Towards a protocol for the curation and dissemination of vulnerable people archives

Silvia Calamai University of Siena Italy Chiara Kolletzek Lawyer and Record Manager, Bologna, Italy Aleksei Kelli
University of Tartu
Estonia

silvia.calamai@unisi.it chiara.kolletzek@live.it aleksei.kelli@ut.ee

Abstract

This paper aims at introducing possible guidelines in defining a protocol for the curation and dissemination of speech archives, which appear to have $-de\ jure$ – the highest restrictions on their curation and dissemination. This case study has been undertaken because of the discovery of the Anna Maria Bruzzone archive, containing the voices of people with mental disabilities recorded in 1977 in an Italian psychiatric hospital.

1 Introduction

This paper presents a coherent reflection on the possibility of defining a protocol for the curation and dissemination of speech archives which appear to have – $de\ jure$ – the highest restrictions on their curation and dissemination since they contain the voices of insane people. This case study has been undertaken because of the discovery of the Anna Maria Bruzzone archive¹. Bruzzone's interviews were recorded long before the Italian Data Protection Code (IDPC) was issued (2004), so that the informants were not explicitly asked to give their authorization for the use and dissemination of the recordings, although during the interviews the recording device was always kept visible.

The archives are covered with several rights (for further discussion, see Kelli et al. 2015). Firstly, speech itself could be protected as copyrighted work. Secondly, individuals who speak could have performer's rights. Thirdly, the person who created the archive has database rights. Lastly, interviewees' personal data have to be protected from unauthorized use and dissemination. Due to the focus of this article, the analysis is limited to personal data protection.

In the paper, some legal issues affecting the use and re-use of the archive are presented and discussed. The model envisaged aims at finding a balance between the rights of the recorded people (and their heirs) such as privacy, and the right of information and the protection of memory. The focus is on the General Data Protection Regulation (GDPR) which is applicable in all EU member states from 25 May 2018. National laws may specify its application, especially the provisions concerning specific areas of personal data processing (e.g., research: see Kelli et al. 2018).

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

¹ In the near future, the Archive will be part of the CLARIN Infrastructure and metadata description will be pursued according to COALA. COALA generates corpus and session CMDIs according to the media-corpus-profile and the media-session-profile for the Component Registry, by converting five CVS tables to the CMDI format. A mobility grant under the H2020 project CLARIN-PLUS allowed the first author to prepare a feasibility study on this topic (Bayerisches Archiv für Sprachsignale c/o Institut für Phonetik, Universität München; 4-7 December 2017).

The paper has also benefited from the CLARIN workshop "Hacking the GDPR to Conduct Research with Language Resources in Digital Humanities and Social Sciences" (Vilnius, 7 December 2018), which aimed at bringing together legal experts and researchers from the Digital Humanities and Social Sciences disciplines working with Language Resources in order to exchange views and explore ways of creating and using LRs under the GDPR regime.

Silvia Calamai, Chiara Kolletzek and Aleksei Kelli 2019. Towards a protocol for the curation and dissemination of vulnerable people archives. *Selected papers from the CLARIN Annual Conference 2018*. Linköping Electronic Conference Proceedings 159: 28–38.

The paper is organised as follows: in § 2 Bruzzone's speech archive is described, in § 3 the topic of personal data and special categories of personal data is addressed, while in § 4 and in the Conclusion the possibilities of finding a balance between research, dissemination and protection of privacy are discussed. Finally, the English translation of the informed content is provided in the Appendix.

2 The speech archive of Anna Maria Bruzzone

Anna Maria Bruzzone's book *Ci chiamavano matti. Voci da un ospedale psichiatrico* (Einaudi, Torino 1979) contains the testimonies of thirty-seven patients in the Arezzo psychiatric hospital collected in 1977 (see Fig. 1).



Fig. 1 The book and the original recording device used for the fieldwork.

The book – out of print – testifies the patients' miserable lives inside and outside the hospital and sheds light on the atrocity of their everyday condition by letting them speak for themselves. The author wrote it after a two-month stay in Arezzo, when she spent almost every day in the hospital, attending the general meetings and participating in the lives of the inpatients, in a continuous dialogue of which only a part is collected in the published interviews. The oral recordings on which the book is based were believed to be lost forever. After a long and strenuous search we have been able to locate the original tapes (see Fig. 2), which were donated to the Department of Educational Sciences, Human Sciences and Intercultural Communication of the University of Siena (UNISI) – Arezzo.



Fig. 2 The original tapes of Anna Maria Bruzzone's archive.

This discovery is extremely important because the digitisation and cataloguing of this archive would produce the first digital oral archive related to an Italian psychiatric hospital – which was located in the same buildings as the UNISI Department, where the Historical Archive of the Arezzo psychiatric hospital is also housed.

Reading a testimony and listening to it from the voice of the interviewee are not the same thing and Bruzzone herself was well aware of this (Bruzzone 1979: 22). Furthermore, the published texts are not the exact transcriptions of the original testimonies. In fact, after producing the first, complete transcriptions, Bruzzone had to edit them to make them suitable for publishing. In addition to editing out the speeches so that the interviewees' voices could flow without interruptions, she had to make other cuts and adjustments in order to make the text clearer or more readable, and she even had to give up on publishing some of the testimonies because otherwise, the book would have been too long (see Fig. 3). As she admits, this task was a hard, painful one to her (Bruzzone 1979: 25). Therefore, having the original tapes at our disposal is of fundamental importance, as it allows us to re-connect the published testimonies to the original ones.

The archive consists of 36 tapes accompanied by the handwritten and the typewritten transcriptions of all the interviews. In addition to the complete transcriptions, different versions show all the work of editing made by A.M. Bruzzone so that the interviews could be prepared for publishing (see Fig. 3-4).

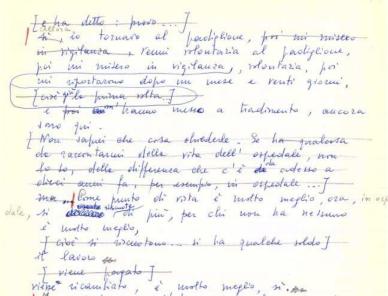


Fig. 3 The handwritten transcription.

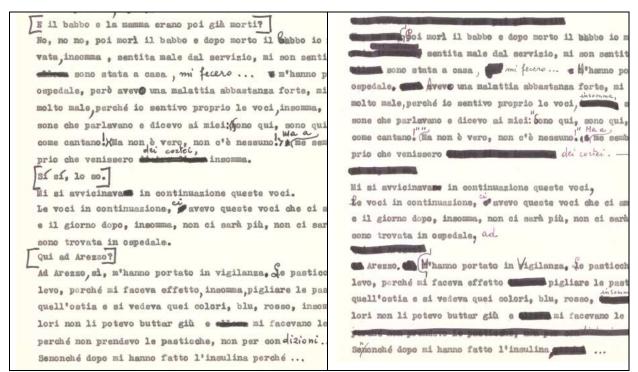


Fig. 4 The typewritten transcription (two versions of the same passage in the interview).

This opens up the possibility to understand, document and examine the changes undergone in an interview from the moment it was recorded on tape to its publication in the book, through the comparative study of all the available documents: the original audio recording, the first, handwritten transcription, the typewritten transcription, the edited version and, finally, the one published in the book. Moreover, it is now possible to associate the oral life stories with the medical diagnosis of every single inpatient (preserved in the Historical Archive of the Arezzo psychiatric hospital), since the real names and not the pseudonyms were been found in the box of every single tape.

3 Personal data and special categories of personal data

The curation and dissemination of archives of vulnerable people are subject to the regulation of personal data. The GDPR defines personal data as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". Article 29 Working Party² (WP29) explains that "it is not necessary for the information to be considered as personal data that it is contained in a structured database or file. Also information contained in free text in an electronic document may qualify as personal data" (2007: 8).

The critical issue here is how to interpret the concept of 'identifiable'. The absolute and relative approaches described in the literature are displayed in Fig. 5 (from Spindler, Schmechel 2016).

_

² According to the Data Protection Directive, the Working Party on the Protection of Individuals with regard to the Processing of Personal Data (WP29) is composed of a representative of the supervisory authority or authorities designated by each Member State and of a representative of the authority or authorities established for the Community institutions and bodies, and of a representative of the Commission. The GDPR replaces the Data Protection Directive.



Fig. 5 Absolute and relative approach in the identification of the data subject.

Some authors have emphasized the context-dependency of identifiability (Oostveen 2016: 306). In the analysed case, the individuals are identifiable and no further analysis is required.

The situation concerning speech archives becomes even more complicated for several reasons. Firstly, the human voice is considered biometric data (see González-Rodríguez et al. 2008; Jain et al. 2004). Biometric data is defined as "personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person" (GDPR art. 4). Secondly, the archive under consideration concerns health data³. Biometric and health data both belong to the special categories of data (sensitive or delicate data). According to the GDPR, special categories of personal data are "data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation" (Art. 9). Interviews with psychiatric patients relate to special categories of personal data; they took place inside the psychiatric hospital, and they explicitly and directly identify the subjects as 'patients' or - more often – as 'crazy' (with a label that also has serious consequences for the inpatient's family). Thus, for the data subject, the interviews highlight data about health, but also about sex life, racial/ethnic origin, religious, philosophical or other beliefs. All these data must be processed with special protection measures.

Another relevant question is whether curation and dissemination of archives is processing personal data. The GDPR conceptualizes the processing in a very broad manner so that it catches almost all data-related activities. According to the GDPR, processing *inter alia* covers collection, structuring, storage, adaptation retrieval, use, dissemination, erasure or destruction. It means that the curation and dissemination of the archive constitute the processing of personal data. The potential options for using speech archives are analysed in the next section.

The challenge: how to strike a fair balance between research, dissemination, and 4 protection of privacy

The primary challenge for historical and linguistic research on past speech archives is represented by finding a balance between two socially relevant interests: the protection of personal data (the right to privacy) and the transmission of knowledge and freedom of research. Privacy and data protection do not exist in isolation. On the one hand, the Charter of Fundamental Rights of the European Union (Charter) protects private and family life and personal data (Art. 7-8). On the other hand, freedom of expression, information and science are also protected (Art. 11, 13). Even the GDPR itself expresses the following principle: "[t]he processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle

impact on the data subject's rights.

³ Early EU case law determined that even "[r]eference to the fact that an individual has injured her foot and is on half-time on medical grounds constitutes personal data concerning health" Case C-101/01). The analysed case has more intensive

of proportionality" (Recital 4). Although the principles above can be used as guidelines, there is a need to search for possible solutions. The following routes are briefly described: duration of the data subject's rights, anonymisation, consent and research exemption.

The GDPR does not apply to the personal data of deceased persons (Recital 27). It means that EU member states can regulate the issue. WP 29 has correctly pointed out that data on the dead can relate to the living and be protected personal data (2007: 22). Therefore, this option is not a solution to the problem.

The GDPR also does not apply to anonymous data, which means the natural person is not identifiable (Recital 26). It is explained in the literature that "there is a strong incentive to anonymise data. Through anonymisation the data are placed outside the scope of data protection; by making data non-identifiable, the controller is relieved of the burden of compliance with data protection's rules and limitations" (Oostveen 2016: 307). WP29 in its opinion on the anonymisation techniques emphasizes "the potential value of anonymisation in particular as a strategy to reap the benefits of 'open data' for individuals and society at large whilst mitigating the risks for the individuals concerned. However, case studies and research publications have shown how difficult it is to create a truly anonymous dataset whilst retaining as much of the underlying information as required for the task" (2014: 3). WP29 describes the problem very well. Anonymisation of data without destroying its informational value is almost impossible. In fact, anonymisation may not correspond to the needs and scope of historical research, which may be interested in, among others, the analysis of the relational structures of individuals. In short, historians are interested in "names and faces".

There are two potential legal grounds to use the archives (GDPR Art. 6): 1) consent; 2) processing in the public interest.

It is explained in the GDPR that it is not always possible to fully identify the purpose of processing for research purposes. Therefore, data subjects may give their consent to certain areas of scientific research (Recital 33). The GDPR defines consent as: "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her" (Art. 4 (11)). There are several additional consent requirements which can be summarised as follows (GDPR Art. 7):

- 1) the controller has to demonstrate that the data subject has consented to processing;
- 2) consent is presented in a manner which is clearly distinguishable from the other matters;
- 3) consent is in an intelligible and easily accessible form, using clear and plain language;
- 4) the data subject has the right to withdraw his or her consent at any time:
 - 4.1) the withdrawal of consent does not affect the lawfulness of prior processing;
 - 4.2) the data subject is informed of the right to withdraw;
 - 4.3) it is as easy to withdraw as to give consent.

Processing of special categories of personal data requires explicit consent (GDPR Art. 9 (2)a).

The consent requirements are visualised in the following graph:

⁴ It is also necessary to bear in mind that anonymisation itself is a further processing of personal data which must meet the GDPR requirements (WP29 2014: 3).

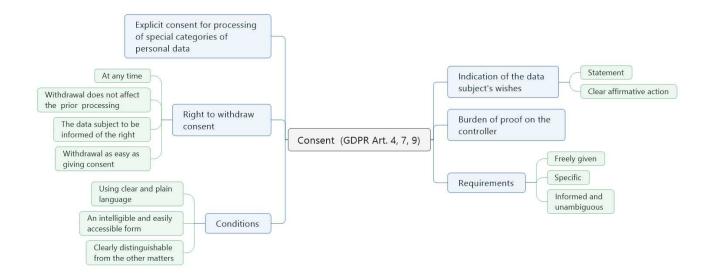


Fig. 6 Consent requirements in the GDPR regime.

The last option is to process personal data on the grounds of research exemption. GDPR prohibits the processing of special categories of personal data unless special grounds exist (Art. 9 (1)). Processing of special categories of personal data is allowed if it is necessary for scientific or historical research purposes. It must be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject (GDPR art. 9 (2) j) (For further discussion on processing personal data without consent see, Kelli, et al. 2018).

Given this general framework, we can return to the case study under investigation, since the re-use of Bruzzone's interviews for historical purposes could be a prime example of establishing a "legal chain" for personal data processing, which can be summarized as follows. Firstly, throughout the inspection of medical records – if any – preserved in the Psychiatric Hospital Archive, the research group identified the real names of the patients and matched them with the pseudonyms, as attested to the volume (Bruzzone 1979). Secondly, the research group tried to go back to the interviewees, even contacting all the network of the people – physicians, nurses, social workers, ordinary citizens – involved in the recent history of the Psychiatric hospital.

This reconstruction helped to obtain detailed and clear informed consents (for further discussion on the consent see WP29 2017; GDPR art. 7, 9 (2) a), describing the aims, the scope and the positive spill-over effects of the dissemination of such an oral archive (see Appendix, with the translated consent form). If the consent form is obtained, the oral archive could be finally enjoyed by the research communities and the entire civil society.

At present two former patients were discovered. One of the two – who is now living an ordinary life – signed the consent form, gave additional interviews to the research group, and also actively collaborated with the project. The other one is now in a nursing home and has a legal guardian who agreed to sign the informed consent. Reasonably, it might be concluded that all the other voices belong to deceased persons. A plan of communication and dissemination was therefore set up in order to reach possible relatives and right holders, and two events were organised in late 2018 in order to disseminate the research project. One was included in Bright 2018 The European Researchers' Night at University of Siena (in Arezzo), while the other took place in an extra-academic context, in order to reach a different public and to involve different networks (see Fig. 7 and 8 respectively).

_

⁵ Carried out by a linguist (the first author), a philosopher of science (Marica Setaro), an oral historian (Caterina Pesce), and an archivist (Lucilla Gigli).



Fig. 7 The European Researchers' Night at University of Siena

Fig. 8 The poster of the dissemination event ("Shared voices") in a meeting place in Arezzo

There was much media coverage of the events. The regional television news interviewed Roberto (the former patient) who listened live to his 1977 voice (see Fig. 9)⁶.



Fig. 9 The boardcast service ("Voices from the madhouse").

5 Conclusions

The analysed archive is subject to the GDPR since it contains special categories of personal data. The curation and dissemination of the archives is the processing of personal data which requires legal grounds and other measures assuring GDPR compliance. Personal data relating to the deceased data subject could still be protected due to its links to living individuals. One option would be data anonymisation which is not always an option for historical research. An additional option is to acquire explicit informed consent. If it is not possible to obtain consent, the research exemption might be applicable. The use of research exception requires the introduction of safeguards protecting the data subject's rights (e.g. pseudonymization, limited access, and so forth).

_

⁶ The broadcasting service can be retrieved at the following url: https://www.rainews.it/tgr/toscana/video/2018/09/tos-arezzo-manicomio-bruzzone-88ee6948-a3ab-4135-a034-fdec303b90cd.html?wt_mc=2.www.fb.tgrtoscana_ContentItem-88ee6948-a3ab-4135-a034-fdec303b90cd.&wt (01.04.2019)

References

- [Case C-101/01] Case C-101/01. Criminal proceedings against Bodil Lindqvist. 6 November 2003. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1521039149443&uri=CELEX:62001CJ0101 (14.4.2018);
- [Charter] Charter of Fundamental Rights of the European Union. 2012/C 326/02. OJ C 326, 26.10.2012, p. 391–407 (BG, ES, CS, DA, DE, ET, EL, EN, FR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV). Available at http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT (14.4.2018);
- [Data Protection Directive] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L 281, 23/11/1995 p. 0031 0050. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046&qid=1522340616101&from=EN (29.3.2018);
- [GDPR] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ L 119, 4.5.2016, p. 1-88. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1515793631105&uri=CELEX:32016R0679 (29.3.2018);
- [González-Rodríguez et. al. 2008] Joaquín González-Rodríguez, Doroteo Torre Toledano, Javier Ortega-García (2008). Voice Biometrics. In Handbook of Biometrics edited by Anil K. Jain, Patrick Flynn, Arun A. Ross. Springer;
- [IPDPC] Italian Personal Data Protection Code. Legislative Decree 30.06.2003 No. 196. English version available at: http://194.242.234.211/documents/10160/2012405/Personal+Data+Protection+Code++Legislat.+Decree+no.196+of+30+June+2003.pdf (11.4.2018);
- [Jain et. al. 2004] Anil K. Jain, Arun Ross, Salil Prabhakar (2004). An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology 14(1). Available at https://www.cse.msu.edu/~rossarun/BiometricsTextBook/Papers/Introduction/JainRossPrabhakar_BiometricIntro_CSVT04.pdf (31.3.2018);
- [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, Wroclaw, Poland. Ed. Koenraad De Smedt. Linköping University Electronic Press, Linköpings universitet, 13–24. Available at https://www.ep.liu.se/ecp/article.asp?issue=123&article=002 (20.8.2018);
- [Kelli et al. 2018] Kelli, Aleksei; Lindén, Krister; Vider, Kadri; Kamocki, Pawel; Birštonas, Ramūnas; Calamai, Silvia; Kolletzek, Chiara; Labropoulou, Penny; Gavriilidou, Maria (2018). Processing personal data without the consent of the data subject for the development and use of language resources. CLARIN Annual Conference 2018 Proceedings: CLARIN Annual Conference 2018, 8-10 October 2018 Pisa, Italy. Ed. Inguna Skadin, Maria Eskevich. CLARIN, 43–48. Available at https://office.clarin.eu/v/CE-2018-1292-CLARIN2018_ConferenceProceedings.pdf (28.1.2019);
- [Oostveen 2016] Manon Oostveen. Identifiability and the applicability of data protection to big data. International Data Privacy Law, 2016, Vol. 6, No. 4, 299- 309;
- [Spindler, Schmechel 2016] G. Spindler, P. Schmechel (2016). Personal Data and Encryption in the European General Data Protection Regulation. JIPITEC 7 (2), 163-177. Available at https://www.jipitec.eu/issues/jipitec-7-2-2016/4440/spindler_schmechel_gdpr_encryption_jipitec_7_2_2016_163.pdf (14.4.2018);
- [WP29 2017] WP29. Guidelines on Consent under Regulation 2016/679. Adopted on 28 November 2017 [adopted, but still to be finalized]. Available at http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=615239 (2.4.2018);
- [WP29 2014] WP29. Opinion 05/2014 on Anonymisation Techniques. Adopted on 10 April 2014. Available at http://collections.internetmemory.org/haeu/20171122154227/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf (20.8.2018).

[WP29 2007] WP29. Opinion 4/2007 on the concept of personal data. Adopted on 20th June. Available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf (26.3.2019).

Appendix. The English translation of the informed consent form

To the relatives of the former patients of the Neuro-psychiatric hospital in Arezzo

Dear Sir/Dear Madam,

We are a research group at Siena University and we are writing to you because we have succeded in finding the compact cassettes collected in the summer of 1977 containing the interviews of the patients in the Neuro-psychiatric hospital carried out by Anna Maria Bruzzone (and partially published in the volume *Ci chiamavano matti*, Einaudi 1979). These cassettes contain invaluable and unique information and we hope that in the near future it will be possible to listen to these precious and touching life stories, and thus to give voice to voiceless people.

These cassettes were digitized and were located in the Arezzo archive. In order to make them accessible for to scientific research it is necessary to provide an informed consent from the relatives of the patients. Obviously, the single interview will be disseminated anonymously or under the pseudonym used by Bruzzone herself, with all due caution. We do want first to honor your wishes, but at the same time we also want to fulfil our research duties, i.e. to ensure that such information is not lost and that a relevant piece of history – non only for the city of Arezzo – could pass to future generations.

Therefore we need your consent in order to work – as researchers – on the retrieved speech archive, which is now bound by the "Soprintendenza Archivistica e Bibliografica della Toscana". Needless to say, <u>all our activities are without interest for financial gain and take place at University (as part of teaching, research, and public engagment)</u>. If you agree with the spirit that underpins our initiative, we kindly ask you to sign the attached informed consent. We also would like to offer you a digital copy of the interview containing the voice of your relative.

We are at your disposal for additional information.

Contact information:

Silvia Calamai, coordinator of the Scientific Committee of the Historical Archive of the neuro-psychiatric hospital silvia.calamai@unisi.it 0575926439

Lucilla Gigli, archivist at the Historical Archive of neuro-psychiatric hospital lucilla.gigli@unisi.it 0575926264 - 0575926292

Authorisation to use		
I undersigned		
Born	on	and resident in
First-degree relative of		
(or)		
Second-degree relative	of	
	Authorise	

The research group led lead by Silvia Calamai, associate professor at Siena University, Department of Education, Humanities and Intercultural Communication

1- to the re-use of my relative's interview for teaching and research activities:				
[yes] [no]				
2- to the consulting of the interview by third parties under the follows	ng conditi	on: (choose only one option)		
[] free consulting				
[] consulting after years from the present date				
[] consulting subordinate to my approval				
I relinquish all rights on a royalty-free basis for				
- Teaching and research activities	[yes]	[no]		
- Publishing of the partial or full interview	[yes]	[no]		
Any other use shall be subject to my additional authorisation.				
My details are as follows:				
Postal address	-			
E-mail address	_			
Phone number				
All personal data will be used and processed by Siena University solely for the present purposes, and their protection and confidentiality will be ensured in accordance with General Data Protection Regulation (art. 5, art 15ff).				
Place and date		Signature		