

# Structuring a Digital Archive of Survivor Testimony in Times of Emergency

Yael Netzer<sup>1,2</sup>, Renana Keydar<sup>1,2</sup>, Keren Shuster<sup>2</sup>

<sup>1</sup>Center for Digital Humanities, the Hebrew University, <sup>2</sup>Edut 710

## Overview

Early morning on October 7, 2023, Hamas invaded southern Israel from Gaza. Simultaneously, thousands of rockets were fired from Gaza toward Israel. In the ensuing hours and days, more than 1100<sup>1</sup> people were killed, thousands were wounded, houses were bombed and burned, and about 250 people were kidnapped to Gaza, alive or dead. The attack took place along a 60 km border and hit more than thirty towns and villages, outdoor music festivals, and bypassers on the roads. The surviving civilians were forced to flee their homes and are internally displaced in temporary residences.

As the scope and scale of the atrocities began emerging, there was an instant surge of initiatives led by civilians to document what had happened.<sup>2</sup> On October 11, 2023, just 4 days after the attacks, we joined a small group of documentary cinematographers and scholars named “Edut 710” (Testimony 710), who began conducting video interviews with survivors of the attacks.<sup>3</sup> Inspired by Fortunoff Video Archive for Holocaust Testimonies,<sup>4</sup> Edut 710’s aim was to create a ‘historical archive of testimonies’. Drawing on our experience in forming digital archives and collections<sup>5</sup>, and in applying distant reading methods for the analysis of testimonies<sup>6</sup>, we believed that a DH-based approach to data-modeling from the start will help secure the sustainability of this grassroots initiative.

In our presentation, we share our experience from building a collaborative digital archive in times of emergency. We detail the challenges we faced and highlight the importance of harnessing best practices and skills to a developing digital archiving effort, managing the needs and overcoming the obstacles of a rapidly growing, distributed organization of more than 200 volunteers.

1 <https://www.nli.org.il/en/at-your-service/announcements/7-october-documentation-initiative> [Last visited 15.08.2024]

2 <https://www.edut710en.org/> [Last visited 15.08.2024]

3 <https://fortunoff.library.yale.edu/> [Last visited 15.08.2024]

4 <https://verbit.ai/> and <https://dixilang.com/> [Last visited 15.08.2024]

5 <https://oona.net/> [Last visited 15.08.2024]

6 Renana Keydar, “Listening from Afar: An Algorithmic Analysis of Testimonies from the International Criminal Courts,” *Illinois Journal of Law, Technology & Policy* 2020, no. 1 (2020): 55–83.

## Research objectives

At the center of our project is the testimony of survivors. This is the main unifying organizing entity in terms of project management and of modeling the knowledge for the archive. Other entities are determined with respect to it: staff members, locations, dates, organizations, etc.

Our research objectives were developed in parallel to the unfolding of disaster and the emergence of the un-planned documentary project. Two fundamental assumptions guided the methods used: first, the importance of testimonies for historical documentation, but also for possible legal use. From these assumptions derived the importance of keeping provenance, registering all information and safeguarding verification and reliability of content.

Our main concern was with modeling the domain, in terms of knowledge and data representation and creating an ontology: a testimony, an interview, the actors, the mandatory information and more. Since we are dealing with events that took place over a limited space and time (several dozens of locations and a limited timeframe of about 48 hours of the Hamas attack), this created a very distinct, closely knit domain with specific recurring entities. On the other hand, we are dealing with an event whose contours are still very much unknown (the number of wounded and killed, the state of the abductees and more).

This work, therefore, deals with understanding, testing and applying best practices of archive building, in 'real time' and on the go in a collaborative context (non-academic, non-DH, non-archiving community), in time of unfolding disaster.

However, given the project's main objective of gathering the testimonies of traumatized survivors very shortly after the atrocity, our main commitment is for the well being of the witness and for facilitating the best conditions for them to tell their story. This means that from a DH perspective, the data comes second in this project.

## Challenges

As the work on this project continues, we defined the challenges of constructing the testimony archive under conditions of emergency, which will be elaborated further in our presentation:

**Knowledge collection and representation:** there is a tension between the project's vision on conducting interviews with survivors, *i.e.*, with minimal intervention, and the need to gather crucial and structured information for the archive. In order to overcome this gap without impeding on the ethical guidelines or distracting the interviewer's attention, we designed easy forms for interviewers to be able to complete while conducting the interview.

**Content management and storage needs:** the project requires storing, management and archiving systems linked together. As storage may amount to tens or hundreds of terabytes (after two months since the event, we have 10 terabytes of about 400 testimonies), and is an ongoing exchange process (photographers, editors, transcribers, etc.), it also requires management of easy and reliable access to tens of users with a variety of permissions. Storage, management and archives must keep integrity and privacy of collected information about witnesses or staff. This emphasized how crucial it is to manage knowledge unattached to a platform.

**Legal and ethics issues:** When working with survivors, it is crucial to maintain their privacy and handle sensitive information with the utmost care, ensuring that their stories are shared respectfully and ethically. Additionally, dealing with copyright issues requires a careful balance, respecting the rights of both the subjects being documented and the documentarians themselves, to protect the integrity and legality of the work.

**Short-term publication needs vs. long-term archiving:** This project was initiated as a historical archive. However, it also aims to share the outputs of the interviews in real-time, on a publicly accessible website, without compromising on the creation of infrastructure for long-term archiving. This means that we continuously work on creating good data, by explicitly detecting entities, people, organizations and not relying only on free text, but using controlled fields for metadata.

**Working with AI:** State-of-the-art, AI-based tools assist with transcription and translation (we use commercial applications, Verbit, Dixilang<sup>7</sup> and Ooona<sup>8</sup> which were donated for the project's use). While using AI is time-saving and allows for the scaling up of the project, it can also create noise and error. We built workflows of human control to safeguard quality.

## Lessons and insights

The paper will offer insights and lessons from the on-going experience of developing real-time, best practices, including what we believe is a “must” for long-term digital archive building, namely aspects that should not be compromised, and what is a “privilege,” given circumstances, resources and time constraints. One major take-away we offer in the paper is that the data structure must serve as the backbone, and data modeling should be agnostic to platform or software. We also highlight, discuss and introduce open, standard knowledge representation practices.

<sup>7</sup> Yael Netzer and Keren Shuster, “Cataloging Digital Personal Archives”, in: Aslberg Book, The Association of Israeli Archivists (AIA), forthcoming 2023.

<sup>8</sup>[https://yuval-harpaz.github.io/alarms/oct\\_7\\_9\\_eng.html](https://yuval-harpaz.github.io/alarms/oct_7_9_eng.html) [Last visited 15.08.2024] (To date, there is no official publication of all names).