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

Silvia Calamai, Aleksei Kelli, Chiara Kolletzek
Rationale of the talk

- How to define a protocol for the curation and dissemination of speech archives, which appear to have — *de jure* — the highest restrictions on their curation and dissemination

- How to find a balance between the rights of the recorded people (and their heirs) and the right of information and the preservation of memory
The speech archive of Anna Maria Bruzzone

(1925-2015)

A.M. Bruzzone, Ci chiamavano matti. Storie da un ospedale psichiatrico, Torino, Einaudi, 1979
What we discovered

- 36 tapes accompanied by the handwritten and the typewritten transcriptions of all the interviews.
- In addition to the complete transcriptions, there are different versions that show all the work of editing made by A.M. Bruzzone so that the interviews could be suited for publishing.

What is now possible to investigate

- understand, and examine the changes undergone by 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.
- 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 name and not the pseudonym has been found in the box of every single tape.
The handwritten transcriptions
Il babbo e la mamma erano poi già morti?
No, no no, poi morì il babbo e dopo morì il babbo e mi non trovò, insomma, sentita male dal servizio, mi non sentita male. Un anno sono stata a casa, mi fecero... n° m'ho portata in ospedale, perché avevo una malattia abbastanza forte, mi sentivo molto, molto male, perché io sentivo proprio le voci, insomma, sentivo per non che parlavano e dicevano ai miei genitori, perché qui, sentite come cantano? Ma non è vero, non c'è messa... NON sembrava proprio che venissero. 

Si, si, lo so.

Nei canti, in continuazione queste voci.
Le voci in continuazione, avevo queste voci che ci ammassavano, e il giorno dopo, insomma, non c'era più, non c'era più, e mi sono trovata in ospedale.

Quale ad Arenzo?

Ad Arenzo, sì, m'ho portata in vigilia e le pastiche non le ho dato perché mi faceva effetto, insomma, pigliare le pastiche in quell'estasi e si vedeva quel colore, blu, rosso, insomma, quel colore non li potevo buttare giù e mi facevano le pance perché non prendevano le pastiche, non per quest'altro...

Sempreché dopo mi hanno fatto l'insulina perché...
The curation and dissemination of vulnerable people archives is subject to the personal data regulation, since they contain personal data.
Personal data

- personal data = “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” (GDPR)
Special categories of personal data

- Speech archives
  - Human voice is considered biometric data (González-Rodríguez et. al. 2008; Jain et. al. 2004).

- Speech archives with vulnerable subjects
  - health data

Biometric and health data both belong to the special categories of data (sensitive or delicate data).
The challenge

- How to strike a fair balance between research, dissemination, and protection of privacy

Two socially relevant interests:

- the protection of personal data
- the transmission of knowledge and freedom of research.
The GDPR does not apply to the personal data of deceased persons (Recital 27).

The GDPR also does not apply to anonymous data (= the natural person is not identifiable (Recital 26)).

Right-holders (e.g. relatives)

Anonymisation may not correspond to the needs and scope of historical research (historians are interested in “names and faces”).

Any solution?

But…
The ‘legal chain’ for the re-use of the Archive

- **identification of the real names** of the patients and their matching with the pseudonyms, as attested in the volume (Bruzzone 1979).
- **go back to the interviewees**, also engaging the network of all the people – physicians, nurses, social workers, ordinary citizens – involved in the recent history of the Psychiatric hospital.
Dissemination via TV news

Acquiring informed explicit consent

- The ‘legal chain’ could help to assess the possibility and feasibility to obtain detailed and clear informed consents, describing
  - the aims,
  - the scope
  - the positive spill-over effects of the dissemination of such an oral archive.
- If the consent form will be obtained by the right-holders, the oral archive could be finally enjoyed by the research communities and the entire civil society.
The relationship with CLARIN Infrastructure

- Anna Maria Bruzzone Archive will be part of the CLARIN Infrastructure and metadata description will be pursued according to COALA (Schiell, Jackl 2014).

- 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).
SELECTED REFERENCES


- Manon Oostveen. Identifiability and the applicability of data protection to big data. International Data Privacy Law, 2016, Vol. 6, No. 4, 299–309;