

Digital Video in a Genocidal Age – The Holocaust in 52,000 Acts
Professor Douglas Greenberg, USC Shoah Foundation Institute

Douglas Greenberg, professor at the University of Southern California discussed the visual historical archive that was created to document the stories of Holocaust survivors.

Professor Greenberg opened his talk by explaining how the visual historical archive originated. He recounted that, during filming of Schindler's List in Krakow, Poland, Steven Spielberg was approached by Holocaust survivors wanting to tell their own stories. This led to collaboration between Spielberg, the Shoah foundation and others, and the visual historical archive began to take shape. Professor Greenberg noted that the initial idea was to make 50,000 short documentary films available for educational and scholarly pursuits.

He continued, explaining the challenges that each documentary film team faced as they gathered data for the archive. First, they had to locate survivors and find those willing to tell their stories. Instead of soliciting testimony, they used Steven Spielberg's celebrity status to air ads on television, seeking out interested Holocaust survivors from all over the world. The teams collected video testimonies in 56 countries and in 32 languages. By the end of the data collection, the visual archive project had 120,000 hours of video, approximately 8 petabytes of uncompressed data. These films documented 52,000 survivors' stories.

Once the data was collected, the team faced a new challenge, of searching and organizing the many hours of collected video. Professor Greenberg explained the complexity of sorting the video clips, noting that each video was viewed in its entirety in order to categorize specific keywords, such as town names, family names, or more general topics. He also described the intricacy of stratifying keywords when the videos spanned 32 different languages. Professor Greenberg recalled that at the time the project began, there were no robust search and indexing functions available to parse through video data.

Professor Greenberg asserted that due to the paucity of video navigation tools, the team decided to build software that would enable them to catalog and index video data. Cataloging would allow users to search for whole videos using keywords and indexing would facilitate searches within a particular video based on keywords. The software was eventually built, allowing translators to index approximately 50,000 keywords in English.

While the software was robust enough for users to catalog and index data, Greenberg explained that the team was left with another challenge; creating a comprehensive indexing system. Because topics could often be categorized under various keywords, a

consistent categorization method was needed. For example, if a user searched the term “starvation”, but relevant results were indexed as “hunger” then the user would never find this information. They also ran into indexing issues because of the variety of spellings and the recurrence of location names throughout a particular region. Professor Greenberg argued that the work was both technical and scholarly, as it required a great deal of historical context on the part of the translators to understand the survivors’ viewpoint and document it accordingly.

Professor Greenberg then posed a question to the audience. He asked whether the Shoah Foundation and other partners should make the archive available on the Internet. He explained that the application was available on “Internet2,” which is a private network shared amongst research institutions. The Shoah Foundation initially made the decision not to make the application available on the regular Internet to dissuade random users from downloading and editing clips of survivor testimony. Greenberg asserted that Internet2 is not only more secure, but has more bandwidth, which is appropriate for streaming the large video files. The actual video data is kept on servers maintained by the Shoah Foundation in order to protect the archive’s sensitive information.

He then demonstrated the visual historical archive for the audience. After logging in, Professor Greenberg navigated to the Quick Search functionality, which operates similar to a Google search. He typed “Firestone” in the search box, yielding every instance of the word in the entire archive, and any related roles that a person associated to the term had in the making of the archive (if any). He also showed how users can view a thumbnail and detailed biographical data of each survivor that includes basic demographic data and more specialized information, like camps to which people were taken. The search functionality is extremely dynamic because it searches all 52,000 testimonials and allows users to filter the results.

Professor Greenberg argued that the utility in having access to this material was first for scholarship on the Holocaust. He explained that when a scholar was trying to write a book about a particular labor camp, the scholar was confronted with a dearth of formal information. However, the scholar utilized the visual archive and retrieved data from 80 different testimonies, which he combined with existing research to reconstruct the story. Greenberg also noted that the archive could be used as an educational resource for both K-12 and undergraduate and graduate studies. He spoke specifically about one of his classes at USC where a student used the archive to construct a story about love and sex during the Holocaust. He explained that while this type of scholarship was not related to mainstream Holocaust research, it had been made possible by the expanse of information in the archive. Another student of Professor Greenberg’s focused on the role of women in ghettos.

Professor Greenberg concluded by raising issues around the legitimacy of the digital video archive. He posed another question, asking what was the proper way to cite video. Moreover, as the Internet and sites like YouTube continue to make videos readily available, he asked the audience whether it should be a point of conflict for a student to upload one of the testimonial videos onto such a site. Finally, he posed the question of

what happens to high quality, educational, scholarly information when it gets on the Internet. He wondered if data lost its credibility or if Facebook, his daughter's primary mode of communication, was really a benefit to information sharing.

Professor Greenberg then opened the floor to questions. Several of the questions focused on the relevance of the archive in the new digital world. One of the audience members asked whether Professor Greenberg and his colleagues had considered an application that taps into the Web 2.0 culture of today so that all victims of human rights abuses could be connected. In other words, he asked if there was a value in a survivor-generated archive. Professor Greenberg argued that the visual is founded in deep scholarship and cautioned that a wikipedia version could raise issues of ethics and credibility. He did however consent that, because of the ubiquity of video technology and the ease of today's video production, that perhaps even the visual historical archive would eventually go in this very direction.

Another audience member asked about the challenges in trying to use this type of technology with undergraduate students. Greenberg explained that the archive should be used as one source of knowledge, not the sole resource on the historical events that occurred between 1933 and 1945. He explained that it was a professor's job to frame the archive materials within the historical context.

When asked whether students were allowed to build test reels, Greenberg explained that it is not practical for the Shoah Foundation staff to continually collate videos for student purposes. However, he is considering allowing students in his next undergraduate course to upload their research to YouTube, which he hopes will be a good way to demonstrate to his colleagues and the administration at USC about what is possible when information is easily accessible.

Another audience member asked why the archive streams video at 3 mb/second when a slightly lower quality facsimile of the video could be transmitted at a much more reasonable compression rate. Greenberg responded that when the project initially started, in 1992, they wanted user to view the video on a television screen, which is why they stayed away from greater compression. However, Greenberg contended that due to advances in technology, quality is no longer an issue. He also mentioned that there is a fine line between protecting integrity and providing access and he feels that we are moving towards access, as more technologies are more readily available.