# 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 eresearch; applying search technology in audiocollections

## Interviewproject: life-story- interviews

- Semi-structured topiclist
   jouth 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

**Digital** database

Personal data

**Registration number** 

Military education and career

Personal documentation

Summery 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, standarized topiclist, sometimes combined with survey/mix method)



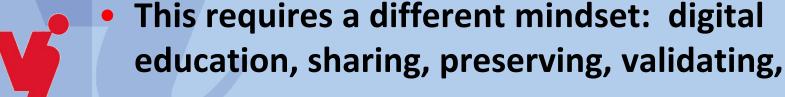
Two kinds of interview-data: research-generated and process-generated

# 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 anonimity
- 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 facilitaite 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





# 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
- Independece from commercial software developers
- Speeding up of automatic speech retrieval techniques
- New judicial arrangements with regard to informed consent and copyright
  - Academis 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 VeteranTapes:



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

Henk van den Heuvel<sup>1</sup>, René van Horik<sup>2</sup>, Eric Sanders<sup>1</sup>,

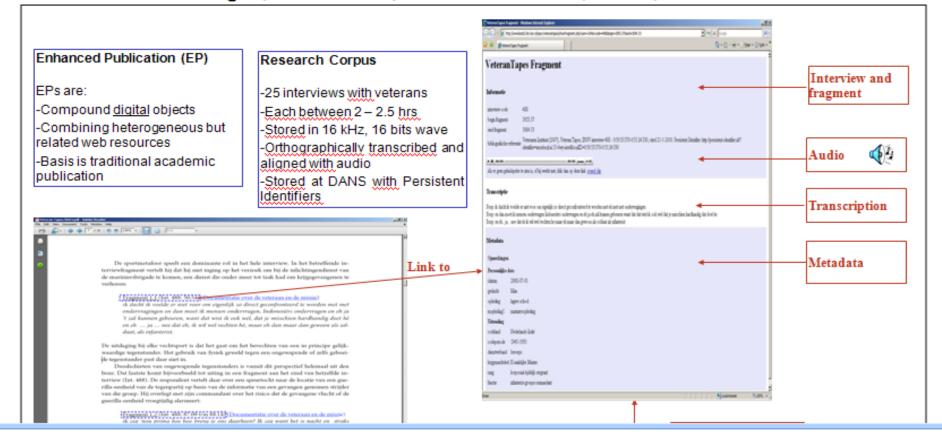
Veteraneninstituut.nl

Stef Scagliola<sup>3</sup>, Paula Witkamp<sup>2</sup>

Veteraneninstituut.nl

The Hague, Netherlands; <sup>3</sup>Veterans Institute, Doorn, Netherlands







Proposed 1 to the 20th AT 00 the BELLE Documentatic event de veterann en de missie) if, say mos prima fore fore breigt je ons destricor? It say want het is macht en stroits richt je dan 'e platjoeten en hi dan vetere ne niet macht er e moeten soeken. Te dat weten we, we hebben op de kaart alles ei uitgezocht, maar we houden hem gewoon met een touw ann zijn nek vast'. Weet u wel, dus een touw, dat hij niet weg kon lepen. Dan moest ie dus verder de weg wijzen dus dat ie ek dat die niet hard weg kon lopen. Ik zeg 'ia old'. Ik nee 'maar hii moet wel niin mond ch weer houden natuurliik als wii in de

Fragment link is made by fragment fitter tool:

#### Fragment fitter tool

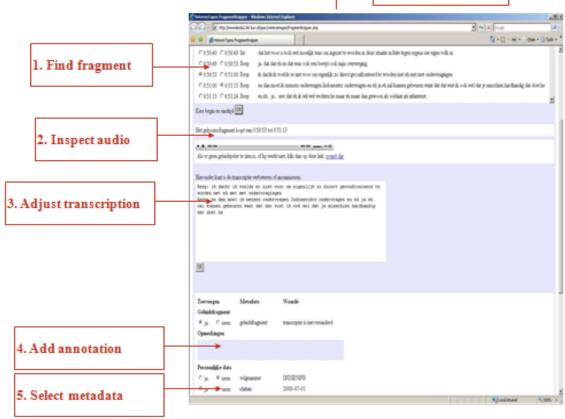
- -Web application
- Used by researcher to:
- 1. Find the interview fragment
- 2. Inspect audio segment
- Anonymise & improve transcription
- 4. Add annotation
- 5. Select metadata

#### 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



FOUNDATION

#### 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. 

SURF

#### Links online

- Demonstration enhanced publication
- link <u>film SURF</u> 0.00- 1.07 / 2.44 3.54
- http://www.watveteranenvertellen.nl



# Thanks!



