



HAL
open science

inTEIrvIEWS: an ODD for qualitative interviews in history and social sciences

Marie Puren, Florian Cafiero

► **To cite this version:**

Marie Puren, Florian Cafiero. inTEIrvIEWS: an ODD for qualitative interviews in history and social sciences. TEI Conference and Member's Meeting 2021, Oct 2021, Virtual Conference, United States. hal-03402606

HAL Id: hal-03402606

<https://hal.science/hal-03402606v1>

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

INTEIRVIEWS

An ODD for qualitative interviews in history and social sciences

Marie Puren ^{1,2} and Florian Cafiero ^{3,4}

October 25th, 2021 - Virtual TEI Conference "Next-Gen TEI"

1- EPITECH Paris

2 - Centre Jean Mabillon - Ecole Nationale des chartes /PSL

3- Groupe d'Etude des Méthodes de l'Analyse Sociologique de la Sorbonne - CNRS / Sorbonne Université

4 - Medialab - Sciences Po Paris



Sociology

Anthropology

**Qualitative
interviews**

Information science

Contemporary history

Public Health

Political science

WHY ENCODING QUALITATIVE INTERVIEWS?

MAKING THE MOST OF A PRECIOUS RESOURCE

- Take a lot of time, and sometimes money, to realize and transcribe.
- "One-time-shot" : most people interviewed will not accept to take time to answer researchers n times.
- Most of the time, much more information collected than what is used in the first place - by lack of time for systematic analysis; because some of the answers collected do not answer the initial scientific question etc.
- The transcripts, if encoded properly, could be harvested more easily by the researcher, and help the realization of more quantitative analyses.
- They also could be used again by other researchers.

ADDRESSING THE "REPRODUCIBILITY CRISIS"

- Human and social sciences among the disciplines targeted by accusations of "lack of reproducibility".
- More broadly, lack of confidence in qualitative research by some actors.
- Could help peer-reviewers to assess the quality of the interview analysis : most of the time, authors declare they transcribed all the interviews, but reviewers do not have access to them, and would not have the time to browse them to look for relevant information.

WHY TEI ?

A SUSTAINABLE AND EQUITABLE SOLUTION

- Old standard with more and more users.
- Can be used by anyone, with free programming tools
- This last point is crucial, as existing solutions are expensive, which increase inequalities among students, researchers, universities etc.

A STANDARD ADAPTABLE BY EACH RESEARCH COMMUNITY

- Can be tailored to one's need : TEI enables to combine elements and attributes as needed, and to add new elements and attributes if needed.
- Within a given community / discipline, it is possible to agree on an available ODD customization that will ensure the interoperability, shareability and reusability of the TEI files.
- Already adopted by many scholars in the humanities - thus easy to find help, courses, summer schools etc. on the topic for researchers in the field.

OUR ODD

ODD available on Github : https://github.com/mpuren/InTEIreviews/blob/main/tei_interview.xml

- We first thought of asking for a new element, but ended up finding a way to comply to the existing standards.
- Combination of the elements and attributes declared by the modules “Transcription of Speech” and “Language Corpora”

- General metadata : time, place, actors of the interview etc.
- Indication of who is the speaker (<u> with attribute "who")
- Comments on the speakers' behavior and tone, and on events in the environment of the interview.

- Adding an **attribute "purpose"** to the list of already existing attributes born by <u>

```
<u who="interviewer" purpose="launching_topic">  
  what do you think of this video ?</u>
```

- Sharing one's interpretations about relevant passages of the transcription : **<seg> element bearing an "xml:id" to delimit these passages + <standOff> container.**
- Indication of passages deleted to respect ethics code or anonymity when relevant with the **element ** containing **elements <gap> and <desc>** to share as much of the information possible.

```
<del>James<gap reason="anonymised" quantity="1"
  unit="words"/><desc>Personal data</desc><supplied reason="pseudonymised"
  >J</supplied></del>
```