

Authorship and ownership in the digital oral archives domain: The Gra.fo digital archive in the CLARIN-IT repository

Silvia Calamai

University of Siena, Italy

`Silvia.calamai@unisi.it`

Francesca Biliotti

University of Siena, Italy

`francesca.biliotti@unisi.it`

Aleksei Kelli

University of Tartu, Estonia

`aleksei.kelli@ut.ee`

Abstract

The paper addresses the problem of authorship and ownership with relation to a digital oral archive created through the digitisation of several analogue archives. The case study is provided by the *Gra.fo* digital archive (*Grammo-foni. Le soffitte della voce*, Scuola Normale Superiore & University of Siena, Regione Toscana PAR FAS 2007-13), a collection of around 30 Tuscan oral archives that is in the process of being documented in the CLARIN-IT repository¹.

1 Introduction

Today, thanks to new and accessible technologies, oral recordings are enjoying a resurgence: on the one hand, technological process has brought recording tools within everybody's reach; on the other, many existing analogue archives are being digitised in order to ensure their preservation. Both in the recording of new audio data and in the digitisation of already existing ones, three aspects must be taken into careful consideration. Firstly, long-term preservation of data and metadata is essential for the persistence of the data derived from a research project beyond its limited timespan. Secondly, the choice of data and metadata formats is crucial in order to make data findable, available, interoperable and reusable. Thirdly, from a legal perspective archives are covered with several rights. Oral recordings containing original contributions constitute copyright protected works. Persons involved in these recordings have *inter alia* related rights (the rights of performers) and are entitled to personal data protection. Archives are also protected as databases (see Kelli et al. 2015).

The case study for the present paper is provided by the project *Grammo-foni. Le soffitte della voce* (*Gra.fo*; Scuola Normale Superiore & University of Siena, Regione Toscana PAR FAS 2007-13). *Gra.fo* discovered, digitised, catalogued and disseminated via a web portal (<http://grafo.sns.it/>) nearly 3000 hours of speech recordings stemming from around 30 oral archives collected by scholars and amateurs in the Tuscan territory. The *Gra.fo* digital archive is a heterogeneous collection of archives stemming from different disciplines and preserving a multitude of types of documents. It is also a highly hierarchised set: an archive of archives that in turn are made of subdivisions that can be further partitioned into subdivisions. As a collection of archives preserving valuable linguistic resources, the *Gra.fo* digital archive is in the process of being documented in the CLARIN-IT repository. At the beginning of this documentation process, issues concerning searchability, granularity and consistency of the metadata

This work is licenced under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

¹ The authors wish to thank Pawel Kamocki, Chiara Kolletzek, and Irene Mecatti for their legal advice. Any errors or shortcomings are entirely ours.

descriptors were expected and described in Calamai, Frontini (2016). Simultaneously, several crucial legal issues are being raised which concern the relation between the original archives (constituted by the recordings originally collected by the researchers) and the digital archives (made of the digital files derived from the original recordings), and the ownership and authorship applicable to these two entities. The paper introduces the description of the *Gra.fo* archive in the CLARIN-IT repository (§2), posing the question of what should be the reference – the original or the digital archive – (§2.1) and trying to answer it (§2.2); §3 is devoted to the discussion of the issues of authorship (§3.1) and ownership (§3.2) before coming to some conclusions on the matter (§4).

2 The starting point

The CLARIN-IT repository (<https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/>) is managed by the Institute of Computational Linguistics “A. Zampolli” of the National Research Council (ILC-CNR). For the Italian SSH community, *Gra.fo* represents a case study allowing the testing of the system in preparation for the documentation of other oral archives. Therefore, every decision made for the *Gra.fo* archive – arduous though it may be – is an important benchmark for future developments.

The *Gra.fo* digital archive enters the CLARIN-IT repository as an independent collection. The different archives comprised in the *Gra.fo* archive are described as single items within the collection, while no description is provided for the single oral documents constituting the archives. Such solution corresponds to the intermediate alternative put forward by Calamai and Frontini (2016) to the maximum-granularity option and the minimum-granularity option.

2.1 What shall we describe?

The description of the archives unveils the crucial problem of what should be the object of our description: the digital archives produced within the framework of *Gra.fo* or the original ones produced by the researcher(s) who collected the recordings in the first place. Upon close inspection, most of the metadata descriptors could be interpreted in one way or the other leading to opposite descriptions. For example, when indicating the contact person of a given archive, one could refer to the contact person of the original archive (i.e. mostly, its author or its owner, depending on the circumstances) or to those of the digital archive (i.e. the *Gra.fo* scientific coordinators); when stating the author of an archive, one could indicate the ‘creator’ of the original archive (e.g. the researcher) or that of the digital one (i.e. the *Gra.fo* consortium); and so forth.

Such dichotomy could presumably apply to every archive that is transferred from the analogue to the digital domain. Yet in *Gra.fo* the picture is made even more complex by the fact that the digital archives and the single digital oral documents accessible via the *Gra.fo* portal do not mirror the original ones, as they are the result of a meaningful interpretative activity, and can be therefore considered as derivative works. As described in Calamai, Biliotti, Bertinetto (2014), in speech recording fieldwork, a document (e.g. an interview, a narrative, etc.) can be distributed over various carriers or portions of carriers, so that one and the same carrier may contain various unrelated documents while more than one carrier can refer to one and the same document. This led *Gra.fo* to consider the documental unit as independent from the carrier, which is viewed as a mere container, and to create new digital oral documents corresponding to the single communicative events (e.g. interviews, narratives, etc.) contained in the original recordings. After being edited from digital copies of the original recordings, these new digital documents are extensively described, transcribed (in some cases), and made available to the end user.

2.2 What we think we should describe

Our viewpoint is in line with the recently emerging idea that digitisation does not produce a copy of the physical reality; it rather produces a new reality that – as such – deserves recognition and a proper treatment (Sheridan 2017). In the *Gra.fo* project, the digitisation often produces something different from the original analogue document and the ‘final object’ can be seen as the outcome of an interpretative process (digitisation is carried out by a technician who knows nothing about the content of the tapes, while the digital object that can be accessed via web is created by an expert cataloguer). Because the digital archives accessible via the *Gra.fo* portal do not mirror the original

ones, we believe that, in describing the *Gra.fo* archives in the CLARIN-IT repository, the digital archive should be the reference rather than the original, analogue one, provided that the source of the digital archive is clearly mentioned. Accordingly, when stating the size of an archive one should certainly indicate the number of digital oral documents it contains (rather than the number of open reel tapes or compact cassettes), since these are the documents that the user will find in the *Gra.fo* portal. Similarly, when stating the date of release of a given archive, one should refer to the date when the archive was made public in the *Gra.fo* portal. When indicating a contact person, one should always refer to the *Gra.fo* scientific coordinators, mentioning the contact person of the original archive only when that is deemed appropriate or useful for some reasons (sometimes the latter could be a mere depository and might not be willing to be contacted for issues related to the archive). Any relevant information concerning the original archive (who collected it, when, etc.) will be provided in the “Description” box in the metadata record.

3 Who are the author and owner of the *Gra.fo* oral archive?

With respect to the metadata issues addressed in §2.2, one exception exists: for the metadata descriptor related to the author, we think that both the author of the digital archive and that of the original one should be mentioned. The reasons for this are various. Firstly, the work of interpretation, editing, description and transcription carried out by the *Gra.fo* consortium certainly deserves recognition: the digital archive is not a mere copy of the analogue archive since it has new creative input in it. Therefore, the digital archive can be considered a derivative work of the analogue archive (see Art. 4 Law 22 April 1941, n. 633). Secondly, the policies of the *Gra.fo* portal clearly state that all its contents are the fruits of the *Gra.fo* staff’s labour. However, without the original archive, *Gra.fo*’s work would simply be non-existent. Thus, the contribution of the researchers who collected the original recordings should be recognised as well, since the original researchers are the authors of the corresponding original archives. The following paragraph is devoted precisely to this thorny issue.

3.1 Authorship

According to the Berne Convention for the Protection of Literary and Artistic Works “the author shall have the right to claim authorship of the work” (Berne Convention Article 6bis).² The question whether oral archives – mainly represented by interviews and answers to questionnaires – fall within the realm of *Diritto d’Autore* law is increasingly a matter of discussion among lawyers and scholars dealing with oral archives and an in-depth analysis of this issue will be provided in the long version of the paper.

While there is no doubt that the authors of the digital archive are the researchers who worked for the *Gra.fo* project, identifying the authors of the original archives is a more complex task. In the domain of oral archives, the author is not identifiable in a straightforward way, nor is the only subject holding rights over an archive. In fact, at least four entities are entitled economic rights over an oral archive: the informant(s), the researcher(s) who collected the document, the individual/organisation commissioning the research, the individual/organisation at whom/which the archive is deposited (Le Draoullec 2006, Stéphan 2013). Discerning which of these should be given the status of ‘author’ might not be straightforward. Such issue becomes extremely complex when it comes to ‘collective’ archives created in the context of some geo-linguistic enterprise. Archive *Carta dei Dialetti Italiani* (one of the greatest endeavours in Italian dialectology research: see Calamai, Bertinetto 2012) is a convenient example: one scholar conceived and directed the enterprise, each region had its own research team that was directed by a coordinator, many different researchers carried out fieldwork, and many speakers were interviewed. Thus, who should be recognised as the author? According to the Italian *Diritto d’Autore* Act (Art. 7), we may consider Oronzo Parlangei, who was at the time the scientific coordinator of the enterprise, the author of *Carta dei Dialetti Italiani*. Looking at other

² With a reference to the doctrine of moral rights, it is suggested that the Berne Convention might conceptualise the author as the natural person who created the work rather than any legal person or entity (Adeney 2006, 115). The concept of authorship is, however, controversial and copyright laws of continental Europe and common law countries diverge. Continental law countries usually define the author as a natural person. Common law countries accept legal entities as authors as well (e.g. in case of work made for hire) (Ricketson, Ginsburg 2006, 358-363).

experiences of archives preservation and dissemination in the world is certainly useful, but the guidelines derived from each experience clash with the fact that every country has its own relevant legislation.

Laws about authorship differ greatly from one country to the other. In the American tradition, according to MacKay (2016, pp. 75-76), “the speakers in the recorded interview automatically own their own words from the moment they are spoken, until or unless transferred to another entity through a legal release agreement”. According to French norms, instead, researchers are the authors of their recordings, or they might be co-authors in the case a speaker participates in the exchange creatively (Stérin, 2016). According to the Italian *Diritto d’Autore* Act (Art. 2), authorship is applicable to original, creative works produced in literature, music, visual arts, architecture, drama and cinema. Nevertheless, the list is merely illustrative, and not complete. At the beginning of the project, a rather conservative approach was adopted and authorship was recognised only to authors of oral improvised poetry where the poet creates something totally original (the opposite case is represented by answers to questionnaires where the respondent’s task is limited to the translation of linguistic elements into his/her own dialect). The topic will be deeply explored in the long version of the paper.

3.2 Ownership

Oral archives have a complex life. Those who spent their lives doing fieldwork might have guarded their archives jealously. This is often the case for the researchers who financed their own research and are, therefore, also the owners of their archives (e.g., among the *Gra.fo* archives, Archive ‘Vanna Brunetti’). Those who received funding for their research might have consigned their recordings to the funding organisation (e.g. Archive ‘Duse Lemetti – Gruppo Vegliatori’ was entrusted to the Municipality of Galliciano, which financed the research), or to an organisation that guaranteed the physical conservation of the materials (e.g. Archive ‘Angela Spinelli’ was consigned to Biblioteca Lazzarini). Some persons might have come to own an archive through their friends (e.g. Archive ‘Edda Ardimanni’ is made of recordings that were collected by different – unknown – researchers and then donated to Edda Ardimanni). Yet others might have inherited an archive from a deceased relative (e.g. Archive ‘Anna Maria Bruzzone’ belongs to one of Bruzzone’s heirs). Therefore, owners may correspond to the researcher who collected the recordings, to the organisation funding the research, to the organisation guarding the archive, or even to other persons that were not involved in the research at all (researchers’ heirs or friends). In *Gra.fo*, the archives’ owners were asked to sign a legal agreement for granting their rights over the recordings to the *Gra.fo* project so that these could be digitised, catalogued, transcribed, and disseminated through the web portal. Thus, the original recordings were retained for the time necessary to work on them and then returned to their legitimate owners. On the contrary, the portal and everything it contains (including the edited audio files and the relative descriptions) belong to Scuola Normale Superiore, University of Siena and Regione Toscana. This analysis takes us to the following conclusion: as for the digital archives, the authors are the researchers who worked at *Gra.fo*, while the owners are SNS, UNISI and Regione Toscana.

4 Conclusion

Among Italian scholars, issues related to rights over oral recordings are now taken into great consideration if compared to the past, when little or no attention was given to legal and ethical issues. Undoubtedly, Italy lacks such a reference point as the French “Questions éthique et droit en SHS” (<https://ethiquedroit.hypotheses.org/>). However, the need to agree upon some guidelines is increasingly felt and 2015 saw the launch of the AISO’s (Italian Oral History Association) *Good practices in oral history* (<http://aisoitalia.org/?p=4795>), the culminating point of a long-standing process of reflection and discussion carried out by oral historians, anthropologists, and legal experts. What the conclusions of the meetings of the AISO’s Good practices working group seem to suggest is that there is no general rule for establishing who the author of an oral document is and, moreover, the identity of the author cannot be decided *a posteriori*: only the agreements made between interviewer and interviewee in the context of the interview can tell us who the author of that document is (Sinello 2015). The problem is crucial for those undocumented archives created in the 60s and 70s (when authorship and privacy were not

common issues among linguists and historians) that now demand hard work from the curators' part to reconstruct their story. Therefore, the inclusion of the *Grafo* archives in the CLARIN-IT repository appears not only as a metadata 'translation', but also as a refined reflection on authorship, ownership, and on the relationship between original source and digital objects.

References

- [Adeney 2006] Elizabeth Adeney. *The Moral Rights of Authors and Performers. An International and Comparative Analysis*. Oxford University Press, 2006.
- [Berne Convention] Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886, completed at PARIS on May 4, 1896, revised at BERLIN on November 13, 1908, completed at BERNE on March 20, 1914, revised at ROME on June 2, 1928, at BRUSSELS on June 26, 1948, at STOCKHOLM on July 14, 1967, and at PARIS on July 24, 1971, and amended on September 28, 1979 <http://www.wipo.int/treaties/en/text.jsp?file_id=283698#P85_10661>
- [Calamai, Bertinetto 2012] Silvia Calamai, Pier Marco Bertinetto. 2012. Per il recupero della Carta dei Dialetti Italiani. T. Telmon, G. Raimondi, L. Revelli (eds.) *Coesistenzae linguistiche nell'Italia pre- e postunitaria. Atti del XLV Congresso internazionale di studi della Società di Linguistica Italiana (Aosta/Bard/Torino 26-28 settembre 2011)*, 2 voll., Roma, Bulzoni: 335-356. [ISBN: 978-88-7870-722-1].
- [Calamai, Biliotti, Bertinetto 2014] Silvia Calamai, Francesca Biliotti, Pier Marco Bertinetto. 2014. Fuzzy archives. What Kind of an Object Is the Documental Unit of Oral Archives?. M. Ioannides et al. (eds.) *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection EuroMed 2014*, LNCS 8740, pp. 777–785.
- [Calamai, Frontini 2016] Silvia Calamai, Francesca Frontini. 2016. Not quite your usual kind of resource. *Grafo* and the documentation of Oral Archives in CLARIN. *CLARIN Annual Conference 2016*, Oct 2016, Aix-en-Provence, France. <https://www.clarin.eu/sites/default/files/calamai-frontini-CLARIN2016_paper_14.pdf>.
- [Kelli et al. 2015] Aleksei Kelli, Kadri Vider, Krister Lindén (2015). The Regulatory and Contractual Framework as an Integral Part of the CLARIN Infrastructure. 123: Selected Papers from the CLARIN Annual Conference 2015, October 14–16, 2015, Wrocław, Poland. Ed. Koenraad De Smedt. Linköping University Electronic Press, Linköpings universitet, 13–24 (<http://www.ep.liu.se/ecp/article.asp?issue=123&article=002>).
- [Le Draoullec 2006] Ludovic Le Draoullec. 2006. L'utilisation des corpus oraux à des fins culturelles : quels contrats mettre en œuvre ?. *Bulletin de l'AFAS* (URL: <http://afas.revues.org/622>; DOI : 10.4000/afas.622).
- [MacKay 2016] Nancy MacKay. 2016. *Curating Oral Histories, Second Edition: From Interview to Archive*. Routledge.
- [Ricketson, Ginsburg,] Sam Ricketson, Jane C. Ginsburg. *International Copyright and Neighbouring Rights: The Berne Convention and Beyond. Second Edition. Volume I*. Oxford University Press, 2006.
- [Sheridan 2017] John Sheridan. 2017. *Creating the disruptive digital archive*. Digital Preservation Coalition (<http://www.dpconline.org/blog/disruptive-digital-archive>).
- [Sinello 2015] Rachele Sinello. 2015. *Le vite degli altri. Verso la definizione delle linee guida italiane per la storia orale*, Tesi di laurea, Corso di Laurea triennale in Lettere, Università Ca' Foscari di Venezia, A.A. 2014/2015, relatore prof. Alessandro Casellato.
- [Stéphan 2013] Lena Stéphan. 2013. *Les archives sonores : conservation et valorisation du patrimoine oral. Mémoires Master "Archives numériques"*, École Nationale Supérieure des Sciences de l'Information et des Bibliothèques.
- [Stérin 2016] Anne-Laure Stérin. 2016. *Le chercheur mène un entretien enregistré* (<http://ethiquedroit.hypotheses.org/1133>).