Dartmouth College Presentation
October 26th, 2012

Sam Gustman
Chief Technology Officer and Associate Dean
USC Shoah Foundation and USC Libraries
sgustman@usc.edu
http://sfi.usc.edu
Audio-Visual Testimonies

Schindler’s List

After filming *Schindler’s List*, Steven Spielberg established the Survivors of the Shoah Visual History Foundation to record and preserve the testimonies of Holocaust survivors and other witnesses.

**Initial Goals:**

- Give a voice to Holocaust survivors and other witnesses
- Record each interviewee’s full life story
- Create and preserve an archive for history and education
- Provide interviewee with a copy of their testimony
Audio-Visual Testimonies

52,000 = 105,000
video testimonies hours

<table>
<thead>
<tr>
<th>57 countries</th>
<th>34 languages</th>
<th>2,000 interviewers</th>
<th>1,000 videographers</th>
<th>70 regional staff</th>
</tr>
</thead>
</table>

7 years of manual indexing

| 7.5M clips | 62,000 indexing terms | 1.2M people | 46,000 locations | 500,000 images |
## Holocaust Testimonies by Experiences

<table>
<thead>
<tr>
<th>Experience</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jewish Survivors</td>
<td>49,307</td>
</tr>
<tr>
<td>Rescuers and Aid Providers</td>
<td>1,136</td>
</tr>
<tr>
<td>Sinti and Roma (Gypsy) Survivors</td>
<td>407</td>
</tr>
<tr>
<td>Liberators and Liberation Witnesses</td>
<td>362</td>
</tr>
<tr>
<td>Political Prisoners</td>
<td>262</td>
</tr>
<tr>
<td>Jehovah’s Witness Survivors</td>
<td>84</td>
</tr>
<tr>
<td>War Crimes Trials Participants</td>
<td>62</td>
</tr>
<tr>
<td>Other WWII experiences</td>
<td>41</td>
</tr>
<tr>
<td>Survivors of Eugenics Policies</td>
<td>12</td>
</tr>
<tr>
<td>Homosexual Survivors</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>51,679</strong></td>
</tr>
</tbody>
</table>
Other Genocide Testimonies by Experiences

Rwandan Genocide:
- Survivors: 14*
- International Aid Providers: 2

Cambodian Genocide:
- Survivors: 3

Armenian Genocide:
- Survivors: 1*

Total: 18

*50 Rwandan genocide testimonies from KGM will be added; ca. 400 Armenian genocide testimonies from AFF will be added.
Structure of a Testimony

• **Introduction**
  - Paper slate
  - Interviewee introduces himself/herself
  - Interviewee introduced by Interviewer on camera

• **Video Interview**
  - Life before the persecution
  - Life during persecution
  - Life after the persecution
  - Photos and artifacts
  - Family
  - Filmed site visits [162 interviews]
Cataloguing/Indexing

- Index terms structured in a thesaurus and include over 62,000 terms in compliance with NISO Z39.19 standard.

- Cataloguing provides biographical information for testimony level access:
  - Interviewee's names
  - DOB/POB
  - Experience information
  - Religious affiliations
  - Language of interview
  - Place of interview
  - Date of interview

- Video indexing provides access to each minute of video as if it was a web page using index terms:
  - Geographic locations
  - Types of places (camps, prisons, ghettos, etc.)
  - Dates
  - Years
  - Historic events
  - People
  - Organizations
  - Experiences (pre-, during, and post-WWII)
  - Psychological responses
  - Movements/Migrations

- Information can be converted into XML or MARC formats.
Automatic mass digitization systems processing 500 hours of testimony per week.

Digital files made for each tape at once in:
- MJPEG2000
- MPEG-2
- Flash
- Windows Media
- Quicktime

Quality assurance systems for all content

Entire archive exists already in MPEG-1 (3 mbps) digital format.
Digital Preservation

- 40 Petabyte capacity (only 8 Petabytes used)
- SHA-1 based fixity for all files
- Automatic media retirement based on age
- Mirroring of all content
- All data is compliant with the Library of Congress Audio and Video Conservation Systems
- Connected to 9th fastest academic super computer in USA for mass processing
Access to Testimony

1. Website
   - http://sfi.usc.edu/
   - Featured clips and whole testimony to browse.

2. YouTube
   - http://Youtube.com/user/USCShoahFoundation
   - Limited search and browse through 96 full-length testimonies, 30 clips, and 163 related videos.

3. Visual History Archive
   - http://vha.usc.edu
   - Watch and search within 52,000 Institute testimonies. Available at 36 institutions worldwide.
   - http://vhaonline.usc.edu
   - The VHA available on the internet with access to 1,100 interviews only, and data for all 52,000 interviews.
IWitness

IWitness is an online application that gives educators and students access to search, watch, and learn from more than 1,000 video testimonies of Holocaust survivors and other witnesses. Provides students ages 13-18 and their teachers with a searchable collection of more than 1,000 video testimonies of Holocaust survivors and other witnesses from the Institute's archive.

Digital and multi-literacy ISTE compliant interface;
University Access – Visual History Archive
Access to Entire Archive via Internet

vha.usc.edu
Watch and search within 52,000 Institute testimonies.
Available at 41 institutions worldwide.
Subscription Model – VHA sites

- Multiple 10gb connections to the Internet, Internet2, Pacific Wave, CENIC, Los Nettos and National Lambda Rail.

- Local cache required.

- Local Web site developed for context.
  e.g. http://www.vha.fu-berlin.de/

- Secure IP address based global access.
Visual History Archive Sites
by Year of Initial Access

2003
University of Southern California

2004
University of Michigan

2006
University of North Carolina at Chapel Hill
Freie Universität Berlin, Germany

2007
University of Minnesota
Monash University, Australia
Duke University
North Carolina State University
United States Holocaust Memorial Museum
University of California, San Diego

2008
Columbia University
Syracuse University
Brown University
University of South Florida
Yad Vashem, Israel
Technische Universität Berlin, Zentrum für Antisemitismusforschung, Germany

2009
Texas A&M University
Stanford University
Central European University, Hungary
Charles University in Prague, Czech Republic
Clark University
University of North Carolina at Greensboro

2010
Rutgers, The State University of New Jersey
University of Minnesota Duluth
Jewish Holocaust Centre, Australia
Universität Salzburg, Austria
New York University

2011
Northwestern University
Royal Holloway, University of London, United Kingdom
University of Massachusetts Amherst
McMaster University, Canada
Aristotle University of Thessaloniki, Greece
University of Athens, Greece
Leibniz Universität Hannover, Germany

2012
University of Pennsylvania
History Meeting House, Poland
University of Haifa, Israel
Jewish Museum Prague
L.A. Museum of the Holocaust
Stiftung Topographie des Terrors, Germany
Visual History Archive Statistics

- Average time spent in VHA: **49 minutes** per session
- Top 1,000 VHA users spend **55 minutes** per session
- Number of searches in the VHA: **240,867**
- Of the 52,000 testimonies, **35,055 were watched 187,085 times**
- Courses across all VHA sites: **325**
- Complete dissertations: **35***
- Published articles and books: **73***

* Which we are able to track
Testimony and Access

Duty of care

Addressing the concerns of:
- Intellectual property
- Privacy
- Context
- Potential misuse

Widening access to benefit:
- Teaching
- Research
- Learning
- Policy-making
Digital Repository Core Functions

- Digitization
- Cataloging
- Digital preservation
- Digital Library access
- File-server services for large assets
• Automatic and mass digitization systems for audio/video, paper and still images.

• Local and remote digital file ingest systems used to bring already digitized files into the preservation and access systems from anywhere in the world.

• Rendering and derivative services are available for Linux based software on the USC super computer.

• Quality assurance systems for all content.
- Patented cataloging and indexing system can be applied to almost any data set.
- Managed infrastructure for manual indexing.
- Import services for existing metadata.
Digital preservation

- 40 Petabyte capacity (only 8 Petabytes used)
- SHA-1 based fixity for all files
- Mirrored
- Automatic media retirement based on age
- Oracle SL8500 T10kC based with samfs/qfs
- Nirvanix mirrored cloud services between USC and Las Vegas.
Digital Library access

- Multiple 10gb connections to the Internet, Internet2, Pacific Wave, CENIC, Los Nettos and National Lambda Rail.
- Web site development.
- Secure IP address based global access.
- Subscription pricing model available.
File-server services for large assets

- Up to Petabyte size personal directories. Up to 10gb connectivity.
- Mac OS, Unix, Windows and Web compatible directories.

**Educational pricing:**
- $70/TB/Month for Nirvanix and Isilon based storage
- $1000/TB/20 years preservation storage
- Support for CIFS, NFS, iSCSI, Symantec Net Backup and Back Up Exec, Commvault
- Personal “Dropbox” like functionality through Oxygen Cloud and Gladinet coming soon.
Nirvanix Data Security

• Authentication to access Cloud File System
  – Username
  – Strong Password
  – 128 bit token based authorization key

• 256 bit AES Encryption
  – Data encrypted before transfer
  – Safe Harbor laws for European Union

• SSL for data transfer

• File obscurification
  – De-identify physical file, removing filename, owner, and type

• User defined data location policies
  – User selects Cloud Storage Data Centers

• SSAE 16 Type II certified
• 3rd Party Security audits performed
• Continuous Penetration Testing to ensure formidable cloud services worldwide
• Real-time global storage monitoring and dashboards
• SNMP for failure notifications
• Biometric hand geometry readers on doors
Target Customers

• Researchers (NSF, NEH and NIH for data management plans http://research.usc.edu/files/2011/05/NSF_Data_Mgmt_rev_AG.doc)

• Small Archives

• Large Archives

• Commercial Entities (VAR program)
## Current Customers

<table>
<thead>
<tr>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academy of Motion Pictures and Sciences - Academy Film Archive</td>
</tr>
<tr>
<td>Albanian Human Rights Project</td>
</tr>
<tr>
<td>Brown University</td>
</tr>
<tr>
<td>USC Dornsife College of Letters, Arts, &amp; Sciences</td>
</tr>
<tr>
<td>USC Digital Library</td>
</tr>
<tr>
<td>USC Information Sciences Institute</td>
</tr>
<tr>
<td>USC Keck School of Medicine</td>
</tr>
<tr>
<td>USC Libraries</td>
</tr>
<tr>
<td>USC Maternal, Child &amp; Adolescent Center for Infectious Disease and Virology (MCA)</td>
</tr>
<tr>
<td>USC School of Pharmacy</td>
</tr>
<tr>
<td>USC Viterbi - Computer Science</td>
</tr>
<tr>
<td>USC Viterbi - Signal and Image Processing Institute, Electrical Engineering</td>
</tr>
</tbody>
</table>