sources for the specialists. These documents, indeed, can help archivists to reveal the sedimentation process of the archive itself and to preserve the authentic context of the documentary production. They also appear to be very useful for the community of researchers and scholars. This happens more frequently with oral archives which include ‘non-conventional sources’, thus bringing together audio documents, fieldworks notes, correspondence, slipcases, analogic compact cassettes or open reels. At the cross-road of two disciplines, Archival Science and Grey Literature, this paper aims to argue the applicability of the concept of grey literature to this wide range of documentary materials, by showing the experience of Archivio Vi.Vo, a regional project aiming at building a model for archiving, preserving, managing and disseminating audio documents.*

### **1. Introduction**

#### **1.1. The background: audiovisual heritage**

The audiovisual heritage of the twentieth century is at risk of being lost. Both analogue and digital documents are in danger: the former because of carrier degradation and playback device obsolescence, the latter because of the hardware and software obsolescence as well as the rapid evolution of standards and formats. On the other side, knowledge conveyed by oral culture may be as disseminated, reliable and accessible as written resources, counteracting the negative effect of technological obsolescence. Looking ahead and careful planning can minimize the repercussions on memory preservation.

#### **1.2. Setting the problem: Archival Science, oral archives and Grey Literature**

Why would Archival Science matter to Grey Literature? It appears that this question has a relevant place in the domain of digital archives, as we try to argue in the present paper. Building a digital archive entails the creation of different types of digital documents: i) documents of the archive itself; ii) documents about the archive. These are deliverables, technical papers, sketches and other different typologies of documents falling uncontroversially into different types of grey literature. They all appear to be extremely useful for supporting these projects as well as studying the evolution of the archive itself. In addition, they often play a role of primary sources for the specialists and prove to be useful in helping the archivists to reveal the sedimentation process of the archive itself and preserve the authentic context of the documentary production. Finally they also appear to be beneficial for the community of researchers and scholars in their studies.

The picture become even more complex in the realm of audio documents and oral archives, that preserve, organize and make available such documents. Oral archives appear to be rather complex and delicate, mainly because of the multifactor ‘fragility’

that characterizes audiovisual heritage (see Calamai, Casellato and Stamuli, forthcoming).

In fact, the archival treatment of audio documents deals with 'oral sources'. Those kinds of sources are at the crossroads of several domains of knowledge and disciplines, each with its own jargon: oral history, linguistics, anthropology, ethnomusicology, digital humanities. In archival science, 'oral sources' are considered 'new sources' or 'not conventional sources' (Carucci, 1993; Mulè 2005). What makes audio documents different from prototypical archival and documentary records is mainly their hybrid nature, characterized by a dichotomy between the 'content' (the audio signals) and the carrier, that can potentially contain more than one audio recording, even different kinds of audiovisual records. The digitisation process of analog audio recordings for their preservation (Miliano, 1999) provides additional material that describes the operations undertaken. Doing so is essential for reporting the history of the document's transmission.

Given this background, the research question raised in this paper can be summarised as follows: "how much grey" is this wide range of documentary materials?

### 1.3. The Case Study: Archivio Vi.Vo.

*Archivio Vi.Vo.* is an Italian regional project supported by Regione Toscana, whose aim is to explore methods and services for long term preservation and secure access for audiovisuals (Calamai et al. 2020; Stamuli et al., in press). It constitutes a pilot study dealing with the description and cataloguing of Caterina Bueno's<sup>1</sup> audio recordings (digitized during the PAR-FAS project *Grammo-foni Le soffitte della voce*, 2010-13) composed of about 450 carriers (audio reels and compact cassette tapes corresponding to nearly 700 hours of recording). The services for long-term preservation and access are developed within the framework of CLARIN-IT, the Italian counterpart of the European Research Infrastructure for Social Sciences and Humanities, CLARIN-ERIC. The Italian node of the European data-center federation, ILC4CLARIN, offers long-term solutions and technology services for sustaining, archiving, deploying, connecting, and analyzing linguistic data. By means of *Archivio Vi.Vo.*, Caterina Bueno's audio recordings and its metadata will be systematized under the CLARIN-IT domain, through an experimental and cross-disciplinary approach to preservation, management and access to audio-video data, aimed at adopting the model and the high-performance computing and archiving services of the GARR network infrastructure, built along the CLOUD paradigm. This model will be disseminated to the scientific community interested in managing audiovisual records and will make it possible for the general public to access the oral sources produced in the territory. The project is thus meant to support and advance activities regarding this topic in the CLARIN-IT consortium and the CLARIN infrastructure.

## 2. Research approach and first outcomes

One of the main objectives of *Archivio Vi.Vo.* is the creation of a model for dealing with issues of preservation, archiving, management and dissemination of audio documents, adoptable by institutions, archives and researchers. In order to achieve this objective, grey literature assumptions have a prominent role.

Starting from the very beginning of the archive's creation, the developers should collect, analyze and make accessible to professional and non-professional audience (1) the overall set of different documents related to the production of the audio records, and

---

<sup>1</sup> Caterina Bueno (1943-2007) was an Italian ethnomusicologist and singer, highly appreciated for its cultural value. Her work allowed the collection of many Tuscanian and central Italy's folk songs that have been passed down orally from one generation to the next until the 20th century.

(2) the documents concerning the design and development of the platform for the long-term preservation and access to the audio archive.

Concerning the former, in the arrangement phase of the archive itself, archival science prescribes to carry out a careful analysis and survey of the archival documentation. In fact, if we conceive the archive not as a dead entity, but as an evolving reality we need to reveal the sedimentation process and give the documents their authentic context of production, shaped by the activities of the producer of the archive itself. This is particularly important for 'not conventional sources' as oral sources are. In Caterina Bueno's sound archive, everything reveals the archive sedimentation process and lends the appropriate and authentic context of production to the audio records: her fieldwork notes, her correspondence, the slipcases of compact cassettes and of the open-reels that she wrote, her diaries, the newspaper clippings she collected, the numbering labels that she used, etc.

*Archivio Vi.Vo.* makes all this data fully retrievable, thus making audio records fully understandable by adopting both bibliographic and archival recommended standards.

The second kind of grey literature documents considered here belongs to the well-known typology of "technical documents" (Pejšová, 2011). The production of these documents is strictly related to the software engineering field, and in particular, to the adopted software process model (sometimes called Software Development Life Cycle model). There are no universal software engineering methods that are suitable for all systems and all companies (Sommerville, 2016) and the software process model has to be chosen considering the kind of application will be developed. The complexity of the *Archivio Vi.Vo.* infrastructure prevents a development from scratch, but, as most of the complex business systems, it has to be developed configuring existing systems and integrating new functionalities. For this reason, within the project, we adopt the process model that Sommerville defines as "integration and configuration", in order to re-use and adapt existing software. Each task of this model has different outcomes: products or deliverables. The latter can be reports, models, sketch, UML schemas, system and user documentation (which are precious for the development, maintenance and possible extension of the software since the early stage of its development, as well as for other similar projects), *i.e.* a vast kind of material that will be published within the infrastructure itself and offered to the scientific community as grey literature.

From this respect, special attention is paid to the technical reports produced inside the project: that is, the description of Caterina Bueno's archive, its archival structure, the metadata scheme, the glossary (in order to have a common vocabulary inside the research project), and the legal documentation related to privacy and copyright. Conversely, certain members of the project staff are producing a rather different type of technical reports, which are useful inside the project in order to document the activities of every single post-doc researcher, for preparing the hand-over from one researcher to another, for communicating accurately through a cross disciplinary project and to express to others the path on which the project is moving on. According to GL, different levels of accessibility and different types of public can thus be envisaged: i) open, for the general public (e.g., the glossary, the structure of Caterina Bueno's archives), and ii) restricted, for internal use (e.g., the final reports of each post-doc researcher involved in the project).

### 3. Final Remarks

As Serini (2003) points out, among the archival documents there are numerous materials which, according to their publication status and the interest they can have among specialists, can be quite rightly grey literature. The cross-disciplinary approach adopted by the *Archivio Vi.Vo.* project will bring to the forefront the technical documentation and the archival materials as 'grey literature', as data distinguished by content and type, but united by the objectives of research, organization and dissemination. According to Serini, what makes a wide range of documents 'grey literature' is the combination of different, but interwoven, layers such as their production (characteristic of a particular research endeavor or to an organization) and the type of distribution.

*Archivio Vi.Vo.* aims at producing a digital ecosystem with audio recordings at the heart. The unconventional documentation produced in the realm of audio archives will be treated in the light of grey literature, conceived as a product in relation to a service, a vehicle of direct communication, participatory democracy, transparency and organisational clarity, and the recovery of memory (Sardelli 2000).

At the same time, the building itself of such an ecosystem produces a flood of different and diverse objects which might be undisputedly inserted inside the GL approved taxonomy. On the other hand, it cannot be excluded that the GL taxonomy might also benefit from the work carried out by *Archivio Vi.Vo.* given the cross-disciplinarity and multimediality of audio archives.

### Bibliography

- Bologna, M. (2015). *La sedimentazione storica della documentazione archivistica*. In: Giuva L and Guercio M. (Eds) *Archivistica. Teorie metodi, pratiche*, 212-235. Roma: Carocci.
- Calamai, S., Stamuli M.F., Casellato, A., (in press), Collaborative best practices: An Italian *Vademecum* on the conservation, the description, the use and the re-use of oral sources. *Sonorités. Sonorités. Bulletin del l'AFAS, Association française des archives orales sonores et audiovisuelles. Questions de droit, d'éthiques et de terrain: qu'est-ce qui change dans les pratiques de recherche? Law, ethics and fieldwork: how are research practices changing?.*
- Calamai, S., Pretto, N., Monachini, M., Stamuli, M.F., Bianchi, S., Bonazzoli, P. (2020) *Building a Home for Italian Audio Archives*. In Navarretta, C. and Eskevich, M. (eds.) *Proceedings of CLARIN Annual Conference 2020*, p. 112-116. Virtual Edition, 2020. Retrieved from: [https://office.clarin.eu/v/CE-2020-1738-CLARIN2020\\_ConferenceProceedings.pdf](https://office.clarin.eu/v/CE-2020-1738-CLARIN2020_ConferenceProceedings.pdf); last visited December 28, 2020.
- Carucci, P. (1993) Prefazione. In Barrera G., Martini A. and Mulè A. (Eds.) *Fonti orali. Censimento degli istituti di conservazione*, 13-23. Roma: Pubblicazioni degli Archivi di Stato. (Quaderni della Rassegna degli Archivi di Stato, 71).
- International Council on Archives ICA, Brothman B. (1999) *ISAD(G): General International Standard Archival Description Second Edition Adopted by the Committee on Descriptive Standards Stockholm, Sweden, 19-22 September 1999*. Ottawa: International Council on Archives. Retrieved from: <https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>; last visited March 25, 2020.
- Miliano, M. (Ed.). (1999). *IASA cataloguing rules (IASA 1999)*. International Association of Sound and Audiovisual Archives IASA Editorial Group. Web edition. Retrieved from: <https://www.iasa-web.org/cataloguing-rules>; last visited April 6, 2020.
- Mulè A. (2005) *Le fonti orali in archivio. Un approccio archivistico alle fonti orali*. In: Ghezzi A. G. (Ed.) *L'Archivio: teoria, funzione, gestione e legislazione*, 284-296. Milano: Pubblicazioni dell'ISU Università Cattolica.
- Serini, P. (2003). *Attualità della letteratura grigia. Il ruolo delle biblioteche nella sua valorizzazione*, Biblioteche oggi - gennaio-febbraio 2003, Retrieved from: <http://www.bibliotecheoggi.it/2003/20030106101.pdf>; last visited March 26, 2020.
- Stamuli, M.F., Calamai, S., Monachini, M. (in press) *L'Archivio Vi.Vo: una piattaforma per la conservazione e l'accesso alle fonti orali in Toscana*. In: Brunetti, D., Robotti, D., Salvalaggio, E. *Documenti sonori. Voce, suono, musica in archivi e raccolte*, Torino, Centro studi piemontesi (Archivi e Biblioteche in Piemonte, 5).
- Pejšová, P. (2012). A linked-data vocabulary of grey literature document types: Version 1.0. *Libraries In The Digital Age (LIDA) Proceedings*, 12. Retrieved from <http://ozk.unizd.hr/proceedings/index.php/lida/article/view/87/55>;
- Sardelli, A. (2000). *Il fascino discreto della letteratura grigia: ridefinizione e ambiti di applicazione*. La letteratura grigia: politica e pratica. 3° Convegno nazionale, atti a cura di V. Alberani e P. De Castro, Roma, ISS, p. 15-20.
- Sommerville, I. (2016). *Software Engineering* (10 ed.). Pearson.