When social science meets digital humanities: the case of sharing & documenting oral history data

Louise Corti
Director, Collections Development and Producer Relations
UK Data Service
University of Essex

UK Data Service

Oral history – approaches to 'data'

Note distinct disciplinary differences for:

- Collecting and documenting oral history
- Analysis of 'data': oral and transcribed sources

Scholarly and pedagogical approaches

Does it matter?

UK Data Service

Disciplinary differences

Approach	Social science	Humanities/history
Recording process	Technical advice Agreed interviewing approach	Technical advice Agreed interviewing approach
Transcription	Varies widely	Orthographic and more
Analysis	Linear coding in CAQDAS for text, less attention to audio	Oral sources; linguistic annotation
Documentation	• Interview level metadata • Speaker tags • Some nonverbal mark up	• Interview level metadata • Speaker tags • Nonverbal mark up • Orthographic transcription • Time codes • Alignment or oral and text
Archiving	Audio often not shared	Sound archives
Publishing	QualiBank; DANS	Spoken language archives

Differences that (seem) to have impact on data management and sharing

- Purpose of data collection and analytic intentions
- Transcription methods
- Metadata standards
- Legal and ethical approaches

While different approaches, some core needs are shared

UK Data Service

Oral history transcription

Why?

- Re-live a certain research event

- Draw interpretation from the data

What?

- First level of manipulation of reality - Interview setting
- Second level of manipulation of reality - Transcription

"Transcription as a highly interpretative process and also an "act of power" (Bucholtz 2000: 1463)

UK Data Service

Fidelity vs legibility

Maximising fidelity to oral conversation

- conversation analysts, phoneticians, sociolinguists, dialectologists
- But, the fidelity of phoneticians is not the same as the fidelity of conversation analysts!

Maximising legibility of sources

- oral historians, historians, sociologists

What might we miss?

- Suprasegmental features (voice tone and quality, speech rate)
- Non-verbal communication (posture, gestures, movement and body position, kinesics, haptics, proxemics)

UK Data Service

Short Summary

This is a short passage from an interview conducted by Stef Scagliola on 15 April 2007 at the Imperial War Museum with major general Julian Thompson on the use of oral history in military history. As curator of the Dutch Veterans Interviewproject initatied by the Dutch Veterans Institute in the Netherlands, she was visiting the Imperial War Museum to do research on the requirements and best practices for a large scale oral history collections related to military topics. Julian Thompson agreed to be interviewed about the benefits and risks of using oral sources for research and on how he used the collections of the IWM for his publications. In the interview he stresses four major points. Firstly, that almost all history starts with oral accounts of events and that the trustworthiness of written accounts is often overestimated, secondly that the author of a book needs to know the subject well in order to distinguish facts from fiction, thirdly that despite these merits, oral history is not regarded as a serious discipline in the academic realm. An other important value, is it social character.

Transcript on the basis of light editing

J.T. - The problem which I recognize, and people here will be the first to admit it, with oral history, there are a number of problems, is that if you interview people who are old, their memories sometimes are bad. Quite often they have had a chance to discuss with other people what went on, so they have come to an agreed story.

Algemene gegevens interview

Titel	Erfgoed van de Oorlog, Bystander Memories, interview 01
Geïnterviewde	[echte naam]
Interviewer	[echte naam]
Plaats	
Tijd	
Trefwoorden	
Samenvatting	
Transcriptie gemaakt door	
Eigenaar van het bestand	

00:00:00

[image: 00:00:00]

IV: En kropen die mannen dan gewoon in de schuur, verstopten ze zich..

GI: Ja, bovenaan was een scheve kap, en daar kropen ze met een ladder naartoe, moesten wij een ladder daar neer zetten en dan kropen ze er naartoe, en dan werd dat dicht gepakt, en dan de ladder weer weg natuurlijk dan een paar uur zaten ze daar stil, der was er zelfs een uit Eindhoven, die was hier gekomen van Eindhoven naar zijn familie, om ja, hier veiliger te zitten, maar die moest hier ook onderduiken, die moest hier ook wegkruipen voor die razzia die er gehouden werd.

IV: Nelly, jullie hadden ook een appelboomgaard. Wat gebeurde daar mee?

GI: Ja, wij hadden een heel grote aard, appelboomgaard daar en peren natuurlijk. En die hadden wij, die moesten geplukt worden natuurlijk, en die hadden wij verkocht aan de zusters Ursuline van Venray, maar aangezien de mannen niet durfden te komen plukken vanwege die razzia allemaal, hebben de nonnen zelf, de zusters zelf, kwamen de appels plukken,

I: *eh, eh,* ik wilde u eens vragen: wat vindt u van deze werkplek?

R: *ja,* wat vind ik van deze werkplek... dat is een moeilijke vraag, *da da da daar* daar moet ik even over nadenken. Ik vind het een hele aardige omgeving, ja, een hele aardige omgeving. Leuke collega's, inhoudelijk interessant werk, enzo. Het gebouw is wat minder, maar ja, je kunt niet alles hebben in het leven, hé?

I: kunt u daar wat dieper op in gaan, wat bedoelt u met leuke collega's?

R: wat ik met leuke collega's bedoel?

la ˈbɔːrea e il faˈvɔːnio ‖

una ˈvɔl̪ta | al̪a ˈbɔːrea ˈvenːe ˈvɔʎːa di ˈpren̪der maˈriːθo ‖ an̪dɔ d̪al faˈvɔːnio e ʎːi ˈdisːe ‖ ˈvwɔ(i̯) ˈɛsːer(e) [il] ˌmi(o) ˈspɔːzo ‖ il faˈvɔːnio ˈɛːr(a) un̪ ˈtiːɸo atːaˈkːaːθo a(i̯) hwaˈtːriːni | e l̪ːe ˈdɔnːe non ʎ an̪daˈvan(o) a ˈdʒeːnio ‖ le ˈdisːe ‖ ˈnɔ | perˈke nːon ˈai̯ neˈaŋk(e) un̪ ˈsɔldo di ˈdɔːθe ‖ la ˈbɔːrea | ˈpun̪ta sul ˈviːvo | si ˈmiːz(e) a sːoˈfːjaːre hon̪ ˈtutːe le ˌsu(e) ˈfɔrtse ‖ soˈfːjɔ pːer ˈtre ˈdʒorni | e nːeviˈhɔ ˈfːitːo ˈfitːo ‖ ˈkwan̪d(o) ˈɛbːe fiˈniːθo di ˈsten̪der(e) il su(o) arˈdʒen̪to (i)n̪ˈtorno | ˈdisːe ‖ ˈɛkːoθi la ˈmi(a) ˈdɔːte | tu kːe dɛːˈʃeːvi he nonʲ t͡ʃe l ˈɔ ‖ e an̪dɔ a rːiɸoˈsarsi del̪ːa faˈθiːha ‖ il faˈvɔːnio skroˈlːɔ l̪ːe ˈspal̪ːe | e sːi ˈmiːz(e) a sːoˈfːjaːre ˈluːi ‖ la hamˈpaɲːa e (i) ˈmon̪ti resˈtaːrono ˈsotː(o) um ˈfjaːθo ˈhaldo he ˈʃːɔl̪se fin l ˈul̪timo ˈfjɔkːo di ˈneːve ‖ la ˈbɔːrea | riɸoˈsaːθasi per ˈbeːne | ˈviːde he ˈdːel̪ːa ˈdɔːθe non resˈtaːva ˈɸju ˈnːul̪ːa ‖ ˌdov ˈɛ an̪daːθa ˈθutːa la ˌθuˈ(a) ˈdɔːθe | la han̪͡tsoˈnɔ il faˈvɔːnio ‖ in̪ˈsomːa | mi ˈvwɔ(i̯) an̪ˈkoːra per maˈriːθo ‖ la ˈbɔːrea ʎ(ː)i risˈpoːse ‖ ˈnɔ | ˈnom voˈrːɛ(i̯) ˈma(i̯) ˈɛsːere [la] ˌθuˈ(a) ˈspɔːza | perˈke (i)n un ˈdʒorno ˈsɛ(i̯) haˈɸaːʃe di man̪ˈdarmi m ˈfuːmo ˈθutːa la ˈdɔːθe ‖

Data capture - practice consistency

Example: Transcription of text from audio
- Use a uniform layout throughout – use a template
- Provide guidance of how you would like the data transcribed
  - Disciplinary preferences - capture verbal, non-verbal etc?
  - Indicate speakers
  - Implications of various technologies – video, multiple camera, web

Example: recording equipment
- Specify expected formats
- Standard equipment settings, possible autocapture of metadata
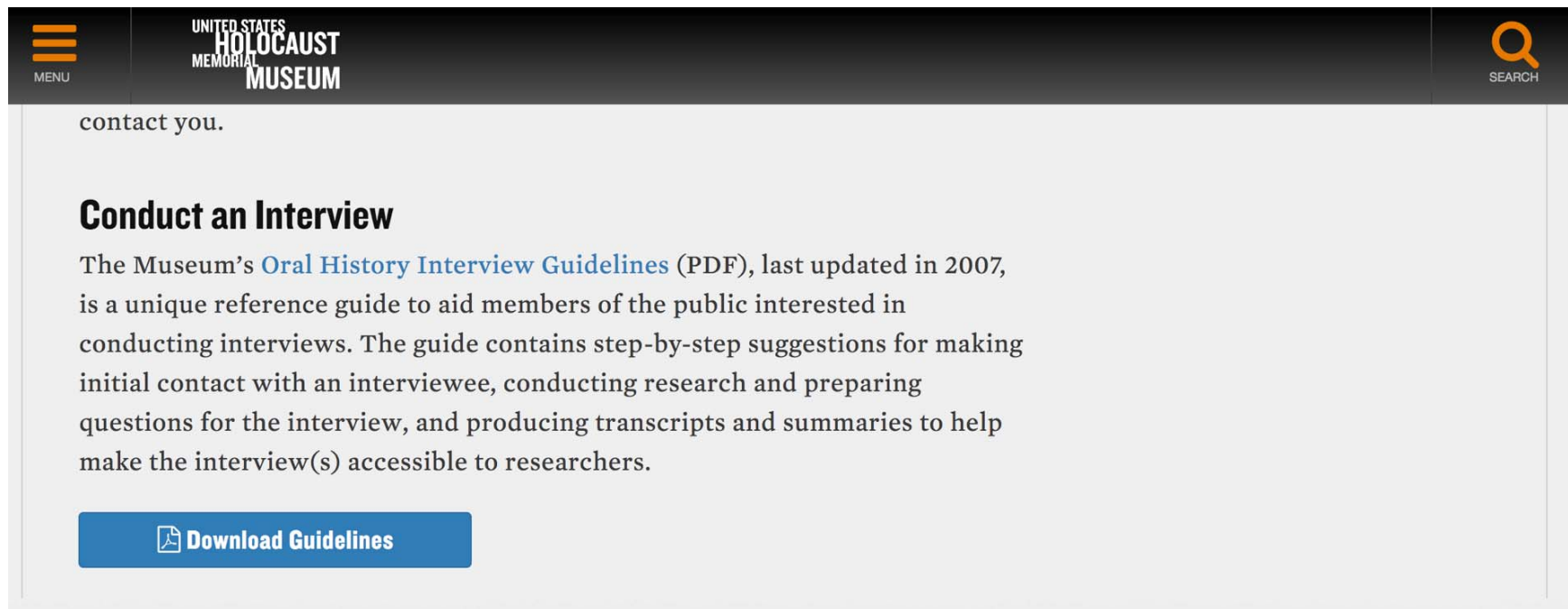
Example: digitisation of audio
- Specify the expected output
- Equipment settings, autocapture of metadata

UK Data Service

Regardless of who does the work

Rules help to make data cleaner and easier to share

Templates are always good!

Conduct an Interview

The Museum's Oral History Interview Guidelines (PDF), last updated in 2007, is a unique reference guide to aid members of the public interested in conducting interviews. The guide contains step-by-step suggestions for making initial contact with an interviewee, conducting research and preparing questions for the interview, and producing transcripts and summaries to help make the interview(s) accessible to researchers.

📄 Download Guidelines

Suggested formatted transcript template rtf

Study Name:
Depositor:
Interviewer:

Interview number:
Interview ID: Firstname Lastname
Date of interview:

Information about interviewee
Date of birth:
Gender:
Marital status:
Occupation:
Geographic region:

I: Just one or two factual details first of all before we go on to your health and that....
        how old are you?

FL: I'm 58 in June.

I: What schools did you go to? Can you remember that far back!

FL: Oh... the last school was at Longside, aye, ken Longside?

I: No, where is that?

FL: Peterheid. Village... That was the last school.

I: Uh-huh, so you lived in Peterhead...

FL: No, Longside.

I: Longside. And, do you work at all? At the moment?

FL: Just look after my grandchildren. Like that... well, my grand-daughter comes in at
        night., well, her mither cleans the school... and I look after my grandson whose
        mother works... aye, that kind of thing. I take him in an' keep an eye on him.
        Well, he's at the school but I give him his dinner an' look after him at night till she
        comes an' picks him up.

I: Just one or two factual details first of all before we go on to your health and that....
        how old are you?

FL: I'm 58 in June.

I: What schools did you go to? Can you remember that far back!

FL: Oh... the last school was at Longside, aye, ken Longside?

I: No, where is that?

FL: Peterheid. Village... That was the last school.

Collection-level header

Data- level header

Speaker tags

UK Data Service

Publishing oral history - requirements

- Collection level metadata
- Item level data list – finding aid
- Interview level metadata
- Interview summary
- Video/audio

Added value

- Mark-up/annotation of text, TEI, e.g. correct typos <>
- Alignment of AV and text
- Enable fine-grained citation

British Library oral history – pdf transcript not syncronised with audio

Elisabeth Frink Page 1
C466/12/01 F2986 Side A

F2986 Side A

Interview with Elisabeth Frink, December 28th, l992.

Let's start, Liz, right back with your childhood, where you were born, and your earliest memories.

I was born in Suffolk, in a house that my father had been born in, before me, which was nice. It's an old house. And my grandfather was a Canadian. And he'd come over to go to Cambridge, and to school, and finally Cambridge University. Never went back to Canada, because he liked England so much, and settled in Suffolk. And so Dad was born in, and consequently we all, my brother and myself were.

Was your grandmother English?

Yes. She was English, yes. My childhood memories are very good, really, because they're all to do with country things, and my dog, and my pony, for quite a while. And ...

So you learned to ride very young?

I did learn to ride very young. I was never quite sure whether I liked it, because, you know, one had very silly ponies, usually. People were a bit daft in those days, they'd stick their

<table>
<tr><td colspan="4">The British Library   Oral History</td></tr>
<tr><td colspan="4">Interview Summary Sheet</td></tr>
<tr><td>Ref no:</td><td colspan="3">C1345/22</td></tr>
<tr><td>Collection title:</td><td colspan="3">Unheard Voices: interviews with deafened people</td></tr>
<tr><td>Interviewee's surname:</td><td>Beal</td><td>Title:</td><td></td></tr>
<tr><td>Interviewee's forename:</td><td>Christine</td><td>Sex:</td><td>Female</td></tr>
<tr><td>Occupation:</td><td></td><td>Date and place of birth:</td><td>1957, Southampton</td></tr>
<tr><td>Date(s) of recording:</td><td colspan="3">13th March 2009</td></tr>
<tr><td>Location of interview:</td><td colspan="3">The City Inn Hotel, Bristol</td></tr>
<tr><td>Name of interviewer:</td><td colspan="3">Sarah Smith</td></tr>
<tr><td>Speech to Text reporter:</td><td colspan="3">Mirella Fox</td></tr>
<tr><td>Type of recorder:</td><td colspan="3">Marantz PMD660 on compact flash</td></tr>
<tr><td>Recording format :</td><td colspan="3">WAV 16 bit 44.1 kHz</td></tr>
<tr><td>Total no. of tracks:</td><td>1</td><td>Mono or stereo:</td><td>stereo</td></tr>
<tr><td>Total Duration:</td><td colspan="3">1 hr. 8 min.</td></tr>
<tr><td>Additional material:</td><td colspan="3"></td></tr>
</table>

UKDS QualiBank – searching content of data

QualiBank

Rich item-level metadata needed for on screen navigation and context

Interview with Mr. Keble

External resources | Audio | Images | Collection documentation   Access this collection from Discover

DETAILS

Collection ID (SN):	2000
Title:	Family Life and Work Experience Before 1918, 1870-1973
Principal investigator:	Thompson, P., University of Essex. Department of Sociology
	Lummis, T., University of Essex. Department of Sociology
Sex:	Male
Age group:	65-74
Socio-economic status	Semi-routine
Region:	East of England
Licence:	Creative Commons Attribution 4.0 International License.

CC BY

+ Additional metadata

INTERVIEW TRANSCRIPT

I: I'd like to start, if I may, by asking you your birth date.

R: November 9th, 1902.

I: Could you tell me how many children there were in your family?

R: There were 11 of us. I was the eldest.

RELATED RESOURCES

Related external resources

XML TEI Transcript

XML QuDEx metadata

Web Resource: Abstract of the Edwardians book

Related audio

Audio extract from Interview with Mr. Keble

Audio extract from Interview with Mr Keble

Related images

Image relating to the Edwardians

Image relating to the Edwardians

Related collection documentation

User Documentation: User Guide

School Leavers on the Isle of Sheppey - TEI

```
1    al version="1.0" encoding="utf-8"?>
2  ▽:  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3      xsi:schemaLocation="http://www.tei-c.org/ns/1.0 http://www.tei-c.org/release/xml/tei/custom/schema/xsd/tei_all.xsd"
```

ESSAY

Reflections

R: [Name omitted]. I am 75 years old and I live on the Isle of **Sheppey** with my dear wife Elsie.

R: Looking back on my life it's been very enjoyable even school wasn't ~~too bad~~ wasn't very brainy in most subjects but I was in music. At
the end of the fifth year achieved a Grade 1 C.S.E. which was an O level in those days. After the fifth year I stayed on for a {illegible} year
in the 6th form, one thing I never liked about the sixth form was Hutley! Err " **excuse** me, Mr Hutley, well anyway he was just like one of
those square bashers you get in the army, whenever you was out of line " **watch** out Hutley's about"

> **Original Text**
> whatch

R: After my year was up I had gained a couple more qualifications. I then left school and went into the Royal Marines Band where I literally
saw the world and also gaining a lot of experience.

R: I came out of the **Marines** when I was 55 and I had gained myself quite a bit of cash I came out of the **Marines** and bought a house and
retired.

R: From then on it became fun just like when I was 15 and I use to go skateboarding I thought I would have another go at it, but now they
are different they run on nuclear energy and they are very expensive.

RELATED RESOURCES

Related external resources

XML TEI Transcript

XML QuDEx metadata

```
45       twice I stayed that way till I retired last month. I did get married to someone who
46       I met on a <sic>archaeolical</sic><corr>archaeological</corr> site who been in a similar kind of position we had very
47       similar tastes in most things and we both shared nearly the same views of life.</u>
48     </body>
```

Document or paragraph citation

Citation

A unique citation reference has been generated based on your selection.

of Essex. Department of Sociology: "Interview with Mr. Keble" in "Family Life and Work Experience Before 1918, 1870-1973" 7, UK Data Service [distributor], 2009-05-12, SN:2000, Para. 5-9. http://dx.doi.org/10.5255/UKDA-SN-2000-1, https://discover.ukdataservice.ac.uk/QualiBank/Document/?cid=q-cdef4a1a-8bfe-4f24-8620-9753c9021760

Select the text above

You can copy and paste this citation as required in your outputs. This citation includes a URL which will link directly back to this page, where the cited text will be highlighted.

Show preview of citation URL in action

Need for rich metadata

- Key is identification of data objects
    - Type: Interview transcript or audio recording etc.
    - Relationship between data objects
    - Descriptive categories at the object level, e.g. interview characteristics, interview setting

- Capacity to capture rich annotation of parts of data e.g. codes and memos

- Enable alignment of audio and text

- Enable fine-grained citation (dip into context)

UK Data Service

Documentation of oral history *collection*

Family Life and Work Experience Before 1918, 1870-1973

- 450 unanonymised life story interviews – Zip bundle RTFs
- Data list (Excel and PDF)
- Citation file and Read file (information on data preparation)
- User guide - research report, interview schedule etc.

DOCUMENTATION

Title	File Name	Size (KB)
Data list	2000ulist.xls	148
File Manifest for SPSS format zip file	2000_file_information_spss.csv	46
File Manifest for Stata format zip file	2000_file_information_stata.csv	47
File Manifest for tab-delimited format zip file	2000_file_information_tab.csv	46
File Manifest for XML format zip file	2000_file_information_xml.csv	46
User guide	2000uguide.pdf	8129
Data list	2000ulist.pdf	113
Study information and citation	UKDA_Study_2000_Information.htm	6
READ File	read2000.htm	6

Data list

SN 2000 Family Life and Work Experience Before 1918, 1870-1973
Thompson, P.

Interview ID	Title	First Name(s)	Second Name	Occupation	Occupational class	Place of Interview	Date(s) of Interview	Name of interviewer	Interview Notes	No. of Pages	Text File Name
1	Mr	~	Keble	Postman	Semi-skilled manual	Colchester	~	~		55	2000int001
2	Mr	~	Stinchcombe	Farm labourer	Semi-skilled manual		~	~	Includes 29 page pilot interview	109	2000int002
3	Mrs	~	Duckers	Cleaner	Semi-skilled manual	London	19.02.1968	~	Includes 22 page pilot interview	78	2000int003
4	Mr		Knifton	General Omnibus Co.	Skilled manual	London	27.02.1968	~	Includes 14 page pilot interview	74	2000int004
5	Mr	John	Troy	Farm Worker	Semi-skilled manual	London	31.05.1968	~	Includes 21 page pilot interview	85	2000int005
6	Mrs	~	Craig	Tailoress	Skilled manual	London	08.03.1968	~	Includes 18 page pilot interview	61	2000int006
7	Mrs	Clara	Wilson	Head laundress	Semi-skilled manual	London	??.??.1968	Elizabeth Sloan	Includes 26 page pilot interview	73	2000int007
8	Mr	James	Rook	Porter	Unskilled manual	London	~	~	Includes 20 page pilot interview	82	2000int008
9	Mr	Edward William	Wiffen	Bottler - mineral water	Semi-skilled manual	Colchester	~	~		71	2000int009
10	Mr	Harry	Taylor	Wheelwright	Skilled manual	Kelvedon	~	~		96	2000int010
11	Mr	Norris	Thompson	Manager/owner of family shop	Employers and managers	Wilcot	~	~		116	2000int011
									Interview includes a		

UK Data Service

DDI catalogue record; citation

Family Life and Work Experience Before 1918, 1870-1973

Documentation | Related Studies | Publications | Syntax   Access online | Download | DDI XML

TITLE DETAILS

SN:	2000
Title:	Family Life and Work Experience Before 1918, 1870-1973
Alternative title:	The Edwardians
Persistent identifier:	10.5255/UKDA-SN-2000-1
Depositor:	Thompson, P., University of Essex. Department of Sociology
Principal investigator(s):	Thompson, P., University of Essex. Department of Sociology
	Lummis, T., University of Essex. Department of Sociology
Sponsor(s):	Economic and Social Research Council

CITATION

The citation for this study is:

Thompson, P., Lummis, T. (2009). *Family Life and Work Experience Before 1918, 1870-1973*. [data collection]. *7th Edition*. UK Data Service. SN: 2000, http://doi.org/10.5255/UKDA-SN-2000-1

Select the text above to add data citation in your outputs.

Select citation format: APA   **XML citation formats:** CSL EndNote

What to keep from analysis software

- Discipline-specific customs/ planned data analyses
- Proprietary nature of software
  - CAQDAS , e.g. NVIVO, Atlas-ti, Max-QDA
- What to keep?
  - ✓ Clean transcripts, anonymised where needed
  - ✓ Final coding frame – any open format
  - ✓ Data list – data items categorised (finding aid)
  - ✓ Export of any quantitative (coded) data tables
  - ✓ Data ordered time-stamped memos (research notes), like a research diary

Bridging the gap

- Minimal set of standards for capturing and preserving oral history data that meet the needs of all disciplines

- Additional guidance for discipline or analysis needs

- Minimal set of standards for publishing oral history data

  - Time codes

  - speakers turns

  - non verbal, pauses and silences etc

  - interruptions, false starts,

  - spelling corrections

  - comments

  - customized elements (according to different fields of knowledge)

UK Data Service

Work being done by CLARIN OH group

OPEN VOOR REVIEW, PLEASE ADD COMMENTS	journalist	archivist	historian	linguist	social scientist	educators, artists, community leaders
Motives for needing a written representation of an interview	identify cues/quotes from content for article	open up collection, align with video/audio	triangulate what is discussed in interview with other sources/ assess representation of identity or other mental schemes through reading and analyzing	documenting the verbal and non verbal characteristics of the language that is used	grasping representations of the real world in a systematic way to offer evidence based research findings or policy advice	awareness of local history and community enegagement and the power of testimony
Intended audience that reads or checks the transcript	General public, peer researchers	Peer researchers, general public	General public, peer researchers	Peers, partly general public	Peers, policy makers	schoolkids, elderly, students, community members
Practices with regard to transcription:						
Jotting notes down during the interview	yes	yes	yes, additional to recording	yes, additional to recording	yes, additional to recording	yes
Gested: summarize while transcribing	no	yes	no	no	no	no
Pragmatic/light editing: a corrected readable version of what is uttered	no	yes	yes	no	no	no
Mix of pragmatic and conversation analysis: readable version including comments about para-verbal elements according to needs	no	no	yes	no	yes	no
Classic Conversation Analysis: everything that is uttered, with special symbols according to conventions for para-verbal elements	no	no	no	yes	yes	no
speaker turn with white line in between	no	yes	yes	yes	yes	maybe
anonymization	no	no	seldom	???	yes	no
time codes	no	yes	seldom	????	seldom	maybe
symbols for non-verbal components	no	no	no	yes	yes	no
metadata added to header	no	yes	yes	yes	yes	yes
Adherance ot rules/standards?						
Transcription rules - set or follow rules and templates?	no	yes	no	yes	sometimes	no
Emphasis on quality of audio	maybe	yes	yes	yes	yes	no

Transparency requirements?

- Availability of high quality oral sources to validate transcriptions

- Good documentation of fieldwork

- Rich item-level metadata

  - Interview attributes, where, who, coverage?

  - Interviewee attributes

- i.e. process transparency

- What about analytic transparency –

  - Memos in CAQDAS to document analytic choices?

  - Export of coding tree from CAQDAS - approach

UK Data Service

Contacts and thanks

Thanks to slides on transcription borrowed from colleagues Stef Scagiola and Silvia Calamai from our recent oral history workshop

UK Data Service

@ukdataservice

@LouiseCorti

corti@essex.ac.uk

Ukdataservice.ac.uk

www.data-archive.ac.uk

UK Data Service