Sound Archives Accessibility

Silvia Calamai, Veronique Ginouvès, and Pier Marco Bertinetto

Abstract

The paper analyses the conflicting issues that arise when dealing with Intangible Cultural Heritage (ICH) held in audio digital archives, when the demand for open access conflicts with ownership rights and ethical issues. It describes two case studies in order to evaluate the procedures used for doing research on oral materials while respecting the rights of others. The first refers to the activities carried on at the *Phonothèque de la Maison méditerranéenne des sciences de l'homme*, a French sound archive; the second refers to the solutions envisaged by an Italian research project, *Grammo-foni. Le soffitte della voce* (*Gra.fo*), jointly carried out by Scuola Normale Superiore of Pisa and the University of Siena.

S. Calamai (✉)
Dipartimento di Scienze della formazione, Università di Siena, Scienze umane e della Comunicazione interculturale viale Cittadini 33, Arezzo 52100, Italy
e-mail: silvia.calamai@unisi.it

V. Ginouvès
Phonothèque de la Maison méditerranéenne des sciences de l'homme, USR 3125 AMU/CNRS, 5 Rue Château de l'Horloge, Aix-en-Provence 13090, France
e-mail: veronique.ginouves@univ-amu.fr

P.M. Bertinetto
Scuola Normale Superiore, Pisa, Piazza dei Cavalieri, 7, Pisa 56126, Italy
e-mail: piermarco.bertinetto@sns.it

1 Introduction

This paper addresses the impact of the computational era on a very peculiar 'public space': web portals containing digital audio archives.[1] Digital audio archives are the final outcomes of several disciplines, from oral history to linguistics, from anthropology and ethnography to social sciences. They usually contain a significant variety of research data referring to different textual *genres* (e.g., a sequence of interviews on a particular topic; answers to a *questionnaire*; speech *corpora*; spontaneous or semi-spontaneous speech). But they may also contain public events of many kind (e.g., performing art events like for instance folk drama or folk poetry performances, but also political meetings and assemblies). Although they most frequently arise in academic communities and networks, digital audio archives are also created by heritage communities, informal groups and individuals who are interested in their preservation and accessibility.

The concern about digital audio archives is particularly relevant since it addresses issues such as ownership, distributed and entangled responsibility, open access and privacy. The internet appears to be a significant extension of the public space; nevertheless, the distinction between private and public is more important than ever. In addition, the development of Information Communications Technologies (ICTs) modifies our relationships to cultural heritage and archive maintenance. It 'democratises' the access to the data, since it resides and tends to multiply in a throng of repositories and sources. As a consequence, the world of knowledge has become a world of abundance where all pieces of information are always at everybody's disposal, but at the same time the quantity of available contents exceeds by far, more than ever, our cognitive abilities (Ganascia 2015: 67–68). Given this background, the domains of audio documents stemming from fieldwork and oral data collection—both of which contribute to the creation of audio archives—represent an interesting and under-investigated scenario, where at least three intertwined concerns emerge:

- Use and re-use of research data;
- Ethical questions involved in the re-use of research data;
- Legal questions stemming from online diffusion.

These three issues represent a cross-curricular area concerning researchers, scholars, archivists, librarians, public and research institutions. Research data archiving, accessibility and re-use are nowadays at the centre of scientific debate, among different scientific communities around the world. In this respect, the *data deluge* described in the monographic volume of *Science* 331 (2011) appears to be paradigmatic of the renewed attention towards data collection, curation, and access.

---

[1] 'Oral archives', 'sound archives', 'audio archives', 'speech archives' are considered as synonymous in the present paper, although they may refer to different traditions, according to different branches of knowledge.

While available data are exponentially growing, it is crucial for many disciplines to decide which data to preserve and which to dismiss, how to access the archived data and how to reuse them in a consistent, sustainable, ethically-correct way. This need has been strongly felt by physicists, who in 2009 created a working group called 'Data Preservation in High Energy Physics' (DPHEP). More complex and even contradictory appears to be the debate in the domain of the social sciences:

> Although information overload has always been an issue for scholars, today the infrastructural challenges in data sharing, data management, informatics, statistical methodology, and research ethics and policy risk being overwhelmed by the massive increases in informative data. Many social science data sets are so valuable and sensitive that when commercial entities collect them, external researchers are granted almost no access. Even when sensitive data are collected originally by researchers or acquired from corporations, privacy concerns sometimes lead to public policies that require the data be destroyed after the research is completed—a step that obviously makes scientific replication impossible (King 2011: 719).

Methodological obstacles connected to archiving have been extensively discussed e.g. in Britain (Mauthner et al. 1998; Richardson and Godfrey 2003; Parry and Mauthner 2004; Bishop 2009), France (Descamps et al. 2005; Marcadé et al. 2014) and Finland (Kuula 2010/2011). Communities of practice like, for instance, those of the Presto4U EU project dealing with 'Research and Scientific Collections' and with 'Music and Sound Archives' gathered from all around Europe in order to identify useful research parameters in the digital audio-visual preservation domain, to raise awareness and improve the adoption of these results by technology and service providers as well as media owners (PRESTO4U 2014). Research networks were born, especially in France (e.g. *réseau Quetelet*), whose mission is the preservation of fieldwork surveys consisting of questionnaires. More recently, several research groups have appeared (among them, *beQuali*), whose aim is to collect, digitise, and spread qualitative interview data. It is important to underline that such initiatives can be very useful from a scientific and educational point of view, regardless of which method and research style have been used. First, they show the variety of methods and devices used by different researchers. Second, they can be used as a didactic tool for students and fieldwork novices in order to better explain different methods for collecting and gathering data; for creating a corpus; and for reporting the research work according to the principle of accountability.

The paper is organised as follows. Section 2 presents the relationships between digital audio archives and Intangible Cultural Heritage. In Sects. 3 and 4 two case studies are described in order to evaluate the procedures used for doing research on oral materials so as to respect the rights of others. Both cases represent different but intertwined examples of accessibility in relation to digital audio archives: the first refers to the activities carried on at the *Phonothèque de la Maison méditerranéenne des sciences de l'homme* of Aix-en-Provence; the second refers to the solutions envisaged by the Italian research project called *Grammo-foni. Le soffitte della voce* (Gra.fo). The first one is an institution also devoted to preservation and

conservation, while the second is the outcome of a research call. The final section presents some closing observations associated to the accessibility of digital audio archives.

2 Audio Archives and Intangible Cultural Heritage

Digital audio archives are not peculiar to a single branch of knowledge. On the contrary, they appear to be a virtual space in which different kinds of expertise convene and deal with unusual, original research questions concerning audio preservation, cataloguing, transcription, analysis, data re-using, and access rights management. Oral historians, linguists, and anthropologists have often underlined the urgent need to protect analogue and born-digital audio archives collected by professional scholars and ordinary people interested in languages, dialects, tradition, popular music, and ethnology. In every respect, audio archives are a precious resource: linguists, anthropologists, ethnographers, oral historians have spent years collecting materials that deserve safeguarding and circulation. However thousands of hours of speech recordings collected for different purposes, despite having been digitally preserved, are still inaccessible to the communities for which they have been produced, not to speak of the wider audience. In most cases, audio archives collected in the humanities and social sciences are still in the hands of the original researchers. It can even be very difficult to get the basic datasets documentation and even more difficult to persuade researchers and private citizens to provide open information about their data. Crucially, the UNESCO *Convention for the Safeguarding of the Intangible Cultural Heritage,* Article 2 defines this material as belonging to Intangible Cultural Heritage domains, which include:

- oral traditions and expressions, including language as a vehicle of intangible cultural heritage;
- performing arts;
- social practices, rituals and festive events;
- knowledge and practices concerning nature and the universe;
- traditional craftsmanship.

It is widely known that conflicting issues arise when dealing with Intangible Cultural Heritage, since the demand for open access conflicts with ownership rights and ethical issues (Lixinski 2013; Tucci 2013; Farah and Tremolada 2014). It is therefore urgent to identify the possibility of reaching a balance between two conflicting demands: the need for Intangible Cultural Heritage openness and accessibility vs. the respect of all rights related to Intangible Cultural Heritage, e.g. copyright, intellectual property, privacy. In this respect, special attention must be devoted to the dissemination of oral heritage via new technologies, which requires a thorough reflection not only from the technological point of view, but also from the legal one. In fact, most of the analogue recordings that constitute oral heritage were collected at a time when little or no attention was

payed to the legal aspects related to Intangible Cultural Heritage. Thus the need for the open circulation of documents can clash with some inviolable rights (copyright, right to privacy, right to individual oblivion) that can be claimed by those whose voices have been recorded or even by those who have been simply mentioned. Not long ago, it was impossible to imagine that the recorded voices could be accessed via the internet. In this respect, archivists have a new responsibility: they are the 'guardians' of the witnesses' personal data and e-reputation. In order to develop a set of best practices for dealing with the legal aspects related to handling, cataloguing, using, and disseminating oral heritage documents, it is necessary to analyse the European panorama, emphasising the differences, but also trying to find points of convergence among the countries under civil law and those under common law system, in order to make accessible this common heritage beyond national boundaries.

3 The Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme

The *Phonothèque de la Maison Méditerranéenne des Sciences de l'Homme* (Mediterranean Research Centre for the Humanities; henceforth MMSH) is an archival research centre created in Aix-en-Provence (France) at the end of the 1960s by Philippe Joutard, a contemporary historian, and Jean-Claude Bouvier, a dialectologist, both researchers at the *Centre de recherches méditerranéennes sur les ethnotextes, l'histoire orale et les parlers régionaux* of Aix-en-Provence (CREHOP). The collections consist of deposits made by scholars working with oral inquiries or by associations dedicated to heritage preservation. Wishing not only to preserve their recordings and to have their field interviews published, they strived to make their sources available to the general public. In connection with MMSH researchers, CREHOP holds field recordings collections in the domains of anthropology, sociology, linguistics, political sciences, history, music and literature, all focused on the Mediterranean area. It illustrates fields poorly covered by conventional sources or complements them with the point of view of real actors and witnesses.

In 1997, CREHOP integrated MMSH creating a research and training campus including 11 research laboratories, all based in the South of France and specialising in Mediterranean culture. In 2015, the collection held more than 7000 h of speech/sound recorded from the late 1970s around four main topics:

- Oral literature, ethnomusicology, techniques and know-how;
- Life experiences, oral history, collective memory;
- Language and cultural identity;
- Epistemology and methodology: workshops, seminars, courses.

The audio collections have been digitised since January 2000 and include 6000 h of recordings, listed on an online catalogue, while the audio archives are

editorialised on a scholarly blog called *Les carnets de la phonothèque*, where it is possible to enjoy the so-called 'veille active' organised by the Phonothèque working group, whose aim is also to disseminate the contents of the recordings via the World Wide Web. The next section presents two different examples of dissemination the first refers to the European project *Europeana Sounds*, while the second deals with the procedures envisaged by the MMSH audio archive in order to facilitate each scholar to disseminate his/her research archives.

3.1 Dissemination in Networks: The Example of *Europeana Sounds*

The MMSH audio archive centre takes part in several projects supporting the dissemination of the materials. In 2010, the catalogue has joined the *Portail du patrimoine oral* (*Oral Heritage Portal*), a collective catalogue of audio and audio-visual archives on oral tradition in France. The portal, launched in 2011, contains audio and video documents such as songs, tales, traditional music, life stories, recorded *in situ*. At the moment, nine different databases are accessible through this portal: the MMSH audio archive, the Office of Auvergne's Territories Music at Riom (Auvergne), the Centre for study, research and documentation of the spoken word (Poitou-Charentes-Vendée), the Museum of instruments at Céret (Catalogne), the Occitan centre of music and traditional dance at Toulouse, the Regional centre for traditional music (Limosin), the Archives for spoken word heritage (Bretagne), the Music and oral traditions centre (Normandie), and the Bourgogne Centre for spoken word heritage. In 2011, the MMSH audio archive catalogue was integrated into the portal *Isidore*, which provides access to digital and digitised research data in humanities and social sciences in French-speaking countries internationally.

In February 2013, the MMSH Sound Archives Centre was involved in the *Europeana* and *Europeana Sounds* project coordinated by the British Library, which brings together 7 national libraries, 5 archive and research centres, 2 other public bodies, 4 non-profit organisations, 3 universities, and 3 companies in 12 - European countries. The *Europeana Sounds* project deserves special attention because of its innovative potential for audio archives: not only does it allow access to one million audio documents, but it is also focused on promoting a creative re-use of the recordings. Scheduled to run from February 2014 to January 2017, *Europeana Sounds* is co-funded by the European Commission and the Europeana Sounds consortium. The activities of the project are organised in seven thematic work packages: aggregation, enrichment and participation, licensing guidelines, channels development, technical infrastructure, dissemination and networking, project management and sustainability. The majority of these activities depend on Workpackage 3 "Rights Labelling Guidelines", headed by the Netherlands non-profit organisation Kennisland. It provides legal guidelines for integrating audio content into *Europeana Sounds* based on the current status of the Europeana Licensing Framework, including the results of the rights survey. This survey addressed the barriers to online access and proposed guidelines in order to disseminate online audio data. It involved all the European countries participating

in the project and produced a comparative assessment on how legal issues are faced within the different European research communities. In 2014, among other deliverables, the *Europeana Sounds* consortium published an open access best practice guide on the following theme: *Rights Labelling Guidelines. Guidelines for Contributing Audio Content Into Europeana.*

This guide presents a complete survey of all the obstacles relating to online access, proposing solutions for use concerning audio content. Its main key points can be summarised as follows. First, it is necessary to detect the different types of Intellectual property rights (IPR) which may affect a certain audio work from three different points of view: the composition, the performances and the recordings. As for the composition, copyright protection has a time limit, usually 70 years after the death of the creator. Once such time limit has expired, the work enters the public domain. In the case of audio material, the so-called 'related rights' appear to be very relevant too: they warrant a different term of protection and are given to performers, producers, recordists and broadcasters. Therefore, although a composition may be in the public domain, the related digital object may not enjoy the same status, exactly because of the related rights. The time limit concerning the related rights is 50 or 70 years after the first publication or the first communication to the public. A rather different case is represented by the database rights, whose time limit is 15 years after creation: they turn out to be very relevant in case an institution receives digital sound archives from another institution. Second, after a very careful analysis of Intellectual property rights, in case one or more of these rights applies to a certain digital work, it is important to obtain permission from all the rights' holders before publishing and reproducing it. Finally, the guide provides a detailed account of the *Europeana* licensing framework, in order to facilitate *Europeana*'s activities and, in particular, data ingestion into *Europeana* space.

Although 'spoken word digital audio' objects are specifically mentioned in the *Guidelines* when describing the possible Intellectual property rights involved, the extraordinary variety of practices in fieldwork in oral history, linguistics, anthropology, and sociology certainly requires a more in-depth analysis, in order to both cover unpublished audio archives and consider the ethical issues involved in their dissemination (Zeytlin 2012). This is why the MMSH Sound Archive Centre is at present engaged in a project that focuses more directly on the dissemination of research data in the social sciences and humanities: a working group under the auspices of the DARIAH (Digital Research Infrastructure for the Arts and Humanities) consortium is elaborating a best practice document entirely dedicated to ethical and legal issues. The French version of the text (whose English provisional title is "Good practice guide for disseminating digital resources in the Humanities and Social Sciences. Legal and ethical issues in digital research") has been written by different stakeholders (interviewers, interviewees,[2] researchers, archivists) and is now available on a scholarly blog named *Questions d'éthique et de droit en SHS*. The working group produced several tools: specimens for the legal

---

[2] 'Interviewee', 'informant', and 'witness' are considered as synonymous in the present paper, although they may refer to different traditions, according to different branches of knowledge.

agreement between interviewees and researchers, between researchers and institutions, and between researchers, interviewees, and Heritage institutions.

3.2   Ethical and Legal Issues: An Example from the MMSH Audio Archive

In partnership with the team of the MMSH Sound Archives Centre, and according to the topics explored, the researchers choose the terms of access to their data at the moment in which they create the deposit. The MMSH Sound Archive Centre offers the scholars a set of tools (e.g. legal agreements, classification and cataloguing templates) to help them disseminate their research in accordance with best practices and ethical and legal guidelines. As a result, around 2000 h of audio recordings are directly accessible online.

From the ethical and scientific viewpoint, field recordings cannot be disseminated as standalone materials, i.e. without any contextual information. Audio documents in archives need to be carefully interpreted in order to be understood, and any relevant note, drawing, or diary produced by the researcher before, during and after the fieldwork constitutes a precious resource for correctly interpreting the documents. In this respect, the contextualisation of field recordings is a thorny issue: each recorded document collected during fieldwork has to be used and re-used together with all the different elements of the scientific research from which it originates (Descamps et al. 2005). It is very important to clarify that these recordings are not 'the truth'. Indeed, they refer to the witnesses' representation of the given situation. For this reason, sound archives centres usually make sure that they embed as much contextual information as possible into the digital materials.

At the MMSH Sound Archives Centre, the sound archives can be harvested in Dublin Core (DC) on Isidore, a platform allowing access to the digital data of Humanities and Social Sciences, in Europeana Data Model (EDM) on *Europeana*, or in Encoded Archival Description (EAD) on Calames, i.e. the online catalogue describing archives and manuscripts held by French universities and research libraries and institutions (*Catalogue en ligne des archives et des manuscrits de l'enseignement supérieur et de la recherche*), administered by the *Agence bibliographique de l'enseignement et de la recherche* (Bibliographic Agency for Higher Education—ABES). Affiliation to the national platform, Calames, in 2013 represented a relevant turning point for the MMSH Sound Archive Centre. Most importantly, through Calames, the MMSH Sound Archive Centre has access to IdRef (Identifiants et Référentiels), the French system for reference identification in research, which is linked to the Virtual International Authority File (VIAF) Project and used by Worldcat, the world's largest network of library content and services, dedicated to providing access to library resources on the Web. The matched use of VIAF and Worldcat ensures two intertwined properties first, informants are no longer hidden in archive databases, and second, their contribution is held in academic databases throughout the world. Furthermore, once the informants' names have been identified, they are given an international identifying number and thus benefit from the standard features ensured by the International Standard

Name Identifier (ISNI): uniqueness, stability, visibility, sustainability, interoperability, and independence (Angjeli et al. 2014). Uniqueness comes from the fact that a name is unique and duplication is not allowed. Stability derives from the fact that ISNI is an ISO standard (ISO 27729:2012). As for visibility, ISNI facilitates the process of Search Engine Optimization in order to identify the informants' names. Sustainability is a consequence of the ISNI commitment towards the long-term preservation of the data. Interoperability of all the identified names derives from the fact that ISNI works together with VIAF, IdRef, Open Researcher and Contributor ID (ORCID). Finally, the National Library of France and the British Library, being the coordinators of the ISNI International Authority (ISNI-IA), the ISO registration authority of ISNI, are the guarantors of ISNI's independence.

The example referred to in the title of the present subsection stems from the *repertoire* of tales and songs from the Cevennes area given by a privileged witness, the late Marcel Volpilière, who provided more than 20 h of recorded interviews with three separate researchers who deposited their archive material in MMSH Sound Archives Centre and probably in other centres. This repertoire has also been published in a series of audio cassettes and reissued on CDs and in a book. Identifying this witness like a 'real' author helps us identifying other archives containing documents linked to Marcel Volpilière (e.g. recordings of interviews with Marcel Volpilière, other related documents). In this respect, inclusion of the informants' name is an effective contribution to the information-gathering process. Indeed, one can find things such as: unpublished interviews relating to the life of Volpilière as a Cevenol farmer; legends of Mont Lozère; the importance of the chestnut tree for Volpilière himself and for the community's identity; and fantasies of Cevenol farmers in their daily life. Via inclusion in IdRef, one can:

- provide this witness with a unique, stable and long-term international identifier, covering all interviews recorded by several interviewers in a single village in the Cévennes between 1970 and 1990;
- establish links between his publications and the recordings of his interviews by confronting the various forms of materials that he provided, thus underlining the importance of interoperability;
- give greater visibility to a Cevenol farmer, through *WorldCat Identities*.[3]

Before ethical and legal issues were at the centre of the scientific debate, the voices of the witnesses recorded in fieldwork were not integrated in academic databases. Finding a way to face ethical and legal issues might ensure that the sound archives' voices be listened online as a part of our common cultural heritage. Clarifying these issues is crucial in order to reach the main objectives of dissemination, crowdsourcing, creative re-use, discovery and referencing of sound data, as also the *Gra.fo* project proves.

---

[3] https://www.worldcat.org/identities/lccn-n87107956. Accessed November 10, 2015.

4   The Project *Grammo-foni. Le soffitte della voce* (*Gra.fo*)

The project *Grammo-foni. Le soffitte della voce* (*Gra.fo*), jointly carried out by Scuola Normale Superiore of Pisa and the University of Siena, and financially supported by Regione Toscana (PAR FAS 2007–2013), detected and preserved a large number of audio (speech and music) recordings collected on the Tuscan territory, making them publicly available via a dedicated online archive (GRAFO 2011–2014).

The project included five stages:

- fostering the awareness level on the importance of preserving this valuable (but largely invisible) product of cultural heritage;
- contacting the audio recordings' owners to legally agree for the temporary borrowing of the materials;
- collecting, digitising, and (when necessary) restoring the audio materials;
- systematically cataloguing and partially transcribing the speech documents;
- offering the opportunity for online accessibility of digitised content for a large audience.

This large and still growing repository provided the opportunity to discover audio texts which, until now, have been known to a very limited number of possible users, thus ensuring the safeguarding of a specific type of endangered intangible cultural heritage. Besides, the *Gra.fo* archive offers a vast quantity of (mostly unpublished) documents for further linguistic, economic, social, political, historical, and cultural analysis. Until now, the project digitized more than 2800 h of Tuscan speech, and a large part of the data are accessible for download though the web portal, as explained below.

4.1   The Preliminary Stages: Census and Collection

Besides its wealth in paper documents (Petrucci 1994), Tuscany also is a privileged area for collecting and working with oral documents, as it abounds with both public and private audio archives, collected by scholars as well as amateurs. In the effort to produce a census of the Tuscan audio archives, the already existing censuses (Andreini and Clemente 2007) have been used and integrated with information about oral archives collected for linguistic and dialectological research purposes, such as *Carta dei Dialetti Italiani* and *Atlante Lessicale Toscano*. Subsequently, a priority list was created according to three main criteria:

- relevance and antiquity of the materials (older materials might witness disappeared or disappearing language varieties);
- state of preservation of the materials (priority should be given to those materials which look more damaged and whose content, therefore, is more likely to be lost in the near future);
- geographic representativeness (so that every area of Tuscany can be represented in the archive).

Following the above-reported priority list, the audio archives' owners were directly contacted to illustrate the aims and organization of the project. The *Gra.fo* staff then worked with the interested archives to facilitate the project, in collecting the material, and signing legal agreement for the temporary borrowing and future dissemination of the materials. In addition, the owners of the archives with no proper bibliography or accompanying materials were interviewed in order to explain the motivation and aims of the research that inspired the creation of their own archives. Indeed, unlike other kinds of materials, the motivation behind audio documents is often only known to the researcher(s) who collected them. Such interviews (called 'Tell something about your archive') are crucial, as they provide the key for correctly interpreting and cataloguing the archive and thus offer the user an appropriate guide. In some cases, the owners actively helped in the description of their own archives, and the cataloguing could be directly taken care of by someone who had been active in the actual collection of the recordings.

Both the digitization process and the cataloguing stages fall outside the topic of the present paper: the reader is referred to Bressan and Canazza (2013), Calamai et al. (2013), Calamai and Bertinetto (2014), and Calamai et al. (2014)—where some drawbacks associated with the conversion of analogically recorded speech and music to digits are also addressed. Indeed, as claimed in the aforementioned paper from a documental/ecdotic point of view, the act of 'disembodying' the original information inevitably involves a new reconstruction of the content.

4.2   From the Database to the Website

The *Gra.fo* database uses the MySQL system and consists of 59 interconnected tables, some of which have specific constraints. The tables contain information on the fields created for cataloguing and for the creation of the preservation copies, stored in a specific server archive with Raid 5 configuration. The collaborators devoted to digitizing and cataloguing interact with the database through specific applications, respectively called *Audiografo PP* and *Audiografo CP*, with user-friendly interfaces consisting of drop-down menus, checkboxes and open fields.

The web portal is a technological interface which, by querying the database and the server archive containing the preservation copies, allows the end user to search all documents collected in *Gra.fo* (cataloguing records, .mp3 files, transcriptions and the pdf files of the accompanying materials). The website contains the description of the project, as well as the archives and the cataloguing records. The page devoted to the archives lists their names and descriptions, the subsections names, and the 'Tell something about your archive' interview. As for the search, two distinct types are supported:

- by linguistic area (an interactive map allows the users to click on the area of interest and access the corresponding records);
- by content (i.e.: topic, *genre* and type of document, date and place of the recording, language variety).

The cataloguing record of each document provides the following information:

- name and description of the archive (and subsections) to which the document belongs
- conditions of access (i.e. whether the document undergoes access restrictions for privacy reasons—see *infra*)
- title (and alternative title, if present)
- content
- keywords
- researcher's name
- informant(s) name, sex, date and place of birth, education level and profession
- date, place and setting of recording
- typology
- topic
- *genre*
- language variety
- aim of the recording
- bibliography
- type of carrier
- recording (downloadable in .mp3 format)
- accompanying audio-related material (downloadable in .pdf format)
- transcriptions (downloadable in .pdf format).

In addition, all documents concerning the conventions adopted within *Gra.fo* with respect to digitization, restoring, cataloguing and transcription protocols are available on-line. The website and the cataloguing records are openly accessible but, in order to prevent improper use, user authentication is required for the downloading of .mp3 files, transcriptions and accompanying materials.

4.3 Ethical and Legal Issues

One of the major problems faced by the *Gra.fo* project was the treatment of confidential information. Many archives were recorded before approval of the national law on privacy rights (Personal Data Protection Code-2003),[4] so that the informants were not asked to give their authorization for future dissemination. As a consequence, *Gra.fo* only provides the initials (rather than the full names) of the informants and of the people mentioned in the recordings. Their full names together with other personal information are shielded in the *Gra.fo* repository. Considering

---

[4] Personal Data Protection Code-2003, English version available at http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/2427932. Accessed November 10, 2015.

the extremely different types of oral material collected inside the project, three different types of access are made possible, depending on the presence of confidential data in the documents:

- *Full access via web portal*—with documents that do not contain any confidential information, one can be read the summaries and download the full audio documents, the accompanying audio-related materials and the transcriptions (if available).
- *Partial access via web portal*—documents containing some confidential data (less than 90 % of the total recording time) are edited in two different versions: a full version, only available for consultation in the *Gra.fo* physical location, and a partial version, with edited summary and partially obscured mp3 file, available on the web portal.
- *Access in the Gra.fo physical location*—documents mostly consisting of confidential data (over 90 % of the total recording time) are accessible on the web portal only through an edited summary, while the mp3 file is only available for direct consultation in the *Gra.fo* Laboratory.

If the accompanying audio-related material contains confidential data, it is only available for consultation in the *Gra.fo* laboratory, while the transcriptions (if available), are accessible on the web portal after removal of the confidential data.

The right to individual oblivion, something Laouris calls "the right to digital euthanasia" (2015: 124), is another important issue in the digital era and it is not only a mere technical problem, as Laouris above claims. The *Gra.fo* project takes this issue into consideration in two different clauses of the portal policies.[5] First, it is possible to ask the portal administrator to remove particular data from the web. However, it is undeniable that legal problems cannot be solved by merely technical solutions (Hildebrandt 2015: 179). This is especially true in the domain of oral history and intangible cultural heritage. Let us take a more detailed look at the Italian case. In the 2001 *Code of Conduct and Professional Practice Regarding the Processing of Personal Data for Historical Purposes* no more than five lines are devoted to 'oral sources', namely:

---

[5] See Art. 10—Segnalazioni and Art. 11—"Norme riguardanti la riservatezza" at the following url: http://grafo.sns.it/web/guest/policy: "La pubblicazione dei contenuti del portale è effettuata secondo il principio della buona fede e secondo regole di massima correttezza, diligenza e perizia. Chiunque, nonostante l'applicazione di questi principi da parte di *Gra.fo*, ravvisi la violazione di un diritto di cui sia titolare (es: diritto di autore, diritto all'immagine, diritto alla riservatezza), potrà segnalarlo all'indirizzo grafo@sns.it. Qualora sia accertato che la segnalazione ha un valido fondamento giuridico, i Proprietari del Portale si impegnano a rimuovere tempestivamente il contenuto dal portale, dandone comunicazione al reclamante nel più breve tempo possibile".

interviewees: Oral Sources

1. With regard to processing of oral sources [of information], it will be necessary for the interviewees to give their express consent, whether orally or not, even based on summarized information including at least the interviewer's identity and activity and the purpose of the data collection.
2. If an Archive acquires oral sources, the interviewer will be requested to produce a written statement to the effect that the purposes of the interview have been notified and the relevant consent has been obtained from the interviewees.[6]

Secondly, given this rather thorny framework, the key word of the *Gra.fo* staff has been 'transparence', specifically by making the project' aims explicit—either by a face-to-face communication or by letter—to all persons involved: interviewees, interviewers, archive owners, but also archive curators, and—if possible—the descendants of interviewees and interviewers.

5   Conclusion

The theme of accessibility of digital audio archives, as discussed so far, is quite problematic. It involves both developing a code of conduct with respect to professional ethics, and facing legal issues. Several researchers internationally involved in the domain of audio archives feel the need to better spell out their responsibilities with respect to the Digital Era. The importance of all accompanying materials and contextual information associated to each archive has been emphasized above, with respect to both French and Italian case studies (Sects. 3 and 4). This is the pre-condition for a proper re-use of research data. However, such theme goes beyond the scope of academic groups, universities and research centres. Finding the guidelines for accessibility of audio archives is a cultural operation. There are several reasons for it. First, this involves building, promoting and reinforcing an 'open' culture. In many cases, digital audio archives lodge intangible cultural heritage content. It is thus important to foster everyone's awareness that intangible cultural heritage refers to crucial themes for the European digital agenda, including: authorship, copyright, copyleft, and creative commons. In this respect, folklore data—one of the most prototypical examples of intangible cultural heritage—can be considered as a kind of 'open source' product (Bertolotti 2011: 68). In the words of Roman Jakobson and Pëtr Bogatyrëv, writing in 1929:

> An item of folklore begins its existence only after it has been adopted and sanctioned by the community. As in the development of *langue*, the environment prunes a created work to fit its taste; if the community rejects it, it simply dies out. A community retains only those items of folklore which have a functional value for it. Like *langue*, the work of folklore is

---

[6] The English version of the Code can be accessed at the following url: http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/1565819. Accessed November 10, 2015.

extrapersonal and leads only to a potential existence; it is only a complex of certain norms and impulses, the canvas of the actual tradition, which the tellers revive with the embellishment of their individual creation.[7]

Second, defining the guidelines for accessing audio archives allows the researchers to create the condition for returning their contents to the communities and the individuals that produced them. This reinforces the mutual relationship between interviewee and interviewer that comes about during fieldwork with oral sources, whatever the actual domain in which the given oral sources are collected and investigated. A restitution act has the additional advantage of promoting the engagement and the participation of small communities and private citizens. Scientific communities are increasingly concerned with community engagement and empowerment, in order to enhance good behavioural norms inside the communities themselves. As claimed in Art. 15 of the UNESCO *Convention for the Safeguarding of the Intangible Cultural Heritage*, heritage communities, groups and, where appropriate, individuals are asked to create, maintain and transmit such heritage, and to be actively involved in its management.

Finally, the issue of audio archives accessibility has encouraged the rethinking of personal data protection. This should be viewed as a renewed impulse to re-define the privacy value, considering the need to rethink what people consider really worth of protection (Dewandre 2015: 203). As Oates (2015: 225) claims:

> The online agora is a precious public resource. Currently, it is being colonized by corporations and states, in ways that asymmetrically reassign the power of information and personal data to the elites. What is needed is an understanding that a public agora should be conceptualized and protected in a way that tips the balance away from the elites and toward the citizens.

In the realm of the digital sound archive, ethical and legal issues are no longer themes for bureaucrats. Asking all stakeholders involved in the process of building digital audio archives (from individual researchers, to archives' owners, from interviewees and interviewers and their descendants to public and private institutions) what can be freely accessed on the web (and with what kind of constraints) amounts to setting the ethical issues at the foreground of research. In order to obtain useful answers, and positive attitudes towards web diffusion, it is necessary to clearly and honestly explain the reasons behind accessibility. In other words, communication of research results becomes an essential task for scholars. In this respect, the most crucial topic with respect to accessibility concerns the legal issues related to the digital archives that were produced when the web did not yet exist and legal agreements during fieldwork were the exception rather than the rule. A large amount of such data could run the risk of remaining forever inaccessible on the web, unless adequate and careful balance is found between open access on the one side, and ownership rights and ethical issues on the other side. Digital audio

---

[7] Jakobson and Bogatyrëv (1929), English translation by J.M. O'Hara, at https://scholarworks.iu.edu/dspace/bitstream/handle/2022/1711/13%281%291-21.pdf?sequence=1. Accessed November 10, 2015.

archives can thus offer a valuable contribution in establishing rights, duties and ways to access important pieces of the European Intangible Cultural Heritage.

References

Andreini, A., & Clemente, P. (2007). *I custodi delle voci. Archivi orali in Toscana: Primo censimento*. Firenze, Italy: Regione Toscana.

Angjeli, A., Ewan, A. M., & Boulet, V. (2014, August 16–22). *ISNI and VIAF—Transforming ways of trustfully consolidating identities*. Paper presented at IFLA WLIC 2014 Lyon, France. Retrieved November 10, 2015, from http://library.ifla.org/985

Banque d'enquêtes qualitatives en sciences sociales (*beQuali*). Retrieved November 10, 2015, from http://www.bequali.fr

Bertolotti, G. (2011). Materiali immateriali. In G. Bertolotti & R. Meazza (Eds.), *Beni immateriali. La convenzione Unesco e il folklore*, *La ricerca folklorica: Vol. 64* (pp. 63–70).

Bibliographic Agency for Higher Education (ABES). Retrieved November 10, 2015, from http://www.abes.fr

Bishop, L. (2009). Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues, 44*, 255–272.

Bressan, F., & Canazza, S. (2013). A systemic approach to the preservation of audio documents: Methodology and software tools. *Journal of Electrical and Computer Engineering*. Article ID 489515, 21 pages, doi: 10.1155/2013/489515. Retrieved November 10, 2015, from http://www.hindawi.com/journals/jece/2013/489515/

Calamai, S., & Bertinetto, P. M. (2014). *Le soffitte della voce. Il progetto Grammo-foni*. Manziana, Italy: Vecchiarelli.

Calamai, S., Biliotti, F., & Bertinetto, P. M. (2014). Fuzzy archives. What kind of an object is the documental unit of oral archives? In M. Ioannides et al. (Eds.), *Digital heritage. Progress in cultural heritage: Documentation, preservation, and protection EuroMed 2014, LNCS 8740* (pp. 777–785).

Calamai, S., Biliotti, F., Bertinetto, P. M., Bertini, C., Ricci, I., & Scuotri, G. (2013). The Gra.fo sound archive: Architecture, methods and purpose. In A. C. Addison, G. Guidi, L. De Luca, S. Pescarin (Eds.), *Proceedings of the 2013 Digital Heritage International Congress (DigitalHeritage)* (p. 439). Institute of Electrical and Electronics Engineers.

Calames. Retrieved November 10, 2015, from http://www.calames.abes.fr

Descamps, F., Monnier, F., & Schnapper, D. (2005). *L'historien, l'archiviste et le magnétophone: De la constitution de la source orale à son exploitation*. Paris: Ministère de l'économie des finances et de l'industrie, Comité pour l'histoire économique et financière de la France. Retrieved November 10, 2015, from http://books.openedition.org/igpde/104

Dewandre, N. (2015). Rethinking the human condition in a hyperconnected era. Why freedom is not about Sovereignty but about beginnings. In L. Floridi (Ed.), *The onlife manifesto. Being human in a hyperconnected era* (pp. 195–215). Springer.

Europeana. Retrieved November 10, 2015, from http://www.europeana.eu

Europeana Sounds. Retrieved November 10, 2015, from http://www.europeanasounds.eu/

Farah, P. D., & Tremolada, R. (2014). Desirability of commodification of intangible cultural heritage: The unsatisfying role of intellectual property rights. *Transnational Dispute Management, Special Issue, 11*(2). Retrieved November 10, 2015, from http://ssrn.com/abstract=2472339

Ganascia, J.-G. (2015). View and examples on hyper-connectivity. In L. Floridi (Ed.), *The onlife manifesto. Being human in a hyperconnected era* (pp. 65–85). Springer.

Gra.fo project. Retrieved November 10, 2015, from http://grafo.sns.it/

Hildebrandt, M. (2015). The public(s) onlife. A call for legal protection by design. In L. Floridi (Ed.), *The onlife manifesto. Being human in a hyperconnected era* (pp. 181–193). Springer.

International Standard Name Identifier. Retrieved November 10, 2015, from http://www.isni.org

Isidore. Retrieved November 10, 2015, from http://www.rechercheisidore.fr

Jakobson, R., & Bogatyrëv, P. (1929). Die Folklore als eine besondere Form des Schaffens. In *Verzaneling van Opstellen door Oud-Leertingen en Befriende Vakgenooten (Donum Natalicium Schrijnen)* (pp. 900–913). Nijmegen-Utrecht. (Reprinted from *Selected writings: Vol. 4*, pp. 1–15, by R. Jakobson, 1966, The Hague, Netherlands: Mouton)

King, G. (2011). Ensuring the data-rich future of the social sciences. *Sciences, 331*, 719–721.

Kuula, A. (2011). Methodological and ethical dilemmas of archiving qualitative data. *IASSIST Quarterly*, 12–17. (Original work published 2010)

Laouris, Y. (2015). Reengineering and reinventing both democracy and the concept of life in the digital era. In L. Floridi (Ed.), *The onlife manifesto. Being human in a hyperconnected era* (pp. 125–142). Springer.

Les carnets de la phonothèque. Retrieved November 10, 2015, from http://phonotheque.hypotheses.org

Lixinski, L. (2013). *Intangible cultural heritage in international law*. Oxford, England: Oxford University Press.

Marcadé, C., Guinard, B., Coulais, S., Davy, Y., Desgrugillers, E., Gasnault, F., et al. (2014). *Patrimoine culturel immatériel. Traitement documentaire des archives sonores inédites. Guide des bonnes pratiques*. Retrieved November 10, 2015, from https://halshs.archives-ouvertes.fr/halshs-01065125/document

Mauthner, N. S., Parry, O., & Backett-Milburn, K. (1998). The data are out there, or are their Implications for archiving and revisiting qualitative data. *Sociology, 32*, 733–745.

Oates, S. (2015). Towards an online bill of rights. In L. Floridi (Ed.), *The onlife manifesto. Being human in a hyperconnected era* (pp. 229–243). Springer.

Oral Heritage Portal. Retrieved November 10, 2015, from http://www.portaildupatrimoineoral.org

Parry, O., & Mauthner, N. (2004). Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *Sociology, 38*, 139–152.

Petrucci, L. (1994). Il problema delle Origini e i più antichi testi italiani. In L. Serianni & P. Trifone (Eds.), *Storia della Lingua italiana* (Vol. 3, pp. 5–73). Torino, Italy: Einaudi.

Phonothèque. Retrieved November 10, 2015, from http://phonotheque.mmsh.huma-num.fr

PRESTO4U project. (2014). Retrieved November 10, 2015, from https://www.prestocentre.org/4u

Questions d'éthique et de droit en SHS. Retrieved November 10, 2015, from https://ethiquedroit.hypotheses.org

Réseau Quetelet. Retrieved November 10, 2015, from http://www.reseau-quetelet.cnrs.fr

Richardson, J. C., & Godfrey, B. S. (2003). Towards ethical practice in the use of archived transcribed interviews'. *International Journal of Social Research Methodology, 6*(4), 347–355.

Rights Labelling Guidelines. (2014). *Guidelines for contributing audio into Europeana*. Project Number: 620591. European Union. 44 p. Retrieved November 10, 2015, from http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/EuropeanaSounds-D1.2%20D3.1-Rights-labelling-guidelines_Guidelines-for-contributing-audio-content-into-Europeana-v1.1.pdf

Tucci, R. (2013). Beni culturali immateriali, patrimonio immateriale: Qualche riflessione fra dicotomie, prassi, valorizzazione e sviluppo. *Voci, 10*, 183–190.

Virtual International Authority File Project (VIAF). Retrieved November 10, 2015, from https://viaf.org

Worldcat. Retrieved November 10, 2015, from http://www.worldcat.org

Zeytlin, D. (2012). Anthropology in and of the archives: Possible futures and contingent pasts. Archives as anthropological surrogates. *Annual Review of Anthropology, 41*, 461–480. (Volume publication date October 2012) First published online as a Review in Advance on July 2, 2012. doi: 10.1146/annurev-anthro-092611-145721