

Enhanced Publications: e-humanities practices and multimedia data in the realm of qualitative research

Stef Scagliola

Veterans Institute / Erasmus Studio for
e-research Rotterdam

COST – Physics of Competition and Conflicts

Eindhoven 20-05-2011



Background

- Military historian – Erasmus University Rotterdam history combined with the insights of the social sciences
- Interviewproject Dutch Veterans – 2006-2011
Goal: collect 1000 interviews with a representative number of veterans of all conflicts and military missions in which the Netherlands were involved (1940-2010)
- near future: researcher at Erasmus Studio for e-research; applying search technology in audio-collections



Interviewproject: life-story- interviews

- Semi-structured topiclist
youth – education – preparation for war/military mission – experiences during mission/conflict - return and adjustment – retrospective
- Main themes older veterans: conscription, conventional war, long periods, look back on their lives
- Main themes younger veterans: all volunteer army, ‘peace missions’, maximum of 4 months, relatively young



1000 biographical interviews of $\pm 2,5$ hours

Digital
database

Personal data
Registration number
Military education
and career
Personal
documentation
Summary for every
10 minutes
Thesaurus with
thematic keywords
Explanation of
abbreviations,
linggo, technical
terms

Matches data ministry of Defense

Channel to cultural heritage institutions

Searchable with text

Searchable by attributed key-words

Material for military dictionary,
anticipating automatic speech
retrieval



Interview-data

History: oral history/oral sources (about the past, can be complementary to written sources, sometimes only oral sources)

Social sciences: qualitative data (about current matters, standardized topic list, sometimes combined with survey/mix method)

Two kinds of interview-data: research-generated and process-generated



Problematic aspects of research based on interviews

- Most scholars don't use standards to document their data
- No context available about how the research project was set up
- Data is not shared
- Data is not digitized and perishes
- Reuse is uncommon and hampered by code of anonymity
- No possibility to validate the interpretation of the interviews
- Need for full transcript in order to analyse content in detail, is time-consuming and expensive



ICT offers solutions

- Long term digital preservation
- Tools to give selective access to authorized researchers
- Tools to facilitate re-use of qualitative data
- Tools to compare data from different disciplines
- Tools to give direct access to spoken document
- Tools to enhance qualitative data
- Development of open source software
- **This requires a different mindset: digital education, sharing, preserving, validating,**



What are the basic issues for the implementation of ICT in those area's of the humanities where scholars use qualitative data?

- How to get them to digitize their data
- How to get them to share their data
- How to make them supply context which is necessary to recontextualize data
- they need full transcripts to analyse their data
- they have to migrate their data to new formats in order to keep access to the data
- They need academic credits for all this



Solutions for these issues and an agenda for Information scientists and universities

- Radical adjustment of academic curriculum with regard to collecting, processing, curating and presenting data
- Independence from commercial software developers
- Speeding up of automatic speech retrieval techniques
- New judicial arrangements with regard to informed consent and copyright
- Academic credits for data-curation



Experiment with reuse: The Veteran Tapes

- Combined effort by the Veteran Institute and Data Archiving Networked Services (DANS)
- full text-transcriptions of 25 interviews on various military conflicts/missions (1940 - now)
- diverse group of researchers (theology, anthropology, sociology, psychology, discourse analysis, methodology) have been invited to write an article based on the same set of interviews.
- The project is meant to prove that this type of collections can be of use to social scientists of various disciplines, as long as a set of conditions for the reuse of qualitative data is met.



Results

7 articles from different perspectives:

- Images of friend and foe
- Discours of acts of violence
- Moral dilemmas
- Religion and spirituality
- Coping with shocking experiences
- Images of the local population
- Personal experience versus institutional history



Experiment with enhanced publications

- Combined effort SPEX-university of Nijmegen, DANS, VI
- Objective: Link structure for rich biographical data and links to recorded audio (aligned via ASR)
- Challenges: transcript alignment, NE relation detection, topic classification, result presentation, harmonisation of reference styles, faceted IPR handling
- Further aims: explore the (dis)similarities in user needs for the various disciplines involved, stimulate resource sharing.





The Veteran Tapes:



Research Corpus, Fragment Processing Tool, Enhanced Publications for the e-Humanities

Henk van den Heuvel¹, René van Horik², Eric Sanders¹,

Veteraneninstituut.nl

Stef Scagliola³, Paula Witkamp²

Veteraneninstituut.nl



¹CLST, Radboud University Nijmegen, the Netherlands; ²DANS, The Hague, Netherlands; ³Veterans Institute, Doorn, Netherlands



Enhanced Publication (EP)

EPs are:

- Compound digital objects
- Combining heterogeneous but related web resources
- Basis is traditional academic publication

Research Corpus

- 25 interviews with veterans
- Each between 2 – 2.5 hrs
- Stored in 16 kHz, 16 bits wave
- Orthographically transcribed and aligned with audio
- Stored at DANS with Persistent Identifiers



Link to



Interview and fragment

Audio



Transcription

Metadata

en ik... ja... nu... dat ik, ik wil wel vertellen het, maar ik durf maar dan gewoon als...
daar, als ingeferd.

De uitdaging bij elke rechtspraak is dat het gaat om het bevochten van een in principe gelijk-
waardige tegenstander. Het gebruik van fysiek geweld tegen een ongewapende of zelfs gewo-
ne tegenstander past daar niet in.

Doordrachten van ongewapende tegenstanders is vanuit die perspectief helemaal uit den
been. Dat laatste komt bijvoorbeeld tot uiting in een fragment aan het eind van hetzelfde in-
terview (lat. 488). De respondent vertelt daar over een speurtocht naar de locatie van een ge-
rilla-eenheid van de tegenpartij op basis van de informatie van een gevungen gesomene strijder
van die groep. Hij overlegt met zijn commandant over het risico dat de gevungene vlucht of de
gerilla-eenheid vroegtijdig alarmeert:

Documentatie over de veteraan en de miszie!
ik zeg, nou prima hoe hoe lang je ook daarvoor? Ik zeg want het is macht en strik
trekt je aan 'n platen en hi dan weten we niet waar we moeten zoeken. 'n dat we-
ten we, we hebben op de kaart alles af uitgezocht, maar we houden hem gewoon met een
tegenstander mee. 'Waar is het, dat een tegenstander, dat hij niet weg kan lopen. Dan
moet je dus verder de weg wijzen dan dat is ik dat die niet hard weg kan lopen. Ik zeg
'n ik'. Ik zeg 'maar hij moet wel zijn mond, ik weet, handen, maten, ik wil in de



Fragment link is
made by fragment
fitter tool:

Fragment fitter tool

- Web application
- Used by researcher to:
 1. Find the interview fragment
 2. Inspect audio segment
 3. Anonymise & improve transcription
 4. Add annotation
 5. Select metadata

IPR policy

- Linked fragments accessible for every EP reader
- Citations must be duly anonymised
- Further access through Veterans Institute and DANS

Future work

- Text based search facility
- More interviews using ASR output for text mining
- Improve annotation facility with file uploads
- Standardisation in CLARIN-NL

1. Find fragment

2. Inspect audio

3. Adjust transcription

4. Add annotation

5. Select metadata

ACKNOWLEDGEMENT

The presented work was supported by grants from SURF, the higher education and research partnership organisation for ICT, and from the programme Erfgoed van de Oorlog, which is a programme from the Dutch Ministry of Health, Welfare and Sport.

Links online

- Demonstration enhanced publication
- link [film SURF 0.00- 1.07 / 2.44 – 3.54](#)
- <http://www.watveteranenvertellen.nl>



Thanks!

