

**Prof. Machteld VENKEN**  
**Dr. Inna GANSCHOW**  
**C2DH, UL**

#### **DH Benelux 2024**

Breaking Silos, Connecting data: advancing integration and collaboration in DH  
<https://2024.dhbenelux.org/call-for-papers/>

### **Designing a best practice digital workflow in times of war**

Insights into the launching of the U-CORE project

This paper provides an insight into the ongoing interdisciplinary research aimed at designing a best practice digital workflow for the collection, research, and archiving of audio testimonies related to the Russian full-scale invasion in Ukraine. As the first and largest international documentation project of testimonies on this war, it serves as a continuous guide for similar smaller documentation initiatives. The documentation project, known as '24.02.2022, 5am. Testimonies from the War' originated as an ad hoc collaboration between loosely affiliated institutions in Luxembourg, Poland, Ukraine, and Great Britain in March 2022.<sup>1</sup> It initiated with the development of a joint informed consent procedure and interview question catalogue. Nevertheless, we discovered differences in cultural translation, scholarly practice and disciplinary specialties in the implementation of our jointly composed procedure, which impact further decision making.

The presenters offer an overview of the reflection with the different partners on the composition of a central digital environment for the 400 testimonies. What (meta)data can be synchronised? When and why is that not always possible, i.e., where do we have to explicit the specificity of scholarly intentions, feelings of sensitivity about certain information, and cultural interpretations of metadata among the different teams? The metadata field 'education', for example, was filled in with the content of the study by the

---

<sup>1</sup> The international will of the documentation project '24.02.2022, 5am. Testimonies from the War' (01.03.2022-31.12.2023) consisted of the following parties: Center for Urban History (Ukraine), Institute of Philosophy and Sociology at the Polish Academy of Sciences (Poland), the University of Saint Andrews (United Kingdom) and the Centre for Contemporary and Digital History (Luxembourg). The partners gathered 400 oral testimonies. Similar initiatives currently take place at the Fernuniversität Hagen and the Bavarian State Library (Germany).

team from Luxembourg ('mathematician'), compared to the degree of interviewees by the Ukrainian team ('master degree').

From January 2024 onwards, the research project 'Researching the Collecting, Preserving, Analysing and Disclosing of Ukrainian Testimonies of the War' (U-CORE) aims to consolidate the disparate data sets of the documentation project, which are stored across various institutions, servers, countries, and technologies.<sup>2</sup> This paper highlights novel aspects introduced to bring clarity to the creation of a central digital environment for ongoing and future research on this war.

The first innovation involves using a planned second round of interviews to fill gaps detected during the joint reflection with all teams by employing synchronised follow-up questions, such as for "education". However, the aim to create a unified set of metadata has already appeared to be unreachable, as the teams have different understandings of which questions about metadata may potentially be considered inappropriate. One example is the question about the citizenship of an interviewee. The team from Ukraine does not want to ask that question as some long-term inhabitants of Ukraine are still having Russian passports, which they may not want to reveal.

The second innovation includes protective measures to ensure the security of interviewees during an ongoing war. The digital workflow incorporates encrypted recorders, separate storage for personal data, and a secure, encrypted server. CatDV, a media asset management tool, facilitates ingesting, indexing, transcoding, and archiving. The CatDV Pegasus license allows the development of a multilingual data model (Ukrainian, Russian, and English) and provides two-factor identification.<sup>3</sup> Our data model in CatDV also provides a diversified access to the different sets of metadata. Whereas the access for administrators allows to consult both the pseudonymised and the original version with the real name of the interviewee, the access for certain team members only foresees the limited editing of the

---

<sup>2</sup> The bilateral project is financed by the Luxembourg National Research Fund (INTER) and the National Science Center in Poland (OPUS) as a WEAVE.

<sup>3</sup> More information about CatDV Pegasus can be found here:  
<https://www.squarebox.com/products/desktop/catdv-pegasus/>.

collection. In addition, the user access offers the possibility to listen to selected interviews as well as to search through the metadata by means of filtering. The interviews of the team in Luxembourg were transcribed using HappyScribe, a digital automatic speech to text transcription tool.<sup>4</sup> Upon our request, the company increased its security measures for the data processing of our collection.

The third innovation applies transcription ethics and transcription criticism to the collection. The ethical dilemma arises in deciding whether to pay newly arrived Ukrainian refugees for manual transcription or use automatic speech to text transcription software to help train automatic transcription for Ukrainian and because a computer cannot be traumatised. The paper also offers an initial comparison of outcomes from automatic AI-driven, manual, or mixed transcription methods.

Lastly, the paper reflects on ideas regarding the final data repository and platform for the central digital environment, shared among project partners within the Oral History Technology Watch Working Group of U-CORE. This includes ‘thinkering’<sup>5</sup> with the platform ‘Oral History digital’ launched in 2023 to align with project partners’ decisions on data model architecture.<sup>6</sup>

---

<sup>4</sup> More information about HappyScribe can be found here: <https://www.happyscribe.com/>.

<sup>5</sup> ‘Composed of the verbs “tinkering” and “thinking”, it describes the action of playful experimentation with technological and digital tools for the interpretation and presentation of history’ (Fickers, van der Heijden 2020).

<sup>6</sup> More information about HappyScribe can be found here: <https://www.happyscribe.com/>.