**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #1: Principles & Manual

Lead-beneficiary (Tool #1)	*OFM*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*OFM (Michael Loebenstein)*
Contributor(s)	*OFM (Claudio Santancini)*
Reviewer(s)	*LBI (Ingo Zechner)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.6, 2019-12-31*

Tool #1: Principles & Manual

Table of Contents

**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

1. Introduction

This Advanced Digitization Tool Kit[1] is a comprehensive and tested set of best practice documents intended to inform the digitization of film materials and related non-filmic archival materials in film heritage institutions (FHIs), other collecting institutions, or academic institutions.

The purpose of these documents is to provide information about digitization principles and standards and to guide

- comprehensive research on relevant materials;
- the selection of the most suitable technical elements for digitization;
- the selection of the most appropriate digitization workflow;
- the creation of suitable access formats;
- the creation of appropriate documentation.

The best practices presented in these documents have been developed, discussed and tested by internal and external stakeholders of the VHH project, including VHH Consortium members and Associated Partners, staff of archival institutions providing materials and know-how to the VHH project, as well as by related research projects.

In VHH, these documents are intended as an aid for researchers, research assistants, and staff at the archival institutions undertaking digitization of heritage materials within the project.

These documents are living documents. In VHH this deliverable is to be used together with the following deliverables:

- Deliverable D1.4 *Data Management Plan* (M6)
- Deliverable D3.3 *Ethics Guideline* (M12)

---

[1] Deviating from the spelling in the DoA oft he VHH project, we decided to spell "digitization" with an "z" instead of an "s".

2. Methodology

For this Advanced Digitization Tool Kit, existing practices and processes employed by archival institutions have been surveyed, in particular:

- DIF Frankfurt (Germany)
- BFI, Imperial War Museums (United Kingdom)
- EYE Filmmuseum (The Netherlands)
- Irish Film Archive (Ireland)
- Library of Congress, National Archives and Records Administration, United States Holocaust Memorial Museum (USA)
- Österreichische Mediathek (Austria)
- Cinemateca Portuguesa (Portugal)

Technical Commission recommendations of IASA (International Association of Sound and Audiovisual Archives), FIAF (International Federation of Film Archives) and AMIA (Association of Moving Image Archivists) were consulted and evaluated.

We also considered specifications, recommendations and standards that emerged during our evaluation period, in particular the German DIN Spec 15587 ("Empfehlungen zur Digitalisierung von kinematografischem Film")[2] and the emerging CEN/TC 457 "Preservation of digital cinema movies" standard[3].

This tool kit builds on the results of the "Advanced Digitization Workshop" hosted by VHH Associated Partner USHMM (United States Holocaust Memorial Museum) in Washington D.C. on March 20-22, 2019. VHH Consortium members LBI, OFM, rtds, CERCEC, JLU and TUW participated in this workshop next to professionals from USHMM, VHH Associated Partner NARA (National Archives and Records Administration), Library of Congress, technical provider Colorlab and the FADGI Audio-Visual Working Group[4].

We also evaluated processes, "lessons learned" and outcomes from peer projects such as "Ephemeral Films: National Socialism in Austria"[5] and "I-Media-Cities"[6].

In scope:

- digitization of motion picture film (35mm, 16mm, 8mm, 9.5mm)
- digitization of photographs, documents or other paper-based materials
- processing of materials (in the imaging process / post-processing)

---

[2] https://www.beuth.de/en/technical-rule/din-spec-15587/299067326 (31.12.2019)

[3] https://standards.cen.eu/dyn/www/f?p=204:7:0::::FSP_ORG_ID:2368598&cs=175FF5F54D8693823F5172C2433D57B3E (31.12.2019)

[4] http://www.digitizationguidelines.gov/ (31.12.2019)

[5] http://efilms.ushmm.org/ (31.12.2019)

[6] https://www.imediacities.eu/ (31.12.2019)

- file formats and wrappers
- quality assurance and quality control protocols and procedures
- capture of technical metadata
- documentation of digitization and/or restoration work

Out of scope:

- ingest procedures (developed as internal documents by OFM)
- media asset management systems (to be evaluated in VHH as part of WP5)
- long-term digital preservation of assets (separate guidelines exist[7])

---

[7] Guidelines for the Digital Film Collection of the Austrian Film Museum, English version at
https://www.filmmuseum.at/en/collections/film_collection/digital_film_collection (31.12.2019)

3. Principles of Advanced Digitization

Advanced Digitization is a bundle of practices and processes to ensure that digital heritage materials used for research and access is authentic, is suitable for analysis, annotation and interlinking with various media types, conforms to open standards and offers sustainable preservation pathways to ensure its ongoing accessibility.

By digital heritage materials we refer to files created from analog heritage materials ("turned-digital" materials) of a diverse nature: 2-D objects such as paper records, still photographs and related graphic materials, books or other paper records; time-based audiovisual records such as motion picture film, analog and digital video as well as sound recordings captured and stored on physical media. Many of the principles addressed in the attached guidance documents are also valid for dealing with "born-digital" files created by third parties and acquired in the course of the project.

Advanced Digitization is not an end-in-itself but a means to an end. It is to ensure that digital heritage materials are *fit for the purpose of re-use*. To be fit for re-use, a digitized object must be of high quality and sufficient size, with good-quality descriptive metadata in an open-standard format and usage conditions that allow for re-use.[8]

Advanced Digitization aims at creating digital files which

- originate in the most authentic and complete analogue archival film records;
- are created utilizing imaging technology, resolution and bit depth appropriate to retrieve its image and sound content in as much detail as feasible;
- conform to documented industry standards for file formats and wrappers;
- are interoperable, suitable for automated and manual analysis and annotation;
- are suitable for long-term archival preservation.

This is to ensure that

- digital data is authentic (preserves the specific properties of the original analogue asset);
- the necessary supporting information (metadata) to effectively access and preserve the digital objects and record relationships between them is collected or created;
- information needed to interpret a data object is collected or created.

The Advanced Digitization principles are aligned with the Open Archive Information System (OAIS) model as the most suitable model for long-term preservation and

---

[8] We hereby refer to the criteria articulated in the Europeana Whitepaper "Transforming the World with Culture" (2015), in particular pp. 6.
https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Presidencies%20White%20Paper.pdf (31.12.2019)

accessibility of digital materials. OIAS is an ISO Standard (ISO 16363) which sets the framework a digital archive needs to operate in to guarantee long-term access to data.

The following "Obligatory Responsibilities" defined in the OAIS Model are particularly addressed in the referenced guidance documents:

- Acquisitions (digitization and procuring of "turned-digital" materials) must be actively managed and documented in great detail from the point of entry.
- The party responsible for the digitization process must be able to control management of the materials (for instance, no copy-protection or encryption, and the use of open formats and standards).
- The party responsible for the digitization must have clearly documented collecting policies and data collection strategies: for what purpose and for which type of usage is the data acquired, transformed and stored?
- All occurrences of data processing and data transformation need to be clearly documented in order to be retraceable (documentation) to ensure that users of the archive can easily access data in adequate and unaltered form.

4. The VHH Advanced Digitization Tools: Inventory

As a tool kit, this deliverable is not presented as a single, narrative document, but as a collection of corresponding documents. Departing from the overarching principles established in *Tool #1 Principles & Manual*, users can select the relevant tools for dealing with Advanced Digitization of specific media and document formats.

- **Guideline** documents offer comprehensive background information and best-practice considerations for specific common types of heritage materials scheduled for digitization.
- **Checklist** documents serve as a "back-pocket brief" to help researcher in the field to establish basic facts about digitization processes when dealing directly with third-party content providers.

Each document has its own introduction, table of contents and its own pagination. However, the entire set has a second continuous "VHH ADTK" pagination in the upper left corner of each page. The page numbers provided in the following inventory list refer to these numbers:

- Tool #1 *Principles & Manual* | p. 1-8.
- Tool #2 *Checklist Archival Research* | p. 9-14.
- Tool #3 *Checklist Memorial Research* | p. 15-20.
- Tool #4 *Checklist Holocaust Photos Research* | p. 21-27.
- Tool #5 *Guideline Film Digitization* | p. 28-44.
- Tool #6 *Checklist Film Digitization* | p. 45-51.
- Tool #7 *Guideline Film Access Copies* | p. 52-68.
- Tool #8 *Guideline Text and Image Digitization* | p. 69-92.
- Tool #9 *Checklist Text and Image Digitization* | p. 93-101.
- Tool #10 *Guideline Citavi* | p. 102-111.
- Tool #11 *Quality Control Form Text and Image Digitization* | p. 112-115.

**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #2: Checklist Archival Research

Lead-beneficiary (Tool #2)	*LBI*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*LBI (Ingo Zechner, Sema Colpan, Jakob Zenzmaier)*
Contributor(s)	*JLU (Ulrike Koppermann)*
Reviewer(s)	*OFM (Michael Loebenstein)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.6, 2019-12-31*

VISUAL HISTORY OF THE HOLOCAUST
Rethinking Curation in the Digital Age

Tool #2: Checklist Archival Research

Table of Contents

Introduction

Tool #2 *Checklist Archival Research* is intended for the preparation, organization and execution of on-site research in archives. It is a living document to be used together with the following tools:

- Tool #3 Checklist Memorial Research
- Tool #4 Checklist Holocaust Photos Research
- Tool #10 Guideline Citavi

Research steps

1. Create/use a cloud-based repository

- Create/use a cloud-based file storage system (the Cloud Storage) [in VHH & EDU OneDrive is used]
- Create/use a cloud-based reference database (the Reference Database) [in VHH & EDU the reference management and knowledge organization software Citavi[1] is used]

2. Collect and check secondary literature

- Collect secondary literature
  - Check the Reference Database [VHH/EDU: Citavi] for secondary literature already in the Reference Database [VHH: Citavi]
  - Check the Cloud Storage [VHH/EDU: OneDrive] for additional literature to be added to the Reference Database [VHH/EDU: Citavi]
  - Search for additional literature to be added to the Cloud Storage/the Reference Database [VHH/EDU: OneDrive/Citavi]
- Collect archival references
  - Check archival references already available in the Reference Database [VHH/EDU: Citavi]
  - Check secondary literature (footnotes, appendix) in the Reference Database [VHH/EDU: Citavi] for additional archival references
  - Collect archival references in the Reference Database [VHH/EDU: Citavi] by applying relevant categories

3. Ask memorial sites for relevant documents/holdings

- Create list of memorial sites to be contacted
- Check websites of memorials for information on archival documents held by memorial site
  - Add results to the Cloud Storage/the Reference Database [VHH/EDU: OneDrive/Citavi]

---

[1] Cf. Tool #10 *Guideline Citavi.*

- Send letter asking for information on relevant archival documents (textual documents incl. both originals and copies) plus metadata and conditions of digitization and usage
  - Add answers to the Cloud Storage/the Reference Database [VHH/EDU: OneDrive/Citavi]

4. Create list of relevant archives

- Generate and extract List of Archives from the Reference Database [VHH/EDU: Citavi] and add to the Cloud Storage [VHH/EDU: OneDrive]

5. Collect and check finding aids

- Collect finding aids
  - Check the Reference Database [VHH/EDU: Citavi] for finding aids already in the Reference Database [VHH/EDU: Citavi]
  - Check websites of archives for additional finding aids (database entries and printed finding aids)
    - download database entries (by using export function or create PDF of entry)
    - download printed finding aids and create OCR versions
    - add downloads to the Cloud Storage/the Reference Database [VHH/EDU: OneDrive/Citavi]
- Collect archival references
  - Check finding aids in the Reference Database [VHH/EDU: Citavi] for additional archival references
  - Collect archival references in the Reference Database [VHH/EDU: Citavi] by applying relevant categories

6. Create list of relevant holdings

- Generate and extract List of Holdings per archive from the Reference Database [VHH/EDU: Citavi] and add to the Cloud Storage [VHH/EDU: OneDrive]

7. Check access conditions of archives

- Check websites for researcher registration conditions, directions and opening hours, pulling times, any limitations on the number of pulled items, digitization

conditions (see Advanced Digitization Guide), rights of usage, contact persons (for digitization, for rights of usage and archivists for specific holdings)

- o Add information to List of Archives

8. Document research on-site

- Check research room for additional finding aids (databases, printed finding aids, card catalogues/indices that are only accessible onsite)
  - o Digitize additional finding aids
  - o Add to the Cloud Storage/the Reference Database [VHH/EDU: OneDrive/Citavi]
- Use List of Holdings for documenting research status:
  - o date pulled, date checked, date refiled
  - o to digitize, not to digitize (including short note on reason), digitized: completely or partially (if partially, add reason)
  - o digitization obstacles
  - o digitization Standard applied
  - o related metadata such as cover of folder, label of box, list of content (if applicable)

9. Check holdings for further relevant holdings

- Add information on holdings to the Reference Database [VHH/EDU: Citavi] and repeat relevant steps

VISUAL HISTORY OF THE HOLOCAUST
Rethinking Curation
in the Digital Age

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #3: Checklist Memorial Research

Lead-beneficiary (Tool #3)	*JLU*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*JLU (Ulrike Koppermann)*
Contributor(s)	*LBI (Ingo Zechner)*
Reviewer(s)	*OFM (Michael Loebenstein)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.7, 2019-12-31*

VISUAL HISTORY OF THE HOLOCAUST
Rethinking Curation in the Digital Age

Tool #3: Checklist Memorial Research

Table of Contents

Introduction

Tool #3 *Checklist Memorial Research* is intended for the preparation, organization and execution of research in memorials. It is a living document to be used together with the following tools:

- Tool #2 Checklist Archival Research
- Tool #4 Checklist Holocaust Photos Research
- Tool #10 Guideline Citavi

1. Research focus/Material relevant to the project

1.1. Footage

- Check OneDrive for footage taken at respective camp/subcamp.
- If footage is extant review footage, and create an index of:
  - Location, buildings, people (allies, medics, survivors, captured SS-men, civilians, official visitors, etc.), events, measures taken by allies, etc.
  - Open questions regarding unidentified locations, buildings, people etc.

1.2. Sources

In order to contextualize the indices created in 1.1, the following sources are required:

- Oral History Interviews (audio and video)
- Personal correspondence
- Diaries/personal reports/autobiographies
- Photographs
- Official/administrative records/correspondence by liberators, SS, locals or survivors describing the 'evacuation', the 'liberation', conditions at liberated camp, measures taken at liberated camp, or referring to topics/indices in other ways.

Particularly interesting are documents referring to acts of filming and photography by allied troops or other parties present on-site:

- Trial records referring to the use of film footage and photographs as evidence/exhibits in court
- Local and international press clippings reporting on the 'liberation' and allied operations or on the use of footage/photos in court

Focal point of our research are camps in which footage was taken. In addition to the main camp, this may include research on subcamps. Subcamps in which no footage was taken are not to be neglected entirely, as photographs may nevertheless exist.

2. Before contacting

- Check website for information on archival documents held by memorial site.
- Check website for online search tools (database, catalogue, etc.).
- If documents are available online, add entries to Citavi.
- Check Citavi for entries on documents held by the memorial's archive.

3. When contacting

3.1. Search Tools

- Ask for search tools available on-site (database, catalogue, etc.) or via mail (excel charts, finding aids, etc.).

3.2. Material relevant to the project

- Provide memorial staff with concise research questions based on your preparatory work (step 1.-2.).

3.3. Archival provenance/conditions of use

- Clarify whether documents are copies originating from originals in other source archives or originals held by the memorial.
- If copies: ask for source archives and original signatures.
- Ask for material/digital format: manuscripts, typoscripts, microfilm, copies originating from microfilms, scans (*.jpeg, *.pdf)…
- Ask for conditions of digitization.
- Ask for terms of use.

3.4. Cooperation

- Clarify who's doing what: What groundwork/research can be done by members of staff at the memorial and in what timeframe? What research tasks require VHH-researcher's visit?
- Ask for digital material that may even be shared prior to research visit.

4. Research on-site

- Browse through material provided by archivists, select relevant material.
- Continue research with search tools available on-site.
- Determine what documents originally held by other source archives would need to be digitized off-site / in other archive locations.
- Digitize relevant documents according to VHH-standards or determine documents to be digitized by members of staff at the memorial.

- Check-out the exhibition with regard to the representation of the camp's/subcamps' 'liberation'. If permitted, take photographs for documentation.

5. Follow-up

5.1. Documentation

- Add entries and link documents to Citavi.
- Add documents to OneDrive.

**VISUAL
HISTORY
OF THE
HOLOCAUST**

Rethinking Curation
in the Digital Age

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #4: Checklist Holocaust Photos Research

Lead-beneficiary (Tool #4)	*JLU*
Work Package No. and Title	*WP2 Curation and advanced digitization*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitization standards*
Task Leader	*OFM*
Main Author(s)	*JLU (Ulrike Koppermann)*
Contributor(s)	*LBI (Ingo Zechner)*
Reviewer(s)	*OFM (Michael Loebenstein)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.3, 2019-12-31*

Tool #4: Checklist Holocaust Photos Research

Table of Contents

Introduction

Tool #4 *Checklist Holocaust Photos Research* is intended for the preparation, organization and execution of research on Holocaust-related photographs. It is a living document to be used together with the following tools:

- Tool #2 *Checklist Archival Research*
- Tool #3 *Checklist Memorial Research*

General remarks

- Unlike most textual records photographic records often exist in multiple copies (hard copies and digital copies) in a number of different collections. In other words, several archives may hold the 'same' photograph. In many cases, 'secondary archives' have systematically obtained copies from the 'source archives'.
- In some cases, the 'source archive' may already have obtained the photograph as a copy from another 'source archive' or from a private source.
- In some cases it may be difficult to determine the source because duplicates have spread for decades.
- It may be true that there is no 'original' of a photograph unless the negative survived. However, the quality of each copy may differ even if it has been generated directly from the negative. Moreover, with each generation of reproduction a decrease of quality is to be expected.
- A print may have turned into an original in its own right if its verso has been used for annotations or if it has been included in a photo series or a photo album. Such contextual information and other metadata often get lost with the next generation of reproduction even if it may be crucial for the identification and interpretation of the photographic record.
- Advanced Digitization of photographs aims at
  - identifying the original/s or – if the original/s is/are lost – the best available reproduction/s of a photograph;
  - determining and collecting all relevant contextual information and other metadata;
  - determining whether existing digital reproductions may of sufficient quality;
  - generating new digital reproductions if necessary.

- Metadata provided in the archives' catalogs is of varying detail and accuracy, depending on the information the archives received when obtaining the photographs and research they may have done themselves.
- Depending on the legal situation and terms of usage, the accessibility of the 'same' photos and relating metadata may differ between archives.

**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

Research steps and options

1. Where to begin

- With regard to photographs relating to the Holocaust, the United States Holocaust Memorial Museum (USHMM) comprises the most extensive, searchable and accessible collection and online catalog.
- In addition to photos genuinely held by USHMM Collections, the USHMM Photo Archives holds vast numbers of reproductions of photographs from other 'source archives'.
- Therefore, it is only reasonable to begin your research with a centralized approach and to make the USHMM online catalog your point of departure rather than to consult the presumed source archives individually as it is usually done for textual records.

2. Conducting photo research using the USHMM online catalog

- The USHMM catalog allows researchers to combine their own search terms with structured criteria provided within the catalog: e.g. digital availability, dates, holding institution, type of material (AV, photos, text, etc.), key topics (particular camps, ghettos, events, etc.).
- Thus, depending on your research question and interest, the USHMM catalog enables you to conduct a broad or a rather specific and filtered search.
- Note that the search criterion "holding institution" (archives from which the USHMM obtained the copy) can only be applied to AV records but not to photo research. However, you can choose the name of a particular source archive as your search term, e.g. YIVO, and limit your search to "photographs". Thus your search results comprise all digital copies of photos the USHMM received from YIVO.
- If you didn't include an archive's name in your search term but conducted a search for subject matter, topic or location, information on the provenance and source archive are provided below each photo in the fields "photo credit" and "photo source".
- To avoid duplication of work, please keep track of your search terms and combinations of search terms.

3. Exporting metadata from the USHMM catalog

- If the number of your relevant search results remains within a reasonable scope, you can export the metadata yourself by creating a PDF of the entry.

VISUAL
HISTORY
OF THE
HOLOCAUST
Rethinking Curation
in the Digital Age

- In rare cases and if the number of your relevant search results exceeds a reasonable scope, you may ask a USHMM photo expert to run a search for you based on your search criteria. Members of staff have access to an internal database.

4. Researching photos originating from archives other than the USHMM

- The USHMM online catalog and internal database may either already satisfy your research interest or provide the basis for an in-depth research at 'source archives'.
- For research at 'source archives' refer to Tool #2 *Checklist Archival Research*.

5. Researching photos genuinely held by the USHMM

- Photographs genuinely held by USHMM Collections are attributed to specific collections.
- If a finding aid is available for such collection, download the PDF and save it to the Cloud Storage used in your project. [VHH: OneDrive]
- Please note that not all collections have yet been digitized.
- When preparing for your next visit, remember to request all analog collections you wish to see in advance.

6. Obtaining high resolution scans from the USHMM

- High resolution scans can be requested from the USHMM by clicking the button 'order photo' and submitting a personalized request.
- After your request has been reviewed, you'll receive a TIFF file of the photo and text-file containing metadata.
- Obtaining a digital copy of photos genuinely held by the USHMM is currently free of charge. However, this may differ for photos from other source archives. For those photos, special permissions from 'source archives' may be required in addition. Members of staff handling your request will advise.

7. Quality Control

- Refer to Tool #2 *Checklist Text and Image Digitization* to perform quality control of existing digital copies.

8. Better check twice: additional online catalogs

- Even though the USHMM catalog is generally the most convenient and extensive database, some specialized archives have compiled even more metadata on certain photographs, e.g. the online photo archive of the Buchenwald and Mittelbau-Dora Memorial provides more detailed information on the relevant liberation photos than the USHMM.
- Yad Vashem maintains a vast digital collection which includes photos origination from other 'source archives'. Even though the user interface and search tools differ, it may be consulted in a similar way as described for the USHMM catalog.

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #5: Guideline Film Digitization

Lead-beneficiary (Tool #5)	*OFM*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*OFM (Claudio Santancini, Michael Loebenstein)*
Contributor(s)	*-*
Reviewer(s)	*LBI (Ingo Zechner)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.6, 2019-12-31*

**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

Tool #5: Guideline Film Digitization

Table of Contents

1. Introduction

Tool #5 *Guideline Film Digitization* is intended as a guiding document to inform the Advanced Digitization of film materials in film heritage institutions (FHIs) or other collecting institutions. It represents the results of the "Advanced Digitization Workshop" held in Washington D.C. on March 20–22, 2019 and hosted by the United States Holocaust Memorial Museum.

The purpose of this document is to

- inform the selection of the most suitable technical elements for digitization;
- guide the selection of the most appropriate digitization workflow;
- guide the creation of appropriate documentation.

This tool is a living document to be used together with the following tools:

- Tool #6 *Checklist Film Digitization*
- Tool #7 *Guideline Film Access Copies*

2. General principles

Do not re-scan (unless necessary)

Before engaging in any digitization ascertain whether suitable digital elements already exist and can be (easily) accessed.

Originals first

The digitization process should never damage or endanger the original analog elements.

Different films, different solutions

There is no 'one size fits all' solution when it comes to (film) digitization. Each new digitization project should be carefully evaluated in order to get the best results, considering:

- the source element(s)
- the imaging technologies available
- the available resources and level of commitment of all the players involved
- the final aim of the project (e.g. simple online access rather than long term digital archiving)

Start from the earliest generation

Although this is not always true, in most cases earlier generations of film elements carry the most information for digitization (the best and/or most complete image and/or sound representation) and should therefore be prioritized for digitization.

As a rule of thumb we recommend selecting suitable elements according to this hierarchy:[1]

1. Original camera negative (35mm/16mm) / reversal original (16mm and other small gauge formats)
2. Interpositive (col) / dupe positive (b&w) *[note: may also be called 'lavender', 'intermediate', 'fine grain']*
3. Positive print / projection print (1st generation)
4. Internegative (col) / dupe negative (b&w)
5. Positive print (2nd and further generations and optically printed reduction prints)

---

[1] Cf. https://www.fiafnet.org/pages/E-Resources/Technical-Terms.html (FIAF Glossary of Technical Terms) for all technical terms. (31.12.2019)

Overscan

Whenever possible the digitization should truthfully capture the physical properties of the source elements. The area around the image carries information that – despite the fact that it might simply be discarded when creating access copies – is valuable in at least two contexts:

- *Restoration.* A digital reproduction of physical elements like splices or perforations makes the digital post-processing of image and sound easier (or, sometimes, possible altogether).
- *Preservation.* When accessed in the future, our digital elements must provide as much information as possible not only about the contents of the source elements themselves but also about their entire physical history and appearance at the moment of the digitization.

Avoid restoration[2]

The digital capture should always remain faithful to the source. Avoid processing as much as possible at capture (scanning) stage. "Restored" or otherwise processed versions can only be accepted if also the raw, unprocessed scans are provided and if a complete and clear documentation of the process is made available.

Documentation

Your digitization project must entail the production of detailed documentation stating:

- how and why the source element was chosen for digitization;
- when, where, and by whom the film was digitized;
- what kind of preparation/processing the film went through before and/or after the scanning process;
- technical details (in a human readable form) about the process including equipment used;
- photographic documentation and a written report on any damage inflicted on the film material during the process;
- a list (and/or database entry) of all digital elements created;
- checksums/hashes for all digital elements monitored at every stage of file copying or movement.

---

[2] Restoration, in the words of John Ruskin: "[...] a destruction accompanied with false description of the thing destroyed". John Ruskin, The Seven Lamps of Architecture (1880).

VISUAL
HISTORY
OF THE
HOLOCAUST
Rethinking Curation
in the Digital Age

3. Technical requirements

There are some important issues to be mindful of:[3]

Frame rate

The films capture frame rate must be maintained. That means that every digital still frame or video frame must represent one frame of the physical film. No digital interpolation or re-timing is acceptable. For access purposes this can be done later and only if an untouched copy is secured.

Aspect ratio

Cropping or other adjustments of the image geometry must be avoided while scanning. The exposed image area (width of the camera gate) should always be captured in its entirety.

What flavor of 2K or 4K

When choosing the resolution there is another aspect to consider: some contemporary standard resolutions include the formulas "2K" and "4K" but they refer to image geometries much wider compared to the common analogue cinema aspect ratios (see "Resolution").

The "K" generally stands for the number of pixels (in thousands) of the horizontal resolution. To name an example, the 2K DCI Flat resolution (a very common standard for digital cinema projection) has an aspect ratio of 1.85:1 with a horizontal resolution of 1998 pixels. The resulting vertical resolution is 1080 pixels. Nesting a 4:3 frame – the standard aspect ratio for silent cinema – into a 2K Flat container can only be achieved by means of cropping and/or padding the image, resulting in an image size of 1440x1080 pixels (smaller than Full HD video). Although this could be acceptable for Access Copies it cannot be considered sufficient for preservation.

A scanning resolution of 2K must refer to a container whose ca 2000 horizontal pixels are entirely dedicated to the description of the digitized image and whose vertical dimension is calculated according to the aspect ratio of the frame. In the case of a 4:3 film frame this would result in a resolution of 2048x1536 pixels.

---

[3] For a definition of technical terms see chapter 6, Glossary of Common Digitization Terms.

4. Digitization standards

4.1. Gold Standard

Resolution

- 4K overscan for 35mm elements.
- 2K overscan for 16mm and small gauge (9.5mm / 8mm / S-8mm)

Note: The overscan must include at least part of the perforations but it would be better to represent the film edge to edge (full overscan).

Bit depth

- 10/12bit for log encoding
- 16bit for linear encoding

File format

- DPX or TIFF sequences for vision
- PCM/WAV (48/96kHz 24bit) for sound

Processing

If the source is heavily scratched and the scanning facility offers this possibility (and if the process doesn't harm the original in any way) it is recommended to use a wet gate.

4.2. Silver Standard

Resolution

- 2K overscan for 35mm elements
- 2k overscan or HD (1080p vertical resolution) overscan for 16mm elements

Bit depth

- 10bit for log encoding, 12/16bit for linear encoding for DPX/TIFF
- 10bit minimum for video streams

File format

- DPX or TIFF sequences for vision
- PCM/WAV (48kHz 16bit) for sound

- alternative video file formats such as Apple ProRes, DNxHD, FFV1 or MJ2 acceptable as long as they
  - employ 4:4:4 color space (color footage) or 4:2:2 color space (b&w)
  - are either uncompressed or employ a lossless compression algorithm

Processing

See Gold Standard above.

4.3. Minimum Accepted Standard

Only to be accepted in consultation with project management [in VHH: Deputy Project Coordinator] and on a case-by-case basis.

Resolution

- 2K cropped
- HD (1080p vertical resolution) cropped

Bit depth

- 10bit minimum for video streams

File format

- video file formats such as Apple ProRes, DNxHD, FFV1 or MJ2
- visually lossless" encoded video (if such "lossy" compression is the only possibility, we advise to select the highest possible bitrate as recommended by the vendor)

4.4. Unacceptable

as master materials are:

Resolution

- SD (PAL / NTSC)
- 720p

Bit depth

- 8bit

File format

- files encoded using temporal and/or lossy compression such as MPEG-2, MPEG-4, DV / DVCAM and other legacy vendor-specific codecs

5. Naming conventions

No single naming convention for digital films fits the needs of all potential users and clients. However, there is a mandatory set of conventions that should always be followed:

- Agree on your naming convention prior to digitizing to avoid unpleasant surprises and document it.
- Once a standard has been chosen, ensure it is applied consistently.
- Try to extend your naming convention beyond single files and consider using it for a consistent and orderly organization of your filing system (e.g. a uniform tree-like structure of properly named folders and subfolders).
- Files must include at least

  - a unique identifier to reference the film work represented
  - information on the source of the scan
  - sequential numbering with at least six digits (f.e. _000001.tiff etc.)

- Avoid verbose and/or semantic file names. The file name should include essential information for file identification, but it must not be descriptive (as well as excessively long).
- Information about the file's contents should be provided as separate metadata linked through a clear unique identifier.
- For ensuring the possibility of cross-platform sharing of the data and for long-term preservation purposes blank spaces (' '), symbols or special characters like German umlauts ('ä', 'ö', 'ü'), asterisks ('*'), slash or backslash ('/', '\'), etc. must be avoided.
- Only the use of letters, numbers, dashes ('-') and underscores ('_') is permitted.

6. Minimum documentation convention

The digitization must always be well documented.[4] All the elements constituting the documentation – like scanner logs, descriptive documents, the scanner operators' report, photographic documentation, technical details about the project and its outputs, other metadata, etc. – should be:

- delivered as separate digital files linkable to the project and/or to the specific film element(s) that they describe (f.e. through a unique identifier)
- compliant to the naming convention chosen for the scans
- organized, if possible, in a logical structure that allows a human reviewer to understand its contents
- not necessarily restricted to a specific file format (e.g. *.txt rather than *.pdf for documents, *.json rather than *.xml for logs or *.jpeg rather than *.tif for pictures).

It is however highly recommended to vet the different possibilities in order to maximize compatibility, interoperability and long-term accessibility.

---

[4] Cf. chapter 2, "Documentation" section.

7. Checksumming

It is strongly recommended to ask for and/or provide hash codes for every digital object produced during the digitization project.

The most widely used checksumming algorithm for digital film and video is md5, but other algorithms can be accepted.

In the case of AV streams the best choice is the combination of one file-based checksum file and one frame-based checksum like FFmpeg's framemd5.

8. Delivery standards

Unfortunately, there is no easy and universal solution for a truly cross-platform, hassle-free data sharing.

In the case of digital film, the high storage requirements render virtually impossible cloud and/or network-based data transfer, limiting the choice to the use of physical storage devices. This is nevertheless a multiform choice and can involve a broad range of different (and, sometimes, competing) technologies, formats and standards. None of these suits the requirements of film digitization better than all the others.

The best approach to this problem is to ensure during the preliminary stages of a digitization project that the physical media that will be used for the delivery of the digitized elements (e.g. external HDDs, SSDs, magnetic data tapes, etc.) and their hardware/software set-up (e.g. the file system, the connection interface, etc.) can be correctly managed by everyone involved in the project itself, from the scanner operator to the final user.

VISUAL
HISTORY
OF THE
HOLOCAUST
Rethinking Curation
in the Digital Age

9. Glossary of common digitization terms

Aspect ratio

Defines the geometry of the film image captured by the camera and/or projected on the screen. It's generally expressed with two numbers separated by a column (":"), like for example 1.33:1, where the first number refers to the horizontal dimension (or resolution in the digital domain) and the second to the vertical.

Bit depth

The amount of information stored for each pixel of a digital image. In most of the standards used for digital film/video the pixel is divided into three different channels (e.g. RGB or X'Y'V'). The bit depth generally refers to the number of bits dedicated to each of these channels (e.g. a 10bit RGB image uses 10 bits for the red channel, 10 bits for the green channel and 10 bits for the blue channel, for a total of 30 bits for each pixel). The bit depth determines the number of possible light/color values of a digital image.

Bitrate

The quantity of digital information stored for a digital film/video in a given amount of time. It is often referred to as the number of bits in a second (or, more often, a multiple of bits like Mb, Gb, etc.) and it determines the storage space needed for storing the file. Although there is not necessarily a direct connection between bitrate and quality, in the case of compressed images a higher number often results in better images.

Compression

The process adopted to reduce the size of a file while maintaining as much as possible of its visual and auditive properties. Some forms of compression allow for a full recovery of the original file ("lossless compression"), others discard parts of the images spatial and/or temporal information ("lossy compression").

File format

A file format is a standard used for encoding digital data. In the case of digital film and video it might refer to two different things:

- *The wrapper or container.* This contains the image and sound streams along with metadata and information needed for the correct reproduction of the streams. Common examples are *.mov, *.mp4, *.avi, *.wav or *.mkv.
- *The codec.* The standard used for encoding/decoding the streams. Common examples are Apple ProRes, h264, ffv1, DNxHD, Jpeg2000, etc.

For film digitization the most common formats are DPX and TIFF. These formats store each frame in a separate file, resulting in a sequence of files ("wrapped" by nesting it in a folder).

Frame rate

The speed at which motion picture film is exposed in the camera and/or played back. It is expressed as the number of frames to be reproduced in a second (e.g. 24fps).

Resolution

The number of pixels in a digital image. The resolution is commonly expressed as "number of horizontal pixels x number of vertical pixels" (e.g. 1440x1080) but there are common names such as "Full HD" or "2K" that represent standard resolutions.

Common resolutions for digital film are:

Common Name	Aspect Ratio	Dimensions
HD	1.77:1	1280x720
HD 4:3	1.33:1	960x720
Full HD	1.77:1	1920x1080
Full HD 4:3	1.33:1	1440x1080
DCI 2K - Flat	1.85:1	1998x1080
DCI 2K - Scope	2.39:1	2048x858
2K	1.33:1	2048x1536
Ultra HD (4K UHD or UHD)	1.77:1	3840x2160
DCI 4K - Flat	1.85:1	3996x2160
DCI 4K - Scope	2.39:1	4096x1716
4K	1.33:1	4096x3072

Table 1 Common resolutions

10. References

Austrian Film Museum: Guidelines for the Digital Film Collection (English version):
https://www.filmmuseum.at/en/collections/film_collection/digital_film_collection
(31.12.2019)

FIAF's "Choosing a Film Scanner":
https://www.fiafnet.org/images/tinyUpload/E-Resources/Commission-And-PIP-Resources/TC_resources/Choosing%20a%20Scanner%20v1.1.pdf (31.12.2019)

FADGI's "Digitizing Motion Picture Film":
http://www.digitizationguidelines.gov/guidelines/FilmScan_PWS-SOW_20160418.pdf
(31.12.2019)

IASA's "Guidelines on the Production and Preservation of Digital Audio Objects" (web edition):
https://www.iasa-web.org/tc04/audio-preservation (31.12.2019)

Cintel's "Digital Film Supplement":
http://web.archive.org/web/20060314020114/http:/www.quantel.com/resource.nsf/Files/Digital_Film/$FILE/Digital_Film.pdf (31.12.2019)

ARRI's "Companion to Digital Intermediate":
http://dicomp.arri.de/digital/digital_systems/DIcompanion/index.html (31.12.2019)

Memoriav's recommendations for digital archiving:
https://memoriav.ch/wp-content/uploads/2017/09/Empfehlungen_Digitale-Archivierung_Version1.1_EN.pdf (31.12.2019)

DIN SPEC 15587:
https://www.beuth.de/de/technische-regel/din-spec-15587/29906732 (31.12.2019)

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #6: Checklist Film Digitization

Lead-beneficiary (Tool #6)	*OFM*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*OFM (Claudio Santancini, Michael Loebenstein)*
Contributor(s)	
Reviewer(s)	*LBI (Ingo Zechner)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.7, 2019-12-31*

**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

Tool #6: Checklist Film Digitization

Table of Contents

Introduction

Tool #6 *Checklist Film Digitization* is intended as an aid for the digitization of relevant primary film materials in film heritage institutions (FHIs) or other collecting institutions.

The purpose of this document is to

- inform the selection of the most suitable technical elements for digitization;
- help establish the digitization capabilities of the relevant archive content providers / collecting institutions;
- guide the selection of the most appropriate digitization workflow;
- provide basic quality management tools to aid the decision-making process.

This tool is a living document to be used together with the following tools:

- Tool #5 *Guideline Film Digitization*
- Tool #7 *Guideline Film Access Copies*

1. Research steps

1.1. Create a registry of relevant FHIs

- Ensure list is updated regularly and discoverable to all relevant parties.

1.2. Check finding aids

- Aggregate existing finding aids so they are discoverable to all relevant parties.
  - o Check for finding aids already in your project's reference database [in VHH/EDU: Citavi].
  - o Check websites of FHIs for additional finding aids (database entries and printed finding aids):
    - Download database entries (by using export function or create PDF of entry).
    - Download printed finding aids.
    - Add downloads to the relevant research repository.

1.3. Create list of relevant holdings per FHI

- Create a priority list ("pull list") per FHI and check lists for redundant entries.

1.4. Check access conditions of archives

- Check publicly available information for digitization of film content.
- Enquire specific conditions via email or by telephone.

2. Digitization steps

2.1. Are the required film elements already available in a digital format?

- If yes: check compatibility with selected quality standards.
- If no: proceed to next item.

2.2. Is digitization performed in-house or outsourced?

- If in-house: proceed to next item.
- If outsourced: can digitization be performed by a lab chosen by the project [in VHH/EDU: OFM] / can physical film materials be shipped to Austria and/or other EU countries?
  - If yes: proceed with enquiring conditions of movement of cultural goods.
  - If no: enquire about recommended/vetted vendors.

2.3. What scanning technologies are available?

- Telecine or scanning?
- If Telecine: see 'Minimum Acceptable Standard' procedures in Advanced digitization Guideline.
- If scanning is possible check for:
  - Resolution available (see 'Gold standard' to 'Acceptable Standard' in Advanced digitization Guideline)
  - Overscan yes/no
  - Sound yes/no
  - Bit depth
  - File formats
  - Delivery methods

2.4. What film elements are most suitable for digitization?

- Select best materials according to
  - earliest generation (see hierarchy in Advanced digitization Guideline)
  - best image quality (see condition report or review original)
  - best overall condition (see condition report or review original)

2.5. Select best digitization process based on selected film elements and FHI / vendor capabilities and budget

- Gold Standard
- Silver Standard
- Minimum Acceptable Standard

2.6. Establish quality control procedures

- To assess the integrity and the quality of the scan request if the following is possible to obtain/perform:
  - checksums (request a checksum from provider)
  - technical metadata (request a technical report and/or log file)
  - documentation of the process (request a written report)
  - visual inspection of the raw scans (review of the unprocessed files at provider facilities)

At least a "spot-check" should be performed. A compressed, if possible losslessly compressed file ("proxy") may be used for QC if a review of the raw scan files is not possible due to technical constraints.

2.7. Establish documentation standard and procedures

- Make sure that every step of the process is represented in the documentation as described in the Advanced Digitization Guideline. In particular ensure that the following is provided:
  - Provider-specific information on the element(s) used for digitization:
    - ID (catalog number/item number etc.) of the element(s) used
    - Information on the parentage of the element(s) used ("is positive print from negative XY"; "is dupe neg of element YZ" etc.)
  - written report on the physical preparation of the film elements including any alterations or repair performed (technical restoration).
  - written report on the digitization process, including operator responsible, date, equipment used
  - scanner logs or equipment/software specific metadata created in the process
  - a written report on any alterations, image or sound processing performed in the digital domain

2.8. Check delivery procedures

- Agree on delivery methods before the beginning of the project.
- Agree on file naming convention before the beginning of the project.
- Ensure readability and compatibility of the storage/delivery media and technologies chosen between all parties involved, e.g.
  - is the filesystem readable/writable on all the platforms used?
  - does the delivery method require special software to access the files?
- If checksums are provided with the scans: monitor file integrity upon delivery and at any new file copy/transfer.
- If not: calculate checksums upon delivery at OFM and before any new file copy/transfer.

**VISUAL
HISTORY
OF THE
HOLOCAUST**

Rethinking Curation
in the Digital Age

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #7: Guideline Film Access Copies

Lead-beneficiary (Tool #7)	*OFM*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*OFM (Claudio Santancini, Michael Loebenstein)*
Contributor(s)	*-*
Reviewer(s)	*LBI (Ingo Zechner)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.5, 2019-12-31*

**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

Tool #7: Guideline Film Access Copies

Table of Contents

1. Introduction

Tool #7 *Guideline Film Access Copies* is intended as an aid for the creation of access copies of digitized film materials in film heritage institutions (FHIs), other collecting institutions or research and educational organizations.

The purpose of this document is to

- inform the selection of the most suitable technical elements for the creation of access copies;
- guide the selection of the most appropriate production workflow;
- provide some template settings for commonly used software as a reference.

This tool is a living document to be used together with the following tools:

- Tool #5 *Guideline Film Digitization*
- Tool #6 *Checklist Film Digitization*

2. General principles and caveats

2.1. Introduction and terminology

"There is no 'one size fits all' solution when it comes to (film) digitization". This general principle of advanced digitization of motion picture film is even more valid when discussing the creation of access copies, amplified by the diversity of potential delivery scenarios, use cases, platforms, file formats and containers. Instead of an exhaustive examination of the topic it is therefore intended as a reference tool aimed at helping users and content providers alike to maintain consistency and quality in the production and in the procurement of access copies.

What do we mean by 'access copies' of audiovisual material? In our understanding these are digital files produced from master or intermediate quality files that employ file format conversion and various means of file size reduction to allow for their delivery and/or use for end-users. This is particularly important in the context of the VHH project as digital film master files are per best practice stored at high resolution in uncompressed or mathematically lossless compression formats like DPX, TIFF or FFV1 which are formats generally unsuitable for real-time network delivery and playback on consumer equipment. Moreover, in some cases access copies will have to be produced from files originating in consumer formats such as DVD(-R) or broadcast recordings, introducing yet another level of degradation of the visual and aural properties of the source materials.

The use cases for access copies range from visually lossless "mezzanine" files for automated analysis to heavily (lossy) compressed streaming-ready files for preview and online delivery. For every single case the client and the operator need to decide on the most suitable workflow and output considering the available means and resources, the specific demands of the task and of all the players involved in it: from the provider the source materials to the operator producing the copies, to their final users.

The document is organized in three sections. The first one focuses on the context and philological hierarchies of source elements for producing access copies. The third section offers concrete examples of workflows, while the middle section relates to technical aspects of file conversion.

2.2. File history and hierarchy

When discussing the production of access copies, we take for granted that we are producing a new derivative element of an already existing, higher quality digital element. The Working Papers "Advanced Digitization Guideline" and "Archival Research Guideline: Film Digitization" offer an overview of the production of these master files, and the underlying hierarchy of elements.

**VISUAL
HISTORY
OF THE
HOLOCAUST**

Rethinking Curation
in the Digital Age

For the production of access copies we broadly identify four different categories of digital source elements. In brackets we refer to the terminology proposed by the German DIN Spec 15587 ("Empfehlungen zur Digitalisierung von kinematografischem Film")[1]

- *Raw scans (Raw-Scan Package).* These are the files that are generally produced by the software controlling the film scanner or acquisition system. The term "raw" here doesn't refer to the file format but rather to the fact that the files at this stage remain "untouched", that is free of any digital processing that might change their audio/visual properties other than the ones decided by the operators in preparation for or during the digitization. Not recommended unless the access copies should reflect the unprocessed source film.

- *Master files (Processing Package).* The final output of a digital post-production chain (including here restorations and other digital manufacturing applied following archival requirements) are often one or more elements which represent its final and definitive versions. These are the files used for long-term storage and for the production of derivatives. Master elements may come in the form of separate image and sound files, as folders of single frame images or as consolidated video streams. Recommended.

- *Mezzanine (or "intermediate") files (Distribution package).* These files are sometimes referred to as "mezzanine files" and represent high-quality derivatives off the master files produced in the course of restoration or post-production processes. Intermediates usually are stored as consolidated movie files in a common production format. In case of silent film frame rates (<24fps) intermediate files for some purposes may be re-timed to allow for realistic motion playback on end-user devices. They are often compressed (incl. lossless compression) but preserve most of the audiovisual properties of the scanned film sufficiently to serve as the source element for creating access copies. Recommended **only** if the pre-processing is not an issue or even desirable in your use case.

- *Deliverables (Delivery package).* These are the files produced for sharing/dissemination employing file size reduction and/or loss of quality. Not recommended as a source unless stipulated by the use scenario (f.e. an access copy of a DVD recording, see next point).

---

[1] https://www.beuth.de/en/technical-rule/din-spec-15587/299067326 (31.12.2019)

VISUAL
HISTORY
OF THE
HOLOCAUST
Rethinking Curation
in the Digital Age

2.3. A special case: DVD discs

Producing access copies from DVDs can be considered a special case because they are a complex format[2]. A standard DVD contains at least three different kinds of files, organized in a standard file system:

- VOB files (Video Objects). These files contain the MPEG-2 audio and video streams
- IFO (Information). These are instructions for the correct reproduction of the disc.
- BUP (Backup). Backups of the IFOs

When extracting content of a DVD disc (a process known as "ripping"), a decision needs to be made whether one wants to extract only the AV content or the entire disc structure (including all the chapters, extras and menus).

In the first case what is needed is one or more of the VOB objects that can be found in the "VIDEO_TS" folder inside the disc. Once identified they can be transcoded to the desired target format. Thereby one needs to consider that the video included in DVDs already conforms to a common standard (MPEG-2) so:

- The content might have undergone re-timing, scaling, cropping and other editing and/or conforming processes. It is advisable not to apply further modifications unless very carefully evaluated or necessary in order to fit a specific format convention employed by the end user.
- MPEG-2 is a highly compressed format. Re-encoding these files with high bitrate profiles will only increase the file size with no gain in quality.
- DVDs employ interlacing. It is advisable to employ deinterlacing when producing access copies. This process however can introduce digital artifacts and should only be applied with caution and with sufficient documentation as temporal information is discarded or altered.

A number of software tools exist for the purpose of extracting and transcoding the contents of a DVD[3]. These applications offer a broad range of options and presets for locating AV streams in a disc and for copying them.

When extracting AV elements from of a DVD is not enough and all of the original contents must be retained in the new copy (including menu functionality), a disc image (ISO) should be created. For this purpose, best practice guides exist and should be consulted[4].

---

[2] "An Optical Media Preservation Strategy for New York University's Fales Library & Special Collections", Annie Schweikert 2018. Available here:
https://archive.nyu.edu/bitstream/2451/43877/2/Schweikert_OpticalMediaPreservationNYU_2018.pdf (20.10.2019)

[3] To name one of the best, at least among the open source options, we can suggest HandBrake. See:
https://handbrake.fr/ (20.10.2019)

[4] For example see the Library of Congress format description at
https://www.loc.gov/preservation/digital/formats/fdd/fdd000348.shtml (20.10.2019)

VISUAL
HISTORY
OF THE
HOLOCAUST
Rethinking Curation
in the Digital Age

3. Production steps

3.1. Select the most suitable source material (if you have a choice)

- If the film was digitized for VHH: do you need an access copy of the raw materials? → choose "raw scans" as source.
- If the film was digitized for VHH: do you need an access copy of the processed and consolidated film file → choose "mezzanine", "master" or "delivery" files as source.
- If working from a pre-existing digital file and/or working with a third-party supplier

→ enquire about the highest-quality source available (resolution in pixels; frame rate; bit rate, color space) and request encoding according to VHH standard templates provided in Section 3.

3.2. Establish image geometry (aspect ratio)

As a general rule access copies must faithfully represent the image geometry (aspect ratio) of the original film work they are derived from. Some codecs and file format, video player software implementations and applications as well as hardware devices only allow for the presentation of video in a limited number of aspect ratios. These include in particular 16:9 (1,77:1) and 4:3 (1,33:1); in the case of Digital Cinema (DCI-compliant) a limited set of wide-screen ratios is supported, but silent film full aperture is not.

To faithfully represent the original's intended viewing experience the video content can be nested into a 4:3 or 16:9 composition, thus preserving the original's intended viewing ratio with black bars on top and bottom or pillar-boxing (black pillars left and right) respectively. This operation must be documented and communicated to end users to avoid re-zooming or false anamorphic distortion by the end user (for example interpreting a 1440 x 1080 academy-ratio file erroneously as anamorphic 1920 x 1080 video).

Care must also be taken when transcoding DVD disc content (PAL, and in particular NTSC) to correct the non-square pixels to proper 4:3 aspect ratio (768x576).

3.3. Cropping and masking

While "surplus" information acquired during the scanning stage – including the overscan area outside the camera-gate, film edge or perforation – is crucial information to capture and preserve it may be less relevant in other use cases (such as presentations or cinema projection). In these cases cropping will yield desirable results. It is achieved by discarding the areas of the image that were not meant to be visible by an end-user or audience viewing the film in its historical context (in projection, or at a viewing table). To emulate the original experience of seeing a film during projection we recommend cropping the size of the projection mask to an image that is 4-5% smaller than the camera-gate.

Cropping or masking (masking pixels with black bars) can also be performed to exclude unwanted noise (f.e. lines and fields used to transport non-visible data in SD video) on the edges of the visible image. See also letterboxing and pillar-boxing (see above).

3.4. Establish frame rate

As with image geometry the choice of frame rates for access copies is limited by industry conventions, established technical standards and deliverable specifications. While motion picture films before the advent of sound cinema (and in cases like silent small gauge film well after that point) were shot and presented with a variety of fixed or variable frame rates, digital video and digital cinema implementations dictate a limited set of frame rates. Many applications and device-specific implementations, including the Digital Cinema Initiative (DCI) standard do not foresee frame rates below the sound cinema convention of 24fps.

As a first step one must therefore establish:

- if the film was digitized for VHH: what is the intended / most plausible frame rate to ensure a period-authentic and realistic visual impression / experience of the film work;
- whether the target file format, platform and/or playback device allows for variable frame rate playback.

As a general rule, the VHH-MMSI player is specified to allow for 'frame-rate agnostic' video playback, allowing for the manual setting of the desired playback speed.

Other target platforms and devices may however not permit this. Conforming a source file from a silent-speed film work to a higher frame rate by changing the playback speed will alter the films duration and aural and visual properties. To avoided this distortion, and to ensure that the access copy of the film work can be experienced in its intended, or historically most accurate representation, the video can be digitally processed to simulate the desired playback frame rate.

The most common technique is image interpolation or re-timing. This is achieved by digitally creating new frames (through interpolation or duplication of existing frames) or by dropping excess frames in order to achieve the target frame rate. However, this process introduces artifacts (blurring and jitter) and creates artificial new images which do not correspond to any frames in the underlying film item. Furthermore, most software packages offer only a limited amount of control over where, and how frames are inserted or interpolated. It is therefore advisable to not use 'dark' (undocumented) software implementations in common encoding applications but to use dedicated post-production or effects software applications (see list of recommended software in Section 3).

As such these techniques are not advisable for any usage scenario involving frame-accurate analysis (automated or manual alike) or annotation and description.

Re-timed access copies are of value in the following scenarios:

- access copies for playback in standard format where hardware and software protocols specify a limited set of frame rates, in particular DVD, Blu-ray and DCI (DCP) formats
- access copies for use in screenings or other presentation forms where the approximation of realistic, natural motion and period- and medium-correct representation has priority over frame-accurate analysis

3.5. Select compression method

While lossy compression is discouraged for the production of film scans and master files, it is essential when producing access copies. Files will have to be transferred via the internet and played back on a variety of platforms and devices and this requires smaller file sizes. Another reason is compatibility: almost all current AV formats compatible with common internet protocols, browsers and end-user devices use compression. Additionally – although this is less of a problem with technological advancements and the dropping hardware prices – compression can be applied in order to reduce storage space requirements of both the content providers and the users.

Not all compression methods are alike. While it is out of scope of this Guideline to cover all the possible technologies to compress audio and video, we need to consider two central aspects: the type of compression and the ratio (amount) of compression.

The first choice can be made when choosing the file format (codecs and containers) and the encoding parameters. If it is possible to choose between different formats and parameters, as a rule of thumb, lossless compression is to be preferred over lossy compression and intraframe (spatial) is to be preferred over interframe (temporal) compression. This is however not to be taken as a condition for approving or refusing a copy.

Regarding the amount of compression, our yardstick should be the target bitrate of our output file. The bitrate, for AV materials generally expressed in Kbps or Mbps, gives us a measure of how much digital information is used for representing a section (read: a portion of time) of our video file. Considering that we have three "fixed" elements, resolution, bit depth and frame rate, we can assume that a lower bitrate implies a higher compression (and therefore a smaller file but worse quality of the image and sound). This relation between bitrate and quality can be very helpful when deciding our encoding parameters but it must be evaluated very carefully. Different compression algorithms can yield extremely different results with the same bitrate. It is therefore strongly suggested to run preliminary encoding tests with different compression levels (lower/higher bitrate parameters) in order to evaluate which renders the best visual and/or auditive characteristics while still fitting the storage restrictions.

3.6. Select codec and output file format

When deciding on an output file format with a view to access, the main goal should be to ensure maximum compatibility by verifying the technical requirements prior to encoding, in order to ensure that the output file will be useful to the target users. For this Guideline we can limit ourselves to introducing a few of the most commonly used formats for delivery and access:

- MPEG-2. Although this is an old format it is still of common use, being the standard codec for image and sound in DVD discs and older digital broadcast technologies.
- H.264 (aka Advanced Video Coding) and H.265 (aka High Efficiency Video Coding). These formats are among the most common video codecs today. Their applications range from digital television to on-line diffusion and Blu-ray Discs. Although their specifications allow for high bitrate visually lossless encoding, they are mostly used in contexts in which a good balance between image fidelity and small file size is required.
- Apple ProRes and Avid DNxHD. Compared to the previously mentioned formats, they offer a better image and sound quality but require higher bitrates, thus making them non-applicable for streaming or online progressive download viewing. Their main application is AV editing and post/production, both formats are very common in the production and post-production sector, including archival footage services, and are used for producing mezzanine files for further encoding of access copies.
- JPEG2000. This coded has a big variety of applications but it is of interest to us for its implementation into the DCP (Digital Cinema Package) format. This is the standard format used in theatrical (digital) projection.
- AAC. This sound codec is the successor in the market of the MP3 and is often coupled with H.264 videos. Its good quality, despite it uses a lossy compression, contributed to its wide adoption.
- PCM Audio. This format represents an exception to what we previously wrote about compression for access copies. Thanks to its comparatively low native bitrate, sound recordings are often be stored and disseminated uncompressed, allowing for high quality access copies.

All the above-mentioned output codecs define the way the audiovisual information is transformed, stored in a file (encoded) and interpreted and rendered by playback applications / devices (decoded). The codec alone doesn't complete the description of the format; the encoded audio and video streams are grouped together and stored in what is called a "container" (or "wrapper"). Different containers allow for a different organization of the file: they define which sound and/or image codec can be accepted, if other kinds of data streams can also be stored alongside image and sound (e.g. timecodes or subtitles), if multi-chapter and/or multi-version features can be used and so on. The container also stores important technical and descriptive metadata.

Common AV containers relevant to the project are:

- MOV and AVI (.mov and .avi). These are two different formats grouped together here despite huge differences in structure, compatibility and documentation. They however share many features and are often used in similar contexts. Both are proprietary formats and can wrap a variety of different codecs. They are very common in the context of AV production and post-production and often found in combination with Apple ProRes (.mov) and DnxHD (.avi);
- MPEG-4 and Matroska (.mp4 and .mkv). These are the two most common open formats and regarding many features equivalent to the proprietary ones mentioned above. They are also very common and allow for a flexible choice of codecs. While MPEG-4 is an established and well-documented standard and is used for a plethora of applications and use cases Matroska is gaining ground as an open container format, in particular in the context of wrapping losslessly compressed open codecs like FFV1.

3.7. Establish quality control procedures

- To assess the integrity and the quality of the access file produced perform:
  - verification of the compliance of the file format using a metadata inspector such as MediaInfo or ffprobe[5].
  - visual inspection of the access file, ideally on the target platform
  - at least a "spot-check" should be performed

3.8. Establish documentation standard and procedures

- Make sure that every step of the process is represented in the documentation as described in the Advanced Digitization Guideline. In particular ensure that the following is provided:
  - information on the element(s) used for digitization
  - written report on the conversion process, including operator responsible, date, software used
  - if applicable: any software specific metadata (logs etc.) created in the process
  - ensure that file naming conventions are followed.

---

[5] Cf. https://mediaarea.net/en/MediaInfo (31.12.2019), https://ffmpeg.org/ffprobe.html (31.12.2019)

VISUAL
HISTORY
OF THE
HOLOCAUST
Rethinking Curation
in the Digital Age

4. Common software applications and settings

4.1. Choosing a software application

When dealing with the choice of a software application there are a few very practical points to consider:

- the experience of the user;
- the scope of the project. The production of access copies could describe a straightforward task such as batch-transcoding from one format (or medium) to another as well as a more elaborated process which includes AV-editing, color grading, image and/or sound filtering and operations such as scaling, cropping or de-interlacing, frame interpolation, overlaying of watermarks, etc;
- the existing software/hardware set-up. Some tools might be optimized for a specific platform or require hardware components without which the performance might be affected (or the whole process might not work at all);
- the available budget. Software licenses for AV post-processing tools can be rather expensive and this should be carefully evaluated. Many of the features offered by professional software can be achieved using free and open source solutions.

Depending on the features they offer and on the tasks they are optimized for, we can distinguish three groups of software applications commonly used for the production of access copies:

- Video Editing/post-production tools. This category is mainly composed of professional "industry-oriented" solutions. Most common and most flexible solutions that offer a good balance between usability for entry level operators and richness of features are:
  - o Apple Final Cut Pro
  - o Avid Media Composer
  - o Adobe Premiere
  - o Adobe After Effects
  - o DaVinci Resolve.

In this category we also want to mention the video editor Shotcut[6]: this software is completely open source and multi-platform, customizable and able to handle the vast majority of AV file formats.

- AV Transcoding Tools. These applications are optimized for encoding/decoding of AV files. Common examples are:
  - o Adobe Media Encoder

---

[6] https://shotcut.org/ (31.12.2019)

o Apple Compressor
o HandBrake
o FFMpeg.

FFMpeg in particular is an open source tool that can transcode to/from a large number of different formats. It does not include a graphical user interface and therefore requires experience with command line operations (e.g. Unix and Unix-like shells, Windows cmd and PowerShell and so on). It is however very popular with the archival community and many resources are available to help understanding and using it[7].

A very specific kind of access copies are DCPs – Digital Cinema Packages, the common industry-standard format for distributing digital films to DCI-compliant installations and venues. Although many of the professional applications mentioned offer the possibility to export digital cinema compliant files, we'd like to mention DCP-o-matic[8]. This tool is completely free and open source and can be easily integrated in professional or semi-professional workflows.

- Quality checking tools and AV-players. These programs are not directly used in the production of the access copies but are important for their assessment and evaluation (as well as diffusion!). The landscape of media players is too vast to list in this document but there are at least two applications worth mentioning: VLC media player[9] and mpv[10]. These are both feature-rich and stable, open source and cross-platform solutions that allow to reproduce virtually any AV file format. Another important step for the quality control of access copies, after the visual check, is the metadata inspection. Dedicated software solutions for this task exist, f.e. ffprobe (also part of the FFMpeg project) and MediaArea's MediaInfo[11]. MediaArea also created MediaConch, a tool for conformance checking against a predetermined policy, also realized through metadata extraction and analysis[12].

---

[7] Two very useful resources are: AMIA's *ffmprovisr* (https://amiaopensource.github.io/ffmprovisr/) and Reto Kromer's *FFmpeg Cookbook for Archivists* (https://avpres.net/FFmpeg/) (31.12.2019)
[8] https://dcpomatic.com/ (31.12.2019)
[9] https://www.videolan.org/vlc/ (31.12.2019)
[10] https://mpv.io/ (31.12.2019)
[11] https://mediaarea.net/en/MediaInfo (31.12.2019)
[12] https://mediaarea.net/MediaConch (31.12.2019)

5. Selected templates for Common access copy delivery formats

In this section we would like to present three real-world examples which should cover some of the most common scenarios occurring when dealing with the production of access copies.

The first example is the conversion of a sequence of raw images produced by a film scanner into a mezzanine file (silent).

The second example is an example of compression into an easily shareable output format.

The last scenario is an example of DVD ripping, with deinterlacing.

All the examples are accompanied by ffmpeg recipes.

5.1. Mezzanine file template

Very often we have to deal with big (in terms of both length and storage space) image sequences produced by film scanners which are not easy to share because of software compatibility and high storage and computing power requirements. A good approach is to encode a mezzanine file from these sources, which can then be used for evaluation and/or diffusion. In this example we will show a possible workflow for the production of a FullHD Apple ProRes 422-HQ starting from a DPX, 10bit sequence. The software used will be Adobe Premiere, which allows to natively encode "legal" ProRes files also on Windows platforms. The metadata will be extracted using MediaInfo.

- Input Metadata:
  - Format: DPX
  - Video format: DPX (Raw), Resolution: 2048x1536 (Display Aspect Ratio: 4:3), Bit depth: 10 bit, Scan type: progressive
    - After importing the DPX sequence in our project ('File->Import...') we make sure that it was correctly interpreted by the software (right click on the clip and 'Modify->Interpret Footage...') and then we create a new sequence containing the clip (right click on the clip and 'New Sequence From Clip').
    - C. If everything worked until now (correct framerate, progressive scan type and resolution) we can already proceed with the export. Right click on the sequence and 'Export Media...'.
    - D. In the 'Export Settings' window we have to select our output settings. In 'Export Settings' (the tab on the upper right side) we start by selecting 'QuickTime' under 'Format:' and the path and filename of the output under 'Output Name'. Here we can also uncheck the 'Export Audio' checkbutton (unless, of course, we want to export an audio stream as well).
    - E. In the tab below, under 'Video', we select the 'Video Codec' 'Apple ProRes 422 HQ'. Under 'Width' and 'Height' we select 1920 and 1080 respectively. Here we can also make sure that framerate, scan type and pixel aspect ratio

of the output will match the source by selecting the checkbuttons on the right side of 'Frame Rate', 'Field Order' and 'Aspect'.

- F. Now we can queue the job for batch processing or directly start the render by clicking on the 'Export' button at the bottom.
- G. Output Metadata:

○ Format: MPEG-4

○ Video format: ProRes (422 HQ), Resolution: 1920x1080 (our image size is actually 1440x1080 but we added 240 lines of black pixels on the left and on the right side), Frame rate: 24fps, Scan type: Progressive.

The FFMpeg recipe for this example is:

ffmpeg -f image2 -r 24 -i INPUT -c:v prores_ks -profile:v 3 -pix_fmt yuv422p10le -filter:v "scale=1440:1080,pad=1920:1080:240:0" OUTPUT

'*ffmpeg*' starts the program.

'*-f image2 -r 24*' are the input parameters and inform the software that the source file is an image sequence and its speed is 24fps.

'*INPUT*' must be replaced with the path and filename of the source. In this case this will be recognized as a sequence if the filenames contain a numerical sequence (e.g. from 'image_0000.dpx' to 'image_0250.dpx'). The numerical sequence must be replaced by the escape character '*%*' followed by the number of integers the sequence is composed of and by '*d*' (e.g. 'image_%04d.dpx').

'*-c:v prores_ks -profile:v 3 -pix_fmt yuv422p10le*' are the output parameters. They select the video coded ProRes (with the profile 422-HQ).

'*-filter:v "scale=1440:1080,pad=1920:1080:240:0"*' is the filter chain that creates a 1920x1080 image container (Full HD resolution) and scales the original 2K, 4:3 image down to the resolution 1440x1080 (positioned at coordinates x=240 and y=0).

'*OUTPUT*' must be replaced with the desired target path and filename (with .mov extension).

5.2. Quick output file template

The second example presents the typical situation where a small file must be encoded for reference and fast sharing. This can be achieved by compressing the source and scaling down to a lower resolution. Let's start from the previously encoded ProRes video and produce an HD h264/.mp4 that can easily be uploaded and shared online. For this example, we will be using DaVinci Resolve.

- A. Once a new project is created (with resolution 1440x1080 and 24fps) we can locate and import into the Media Pool (in the 'Media' tab) our source file. In the

project settings we also need to make sure that the 'Image Scaling' parameters are set to 'Scale full frame with crop' for both 'Input Scaling->Mismatched resolution files' and 'Output Scaling->Mismatched resolution files'.

- B. We will now create a new timeline by right-clicking the imported file and then clicking on 'Create New Timeline Using Selected Clips...'.
- C. If no further processing needs to be done (e.g. watermarking, titles, editing, etc.) we can already export our reference file. This is done in the 'Deliver' tab. Here we can select the settings for video codec, audio codec and container.
- D. In the 'Video' tab we select 'MP4' and 'H.264' for 'Format' and 'Codec'. We then set a custom resolution of 960x720 (4:3, HD). We set the 'Quality' encoder parameter to 'Automatic' and 'High'.
- E. For this example – a silent film – we uncheck the 'Export Audio' checkbutton in the 'Audio' tab.
- F. In the 'File' tab we select our destination path and filename.
- G. Now we can click on 'Add to Render Queue' and start encoding.
- H. Output Metadata:

  - Format: MPEG-4
  - Video format: Advanced Video Codec (h264), Resolution: 960x720 (Display Aspect Ratio 4:3), Frame rate: 24fps, Scan type: Progressive

The FFMpeg recipe for this example is:

ffmpeg -i INPUT -an -c:v libx264 -qp 18 -pix_fmt yuv420p -filter:v "crop=1440:1080,scale=960:720" OUTPUT

'INPUT' and *'OUTPUT'* will be replaced with the source and target paths and filenames (the output has a .mp4 extension).

'-an' makes the output silent (no audio stream will be generated).

'-c:v libx264 -qp 18 -pix_fmt yuv420p' are the output parameters. They select the h264 video codec with a 4:2:0 chroma subsampling (better for compatibility with different media players) and a 'visually lossless' compression ratio.

'-filter:v "crop=1440:1080,scale=960:720"' scales the resolution down to 4:3 HD.

5.3. Output file from DVD-Video template

This example is for the extraction of an access file from a (PAL) DVD-Video. A very powerful tool for ripping DVD-Video discs is HandBrake. This is the software used for this example.

- Input Metadata:

o Format: MPEG-PS (or ES)
o Video format: MPEG Video (Version 2), Resolution: 720x576 (Display Aspect Ratio 4:3), Frame rate: 25fps, Scan type: Interlaced (TFF)
o Audio format: ac3 (Dolby Digital), Bit rate: 256kbps, Sample rate: 48kHz, Channels: 2 (LR).

- Once ingested our input (from the folder containing the source .VOB files or the disc directly) we located in the 'Title' drop-down menu the element or chapter we want to transcode (or the range of chapters if it's more than one).
- In the 'Dimensions' tab we set a 720 pixels width and a 576 height (uncheck 'Keep Aspect Ratio'). The 'Anamorphic' setting must be set to 'Custom' and the 'Display Width' to 768 pixels.
- In the 'Filters' tab we only activate the de-interlacing filter setting 'Interlacing Detection' on 'Default' and 'Deinterlace' on 'Yadif'.
- In the 'Video' tab we select the 'H.264 (x264)' encoder and a 'Framerate (fps)' of 25 ('Constant Framerate'). We also set the 'Constant Quality' parameter on the right side on 18.
- In the 'Audio' tab we set the 'Codec' to 'AAC (avcodec)' and the 'Bitrate' to 256 (kbps).
- On the bottom-right side of the main window we click on 'Browse' and then select the desired destination path and filename. In the dropdown menu 'Save as type' we select the extension 'mkv'. Now we can add the file to the Queue and start the render.
- Output Metadata:

o Format: Matroska
o Video format: Advanced Video Codec (h264), Resolution: 720x576 (Display Aspect Ratio 4:3), Frame rate: 25fps, Scan type: Progressive
o Audio format: AAC LC, Sampling rate: 48kHz, Channels: 2 (LR).

The FFMPeg recipe for this example is:

ffmpeg -i INPUT -c:a aac -c:v libx264 -qp 18 -pix_fmt YUV420p -aspect 4:3 -filter:v "yadif" OUTPUT

-'*INPUT*' and '*OUTPUT*' will have to be replaced with the path and filename of the source .VOB file and the path and filename of the desired output (with .mkv extension) respectively.

-The tag '-*aspect 4:3*' will automatically set the pixel aspect ratio to the value needed for the target display aspect ratio of 4:4 (in this case 16:15).

-The tag '-*filter:v "yadif"*' will de-interlace the video.

**VISUAL
HISTORY
OF THE
HOLOCAUST**

Rethinking Curation
in the Digital Age

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #8: Guideline Text and Image Digitization

Lead-beneficiary (Tool #8)	*LBI*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*LBI (Vrääth Öhner)*
Contributor(s)	*LBI (Ingo Zechner, Joachim Schätz) JLU (Ulrike Koppermann)*
Reviewer(s)	*OFM (Michael Loebenstein)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.15, 2019-12-31*

**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

Tool #8: Guideline Text and Image Digitization

Table of Contents

1. Introduction

Tool #8 *Guideline Text and Image Digitization* is intended as a reference document for the preparation, organization and execution of digitization of text documents and images. It is a living document to be used together with the following tools:

- Tool #9 *Checklist Text and Image Digitization*

Both tools are based on results of the "Advanced Digitization Workshop" held in Washington D.C. on March 20–22, 2019 and hosted by the United States Holocaust Memorial Museum.

There are separate tools for the Advanced Digitization of film:

- Tool #5 *Guideline Film Digitization*
- Tool #6 *Checklist Film Digitization*

For research on relevant materials refer to the following tools:

- Tool #2 *Checklist Archival Research*
- Tool #3 *Checklist Memorial Research*
- Tool #4 *Checklist Holocaust Photos Research*
- Tool #10 *Guideline Citavi*

Advanced Digitization is here understood as the entire process from digital imaging to Optical Character Recognition (OCR), including the collection and digitization of relevant metadata.

This Guideline does not aim at perfect results but rather at the best possible results under specific circumstances that are often far from being ideal. Decisions between multiple options should always balance different factors such as image quality, effort, conservatory aspects and others.

VISUAL
HISTORY
OF THE
HOLOCAUST
Rethinking Curation
in the Digital Age

2. General principles

Do not re-scan (unless necessary)

Before engaging in any digitization ascertain whether suitable digital elements already exist and can be (easily) accessed.[1]

Originals first

The digitization process should never damage or endanger the original analog elements.

Different texts and images, different solutions

There is no 'one size fits all' solution when it comes to the digitization of texts and images. Each new digitization project should be carefully evaluated in order to get the best results, considering:

- the source element(s) (in case multiple versions of diverging quality exist)
- the imaging technologies available
- the available resources and level of commitment of all the players involved
- the final aim of the project (e.g. simple online access rather than long term digital archiving)

Overscan

To determine the edges of an object the digital image should not crop the edges of the object captured.

Avoid manipulation of masters

The digital capture should always remain faithful to the source. Avoid processing as much as possible at capture (scanning) stage.

Document decisions

Decisions taken during the digitization process should be well documented.

---

[1] In order to determine if existing digital copies meet digitization standards applied in the VHH-project, please consult the Quality Control Form.

3. Digitization standards

3.1. Archive Standards

Archive Standards for the digitization of Bound Volumes, Unbound Documents and Prints and Photographs listed below follow the FADGI (Federal Agencies Digital Guidelines Initiative) Rating expressed in their Technical Guidelines for Digitizing Cultural Heritage Materials (2016). They are listed for reference purposes and should be consulted from the perspective of determining a Functional Standard (cf. next chapter).

Bound Volumes

	1 Star	2 Star	3 Star	4 Star
Master file format	TIFF, JPEG 2000, PDF/A	TIFF, JPEG 2000, PDF/A	TIFF, JPEG 2000, PDF/A	TIFF, JPEG 2000, PDF/A
Access file formats	All	All	All	All
Resolution	150 ppi	300 ppi	300 ppi	400 ppi
Bit depth	8	8	8 or 16	8 or 16
Color space	Grey Gamma 2.2 SRGB	Grey Gamma 2.2 SRGB Adobe 1998 ProPhoto ECIRBGv2	Grey Gamma 2.2 SRGB Adobe 1998 ProPhoto ECIRBGv2	Grey Gamma 2.2 SRGB Adobe 1998 ProPhoto ECIRBGv2
Color	Greyscale or Color	Greyscale or Color	Greyscale or Color	Color

Table 1 Archive Standard for the Digitization of Bound Volumes

Unbound Documents (including Manuscripts)

		2 Star	3 Star	4 Star
Master file format		TIFF, JPEG 2000, PDF/A	TIFF, JPEG 2000, PDF/A	TIFF, JPEG 2000, PDF/A
Access file formats		All	All	All
Resolution		300 ppi	300 ppi	400 ppi
Bit depth		8	8 or 16	8 or 16
Color space		Adobe 1998 ProPhoto ECIRBGv2	Adobe 1998 ProPhoto ECIRBGv2	Adobe 1998 ProPhoto ECIRBGv2
Color		Color	Color	Color

Table 2 Archive Standard for the Digitization of Unbound Volumes

Prints and Photographs

	1 Star	2 Star	3 Star	4 Star
Master file format	TIFF	TIFF	TIFF	TIFF
Access file formats	All	All	All	All
Resolution	150 ppi[2]	300 ppi[3]	400 ppi	600 ppi
Bit depth	8	8	8 or 16	16
Color space	Grey Gamma 2.2 SRGB Adobe 1998 ProPhoto ECIRBGv2	Grey Gamma 2.2 SRGB Adobe 1998 ProPhoto ECIRBGv2	Adobe 1998 ProPhoto ECIRBGv2	Adobe 1998 ProPhoto ECIRBGv2
Color	Greyscale or Color	Greyscale or Color	Color	Color

Table 3 Archive Standard for the Digitization of Prints and Photographs

---

[2] The original FADGI recommendation of 100 ppi has here been replaced with 150 ppi.
[3] The original FADGI recommendation of 200 ppi has here been replaced with 300 ppi.

3.2. Functional Standard

The functional standard is based on an adaption of Archive Standards to the local conditions given. The goal is to create digital copies that can be recognized as error-free as possible in the subsequent OCR process. (This means as few digital artifacts as possible and as much contrast as possible between printed letters and paper.)

Archives and libraries should be able to make digital copies that are at least equivalent to the 2 Star rating listed above. In most libraries and archives, a flatbed scanner (often in combination with a photocopier) should be available. If the library/archive permits to modify the settings of the scanning unit and to store the scans on an external storage medium, effectiveness and quality of digitization will improve significantly. If not, even in the case of a large number of pages in bound volumes, it is essential to use a digital camera (eventually a smartphone) on a stand (eventually also a book cradle).

Using your own equipment, digital cameras with a resolution of at least 18MP on a stand and even smartphones (preferably also on a stand) that are able to capture images in RAW format may be able to achieve a 3 Star rating (see chapter "4. Specific solutions").

In order to decide which option to choose, the required resolution for the most common document formats is listed below.

DIN Formats	in mm	300 ppi	150 ppi	Required Resolution
A3	297 x 420	3508 x 4961	1754 x 2480	18MP / 5MP
A4	210 x 297	2480 x 3508	1240 x 1745	8MP / 2,2MP
A5	148 x 201	1748 x 2480	874 x 1240	4,5MP / 1,1MP

Table 4 Required resolution for the most common document formats

3.3. Minimum Acceptable Standard

The Minimum Acceptable Standard corresponds to the FADGI 1 Star rating for Bound Volumes and the FADGI 2 Star rating for Unbound Documents and for Prints and Photographs.

In case the documents you are copying are printed in the German Fraktur font, experience shows that – independent from the FADGI rating chosen – you will need a resolution of 600 ppi to achieve reliable results in the subsequent OCR-process.

3.4. Non-acceptable

Non-acceptable are all procedures that do not provide the necessary resolution and that do not allow the data to be stored in any of the formats listed above (TIFF, JPEG 2000, PDF/A). This applies to the generation of lossy *.jpg or *.pdf file formats. The delivery of digital reproductions of objects on physical hard copy – for example as a print-out – is also unacceptable.

3.5. How to choose a digitization standard

The basic rule is: The higher the rating, the better. However, that may not always be true if you consider other factors than image quality:

- Existing copies: Is it worthwhile to rescan if 2 Star rated digital copies are already available and the rescan would only achieve a 3 Star rating?
- Effort: How long will it take and how much will it cost to achieve a higher rating?
- Conservatory aspects: Will rescanning or a certain scanning procedure endanger the original materials?

Using the "Preferred Setup" described below for Unbound Documents as well as for Prints and Photographs will most likely deliver better results than a photocopier. (Bound Documents are a different story, because the results depend on how flat the book can be placed.) On the other hand, the required equipment for the Preferred Setup is quite heavy and the process is time-consuming.

For documents and photographs in black and white, a photocopier is usually sufficient to achieve a 3 Star rating. However, documents in color exceed the capacity of photocopiers to achieve more than a 2 Star rating, not because of resolution, but because of bit depth and color space.

4. Specific solutions

4.1. Preferred Setup

The field camera kit recommended by Tom Rieger (Manager, Digitization Services Section, The Library of Congress) comprises:

- a Sony Alpha 6300 digital camera
- a lens free of distortions (the ZEISS Loxia 2.4/25)
- a suitable stand (the Novoflex Magic Studio Repro Stand, which clips to the edge of a table)
- LED copy lights (Viltrox 2 Sets Photography LED Video Light Lamp, Bi-Color 3300K–5600K, HD LCD Display Screen, CRI 95 for DSLR Table Photo Studio)
- Color Target

We recommend adding a laptop computer with Capture One Pro software for tethered shooting to this field camera kit. Moreover, make sure to bring a 50cm ruler in a neutral color that is compatible with the respective archive's requirements.

To operate this preferred setup, the following steps are required:

- Clip the Repro Stand to the edge of the table you are working on. Use the screw in the middle of the bottom plate to secure the connection between stand and table if necessary.
- Name and create new 'sessions' aligned with the records your digitizing, e.g. name session according to the series and box you are digitizing. This ensures that the digital folders created by the software mirror the archival structure and coherence.
- Attach the camera to the round, metal spacer first, then connect the spacer with the stand. Connect the camera to your computer (via USB-C Tether cable) and open the Capture One Pro application on your computer.
- Switch the camera to "Manual" mode (M).
- Turn on the camera and click on the Camera Icon in the Capture One Pro "Session" window.
- Choose "Live-Image" in the "Window" menu. A new window will open up and now you should be able to see what the camera sees on the screen of your computer. You are now also able to control the camera's settings directly in Capture One.
- Set the file format to "RAW" in Capture One or in the camera's File Format Menu.
- Set the ISO (light sensitivity) of the camera to the base value of the camera's sensor (for the Sony Alpha 6300 this value is ISO 100).
- Manually select an aperture of 5.6 or 8 by turning the aperture dial on the lens.
- Set the appropriate shutter speed. To determine which shutter speed is correct, either use the histogram in the "Session" window of Capture One after you have taken a test shot (the histogram should fill the entire spectrum from black to white, i.e. from left to right) or simply match (based on your own visual perception) the

VISUAL
HISTORY
OF THE
HOLOCAUST
Rethinking Curation
in the Digital Age

brightness of the live-image on your computer's screen with the brightness of the object you are about to digitize.

- To determine the correct white balance, try entering different Kelvin values in the "Color Temperature" panel in the "Live-Image" window of Capture One or in the camera's white balance (WB) menu (e.g. 3200 Kelvin for tungsten or 5500 Kelvin for natural light). The correct white balance is the one where the color temperature on the screen is closest to the color temperature of the object. If you use the LED copy lights described above, simply set the camera's white balance to the Kelvin value you have selected for the copy lights (from 3300K to 5600K).

- Align the camera on the stand plane-parallel to the plane where you place your objects. Use the Live-Image window to control the alignment process.

- Align the size of the image area with the largest object that you want to digitize (such as an A4 sheet) and make sure that the image does not crop the object (Overscan).

- Chose "grid" in the XX menu and customize the number of bars/squares and their color. Using the 'grid' tool allows you to align the documents in a right angle. The grid will only be shown when using the software is will not be inscribed in the digital copies.

- Manually focus the lens by turning the wheel on the lens barrel until the image on your computer's screen reaches peak sharpness.

- Be aware that the light conditions in the research room may change during the course of the day. Therefore, the selected Kelvin values, shutter speed and aperture need to be checked and adjusted regularly.

- Now you are all set to take an unlimited number of images. To take an image, just press the "Take Image" button in the "Live-Image" window of the Capture One Pro application.
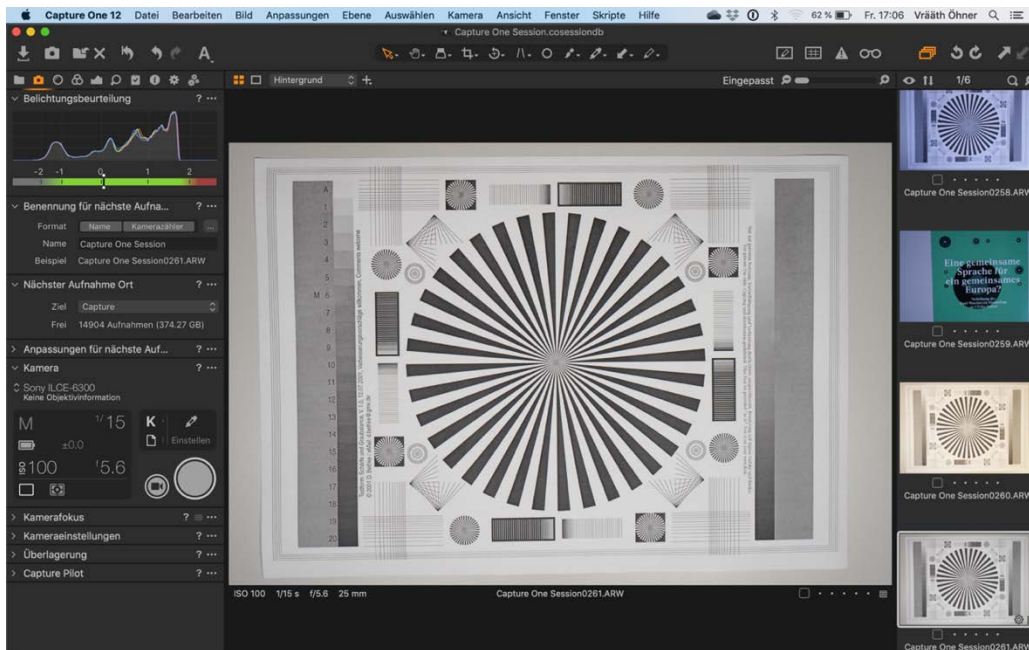
Figure 1 Session window in Capture One Pro. On top left the Histogram of the selected image is displayed. The value indicated on the scale below the Histogram should be close to zero (0).
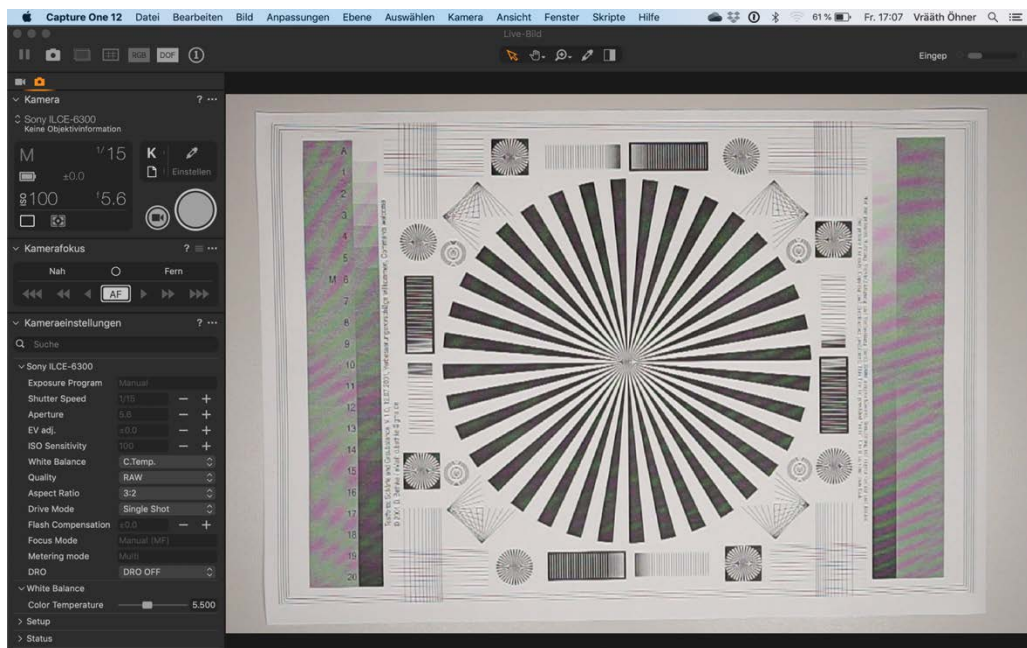

Figure 2 Live-Image window in Capture One. Use this screen to determine brightness and peak sharpness. Don't mind the color moire, it will greatly be reduced in the recorded image. The menu on the left will allow you to control the camera (next page).
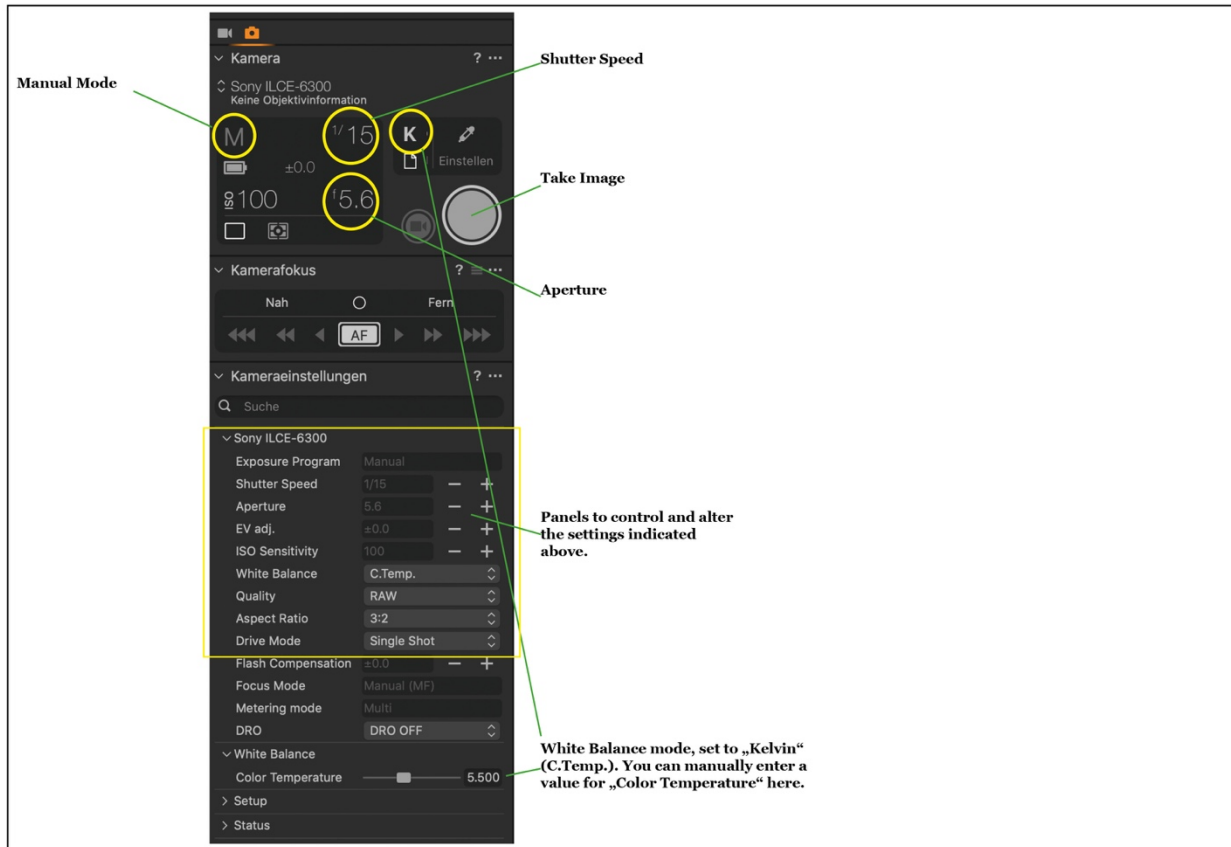
Figure 3 Control menu of the Live-Image window in Capture One

4.2. iPhone/iPad

The Apple iPhone 7 and newer models feature a 12MP resolution sensor and are therefore able to achieve a 3 Star rating for documents smaller than A3 in sufficient light (which allows the use of the base ISO of 25 or up to a maximum of ISO 100).

This is, however, not possible with the standard camera application and the standard recording format for photos (JPEG). The JPEG engine of the iPhone generates pleasing images, but at the same time so many digital artifacts that the recordings fall below the Minimum Acceptable Standard. It is therefore recommended to use the "Pro Camera" app, which allows you to record photos in RAW format (DNG).

To operate this setup, the following steps are required:

- Open the "Pro Camera" Application.

- Go to the "Setup" menu, then choose "RAW" in the "Format" submenu as well as "Manual Mode" in the "Mode" submenu (the latter allows you to manually adjust the brightness of the exposure if necessary).
- Place the iPhone as plane-parallel as possible over the object to be photographed, in such a way that a narrow strip of the object's background remains visible (Overscan).
- Take the photo.

4.3. Android smartphones

For Android smartphones the same applies as for the iPhone. You will need to install the "Adobe Lightroom Photo Editor" app to record photos in RAW format (DNG). The required setup should be very similar to that in the "Pro Camera" app for the iPhone.

4.4. Photocopier

Konica Minolta bizhub C458

This is the photocopier available [for VHH and EDU] at LBI in Vienna.

To operate this setup, the following steps are recommended:

- For bound text materials (e.g. books) select PDF ("Mehrseitendatei", 300 ppi, "Autofarbe").
- For illustrated text materials and archival text documents select TIFF ("Mehrseitendatei", 300 ppi, "Autofarbe"); split TIFFs after scanning 30 pages to avoid large files that may not be transmitted from the photocopier to the server (use Apple's "Preview" ("Vorschau") or Adobe's "Acrobat Pro" to assemble all parts into one complete file).
- When scanning text documents containing German Fraktur font, switch to 600 ppi (to improve OCR results).
- Before scanning bound objects, take the measurements of each object and add the values to the settings of the photocopier (select "Scanformat" | "Ben. d. Format"); measure those parts of the object that will actually be scanned, e.g. the book block ("Buchblock"), not the board ("Buchdeckel"); document measurements in Reference Database (VHH/EDU: Citavi). If scanning loose pages of different sizes, create an individual file for each page (deselect "Mehrseitendatei").

5. Naming Convention

No single Naming Convention for digitized materials fits the needs of all potential users and clients. However, there is a mandatory set of conventions that should always be followed:

- Agree on your Naming Convention prior to digitizing to avoid confusion and document it.
- Once a standard has been chosen, ensure it is applied consistently.
- Try to extend your Naming Convention beyond single files and consider using it for a consistent and orderly organization of your filing system (e.g. a uniform tree-like structure of properly named folders and subfolders).
- Files must include at least[4]

  o a unique identifier to reference the materials represented.
  o information on the source of the scan
  o sequential numbering with at least three digits (e.g. _001.tiff etc.)

- Avoid verbose and/or semantic file names. The file name should include essential information for file identification, but it must not be descriptive (as well as excessively long).
- Information about the file's contents should be provided as separate metadata linked through a clear unique identifier or through a Reference Database [in VHH & EDU: Citavi].
- For ensuring the possibility of cross-platform sharing of the data and for long-term preservation purposes blank spaces (' '), symbols or special characters like German umlauts ('ä', 'ö', 'ü'), asterisks ('*'), slash or backslash ('/', '\'), etc. must be avoided.
- Only the use of letters, numbers, dashes ('-') and underscores ('_') is permitted

---

[4] For example, the identifier may consist of the official abbreviation of the archive, the record group, the series, and the smallest designated archival unit: NARA_RG-111_E1044(A1)_Box-9_Annual-Report-1945_001.

6. Minimum documentation convention

The digitization must always be well documented. All the elements constituting the documentation – descriptive documents (e.g. catalogue entry referring to the record), photographic documentation, technical details about the project and its outputs, other metadata, etc. should be:

- delivered as separate digital files linkable to the project and/or to the specific element(s) that they describe (f.e. through a unique identifier)
- compliant to the Naming Convention chosen for the scans
- organized, if possible, in a logical structure that allows a human reviewer to understand its contents
- not necessarily restricted to a specific file format (e.g. *.txt rather than *.pdf for documents, *.json rather than *.xml for logs or *.jpeg rather than *.tif for pictures)

It is however highly recommended to vet the different possibilities in order to maximize compatibility, interoperability and long-term accessibility.

7. Delivery standards

Unfortunately, there is no easy and universal solution for a truly cross-platform, hassle-free data sharing. If possible, use cloud and/or network-based data transfer instead of physical storage devices [in VHH & EDU: OneDrive].

If physical storage devices are used, make sure that there is always a second copy (preferably kept in a different place) and that external devices are formatted in ExFAT.

8. Postproduction

8.1. Master Copies

Master Copies are the original files produced according to the chosen standard. They are usually stored in one of the Master File Formats (TIFF, JPEG 2000, PDF/A) and may be renamed but should neither be edited nor changed. For production purposes, it is therefore strongly recommended to produce a Production Master by simply copying and renaming the file according to the "Naming conventions".

RAW Files

If the Master Copy is recorded in RAW format, an additional step is necessary. With the exception of DNG (Digital Negative), RAW files are proprietary formats whose future readability cannot be guaranteed. It is therefore necessary to convert the Master RAW files into one of the Master File Formats (TIFF or JPEG 2000) to get an "original" Master file.

To generate a Master from a file that was recorded in RAW format, you need a RAW converter application such as Adobe Lightroom, Capture One Pro, Adobe Photoshop or the like:

- Open the RAW file with the application.
- Export the file in the appropriate format (TIFF or JPEG 2000) without applying any changes.
- Rename the file according to the "Naming conventions".

Production Masters

The Production Master can be created both from the original RAW file or by copying the generated Master file.

Changes made to the Production Master make sense only if they increase the (machine)readability of the recorded document. Such changes include, for example, cropping, aligning, increasing / decreasing brightness or contrast, and setting the correct white balance.

Changes that should be avoided are those that affect the (machine)readability of the document or its further processing. Such changes include excessive sharpening, excessive contrast or brightness adjustment, as well as significant changes to the documents' color balance.

8.2. OCR

Abbyy FineReader desktop application

For standard text use Abbyy FineReader desktop application:

- Choose the option "Convert to PDF" and determine the language(s) used in the document.
- As soon as you open the document in Abbyy FineReader, the automated OCR process will start.
- Check for errors Abbyy FineReader is reporting and also for errors caused by the process itself (e.g. wrong page orientation, wrong division of facing pages etc.).
- Use the tools Abbyy FineReader is offering to correct errors and to fine tune the result.
- Export the document as PDF and/or as PDF/A and save the conversion file as Abbyy FineReader document (*.frdoc) – this will allow you to reassess the file if necessary.

Abbyy FineReader Online

For German Fraktur font use Abbyy FineReader Online:[5]

- Log in and choose "Recognize".
- Upload the file you want to recognize (printed text only; max. 100 MB each; allowed file formats: *.pdf, *.jpg, *.jpeg, *.tif, *.tiff, *.bmp, *.png).
- Make sure that "German (Fraktur font)" is the selected language of your document.
- Select PDF or PDF/A as output format, then start the process. The recognized document will appear on top of the "Documents" list. By simply clicking on the document's name, it will be saved to the selected export destination (usually the location on your hard drive selected for your browser).
- Check for (grave) errors and either try to fix them using Adobe Acrobat Pro or by fine tuning the original file with Adobe Acrobat Pro and redoing the recognition process.

Transkribus

For handwritten text use Transkribus.[6]

---

[5] https://finereaderonline.com/en-us (31.12.2019)
[6] https://transkribus.eu/Transkribus/ (31.12.2019)

8.3. Access Copies

Compression (unless it is lossless) at capture and/or processing stage is generally to be avoided. However, Access Copies created for specific purposes are an exception.

Access Copies are files made available for various kinds of usage with the exception of production purposes. In particular they may be made available through a web platform [in VHH: the VHH-MMSI] or a cloud-based desktop application [in VHH/EDU: Citavi]. Access Copies are created from the Production Master. Here, quality is not the only decisive criterion, but accessibility for users. It is therefore recommended to use only the most common file formats (*.jpg and *.pdf), the most common color space for images (SRGB) and to bring the file size down to a value that guarantees the readability of the document and fast access on older computers or with slow connections or limited storage space [as it is the case in VHH/EDU with Citavi]. This also means that the rating for the Access Copy, regardless of the rating for the Production Master, cannot exceed a 2 Star-rating.

To create an Access Copy for images from the Production Master you need one of the RAW converter applications described above (Adobe Lightroom, Capture One Pro, Adobe Photoshop), for texts it is recommended to use Adobe Acrobat Pro.

Adobe Lightroom or Capture One Pro

- Open the Production Master file with the application.
- Choose "Export" from the "File" menu.
- This will open a new window, where you can select the appropriate settings for the image:
    o storage location
    o file format (JPEG, Quality: 8)
    o color space (SRGB)
    o image size (the exact half size is recommended for the long edge – i.e. 6000px becomes 3000px – and a resolution of at least 150 ppi for large image files).
- Don't apply any sharpening.
- Include the image's metadata (EXIF).
- Select "Export".
- Rename the file by simply replacing the term "Production-Master" with "Access-Copy" and determine the export destination, then press "Save".

Adobe Photoshop

- Open the Production Master file with the application.

- Choose "Image size" from the "Image" menu.
- Enter the new value for the long edge (the exact half size of the existing value), then enter either 150 ppi as the new value for resolution or leave the resolution at 300 ppi.
- Choose "Bicubic sharper" as method for scaling down. Press "OK".
- Now the image should appear noticeably smaller on the screen.
- Choose "Save as" from the "File" menu, rename the file by simply replacing the term "Production-Master" with "Access-Copy" and choose "JPEG" as file format.
- Select the Storage location. Press "Save".
- A new window will open in which the JPEG options can be selected.
- Select "High" (Quality: 8) for "Image options" and "Baseline (Standard)" for "Format options". Press "OK".

Adobe Acrobat Pro

- Open the Production Master file with the application.
- Choose "Save as PDF with reduced size" from the "File" menu.
- Choose the "Adobe 9.0 and higher" option in the dialogue box and click "OK".
- Rename the file by simply replacing the term "Production-Master" with "Access-Copy" and determine the export destination, then press "Save".

8.4. Archival Copies

Archival Copies are files intended for long-term preservation. They are supposed to contain the maximum amount of recorded information and preserve the Master Copy for long-term access. If the Production Master originated from a RAW file, it is advisable to also save the RAW file as an additional Archival Copy (in DNG format).

To create a RAW file in DNG format from the original RAW file, you need a RAW converter application such as Adobe Lightroom or Capture One Pro (RAW into DNG conversion will not work in Adobe Photoshop):

Adobe Lightroom or Capture One Pro

- Open the original RAW file with the application.
- Choose "Export" from the "File" menu.
- This will open a new window, where you can select "DNG" as Image format and "ProPhoto RGB" as Color space in the "File settings" submenu.
- Don't apply any other changes.
- Include the image's metadata (EXIF).
- Select "Export".
- Rename the file by simply adding "Archive-Copy" to the existing file name.

9. Glossary

Bit depth

Bit depth, also known as color depth, is the number of bits used to describe each color or grayscale component of a single pixel. Bit depth is only one aspect of color representation, expressing the granularity with which colors can be expressed: A bit depth of 16 (as in the TIFF or JPEG 2000 file formats) can reproduce variations of grayscale or color more precisely than a bit depth of 8 (as in the JPEG file format). The other aspect is how precisely a range of colors can be expressed (the gamut of a given color space).

ExFAT

Extensible File Allocation Table (ExFAT) is a proprietary file system introduced by Microsoft. ExFAT allows individual files larger than 4 GB and can be used to exchange files between Windows, Mac OS (version 10.6 and later), and Linux operating systems.

Color space

A color space is a specific organization of colors. It allows for reproducible representations of color, in both analog and digital representations. For example, Adobe 1998 RGB and SRGB are two color spaces, both based on the RGB color model. The only difference is that Adobe 1998 RGB (and more so ProPhoto RGB) has a wider gamut than SRGB, which leads to a more precise representation of color.

Gamut

The gamut is the subset of colors which can be accurately represented within a certain color space.

ISO

The ISO value in digital photography indicates a camera sensor's sensitivity to light. The lower the number, the less sensitive to light the image sensor is. ISO is an acronym for International Standards Organization that has issued the standards ISO 6, ISO 2249 and ISO 5800 describing photographic film's sensitivity to light.

OCR

Optical Character Recognition (OCR) is the conversion of digital images of typed, handwritten or printed text into machine-encoded text.

PDF/A

Portable Document Format/Archive (PDF/A) files are an ISO-standardized version of PDF files that are used for the archiving and long-term preservation of electronic documents.

RAW

Raw image files (RAW) contain unprocessed information as captured by the image sensor of a digital camera.

Resolution

The resolution is the pixel density of an image or an image sensor. The pixel density refers to the number of pixels per inch (ppi), both vertically and horizontally.

Tethered shooting

In tethered shooting, the camera is connected to a computer and operated through the computer's user interface.

White balance

In photography and image processing, white balance is the global adjustment of the intensities of the colors (typically red, green, and blue primary colors). An important goal of this adjustment is to render specific colors – particularly neutral colors – correctly. white balance changes the overall mixture of colors in an image and is used for color correction.

10. References

FADGI (Federal Agencies Digital Guidelines Initiative), Technical Guidelines for Digitizing Cultural Heritage Materials. Creation of Raster Image Files (September 2016). http://www.digitizationguidelines.gov/guidelines/digitize-technical.html (17.09.2019)

Library of Congress, Recommended Formats Statement 2019-2020. https://www.loc.gov/preservation/resources/rfs/index.html (17.09.2019)

United States Holocaust Memorial Museum, Basic Guidelines for Digitization (October 2, 2018).

U.S. National Archives and Records Administration (NARA), Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images (June 2004). https://www.archives.gov/preservation/technical/guidelines.html (17.09.2019)

**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #9: Checklist Text and Image Digitization

Lead-beneficiary (Tool #9)	*LBI*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*LBI (Ingo Zechner, Vrääth Öhner, Sema Colpan)*
Contributor(s)	*LBI (Joachim Schätz)* *JLU (Ulrike Koppermann)*
Reviewer(s)	*OFM (Michael Loebenstein)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.13, 2019-12-31*

VISUAL HISTORY OF THE HOLOCAUST
Rethinking Curation
in the Digital Age

Tool #9: Checklist Text and Image Digitization

Table of Contents

Introduction

Tool #9 *Checklist Text and Image Digitization* is intended as an aid for the preparation, organization and execution of digitization of texts and images. It is a living document to be used together with the following tools:

- Tool #8 *Guideline Text and Image Digitization*

The Guideline provides detailed background information and recommendations for many steps addressed in the Checklist.

There are separate tools for the Advanced Digitization of film:

- Tool #5 *Guideline Film Digitization*
- Tool #6 *Checklist Film Digitization*

For research on relevant materials refer to the following tools:

- Tool #2 *Checklist Archival Research*
- Tool #3 *Checklist Memorial Research*
- Tool #4 *Checklist Holocaust Photos Research*
- Tool #10 *Guideline Citavi*

Digitization Steps

1. Determine Advanced Digitization obstacles of relevant documents

- Check relevant documents in the Reference Database [VHH/EDU: Citavi] for digitization obstacles such as bound elements, Fraktur, raster graphics, handwriting…
- Document any digitization obstacles by applying the appropriate term to the entry of relevant documents in the Reference Database [VHH/EDU: Citavi].

2. Determine if digital copies of relevant documents are already available

- Check if digital copies are
  - publicly available online
  - publicly available onsite only
  - not publicly available though already created (and if they can be made available to the project)

3. Check the quality of digital copies that are already available

- MANDATORY: Check if those digital copies available are the best copies that exist (in addition to the Access Copies available to the public there may exist Master Copies that can be made available to the project).
- MANDATORY: Check the digital copies in accordance with the Advanced Digitization quality criteria (for e.g. completeness, readability, OCR errors, metadata available…) and document quality control by filling in a Quality Control Form (use Tool #11 *Quality Control Form Text and Image Digitization*).
- OPTIONAL: Ask for digitization equipment that has been used and digitization standards that have been applied to generate those digital copies (then proceed to step number 5).

4. Check digitization options at archive/library

Provided that digital copies of relevant documents are not available, determine if an archive/library

  a) has an internal digitization pipeline
  b) offers on-demand digitization
  c) provides equipment for digitization by users
  d) allows users to bring their own digitization equipment

e) allows users to borrow items and digitize them outside of the archive's/library's own premises

If a): Check if the internal digitization pipeline can be entered.

- o Are the relevant documents part of holdings that have already been scheduled for digitization?
    - ▪ If no, is there a way of including them into the internal digitization pipeline?
    - ▪ If yes, does the digitization schedule match the project's schedule?
        - - If no, is there a way to change the digitization schedule of the archive/library and to prioritize the relevant documents?

If b): Check scheduling and costs.

If c): Check costs.

If d): Check restrictions on digitization equipment allowed (e.g. flatbed scanner, photo camera, stand, lights…).

If e) Check conditions of borrowing.

5. Determine terms and conditions of re-use of digital copies (licenses etc.)

- Check online sources for information.
- Ask archives for information.

6. Determine Digitization Standards

In all cases except usage of own digitization equipment (4d-e):

- Determine whether the archive/library has a guideline for digitization that can be made available to the project and that provides information on how digital copies are created.
- Determine what digitization equipment is used (technical specifications of e.g. scanner type).
- Determine what Digitization Standards are and can be applied:

    - o What settings can be applied (DPI ≥ 300, color/b&w/grayscale)?
    - o What output formats can be produced and made available: analog or digital (if digital: *.tiff, *.pdf, *.jpg, RAW format etc.)?

- Can bound elements be cut open for digitization?
- Is the verso of photos and/or textual documents also digitized even if it is blank?
- Determine which metadata can be made available:

- o Technical metadata: is the EXIF data included in the file?
- o Is archival metadata such as box label, folder covers, index cards etc. also included in the digitization?

- Determine which delivery options are available:

- o Ensure readability and compatibility of the storage/delivery media and technologies chosen between all parties involved (e.g. is the filesystem readable/writable on all the platforms used?)[1]
- o Does the delivery method require special software to access the files?

- OPTIONAL: Request a test digitization of a document/record.
- [OPTIONAL in VHH: Ask Library of Congress to analyze the test digital copy.]

7. Decide about the reuse of existing digital copies

- Based on the criteria listed in the Quality Control Form, decide if existing digital copies will be reused or which of the options 4.a to 4.e will be chosen.
- Consult with project leader and experts if necessary.

8. Determine scope of documents to be digitized

- Published items:

- o Always include front and back cover (if available), title page, bibliographic information, table of contents, footnotes and maybe index.

- Unpublished items:

- o Always include or at least somehow document box label, title page of folder and any form of table of contents or index if available.
- o If possible, digitize complete folder.
- o If not possible or unreasonable to digitize the complete folder (e.g. excessive scope, clear irrelevance of majority of content, lack of time…), document the scope of the complete folder (e.g. estimated number of pages) and the reason for partial digitization.

9. Select element(s) most suitable for digitization

In case of multiple copies available, select element(s) most suitable for digitization:

---

[1] As specified in Tool #8 *Guideline Text and Image Digitization*, the recommended file system is ExFAT.

- Consider best overall condition.
- Consider best image quality.

10. Select your own digitization setup

In case you decide to use your own digitization equipment:

- Choose best possible equipment, software and settings based on Digitization Standard and specific solution chosen.

11. Apply general and specific digitization rules

- General remarks applicable to all documents:
  - Make sure documents are as flat as possible but consider conservatory aspects
  - Digitize page by page (= each page individually, including verso, even if blank)
  - If there is a significant number of blank pages in a row (say more than 5), a place holder may be used: "THE REMAINING PAGES ARE BLANK"
  - For pages that need imaging but cannot be captured, a place holder may be used: "THIS DOCUMENT HAS NOT BEEN DIGITIZED"
  - Use a ruler respectively a tape measure. In the case of standard sized documents (US letter, A4), a tape measure can be omitted as long as the documents' size is included in the meta data.
- Text documents:
  - If language is read from right to left (such as Hebrew), the document(s) should be imaged in the order of its/their reading direction.
  - Thin and translucent paper (onion skin, etc.) should have white support behind it.
- Photos:
  - Make sure not to crop the white border around the actual photographic content.
- Mixed media (court exhibits, photo albums etc.):
  - If possible, always digitize both, each page and each photo individually.
  - As for photo albums, make sure to capture features characteristic of the medium, including front and back cover and facing pages.
- Saving files:
  - Maintain original folder order within digital imaging order.
  - Back up the digital copies immediately (cloud backup if WIFI is available).

12. Establish and apply documentation standard and procedures

- Make sure that every decision taken is properly documented.
    - o With regard to single documents or collections the decisions ought to be documented in the Reference Database [VHH/EDU: Citavi].
    - o However, decisions applying to an entire archive ought to be documented in the List of Archives.

- In particular ensure that documentation includes:
    - o how and why the material was chosen for digitization
    - o information on digitization process chosen (4a–e)
    - o when, where, and by whom the material was digitized
    - o technical details (in a human readable form) about the process including equipment used
    - o information on any deviation from chosen standard
    - o photographic documentation and a written note on damage inflicted on the material during the process
    - o a list and/or Reference Database entry [VHH/EDU: Citavi] of all materials digitized
    - o in case of multiple copies available, information on the element(s) used for digitization

13. Apply naming conventions

- Consistently apply naming conventions chosen for the project.
- Refer to Naming Conventions List of the project for details.

14. Establish and perform Quality Control

- Determine who will be responsible for quality control. If own digitization equipment is used, digital copies should be reviewed by a second team member.
    - o Compare scope of digital files with scope of selected materials and check if all documents scheduled for digitization have indeed been digitized.
    - o Check documentation of digitization process.
    - o Visually inspect digital files.
    - o Compare selected digitization parameters with digital files delivered.[2]

---

[2] For a list of parameters and digitization standards please see 3.1.–3.4. in Tool #8 *Guideline Text and Image Digitization.*

o Check and, if necessary, adapt naming conventions.

15. Apply OCR

- Apply Optical Character Recognition to convert the digital image files into machine-encoded text.
    o For standard text use Abbyy FineReader desktop application.
    o For German Fraktur font use Abbyy FineReader Online.[3]
    o For handwritten text use Transkribus.[4]

16. Create Access Copies

- Determine the specific purpose of the Access Copies to be created.
- Determine the conditions that the Access Copies will have to meet, in particular
    o which file formats may be used
    o which file size may not be exceeded
- Choose the best possible solution.

17. Create Archival Copies

For long-term preservation create Archival Copies.
    o USE DNG as file format for the original RAW files.
    o USE TIFF as file format for the Production Masters.
    o Use PDF/A as file format for the OCR versions.

---

[3] Cf. https://finereaderonline.com/en-us (31.12.2019)
[4] Cf. https://transkribus.eu/Transkribus/ (31.12.2019)

VISUAL HISTORY OF THE HOLOCAUST
Rethinking Curation
in the Digital Age

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #10: Guideline Citavi

Lead-beneficiary (Tool #10)	*LBI*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*LBI (Alexander Prenninger)*
Contributor(s)	*TUW (Sebastian Hofstätter)*
Reviewer(s)	*LBI (Ingo Zechner)* *OFM (Michael Loebenstein)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.5, 2019-12-31*

Tool #10: Guideline Citavi

Table of Contents

1. Introduction

VHH decided to use Citavi as a Reference Database for non-filmic materials and as a tool to organize archival research.

"Citavi is a software program for reference management and knowledge organization."[1]

VHH uses Citavi for the following purposes:

- To collect scholarly literature essential for the background and interpretation of filmic materials and non-filmic primary sources
- To collect non-filmic primary sources
    - already mentioned in the scholarly literature
    - listed in finding aids
    - discovered in the course of onsite research in archives
- To support and facilitate forthcoming archival research
- To create vocabularies, bottom up directly from scholarly literature and primary sources
- To extract quotes, both textual and visual, from primary sources and scholarly literature to be included in the VHH-MMSI

VHH uses Citavi on cloud basis to allow for live updates during collective teamwork and for continuous backups.

Tool #10 *Guideline Citavi* aims at the specific use of Citavi in VHH. It is based on Citavi 6 (Version 6.3.11.0, Swiss Academic Software GmbH). For general descriptions of Citavi the official Citavi Manual should be consulted.

This tool is a living document to be used together with the following tools:

- Tool #2 *Checklist Archival Research*
- Tool #3 *Checklist Memorial Research*

---

[1] https://www1.citavi.com/sub/manual6/en/index.html (31.12.2019)

2. Basic structure

2.1. Citavi projects

Citavi has been designed for research projects. For each research design a new project has to be created in Citavi.

Projects can be saved locally or in the Citavi cloud.

2.2. Citavi workspaces

Citavi was developed primarily as a tool supporting the writing of research papers. Its user interface supports the collection of primary and secondary sources, the collection of references and the definition of tasks. For these purposes, the interface is divided into three workspaces:

- the Reference Editor
- the Knowledge Editor
- the Task Planner

The Reference Editor consists of three parts:

a) a list of all references
b) the details of a chosen reference
c) the file viewer

The Reference Editor also allows for opening an additional column displaying categories, groups, keywords or imports.

Adding references is supported by "Citavi Picker", a browser add-in for Firefox, Chrome, Internet Explorer, Adobe Acrobat and Adobe Reader. Adding quotes to research papers is supported by a Citavi Word add-in.

In VHH, Citavi's Reference Editor and Knowledge Editor are used. The Task Planner will be not used but replaced by a solution more convenient for VHH purposes (see further down Groups).

3. Adding metadata and digital assets

3.1. Adding references in the Reference Editor

Adding references aka bibliographic and archival metadata to Citavi is possible in different ways:

- In the Citavi interface it is possible
  - to run searches in library catalogues and bibliographic databases (which can be added for one's own specific purposes)
  - to retrieve bibliographic metadata by searching for ISBN, DOI or other identifiers
- In the web browser interface, the Citavi Picker allows for adding webpages and literature via the ISBN or DOI (the picker will appear next the them).
- If automated adding is not supported (e.g. for archival sources or literature without ISBN), references have to be added manually. (In this case choose "Title", then select the relevant document type and add the details.)

3.2. Adding digital assets

PDF documents can be added with the Citavi Picker add-in in Adobe Acrobat and Adobe Reader. Adding is possible either as an attachment to a reference already existing in the Citavi project or as a new reference.

The Citavi Picker also allows for saving webpages as PDF files.

Citavi will automatically create file names for imported files.

These file names have to be renamed according to VHH standard:

- For monographs, edited volumes etc:
  NAME-Surname_Titel_Year_full (if all pages are included)
  NAME-Surname_Titel_Year_xxx-yyy (if only a selection of pages is included)
- For journal and newspaper articles:
  NAME-Surname_Titel_Year_xxx-yyy
- For archival sources:
  [no standard defined until now!]

PDF documents without machine readable text should be transformed into machine readable text by applying OCR before adding them to Citavi.

3.3. Keywords

The automatic import of references via identifiers like ISBN or DOI includes the import of keywords, abstracts, tables of contents and external links.

In VHH imported keywords are preserved without further editing. No additional keywords are manually assigned to references. Instead the supergroup "topics" in categories is used to apply descriptive terms from controlled vocabularies.

3.4. Adding metadata from vocabularies: groups

In Citavi groups have been designed to sort references according to external qualities. Typically, groups are used to create structured bibliographies (e.g. primary and secondary sources, websites etc.)

In VHH groups are used to define properties of documents and to support research in libraries and archives. The following supergroups are used:

Properties of documents

1. Publication status (published or unpublished)
2. Source type (primary or secondary source)
3. Media type (e.g. text, audio, film, photo, software etc.)
4. Media format (e.g. bound item, digital file, DVD, CD etc.)
5. Publication type (e.g. monographs, anthologies, periodicals, radio or TV broadcasts etc.)
6. Narrative type (e.g. autobiographies, diaries, correspondence, finding aids, fictional film, graphic novels etc.)

Research parameters

7. Online status (if on- or offline)
8. Source archive (e.g. NARA, IWM, USHMM etc.)
9. To Dos (e.g. to check at archives or libraries, to scan or run OCR)
10. Accessibility status (primarily of archival sources, e.g. classified)
11. Digitization obstacles (e.g. handwriting)
12. Digitization status (e.g. JPEG, MP3, PDF, TIFF, WAF)

Groups 1 to 7 have to be assigned to every reference in the Citavi project.

3.5. Adding metadata from vocabularies: categories

In Citavi categories have been designed to structure a research paper and create a table of contents. Categories therefore equal the headings of chapters or subchapters in research papers. Both references and quotes can be assigned to categories.

In VHH categories are used in a different way. The VHH_CITAVI-Cloud project differentiates two supergroups of categories

a) Archival Provenance
b) Topics

Ad a) Archival Provenance categories are used to sort primary sources according to archives. Subcategories are used to relate primary sources to record groups, series, subseries, boxes etc.

Ad b) Topics categories are used to develop and apply descriptive terms in and from controlled vocabularies. Primary and secondary references are indexed with categories like places, events, agents, objects etc.

4. Creating new digital assets

4.1. Quotations

Quotations from and comments on references can be extracted and added in the reference manager. The easiest way is to mark text in the PDF viewer and add the marked text as a quote.

Citavi also offers the possibility to annotate text by highlighting in color, to add indirect quotations, comments, tasks or summaries.

Images or tables can be added as image quotations using the snapshot tool in the PDF viewer.

Quotes, comments, summaries etc. can be assigned with groups, categories and keywords. (In VHH only categories are used for quotations.)

Organizing quotations is done in the Knowledge Editor.

In VHH quotations are treated as separate digital assets. While the use of bibliographic items may be restricted due to IPR, quotations may be included in the VHH-MMSI.

5. Transferring metadata

5.1. Export via Citavi interface

Citavi includes several options to export data via the Citavi interface. References can be exported to another Citavi project, to clipboard, to RiS, BibTex, EndNote or Excel.

A second way is to a project bibliography. Citavi allows for creating simple, grouped and annotated bibliographies.

- A simple bibliography is a list of all references or a selection of references in the project. References can be selected by group or category.
- A grouped bibliography affords the definition of groupings (e.g. primary and secondary sources).
- An annotated bibliography contains bibliographic information along with other information (e.g. abstract, keywords, groups, categories, quotes, comments etc.).

5.2. Export via Citavi APIs

Citavi enables expandability of its software in two ways: *macros* and *add-ons*. Macros are single file program snippets that can be run via a built-in source code editor. Macros have access to all data available in the Citavi program in a structured form, so that each item that is create and editable in Citavi by users may be exported (saved to a structured file format). With the full control over the export routines, we can export exactly the items & structures needed in further steps of the VHH project.

Add-ons offer the same access to data as macros via a user-interface embedded in Citavi itself. This makes it possible for non-technical-experts to export the required data via a user-interface.

Citavi macros and add-ons are written in C# (a mainstream programming language from Microsoft). The developers of Citavi published examples and guidelines for developing integrations into their product on the open-source platform GitHub. This makes it easy to develop custom integrations tailored to the VHH project.

6. References

Citavi Manual 6:
https://www1.citavi.com/sub/manual6/en/index.html (31.12.2019)

Peter Meurer, Manfred Schluchter (2017): Writing a Research Paper with Citavi 6. Trans. by Jennifer Schultz. Based on the Duden book "Die schriftliche Arbeit – kurz gefasst" [Research Papers In a Nutshell] by Jürg Niederhauser, Wädenswil.:
www.citavi.com/tutorial (31.12.2019)

Citavi Open-Source Software – URL:
https://github.com/Citavi (31.12.2019)

VISUAL HISTORY OF THE HOLOCAUST
Rethinking Curation
in the Digital Age

Deliverable D2.1

Advanced Digitization Tool Kit
Tool #11: Quality Control Form
Text and Image Digitization

Lead-beneficiary (Tool #11)	*LBI*
Work Package No. and Title	*WP2 Curation and advanced digitisation*
Work Package Leader	*OFM*
Relevant Task	*Task 2.1 Definition of advanced digitisation standards*
Task Leader	*OFM*
Main Author(s)	*LBI (Ingo Zechner)*
Contributor(s)	*LBI (Vrääth Öhner)*
Reviewer(s)	*JLU (Ulrike Koppermann)* *OFM (Michael Loebenstein)*
Dissemination Level	*Public*
Due Date	*M12 (2019-12)*
Version (No., Date)	*V1.5, 2019-12-31*

Introduction

Tool #11 *Quality Control Form Text and Image Digitization* is intended for quality control of existing digital copies that have either been generated by third parties or within the VHH project. It is a living document to be used together with the following Tools:

- Tool #9 *Checklist Text and Image Digitization*

**VISUAL
HISTORY
OF THE
HOLOCAUST**
Rethinking Curation
in the Digital Age

Quality Control Form

Text and Image Digitization

Version (No., Date) *V1.3, 2019-12-19*

Archive	
Record group	
Series	
File	

Digitized by	
Delivery date	
Inspected by	
Inspection date	

File format	
Resolution (DPI)	
Bit depth	
Color space	
Color	
Equipment	
FADGI Stars	

Number of files	
Total size	

VISUAL HISTORY OF THE HOLOCAUST
Rethinking Curation
in the Digital Age

No	Quality Parameter	Yes	No
1	File format satisfactory		
2	Resolution satisfactory		
3	Bit depth satisfactory		
4	Color space satisfactory		
5	EXIF data included		
6	Best digital copy available		
7	Complete copy		
8	Original order preserved		
9	Box label included		
10	Folder cover included		
11	Index included		
12	Verso included		
13	Measures included		
14	Edges not cropped		
15	Text not distorted		
16	Sharpness satisfactory		
17	Color satisfactory		