

# DFG Practical Guidelines

## on Digitisation



## Contents

<b>Introduction to the Practical Guidelines on Digitisation</b>	<b>4</b>
<b>1. Checklist for Applicants and Reviewers</b>	<b>4</b>
1.1 General technical procedures and resources	5
1.2 Technical parameters of digital reproduction	6
1.3 Metadata	6
1.4 Full text generation	7
1.5 Long-term availability	8
1.6 Organisational issues	8
1.7 Citing, persistent addressing	9
1.8 Provision of metadata and digital resources to the public	9
<b>2. Objectives and Selection</b>	<b>10</b>
2.1 Objectives	10
2.2 Selection	11
2.3 Duplicate checking and data matching for retro-digitisation of published texts	12
<b>3. Digitisation</b>	<b>12</b>
3.1 Preparation of materials and conservation review	12
3.2 Technical parameters of digital reproduction	13
3.2.1 General explanations and parameters	13
3.2.2 Material-specific parameters	20
3.3 Metadata	26
3.3.1 Indexing, descriptive metadata	27
3.3.2 Structural metadata for digital facsimiles	28
3.3.3 Description of collections and holdings	29
3.3.4 Exchange and dissemination of metadata	29
3.4 Full text generation	30
3.4.1 Text capture	30
3.4.2 Character encoding	33
3.4.3 Markup of full texts	33
3.4.4 Layout	34
3.5 Long-term availability	34
<b>4. Organisational Issues – In-House vs. Outsourced Digitisation</b>	<b>36</b>
<b>5. Citing Digital Resources, Persistent Addressing</b>	<b>36</b>
<b>6. Provision of Project Results to the Public</b>	<b>37</b>
6.1 Rights, licensing and open access	37
6.2 Minimum requirements for provision systems for digitised material	38
6.2.1 Functionality requirements	38
6.2.2 Minimum technical requirements	39
<b>7. Presentation Standards for Text-Based Works, DFG-Viewer</b>	<b>40</b>
7.1 Basic requirements and architecture	40

7.2 Functionality requirements	41
7.3 DFG Viewer	41

Appendix A: METS/MODS Profile for DFG Viewer Display and Transmission by OAI

Appendix B: METS/TEI Specification for the Display of Digitised Manuscripts

Appendix C: LIDO core elements for publication

Appendix D: Describing Collections and Holdings

## Introduction to the Practical Guidelines on Digitisation

As part of the Scientific Library Services and Information Systems programme (LIS), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)<sup>1</sup> funds projects at scientific institutions, particularly service and information institutions, in Germany. The aim is to set up nationwide efficient research information systems and infrastructures. The results of these projects must be accessible to researchers at no charge and for the long-term.

The Practical Guidelines on Digitisation for the funding area of Scientific Library Services and Information Systems are designed to make it easier for applicants to plan digitisation projects and for reviewers to compare proposals. The Practical Guidelines are not meant to create obstacles but rather to formulate standards in order to ensure that funded projects will be sustainable and viable over the long-term.

To complement the standards, the Practical Guidelines also include additional information, for example on conducting conservation reviews of materials selected for digitisation, collecting metadata, producing digitised material, indexing images, producing full text, or the long-term preservation of digital content.

Section 1 summarises the most important requirements. Deviations from the rules described therein are possible, but must be justified in each case. Sections 2 to 6 provide a general and more comprehensive introduction to the issues and methods relevant to projects that aim to digitise objects from libraries, archives or museum collections. These sections are especially geared toward those who are planning such projects and may not have any detailed previous knowledge. Section 7 specifies the presentation standards and formats required by the DFG for text-based objects.

### 1. Checklist for Applicants and Reviewers

As explained in the following sections, there are numerous choices to make when planning digitisation projects. Rigidly prescribing specific standards would therefore unduly restrict the projects to be funded and hamper their continued dynamic development. However, the guidelines provide information on risks with respect to project planning and realisation. The checklist for applicants and reviewers should therefore be understood and applied as follows:

- The technical concepts of all proposals with digitisation components will be reviewed in addition to the assessment with regard to contents.
- Reviewers will check whether the necessary rights clearances regarding the source material as well as provisions that enable full reusability of project results have been completed and substantiated. (→ [1.8](#))

---

<sup>1</sup> The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) is the central, self-governing research funding organisation that promotes research at universities and other publicly financed research institutions in Germany. The DFG serves all branches of science and the humanities by funding research projects and facilitating cooperation among researchers (<http://www.dfg.de/en>). The DFG also supports projects that improve scientific information infrastructures in Germany. The results of funded projects must be accessible to researchers at no charge and for the long term (<http://www.dfg.de/lis/en/>). It should be noted that an applicant's defined institutional tasks and financing should not be substituted by DFG funding. Projects must therefore exceed an institution's ordinary mission, be of a limited time-frame and topical scope, and focus on outstanding materials with nationwide significance. Conversely, projects cannot be funded if they serve primarily the promotion or conveyance of culture or similar purposes, or if they are commercially oriented.

- Proposals must demonstrate plausibly that the project will be implemented according to the standards listed below.
- Any deviation from these standards must be justified in detail.
- For better comparability, proposals (or progress reports) should provide information about estimated (or actual) costs of scanning. Costs of structural data entries per image and title / unit of description, full text recognition and project-specific storage capacities should be included.<sup>2</sup>
- If it is intended that a digitisation project should exceed the standards outlined below, the need for doing so should be explained in detail if this entails higher costs.
- The technical preparation of a proposal must be comprehensive enough to allow an overall evaluation of technical requirements and procedures of the proposal. While the initial project stage may include the testing of innovative technology, it cannot be used to determine, for instance, how long the digitisation campaign will take, what grade of digital copies should be produced, or how the general workflow should be designed. Any pilot studies necessary to resolve such issues must be concluded prior to the submission of a proposal.

### 1.1 General technical procedures and resources

The proposal must describe the intended workflow in sufficient detail to allow reviewers to assess the following questions:

- Is the funding for staff both sufficient and necessary? To help answer this question, average available resources (working hours, storage capacity of computers used in the workflow) per unit must be stated.
- Are the projected processing times realistic? To the extent that the projected processing times are not immediately plausible, they should be substantiated either by experiences gained in previously completed, similar projects, by published and recognised benchmarks, or by the results of self-conducted pretests.

---

<sup>2</sup> See the statistics sheet for Digitisation at <http://www.dfg.de/lis/en/> > Formulare und Merkblätter > Ergänzende Formulare und Formblätter LIS (German only). There are currently no commonly accepted standards for calculating real costs. The Practical Guidelines therefore recommend that the costs of scanning including structural data entries be calculated as follows:

- (1) For outsourced digitisation:
  - (a) Amount of money paid to service provider per digitised item.
  - (b) Proportionate costs of all flat fees charged by service provider (e.g. for naming and storing files, issuing structural data, DVDs for transferring data from service provider to project location).
  - (c) Proportionate costs for staff members occupied exclusively, or to a calculable extent, with digitisation quality control.
- (2) For in-house digitisation:
  - (a) Proportionate costs of newly acquired digitisation hardware in the narrower sense. Hardware is considered depreciated when the project ends (actual entire duration; e.g. in the case of an ongoing four-year project that was initially proposed as a two-year project, the base is the projected number of items to be digitised throughout the full four-year period).
  - (b) Proportionate personnel costs for all personnel who operate digitisation hardware and issue structural data.
  - (c) Proportionate costs for staff members occupied exclusively, or to a calculable extent, with digitisation quality control.

In both cases, (1) and (2), costs must be calculated per digitised item. Expenses related to the following tasks are not considered costs for scanning:

- (a) Project management (e.g. selecting and fetching materials to be digitised).
- (b) Entry of metadata other than structural data.
- (c) Long-term archiving.
- (d) Indirect costs typically assessed in terms of internal cost accounting.

## 1.2 Technical parameters of digital reproduction

The aim of digitisation is to reproduce the original material as faithfully as possible, according to scientific requirements. For all objects to be digitised within a project, the quality proposed must permit batch processing without human intervention to generate reproductions meant for immediate publication. The following guidelines for minimum digitisation requirements refer only to **digital masters** and are general recommendations for all materials. For material-specific recommendations that exceed the scope of these minimum requirements, see section [3.2.2](#).

- The resolution should be such that archive copies allow for the smallest relevant details to be clearly visible when the file is reduced to one-quarter of its original size. A resolution of 300 dpi is recommended as a general rule. (→ [3.2.1.1](#))
- For storage of the digital master a colour depth of 8 bits per channel, i.e. 24 bits, is sufficient. (→ [3.2.1.2](#))
- Masters should be stored as uncompressed Baseline TIFF files. The more advanced options of extended TIFFs should not be used for digital masters. In addition to TIFF, TIFF-LZW or lossless JPEG2000 may also be used as an image master format. However, to store masters in JPEG2000 format it is important to note that only the public-domain parts of JPEG2000 may be used. For information about the risks associated with JPEG2000 and TIFF-LZW in relation to long-term archiving, see section [3.2.1.4](#).

## 1.3 Metadata

Metadata should be provided in a software-neutral, standards-compliant form, generally XML encoding.

- To ensure the optimally distributed and long-term usability of descriptive metadata, indexing should be based on relevant standards and reference models, and linked with published standard data wherever possible. To record personal, biographical and geographic information, the Integrated Authority File (GND) offered by the German National Library must be used. Other controlled vocabularies such as Iconclass for image classification must enable integration on national and international levels. (→ [3.3.1](#))
- The provision of descriptive metadata for further use in accordance with material-specific standards is mandatory (→ [3.3.1](#) and Appendices A and B):
  - Metadata Encoding and Transmission Standard / Metadata Object Description Schema (METS<sup>3</sup>/MODS) for printed text-based works and archive material
  - Metadata Encoding and Transmission Standard / Text Encoding Initiative (METS/TEI<sup>4</sup>) for manuscripts
  - Lightweight Information Describing Objects (LIDO) for (usually rare) pictographic and three-dimensional objects

The metadata must be valid for the relevant XML schema and must also be checked for semantic correctness.

- Provision of the descriptive metadata through an Open Archives Initiative (OAI<sup>5</sup>) interface is mandatory, either in the institution's own system or through a suitable portal. (→ [3.3.1](#))

---

<sup>3</sup> <http://www.loc.gov/standards/mets>

<sup>4</sup> <http://www.tei-c.org>

<sup>5</sup> <http://www.openarchives.org/>

- The digital copies and metadata must be linked by persistent identifiers in the reference system (catalogue, online search system). Metadata entries and digitisation must be coordinated with or performed by a scientific infrastructure facility. Digitised prints should be listed in the Zentrales Verzeichnis digitalisierter Drucke (ZVDD)<sup>6</sup> and digitised archive material in Archivportal-D.<sup>7</sup> All materials must be presented in suitable subject-specific or multi-subject online applications as well as added to the Deutsche Digitale Bibliothek (DDB)<sup>8</sup> and indirectly into Europeana.<sup>9</sup> (→ [3.3](#))
- It is the responsibility of the grant recipient to ensure that the digitisation units produced in the project can be unambiguously identified, and searched and retrieved separately from other units recorded in the same system. (→ [3.3](#))
- Data should be delivered to these portals in accordance with standard formats via OAI if possible. (→ [3.3.1](#))
- The decision whether to generate structural metadata is always material- and project-specific. If structural metadata is used, it is recommended to consult the structural data list available on the DFG Viewer<sup>10</sup> website. The standards currently recommended are METS or TEI. However, the DFG Viewer should be supported in all cases for text-based materials. (→ [3.3.2](#))
- As a minimum, digitisation projects are expected to present the nature and scope of the selection of objects on a website, preferably with an English-language version. In addition, a standardised collection or inventory level description in XML in a standardised format is required to facilitate the future merging of this information in national or international portals that enable the respective retrieval. This description may be based on the Dublin Core Collections Application Profile<sup>11</sup> or on the same metadata standard in which the object descriptions are made available: METS, MODS, TEI headers, Encoded Archival Description (EAD(DDB)) and LIDO all offer the necessary features. (→ [3.3.3](#) and Appendix D)

## 1.4 Full text generation

All proposals to digitise text-based objects are expected to address the option of full text provision. For printed works dating from 1850 or later, the full text must be generated and simple image digitisation is not adequate. (→ [3.4](#))

Full text includes the characters of the master copy, markup data to identify structural features, and metadata, which is usually part of the same file. (→ [3.4](#))

- Full text can be generated in two ways: through Optical Character Recognition (OCR) or transcription. (→ [3.4.1](#))
- To create some uniformity on which to base the evaluation of accuracy, applicants are requested to state the letter accuracy, i.e. incorrect syllabification and layout errors should be ignored. Random samples should be based on the statistical procedure described in 3.4.1. (→ [3.4.1](#))

---

<sup>6</sup> <http://www.zvdd.de/>

<sup>7</sup> <https://www.archivportal-d.de/>

<sup>8</sup> <https://www.deutsche-digitale-bibliothek.de/>

<sup>9</sup> <http://www.europeana.eu/portal/de>

<sup>10</sup> <http://dfg-viewer.de>

<sup>11</sup> <http://dublincore.org/groups/collections/collection-application-profile/>

- For the purposes of character encoding, it is recommended to save the text in Unicode. UTF-8 should be preferred. (→ [3.4.2](#))
- Unless there are cogent reasons not to, full texts of prints and manuscripts must be encoded and marked up following the model of the Text Encoding Initiative (TEI). (→ [3.4.3](#))
- In some cases it is important for the presentation of full text to preserve the layout of a document for the long term. The Practical Guidelines recommend using a suitable formatting language (e.g. XSLT, XSL-FO, XQuery or CSS), which largely ensures independence from special software. If valid reasons prohibit archiving the format with XML techniques, layout information for text documents may also be archived in PDF according to ISO standard 19005-1. However, as explained in section 3.4.2., PDF files cannot replace the provision of marked-up full text in XML. (→ [3.4.4](#))
- In keeping with the principles of open access and open source, it is expected that machine-readable full texts or the XML on which these texts are based, XSLT scripts and DTDs or XML schema files will be made available for reuse as freely as legally possible. (→ [1.8: 6.1](#))

## 1.5 Long-term availability

There is currently no universal solution for the long-term preservation and archiving of digital content that is suitable for all types of objects. For long-term preservation, files are stored in stable, migratable formats in a technically and organisationally secure storage system. Digital data is archived with a similar storage system but with more extensive technical and organisational measures that cover not only the physical retention of the data but also strategies to make them available for use (access). (→ [3.5](#))

The long-term availability of the results of digitisation projects depends firstly on the choice of data and metadata formats. Secondly, it must be ensured that the digital data remains physically available. It should be noted that in DFG-funded digitisation projects, the costs of project-specific data preservation are expected to be borne by the institution for the duration of the project. DFG funding is not available for these costs. (→ [3.5](#))

It should be noted that the DFG views digitisation projects as endeavours by the entire institution: It is assumed that the department in charge of the project will be supported by the in-house IT infrastructure. We also encourage smaller institutions to take advantage of the expertise and services of larger institutions. (→ [3.5](#))

Proposals must contain convincing statements as to institutional long-term preservation and archiving. Digitised material should be archived redundantly. (→ [3.5](#))

## 1.6 Organisational issues

The decision whether to undertake digitisation as an in-house project or to outsource it is always specific to the project and exclusively the applicant's responsibility. If an institution chooses to outsource it must be able to demonstrate that it is capable of successfully managing the project. Services must be precisely defined by contract. The DFG expects that an appropriate percentage of the invoice amount be withheld for security purposes and not paid out to the business providing the service until a quality check has been performed. (→ [4](#))

## 1.7 Citing, persistent addressing

Digital files must have a unique address so that other objects or databases can link to them. In addition to the customary citation format, which can and should still be given via the navigation software, this requires the specification and online documentation of addressing techniques. The accessibility and citability of the resource as a whole and the work's individual physical pages must be guaranteed. Institutions should implement suitable mechanisms (Persistent Uniform Resource Locator (PURL), Uniform Resource Name (URN), Digital Object Identifier (DOI), Handle, etc.) to ensure the persistence and linkability of a resource, thus reliably providing sources for scientific research. For printed works is recommended to generate URNs via the German National Library. (→ [5](#)).

## 1.8 Provision of metadata and digital resources to the public

The DFG funds the digitisation of research-relevant material in order to make this material available to researchers in Germany and the rest of the world. All projects should be designed to make research results available promptly and for the long term, and to enable full scientific reuse in other research contexts.

- Rights relating to the objects to be digitised must be cleared when project planning begins or at the latest when a proposal is submitted. In particular, any copyrights, personality rights, ancillary rights and archive rights must be taken into account. (→ [6.1](#))
- In keeping with the principles of open access and open source, all results – metadata, digital copies, full texts, or the XML on which these texts are based, as well as XSLT scripts and DTDs or XML schema files – must be made available for reuse as freely as legally possible. If they are already in the public domain, they must be marked as such and not to be licensed. If this is not possible, they should be offered under the most liberal Creative Commons licence possible (CC0, CC BY or CC BY SA)<sup>12</sup> and, in the case of protected material, provided with standardised rights notices. (→ [6.1](#))
- Additionally, images should be provided in a form that allows full scholarly use in other research contexts. For this purpose, either high-resolution derivatives in TIFF format or fully resolved JPEG images with a compression of 90, maximum 80, in combination with the full scan resolution should be made available. (→ [6.1](#))
- For projects involving more than just public-domain material and/or collaboration with commercial partners or publishers, delayed open-access publication (a moving wall) of up to one year after the end of the project may be agreed upon. (→ [6.1](#))
- A restriction on open access may be justified by reasons such as privacy, copyright or archival law, but it may not affect more than 5% of the total material to be digitised and requires submission of a plan for the timeline and technical implementation of material availability.<sup>13</sup>

As a rule, digitised material should be accessible in different ways: (→ [6.2.1](#))

- via the providing institution's website
- via the local and regional library catalogue / the local and regional archive portal / the relevant material-specific online application and the national reference and presentation systems (DDB, Archivportal-D)

<sup>12</sup> See licences in the currently valid version: <https://creativecommons.org/>

<sup>13</sup> The DFG cannot fund the processing of non-open-access material.

- via a locally implemented or externally operated DFG Viewer, if applicable to the material
- via Internet search engines
- via an OAI interface

All objects must be provided in a quality sufficient for academic purposes and outfitted with intuitive navigation features to facilitate easy use by the target community and on typical university equipment. All currently popular browsers must be supported to the extent that this is objectively viable. (→ [6.2.2](#) and [7.2](#))

All DFG-funded provision systems are expected to provide an automatic option for users to give feedback on the digital resource. This function should be set up on the project page or centrally in the digital provision system. (→ [6.2.1](#))

The DFG expects that data made available online as part of DFG-funded projects should include a clear reference to its origin and if applicable a reference to the DFG funding. In the case of digitised images, this is usually done by adding an acknowledgement to the published user copy (for example in JPEG). (→ [6.1](#))

As a rule – assuming the nature of the digital reproductions does not implicitly rule out part of this service spectrum – digitisation projects are expected to provide plans for the following publicly accessible interfaces:

- A stand-alone server that provides the digitised objects along with the tools needed to use them.
- All digital reproductions must be published in a way that creates persistently citable URLs with the finest granularity possible. The proper citation format must be clearly indicated.
- An interface in the technical sense, to allow access to all metadata generated by a project. This can also be implemented by way of a central portal.
- Appropriate measures that enable search engines to find the metadata.

## 2. Objectives and Selection

### 2.1 Objectives

Digitisation has become a vital research tool in the humanities, cultural studies and the emerging field of digital humanities. It has made direct research with sources much easier, while conserving valuable and sometimes fragile originals. Not only does the digitisation of library, archive and museum holdings make copies easily accessible online, it also helps build an infrastructure that turns the Internet into an integral research space for the increasingly digital world of research in the humanities and cultural studies. Only by linking these digital objects with other online resources can the potential of the Internet be fully leveraged. Thus the objective is not only to make these materials available and usable, but also and especially to interconnect the different resources to form a virtual research infrastructure.

While there is a broad base of verifiable knowledge for implementing digitisation projects, these insights must not be applied mechanically: What constitutes essential conservational care when digitising medieval manuscripts may be unnecessarily time-consuming and expensive when processing a bulk of government records from the late 19th century.

Because the majority of previously realised and currently planned projects are still focused on text-based materials, the techniques and parameters pertaining to this area will be discussed below in special detail. However, images have been significantly gaining in importance. The guidelines therefore also cover image-based material, whether they are graphic representations, photographs or digital images of three-dimensional objects. This version of the Practical Guidelines incorporates recent developments in these areas, but is not intended to be a complete guide in this respect. In DFG-funded pilot projects, specific techniques and parameters have been and are being identified, e.g. for OCR-supported full-text generation or digitisation of historical newspapers and rare material (archival material, medieval manuscripts), which have been or will be added to the Practical Guidelines after completion of the pilot phases.

## 2.2 Selection

In general it should be noted that the technical aspects of digitisation can be planned quite well, while the intellectual effort required to select the right items is hard to calculate. A decision must therefore be taken in each case as to whether the inclusion of a greater number of documents will ultimately be cheaper and more efficient than undertaking a complex evaluation and selection process. It is highly recommended to take advantage of existing selections and reference works such as bibliographies and subject databases. The basic selection criteria are relevance to and demand by researchers.

Defining a corpus under the criteria of relevance to, or demand by, researchers is not always easy. In difficult situations, the case for a project may be made by cooperating with a specific research community or institution, which can plausibly formulate its own needs. Successful projects often base on cooperative arrangements in which an academic undertaking, e.g. a research or editorial project in philology or legal history, seeks to establish an online presence and links back to library or archive holdings, thus enabling two-way linkage.<sup>14</sup> Alternatively, relevant subject bibliographies that formulate a canon may be used, or a combination of both approaches may be taken. Moreover, the concepts of “digitisation on demand” and “digitisation on use” (the digitisation of commonly or frequently used material in libraries and archives) may be used.

As a general rule, more specialised digitisation projects that focus on a specific research question should provide interfaces to deliver their data to larger projects with a more formal structure. Conversely, interdisciplinary projects should open up their results to more in-depth and specialised use.

The DFG funds the digitisation of research-relevant material in order to make this material available to researchers in Germany and the rest of the world. **Clearance of rights to the material** must be completed when project planning begins, or at the latest when the proposal is submitted, and must be substantiated in the proposal. In particular, any copyrights, personality rights, ancillary rights and archive rights must be taken into account. Rights clearance often proves to be more time-consuming than anticipated. Sufficient time and personnel should therefore be allocated to this task. If the clearance of rights proves to be difficult, we recommend seeking professional support, for example from the institution’s own legal department. Even during the selection process, it should be noted that the results of DFG-funded digitisation

---

<sup>14</sup> Research projects are generally based on funded infrastructures. However, it is also possible to apply for funding for projects that combine both research and infrastructure elements. Alternatively, relevant specialised bibliographies that formulate a canon, or to a mixture of both, may be used.

projects must be made freely available to the research community by way of open access. Wherever material cannot legally be assigned to the public domain, a free Creative Commons licence must be chosen (CC 0, CC BY or CC BY SA). (→ [6.1](#))

## 2.3 Duplicate checking and data matching for retro-digitisation of published texts

To avoid redundant digitisation it is sensible to check, before submitting a proposal, whether the objects selected for digitisation are already digitally available in Germany or elsewhere. The following requirements apply:

Proposals and reports are expected to mention finalised or ongoing national and international digitisation projects to the extent that they relate to the proposed or ongoing DFG project and the materials it covers.

For large-scope projects (e.g. over 1,000 printed works), proposals and reports should explain how they relate to commercial digitisation offerings that are accessible free of charge. A pragmatic effort should be made to keep the number of duplicate digitisations as small as possible. For printed works published between 1501 and 1800, the VD16, VD17 and VD18 bibliographies should be consulted to check for existing digitisations. URNs and PURLs of digital copies must be reported to these bibliographies. For digitisations of incunabula, the Census of Incunabula for Germany (ISTC) should be consulted.

## 3. Digitisation

Digitisation always includes creating digital images and producing metadata, and in the case of text-based works, possibly also capturing full text and generating structural data and mark-up. In the following recommendations the term digitisation refers to the entire workflow (preparation, digitisation proper, generation of bibliographical metadata, structural data, full text, and long-term safeguarding / digital preservation). No distinction is made between different material types (e.g. printed works, rare documents, objects) in the general sections that follow.

### 3.1 Preparation of materials and conservation review

Preparatory activities in digitisation projects are often underestimated and should be carefully taken into account when planning a project. Are the objects actually available? Are there any conservation-related objections to digitising the originals? Are there sufficient personnel to draw out and prepare the objects? Are employees with academic or bibliographic training available to perform completeness checks or collations, if catalogue entries do not provide this information? Digitisation of incomplete or defective prints should be avoided if possible and the reproduction of an ideal copy should be strived for.

Although the conservation review may be very time-consuming, it should definitely not be omitted. We recommend using checklists to assess suitability for digitisation and drawing up a set of digitisation guidelines on this basis.<sup>15</sup> If reproduction could expose an original to risk or

---

<sup>15</sup> A conservation review should include the following criteria: Risk to ink; risk to paint layer / priming coat; ink / paint corrosion; loose stitching; binding or cover material too stiff / inflexible; binding (partially) fractured; cover material at joint (partially) fractured; spine inlay too stiff; spine binding damaged; cracks, tears or flaws; delicate leather grain on spine; risk to spine gilding; very thick vellum volume with hollow spine; cover (partially) fractured; cover loose (front/back); closure straps too stiff / partially fractured, microbial damage; microform available. There are also some technical problems that make digitisation difficult or impossible: Inside margin too narrow; book block corrugated; leaves / signatures very stiff; unavoidable loss of text; original cannot be laid flat; extreme format.

undue stress,<sup>16</sup> it should be done on the basis of existing microfilm<sup>17</sup>, if possible, or not at all. At any rate, valuable historic materials must be handled with due conservational care, even if this reduces the digitisation throughput and takes more time. Scanning systems and equipment such as book supports or securing aids must comply with the guidelines of the conservation review.

## 3.2 Technical parameters of digital reproduction

The aim of digitisation is to reproduce the original object as faithfully as possible, according to scientific requirements. Digitisation parameters should be selected with regard to image quality, long-term availability, and interoperability.

### 3.2.1 General explanations and parameters

Two types of reproductions are important in digitisation: digital masters, i.e. the source or archive format, and derivatives generated for users, which are usually scaled-down copies in other file formats. Derivatives such as JPEG files should be created from the masters depending on the intended presentation and can be modified as needed. This may be necessary e.g. if the assumed screen resolution on the user side changes or if image formats are used for which the properties are optimised for the desired display (e.g. continuous zoom; smooth transitions between segments of large yet detail-rich objects).

The digital master forms the basis for all further processes. Its production should therefore be given special attention and the relevant recommendations should be taken into account with regard to long-term storage.

The following guidelines for minimum digitisation requirements refer only to **digital masters**.

#### 3.2.1.1 Resolution

The resolution of a file is normally measured in dpi (dots per inch).

The minimum scan resolution chosen for digital copies should allow the details of the original to be fully reproduced at life size.<sup>18</sup>

The basic recommendation for target resolution is therefore **300 dpi relative to the format of the original** (output format of digital copy at 300 dpi).

However, 300 dpi only relates to originals intended to be viewed with the naked eye, such as text-based works, graphic works and photographs. Different guidelines apply to media whose complete visual information is only visible when enlarged. This includes miniatures of any kind and especially photographic transparencies (e.g. negatives or slides).<sup>19</sup>

---

<sup>16</sup> Corbach, Almut: Bestandsschonendes Digitalisieren von schriftlichem Kulturgut. In: Digital und analog. Die beiden Archiv-welten. 46. Rheinischer Archivtag. Ratingen 21.-22. Juni 2012. Beiträge. Ed.: LVR-Archivberatungs- und Fortbildungszentrum. Bonn 2013 (Archivhefte 43), pp. 90 – 102.

<sup>17</sup> In archives it is standard practice to digitise microfilm regardless of conservational considerations. For a discussion of the material-specific parameters affecting the digitisation of microfilm, see section 3.2.2.

<sup>18</sup> One indicator of the required minimum resolution for a digital copy is the resolution of the human eye at a comfortable reading distance. The shortest distance at which a person can observe an object for long periods without fatigue is about 25 cm. The human eye is able to distinguish two lines when light falls on at least two non-adjacent photoreceptors with at least one photoreceptor between them. At a viewing distance of 20 cm, a raster frequency of 60 lines/cm can no longer be perceived as separate lines. This has produced the following resolution requirement, which is also standard in the printing industry:  
A raster of 60 lines/cm requires at least 120 pixels.

120 px/cm x 2.54 cm/in = 304.8 dpi, or approximately 300 dpi

<sup>19</sup> Example: Small-format negatives of 24 x 36 mm have been produced for enlargement on photographic paper. They are not suitable for viewing with the naked eye. The general practical guideline for photographs is that small-format negatives (depending on lens, lighting, film and developer) can be enlarged about 10x, in this case producing a photograph of 24 x 36 cm. To achieve

There are also objects that are not covered by the procedure described above to calculate resolution due to their dimensions.

In photography for collection objects and architectural photography, a resolution of 300 dpi with a DIN A2 output format can be achieved with a modern 50 megapixel digital camera. This resolution is sufficient for a wide range of uses, such as prints of different sizes and online presentation. However, higher resolutions may be required if important details cannot be represented within these parameters (e.g. reproductions of paintings larger than double DIN A0).<sup>20</sup>

Guidelines on advisable deviations from the basic recommended resolution of 300 dpi relative to the original format are given along with the material-specific parameters. (→ [3.2.2](#))

The resolution that can be achieved with a digital capture system is not governed simply by the number of available pixels in a scanning line or sensor. It is dependent on various factors, for example the technical quality of the digital recording device (camera or scanner) and the reproduction capability of the lenses used.

It is possible to determine whether a digital capture system is suitable for a digitisation project by carrying out a test digitisation with standardised test charts (ISO 12233 test chart, USAF 1951 test chart).

The effective resolution of the test images can then be calculated with the help of analysis software or resolution tables, which are often supplied with the chart.

### 3.2.1.2 Colour depth

Colour depth determines the differences between brightness and colour values in a digital image. Because digital technology only uses discrete statuses (yes/no), unlike analogue photography it cannot represent brightness and colour differences continuously with fluid transitions.

- At a colour depth of 1 bit, two different statuses are possible: white and black.
- At a colour depth of 8 bits per colour channel, there are  $2^8 = 256$  brightness levels per colour channel from white to black.
- At a colour depth of 16 bits per colour channel, there are  $2^{16} = 65\,536$  brightness levels per colour channel from white to black.
- Greyscale images with only one channel for brightness have a colour depth of 8 or 16 bits.
- So colour images in RGB mode (with one channel each for red, green and blue) have a colour depth of  $3 \times 8$  bits = 24 bits or  $3 \times 16$  bits = 48 bits.

The advantage of the higher colour depth, of 16 bits per channel, lies in the greater colour differentiation. It means that fewer tone values are lost during subsequent image processing.

---

these target dimensions of 24 x 26 cm as described above in the digital copy of the negative, the negative must be digitised with at least 3000 dpi.

<sup>20</sup> According to the principle that camera pixels / target resolution in dpi equals target dimensions in inches, an area sensor of 5700 x 8600 pixels yields the following output dimensions at 300 dpi:

$8600/300 = 18.7$  in x  $2.54$  cm/in = 72 cm

$5700/300 = 12.7$  in x  $2.54$  cm/in = 48 cm

This is important in the case of digital images that require a lot of processing, for example black and white negatives.

Digital capture systems record colours in RGB mode. The analogue/digital converters in high-performance chip-based scan backs and line scan backs can distinguish between differences in brightness with 16 bits. Most digital small-format cameras and flatbed scanners offer 14 bits.

Digital capture systems generally have a high colour depth of 14 or 16 bits per channel. To retain the information density of the generated image files, the image data is post-processed, ideally with 16 bits per channel, i.e. 48 bits, but at least with 8 bits per channel, i.e. 24 bits colour depth.

To store the final digital master, a **colour depth of 8 bits per channel**, i.e. **24 bits**, is sufficient as most modern output and display devices only support tone reproduction with 8-bit differentiation.

### 3.2.1.3 Digital capture process

#### Capturing technique

There are two main techniques in photographic digital capturing. The first is line scanning, where a trilinear scan line, of a fixed width, scans a defined section, with one line each for red, green and blue. Flatbed scanners and line scan back parts for specialist or medium-format cameras use this technique. The colour information for each pixel is physically generated.

The second common technique is the area sensor, where a defined area is equipped with a certain number of photodiodes.

Line scanners and multi-shot-capable area sensor cameras are mostly used for delicate subjects such as intaglio illustrations, maps and textiles. Line scanners can achieve high resolutions and precise detail with originals up to DIN A0 size. For large two-dimensional originals, double DIN A0 size or more, which are to be captured with maximum detail to ensure readability, preference should be given to line sensors combined with a camera system.

Standard flatbed scanners can be used to digitise two-dimensional media up to DIN A3 size. However, the process is not contact-free as the medium comes into contact with the scanner's glass pane. For this reason it is only suitable for originals to which no conservation concerns apply.

#### Image noise

To reproduce as much detail as possible, the manufacturer's recommendations for the chosen camera technology should be followed. This will also minimise image noise. The following rule of thumb applies: The lower the ISO setting, the lower the noise.

#### Lenses

We also recommend the use of modern high-quality lenses designed for the high resolution potential of digital cameras. Prime lenses are to be preferred over zoom lenses, as they allow imaging errors to be corrected more easily and overhead use will damage the mechanics over time. Zoom lenses always deliver a quality compromise over the available focal length range. Shift lenses reduce losses in image quality caused by the later rectification of falling perspective lines. If the reproduction scale is less than 1:10, special macro lenses designed for this type of

imaging should be used. Subject to the camera system being used, high-quality enlarging lenses may also be used for reproductions on this scale.<sup>21</sup>

### Working area

To prevent blurring, caused by shake, during the digitisation process, digitisation units should be positioned in a location which is not subject to vibration. Wooden floors, for example, are unsuitable because they transmit vibrations easily. If reflex cameras are used they should be operated with mirror lock-up to prevent vibration when the mirror flips up. Tripods and repro stands should be of a suitable size to support the weight of the cameras easily.

### Moiré

In the case of objects with fine, evenly distributed details, using the one-shot technique with a digital camera and area sensor may lead to colour distortion and a moiré effect.<sup>22</sup>

A moiré effect is created in digital imaging when the uniform pixel matrix of a digitally generated image interferes with a regular line structure on the original. This is due to the pixel size of the scan line or area sensor, the reproduction scale, and the line frequency of the original. It is difficult to predict the occurrence of this effect as the parameters involved rarely remain constant.

Colour distortion occurs with one-shot area sensors under the same conditions but is the result of an interpolation error by the camera software. In a one-shot camera system, each pixel is only sensitive to one primary colour: red, green or blue. The remaining colour information for the pixel is interpolated from the colour information for the surrounding pixels. If the colour and brightness differences in the original item are too extreme, colour distortion will result. In cases like these, line scanners or multi-shot digital backs may yield better results as they physically generate the colour information instead of calculating it. However, this does make digitisation more time-consuming as scanning times are longer and absolutely constant lighting is essential during the capture process.

To avoid the moiré effect, a test digitisation should be carried out to identify the most suitable process and required resolution before carrying out digitisation proper.

Objects with which moiré may occur include printed half-tone documents, copperplate engravings and textiles.

### Lighting

The most appropriate lighting should be carefully considered before beginning a digitisation project, with due regard to conservational considerations. Line scan systems require continuous, flicker-free light because the original is continuously scanned for a certain period. The light power and colour must remain constant during scanning to achieve uniform results. With area sensors, a flash gun may also be used. In the case of multi-shot systems the flash power and

---

<sup>21</sup> The reproduction capability of a lens can also be identified from the corresponding modulation transfer function (MTF) diagrams. These diagrams are published by various lens manufacturers. The MTF describes how well a lens can transfer the edge contrast, on an original item, to the photographic image. It, therefore, gives an indication of the resolution capability of the lens. The technical data sheets for the lens also provide information about the brightness and sharpness reduction at the edge of the image. Optical imaging errors are especially noticeable here. To reduce image errors a lens should be stopped down twice to exploit the better reproduction capability of the centre. If depth of focus is not important, for example when photographing flat materials, this optimum working aperture should be used.

<sup>22</sup> Moiré should not be confused with the phenomenon created by the interference of print dots on the rasterised original with the pixels on the monitor.

colour must remain constant from one exposure to another. The same applies to continuous light.

From a conservation point of view, it can be argued that brief exposure to intense light is less harmful than extreme heat, particularly as it is the overall light exposure that counts, i.e. it does not matter whether the same amount of light is distributed over a longer amount of time or falls on the item for a short period. However, strong short-term fluctuations in brightness are problematic for the staff from an occupational-health point-of-view.

### Image processing

Photographs are normally generated in a manufacturer-dependent raw image format (RAW), at maximum size with a colour depth of 14 or 16 bits per colour channel. RAW images represent the original camera data. Image corrections should be undertaken in RAW software if possible. As this file is not changed, it can be corrected again if necessary without loss. Editing rules can also be applied to any number of images as a preset, which significantly increases efficiency.

Adobe RGB, the ECI-RGB v2 colour space recommended by the European Color Initiative (ECI)<sup>23</sup>, or the identical colour space L-Star RGB<sup>24</sup> should be selected as the profile for colour images. CMYK colour spaces, as purely print output profiles, are not suitable. To ensure correct colour matching during subsequent image processing, it is essential to produce a grey or colour scale or colour chart along with the item. If no setting adjustments are to be made, it is sufficient to photograph them just once per series of images, representatively for the whole series.

For brightness and tone value control it is important not to crop the tone value histogram (for black and white) at the sides so as to retain the full range of tone values. Currently, tone values of 95% for black and 5% for white can be printed. The recommended maximum brightness values are:

Black: RGB 16/16/16, greyscale values 90%

White: RGB 232/232/232, greyscale values 10%

For reproductions of two-dimensional items, images with the background removed should be produced such that the whole item is represented with a thin surrounding edge. This is the only way to show that the original has not been cropped in any way. Sharpening (unsharp masking) should only be used in moderation for digital masters. Additional sharpening may be used for different derivatives depending on the application.

Any processing of the image file should be implemented ideally with a colour depth of 16 bits per channel, but at least with 8 bits per channel, i.e. 24 bits colour depth. Once all the necessary corrections have been made, the RAW file is converted into a TIFF file with 24 bit/RGB or 8 bit/greyscale, which serves as the master. If further processing is to be done in the TIFF file, the image can be converted from the RAW file with 48 bit/RGB or 16 bit/greyscale. Once all the necessary processing stages are complete, the final master should be stored as a TIFF with

---

<sup>23</sup> European Color Initiative: <http://www.eci.org>

<sup>24</sup> <http://www.colormanagement.org/de/workingspaces.html>

24 bit/RGB or 8 bit/greyscale and may be considered suitable for archiving. However, the RAW format as a proprietary file format is not suitable for archiving.

For reliable matching of originals with the on-screen representation of the digital image, originals should be viewed under D50 standard lighting for graphic workstations in accordance with ISO 3664:2000. The types of monitors used in the graphic industry provide the most accurate representations in these circumstances. They offer consistent colour rendering over the entire grey axis. The monitors should be calibrated to the appropriate target values, with a colorimeter for a D50 standard lighting environment. Please note that a colour temperature of 5800 K should be set on the monitor instead of 5000 K, because 5000 K on a monitor appears more yellow than 5000 K under D50 standard lighting. The following basic values are recommended:<sup>25</sup>

- Luminance: at least 120 cd/m<sup>2</sup>
- White point: 5800 K with chromatic adaptation
- Gradation: 1.8 gamma or alternatively L\*
- Colour spaces: ISOcoated v2 and ECI-RGB 1.0/2.0

### Digital image postprocessing

Additional post-processing is normally required to optimise the quality of digitally captured images. This should be limited to the necessary correction of colour and tone values. Every effort should be made to avoid object deformations, adding or deleting parts of an object, and special effects such as the use of modifying filters. To increase image integrity, positioning aids or backgrounds may be removed from the image at a later point.

#### 3.2.1.4 File formats

According to current knowledge, image masters of greyscale or colour images should be archived in **TIFF uncompressed**<sup>26</sup>. The TIFF format has been around since the 1980s. It has established itself as one of the most important de-facto standards, and it is expected that all standard programs will continue to support this format. However, this holds true only for so-called baseline TIFFs. The more advanced options of extended TIFFs should not be used for digital masters.

In addition to TIFF, **TIFF-LZW** or **JPEG2000**<sup>27</sup> in its lossless form may also be used as an image master format. However, to store masters in JPEG2000 format it is important to note that only the public-domain parts of JPEG2000 may be used.

In the last few years, the ISO standard JPEG2000<sup>28</sup> and, following the expiry of licences, TIFF-LZW have come into the field of view of memory institutions as efficient compression formats. With regard to long-term archiving, the operators of repositories should carefully weigh the advantages and disadvantages of both formats (TIFF-LZW and JPEG2000). Compression formats are generally more susceptible to image loss and their use should be decided according to a risk-benefit assessment. When selecting a format, its prevalence and market penetration should also be taken into account. Regardless of the fact that large and influential libraries such as the Library of Congress and the British Library use JPEG2000, its prevalence so far is less

---

<sup>25</sup> Cf. recommendations of the Swiss Center of Competence for Media and Printing Technology <http://www.ugra.ch/en>

<sup>26</sup> <http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=686>;  
<http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=1099>

<sup>27</sup> <http://www.jpeg.org/jpeg2000/>

<sup>28</sup> [ISO/IEC 15444-6:2013](http://www.iso.org/iso/15444-6:2013)

than that of the uncompressed TIFF format. The licensing situation is still not fully resolved, but some parts of JPEG2000 have been declared free to use.

The various proprietary raw formats are also unsuitable for archiving master images, especially as they can often only be viewed with the corresponding raw software. The Adobe raw format DNG, which is not dependent on any particular platform or camera, has not entered widespread use and is therefore also unsuitable as an archive format.

JPEG and PNG are generally recommended for publication on the Internet due to their widespread popularity. However, other formats or techniques can also be employed if these are more practical for the presentation of, for example, very large image files. It is important to ensure that only formats and techniques are employed that can be used without special technology or software.

For AV media, a distinction must also be made between archive and user formats. For audio data the Waveform Audio File format (WAV) in connection with Pulse Code Modulation (PCM)<sup>29</sup> has established itself as archiving format. The mp3 format (MPEG-2 Audio Layer III) is the most popular user format. No clear recommendation can currently be made for video formats. However, care should be taken to use only licence-free and open formats. The problem with video files is their enormous size, which is why uncompressed storage is hardly feasible for larger projects.<sup>30</sup>

Digital 3D documentations of digitised or digitally reconstructed 3D objects are stored as 3D models. These models allow to accurately represent the form, texture and visual material characteristics of the original object, or to document and visualise objects that are no longer extant (e.g. destroyed).

Any data format that is taken into consideration should be as robust as possible against damage to the data medium, make efficient use of storage capacity, map the model with a logical structure, permit fast data processing and be in wide use.

OBJ (.obj),<sup>31</sup> an open format, has proven itself as a file format for 3D models. It is supported by many 3D graphics programs and is therefore suitable for the sharing of 3D models across programmes and platforms. Optical material properties (e.g. mirroring, transparency, highlights, etc.) are defined in a separate material file (.mtl), which may also contain information on texturing.

Also noteworthy is Collada DAE (.dae),<sup>32</sup> an XML-based open exchange format for 3D data sets. It makes it possible not only to share models and textures, but also to transmit settings and applied changes from one programme to another.

The X3D (.x3d)<sup>33</sup> format, developed specifically for the visualisation of 3D models within WebGL technology, is suitable for web-based documentation and sharing.

---

<sup>29</sup> Minimum quality: 44.1 kHz sampling rate and 16 bit sampling depth

<sup>30</sup> The following formats are recommended for long-term archiving, although it should be noted that no satisfactory solution is yet available: MJPEG2000/MPEG-4 (ISO/IEC 14496-12:2015, ISO/IEC 15444-12:2015), DPX (standardised as SMPTE 268M-2003, v 2.0), MXF/AAF (standardised as SMPTE 377M and proposed as ISO standard). FADGI is currently working on guidelines for archiving films: [http://www.digitizationguidelines.gov/guidelines/MXF\\_app\\_spec.html](http://www.digitizationguidelines.gov/guidelines/MXF_app_spec.html)

<sup>31</sup> <http://www.fileformat.info/format/wavefrontobj/egff.htm> and [https://en.wikipedia.org/wiki/Wavefront\\_obj\\_file](https://en.wikipedia.org/wiki/Wavefront_obj_file)

<sup>32</sup> <https://www.khronos.org/collada/> and <https://en.wikipedia.org/wiki/COLLADA>

<sup>33</sup> <http://www.web3d.org/x3d/what-x3d> and <https://en.wikipedia.org/wiki/X3D>

## 3.2.2 Material-specific parameters

### 3.2.2.1 Text-based works

For our purposes the term “text-based works” refers to both printed works and rare documents such as manuscripts and archival material.

Image digitisation should always be preferred for the digitisation of historical holdings. Even if a machine-readable full text is available, image digitisation and/or the presentation of the digital facsimile should not be omitted because a great deal of information can only be conveyed visually.

The legibility of the text is the deciding factor in the choice of resolution, and this depends on the size of the type or script. Hence, the required scan resolution depends less on the dimensions of the item and more on the legibility of the individual letters. For text documents where the smallest significant character is 1.0 mm or larger, a resolution of 400 dpi is recommended. A resolution of 300 dpi is only recommended if the minimum character size is 1.5 mm or larger.<sup>34</sup>

Folios / pages of the work are always digitised in their entirety with a thin surrounding edge to indicate that nothing has been cut off from the original.

### 3.2.2.2 Graphic representations

As described under [3.2.1.1](#), the recommended minimum resolution for digitisation is 300 dpi, subject to the outcome of a test digitisation with a standardised test chart.

In the case of small items, however, a resolution of 300 dpi relative to the original format will often be insufficient to reproduce the characteristics of the original in consistently recognisable detail. If the artistic technique employed is not identifiable for example, a higher resolution should be selected. Examples might include copperplate engravings, stamps, portrait medallions and miniature paintings.

The highest resolution that can be achieved depends on the digitisation technique used. Here is a sample calculation, disregarding the lens quality: A stamp can be reproduced with a digital camera with a sensor in 24 x 36 mm format at a scale of 1:1. If the camera resolution is 3800 pixels/cm, this gives an image resolution of 4021 dpi.

Conversely, in the case of large items (DIN A0 or larger), the resolution can be reduced in relation to the original format if the objects are intended to be viewed from a greater distance. Posters are one example, as they are often designed to have an impact from far away. In this case the resolution can be reduced to as little as 150 dpi because as viewing distance increases, so does the size of a dot that can no longer be distinguished by the human eye. However, if large-format items are very detailed, for example large topographical maps or copperplate engravings, a resolution of at least 300 dpi should be used.

Due to the added user value, the full resolution of the chosen capture technique should be used for graphic representations if economically viable.

---

<sup>34</sup> See Federal Agencies Digitization Guidelines Initiative (FADGI): Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files, August 2010: [http://www.digitizationguidelines.gov/guidelines/FADGI\\_Still\\_Image\\_Tech\\_Guidelines\\_2010-08-24.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image_Tech_Guidelines_2010-08-24.pdf), p. 59.

Graphics are always stored in their entirety with a thin surrounding edge to indicate that nothing has been cut off from the original.

### 3.2.2.3 Photographs

A distinction must be made between transparent media (negatives or slides) and media intended for viewing (positive images, for example on paper). The latter are normally contact copies or enlargements of negatives on photographic paper.

#### Transparent media

A photographic negative or slide is the result of a photographic capture and therefore constitutes the original source. Negatives are used to make reproductions for disseminating the picture as a photographic positive or print.

Slides allow the quality of a scan to be visually compared with the original with the aid of a standard illumination light box. For colour and tone value correction, the image file should be available ideally in 48 bit/RGB, but at least in 24 bit/RGB. If it is not intended to make any other modifications to the file, the final result is archived in 24 bit/RGB. If the main priority is to document the condition of the original at the time of scanning as the digital master, a recommended alternative is to store the image with 48 bit/RGB, especially if the original shows signs of ageing, which require extensive correction for the various derivatives.

Negatives require a different approach as they are not suitable for viewing the subject. To make the subject usable, a positive derivative must be generated from the digital master. This serves as a digital print of the negative and reproduces the image content. However, the positive representation of the master is usually too dull and, in the case of colour negatives, not colour-true because of the colour mask. To clearly reproduce the subject in the digital print, considerable image correction is therefore required. If the main priority is to document the condition of the original at the time of scanning as the digital master, a recommended alternative is to store the image with 48 bit/RGB or 16 bit/greyscale. In the case of black and white transparent media, storage in greyscale is sufficient. Storing colour information for black and white originals is only justified if important information about the item is conveyed by colour, for example retouching of the negative.

Well-prepared negatives yield images with a tonal range of up to 12 aperture stops, which must be reproduced during digitisation.

Unlike graphic representations, photographic images are technically generated. The quality of a photograph is therefore dependent on a number of different technical factors: image format, lens, film type, granularity of the emulsion, development, exposure, focus and so on.

To take account of these aspects when choosing the most suitable scan resolution, the following alternative to the procedure described in [3.2.1.1](#) is a possibility:

Apart from the image format, the image parameters are generally not known or documented. It may be possible to identify the film material from clues such as notches in the flat film. Glass negatives, however, do not have this feature.

To do justice to the various qualities within a mixed collection, a resolution of 80 lines/mm may be selected as a starting point to determine the scan size. This corresponds to the resolution of modern fine-grain film. This ensures that the full detail of both historical and more recent

photographs can be reproduced in digital form. In most instances the resolution values of modern films are documented in manufacturers' data sheets. A scan resolution of 4000 dpi thus provides sufficient detail.<sup>35</sup> In terms of the film material, the theoretically achievable resolution would be constant for all output formats.

For small-format pictures a scan resolution of 4000 dpi can be used as the reproduction capability of small-format lenses is usually better than that of medium- or large-format lenses. For larger image formats, the scan resolution can therefore be reduced. A useful indicator is the maximum circle of confusion diameter of a dot, which is still perceived as sharp by the human eye for the particular format. As a general rule for photographic sources, the format diagonal in mm (corresponding to the normal focal length) is multiplied by 1/1500. This gives the following maximum permissible circles of confusion for the various formats:

Small format: 0.03 mm; medium format: 0.05 mm; 9 x 12 cm: 0.1 mm; 18 x 24 cm: 0.2 mm

This in turn yields the following target values:<sup>36</sup>

Medium format:	$4000 \text{ dpi} \times 0.03/0.05 = 2400 \text{ dpi}$
9 x 12:	$4000 \text{ dpi} \times 0.03/0.1 = 1,200 \text{ dpi}$
18 x 24:	$4000 \text{ dpi} \times 0.03/0.2 = 600 \text{ dpi}$

Photographs are always stored in their entirety with a thin surrounding edge to indicate that nothing has been cut off from the original.

### Media intended for viewing

The photographic positive is an end product. Unlike a negative, it is not intended to serve as a starting point for reproduction or enlargement. The quality of the image depends on the quality of the negative. Unlike with negative films, there is no reliable data on the resolution capability of photographic papers (measured in line pairs per mm, lp/mm). However, the resolution capability of the photographic emulsion in the positive is sufficient for sharp image reproduction when viewing with the naked eye. The minimum scan resolution selected for digitised copies should therefore be that which can reproduce the exact image quality of the original when output for print or reproduced onto photographic paper.

As described under [3.2.1.1](#), the recommended minimum resolution for digitisation is 300 dpi, subject to the outcome of a test digitisation with a standardised test chart.

In the case of small items, however, this resolution will often not be sufficient to reproduce the characteristics of the original in consistently recognisable detail. In the first half of the 20th century it was common practice to produce only contact copies even from 6 x 9 cm negatives.<sup>37</sup>

Conversely, in the case of large items (DIN A0 or larger), the resolution can be reduced in relation to the original format if the objects are intended to be viewed from a greater distance. Large-scale enlargements are one example, as they are often designed to be viewed from far

<sup>35</sup> To depict the film resolution digitally in lines/mm, at least two pixels are required per line. Thus, 80 lines/mm corresponds to 160 pixels/mm or 1600 pixels/cm. Multiplied by 2.54 (1 in = 2.54 cm), this gives a scan resolution of 4064 dpi, which can be rounded to 4000 dpi.

<sup>36</sup> These values are also specified by the FADGI. See Federal Agencies Digitization Guidelines Initiative (FADGI): Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files, August 2010: [http://www.digitizationguidelines.gov/guidelines/FADGI\\_Still\\_Image-Tech\\_Guidelines\\_2010-08-24.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf), p. 60.

<sup>37</sup> For non-transparent photographs in 4 x 5 in format, FADGI therefore recommends a resolution of 800 dpi. See Federal Agencies Digitization Guidelines Initiative (FADGI): Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files, August 2010: [http://www.digitizationguidelines.gov/guidelines/FADGI\\_Still\\_Image-Tech\\_Guidelines\\_2010-08-24.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf), p. 62.

away. If the image is viewed too close up, the film grain is clearly visible. In this case the resolution can be reduced to as little as 150 dpi because as viewing distance increases, so does the size of a dot that can no longer be distinguished by the human eye.

However, in the case of portfolios with historical photographic positives where the aim is to document the complete presentation including the mounting card, a reduction of the resolution should be avoided in order to keep captions and details identifiable. With large-format portfolios the mounting card often takes up 50% or more of the total area.

Storing colour information for black and white originals is only justified if important information about the item is conveyed by colour, for example tinted prints.

Due to the significant added user value, the full resolution of the chosen capture technique should be used for transparent and viewing photographs if economically viable.

Photographs are always stored in their entirety within the file and a slight surrounding edge is indicated.

#### 3.2.2.4 Microforms

For microforms (16 mm and 35 mm microfilm, 105x148 mm microfiche, positive and negative, B/W and colour), in terms of parameters the information given in the section on transparent media applies. However, with regard to resolution it should be remembered that these are reproductions of originals for backup or protective purposes. A reduction factor of between 1:7.5 and 1:96 may be used for filming. Wherever possible the resolution for digitisation should be based on the original and not the film.

The digitisation of microfilm is often used as a means of mass digitisation, which can be achieved at low cost. The recommended 300 dpi relative to the original size can only be achieved in very rare cases, even if the film's resolution is theoretically adequate. The limiting factor is the generally prevalent types of scanner, which can digitise whole films semi-automatically in a short space of time. For B/W microfilms and microfiche, resolutions of up to 600 dpi relative to the film can be achieved. The choice of resolution must be based on the technical possibilities of the mass process. Reproduction techniques that allow high-resolution individual scans should only be used in exceptional cases.

If film microforms that conform to the standards of federal security microfilming (see section 2.3) are available and used as a basis for digitisation, it is necessary to check whether it makes sense to create and preserve a master. Microfilming itself is in this case a long-term storage medium and copy master. However, it must always be investigated whether it makes more sense to create digital copies (master and user versions) of the original than to generate digital copies for users from the microfilm.

#### 3.2.2.5 Three-dimensional objects

Due to the advances in information technology over the past few years, the digital capture and sharing of 3D objects in the field of cultural heritage is becoming increasingly important. **Digital 3D documentation** primarily records museum collections, as well as the architecture and its features.

Until recently, this was done mainly by taking photographic images from different angles. But today, a proven selection of 3D capturing and reconstruction methods as well as technologies supporting these processes are available.

In contrast to photography, 3D documentation captures the entire geometry of an object, its surface texture and, where possible, its visual material properties, and combines all of it and integrates it into the digital 3D model.

A decisive advantage over a photograph is that the shape and the surface-light interaction of the object can be captured and reproduced faithfully. The 3D model can be visualised and simulated from any perspective in the original lighting situation, as well as in any new lighting situations and environments. One open-source application for the interactive web presentation of high-resolution 3D models is the 3D Heritage Online Presenter (3DHOP).<sup>38</sup>

In digital 3D documentation, a distinction can be made between **retrodigitisation** as a result of the transfer of a physically existing object to a digital copy, and **reconstruction** of a physically non-existent object in the form of a native digital 3D data set.

The purpose of retrodigitisation and digital reconstruction is primarily the indexing, documentation, safeguarding and archiving, as well as the web-based provision of cultural heritage for researchers and the general public. Digital 3D models have the following advantages:

- Digital 3D models of cultural objects can be made easily available and can be accessed by many researchers at once (even in different qualities)
- Missing (e.g. destroyed) parts can be added in reconstruction
- Different hypothetical versions and/or variants of a 3D object can be simulated
- Digital 3D models can be used by museums for exhibition planning, documentation, acquisition planning etc.
- Digital 3D models can be used for virtual presentation and exhibition to the public (in combination with new presentation techniques such as hybrid exhibits) in order to transfer knowledge and increase attractiveness to visitors
- Digital 3D models are a suitable reference for the restoration of damaged originals / generation of physical replicas on the basis of the digital 3D model
- Digital 3D models may be a substitute for loans (avoidance of damage, insurance costs, legal uncertainty relating to ownership)

On the other hand, digital 3D models present new challenges. To date, no standards have become prevalent in the area of software-independent data formats and of the documentation of digitisation / modelling and results. This constitutes a lack of interoperability. In addition, there is a lack of adequate digital research and information infrastructure, which calls into question the sustainability of 3D data sets. This means that there is a great need for action, which is currently being defined, inter alia, by the Council for Scientific Information Infrastructures (RfII)<sup>39</sup> constituted in November 2014.

For the **retrodigitisation** of 3D objects, different technologies are available, depending on the type of object and its properties (size, material, condition, etc.). Currently, the project Colour and Space in Cultural Heritage (COSCH)<sup>40</sup> is trying to develop an online guide for selecting an adequate methodology and technology depending on the task. However, making early contact

---

<sup>38</sup> <http://3dhop.net/index.php>

<sup>39</sup> <http://www.rfii.de/de/index/>

<sup>40</sup> <http://cosch.info/>

with leading technology developers and users in the area of mass digitisation, such as CultLab3D,<sup>41</sup> continues to be important.

There are various optical digitisation methods.

Laser scanning methods are divided into run length measurement and triangulation. In the first case, the time it takes a laser pulse to travel to the object and back again is measured in order to determine the distance to this point. In the second case, the three-dimensional surface of an object is extrapolated from the distortion of a projected laser line as captured by a camera.

Strip light methods use the projection of defined patterns (mostly parallel lines) on object surfaces and likewise extrapolate 3D surface geometry from distortion. To increase resolution, these patterns can be phase-shifted across surfaces.

In photogrammetric methods, pictures of the object to be digitised are taken from different angles and scanned by pairs of features for at least two overlapping images, so that the depth of the associated image point is obtained by triangulation. Characteristic features can be, for example, differences in intensity or contrast.

The principle of time-of-flight cameras is similar to run length measurement in laser scanners, with the difference that an entire scene is illuminated and photographed at once, and a depth value is calculated for each pixel of the camera sensor.

For objects above a certain size, for instance large statues or buildings, a point-by-point measuring technique such as laser scanning or surface coding is no longer efficient, and photogrammetric techniques demonstrate their strengths.

In retrodigitisation, there is a general relationship between object size and resolution, or to be more exact between the measurement range of a sensor and accuracy. The larger the object or the greater the distance, the less precision with which details are recorded.

Geometry and texture are normally recorded in the same step as geometry acquisition, because there is a direct correlation between the measurement points and the points of the imaging system. Many acquisition methods use a camera sensor as an integral component of image capture to extract depth information. This sensor usually has the best viewpoint for each captured section of the object, so the same sensor position can be used for texture capture.

To capture the texture, sufficient, diffuse and even lighting is required. The right white balance is also important as the mostly artificial diffuse lighting on the object surface must be compensated for. Colour calibration is important in this respect as texture has a major influence on realistic appearance.

In addition to geometry and texture, optical material properties, such as gloss and reflection, can also be captured. Typically, this involves recording every combination of light incident and perspective onto an object, in order to subsequently store the textures relevant for each surface point (visibility test). Later, during interactive viewing of the virtual object, the physically correct textures can thus be applied under any illumination and the original appearance of the material and light-surface interaction can be reproduced. In some cases, an object cannot be captured in its entirety, either because of its size, or because only remains of material are extant. In such

---

<sup>41</sup> <http://www.cultlab3d.de/>

cases, selective capture of material properties and subsequent extrapolation to regions of the same or similar material may be appropriate.

The **hypothetical reconstruction** of non-extant (e.g. destroyed) 3D objects is based on the acquisition and interpretation of the historical source material. Computer-aided reconstruction uses established 3D modelling and animation software (e.g. Autodesk 3ds Max and Maya, Cinema 4D, SketchUp, Blender, etc.) to create 3D models. As a complementary method to retrodigitisation, reconstruction offers the possibility to document and visualise that which no longer exists. Significant in this context is the problem of how to present fuzzy knowledge and hypotheses. Of importance are furthermore urgently required documentation standards for interpretative modelling and visualisation, including, among other things, source references and the mapping of the creative-interpretative processes.

“Digital indexing [...] is aimed at integrating and converging digitised and native digital data into uniform, integrated work environments with the objective of dynamic knowledge integration”,<sup>42</sup> notes the Council for Scientific Information Infrastructures. An important contribution to integration and convergence is made by structured data geared to relevant metadata schemes (e.g. LIDO and CARARE 2.0), controlled vocabularies (e.g. Getty Vocabularies) and standard files (e.g. GND), as well as reference ontologies (especially CIDOC (International Committee for Documentation) CRM). The extension of CIDOC CRM for better capturing of retrodigitisation was implemented in CIDOC CRM<sub>dig</sub>. CIDOC-CRM-referenced application ontologies for knowledge representation of computer-assisted reconstruction are currently being introduced to and discussed by the community of experts.<sup>43</sup>

### 3.3 Metadata

The generation of metadata to enable the location of objects and the contextualised presentation of digital images is a central aspect of digitisation. The DFG assumes that the analogue objects selected for digitisation have already been indexed in recognised digital indexing systems and/or will be indexed in detail as part of the digitisation process. Metadata produced as part of the digitisation project should be made available in a **software-neutral** and **standard-compliant form**, usually **XML coding**. This task should be integrated into the project workflow in such a way that, even if the project is terminated early for any reason, a complete set of metadata will be available in a software-neutral format.

If the objects selected as part of a funded digitisation project are factually suitable for inclusion in a material- or subject-specific portal, the project proposal is expected either to outline what measures will be taken in the project to facilitate a data integration to this portal during and after the project, or to explain why a data integration is not necessary or desirable for reasons of content or time and expense.

Generally speaking, a distinction is made between descriptive metadata (bibliographic description, archival indexing, description of rare (frequently non-textual) objects), structural metadata (text or document structure), administrative metadata (e.g. rights management) and technical metadata (e.g. file types). The following considerations apply only to descriptive and structural metadata.

---

<sup>42</sup> Rfll Berichte No. 1: Begriffsklärungen (2016), p. 13, at <http://www.rfii.de/de/category/dokumente/>

<sup>43</sup> <http://www.digitale-rekonstruktion.info/>

The link between an object and metadata and between an object and digital images must always be assured at metadata level. Metadata may also be embedded in the headers of digital images. Since headers are depicted differently by different software products and in the worst case scenario may even be corrupted this is only an extra option.

Metadata formats reference the digital images in various ways. The container format METS allows associating descriptive metadata via mappings in standard formats (e.g. MODS, TEI), as well as structural metadata including references, to digital images. This is especially suitable for fully digitised text-based works which are indexed with both descriptive and structural metadata. Other formats like **LIDO** include, in addition to semantic elements for object description, specific elements for referencing digital resources for the object (both images and audio or video data). Any number of resources can be linked with an object and given their own descriptive elements. **EAD(DDB)** also allows an object to be linked with an unlimited number of digital resources, including integration of METS containers. It is important that digital resources described within metadata are referenced with globally unique persistent addresses, generally URLs (see “Persistent Addressing” in section 5).

For library stock, digital copies and metadata must be recorded either by **cataloguing the electronic version** or by giving the **PURL of the image files** or a **persistent link in the catalogue** (OPAC, network system). University institutions submitting proposals are expected to at least coordinate cataloguing with their local libraries, or actually have it done by them.

Digitised prints should be listed in the **ZVDD**. Digital copies of other library objects should be listed in relevant material-specific portals (e.g. Manuscripta Mediaevalia for medieval manuscripts,<sup>44</sup> Kalliope for remains<sup>45</sup>). Materials that cannot be added to library indexing systems should be presented in **suitable subject-specific or multi-subject online applications**. For archive material, digital copies and metadata must be listed in Archivportal-D or in the DDB. All projects are expected to add the data to the **DDB** and via the DDB to **Europeana**.

It is up to the funding recipient to ensure that the digitisation units produced by the project are uniquely identified and can be searched and retrieved separately from other units recorded in the same system.

Data should be delivered to these portals in accordance with standard formats, if possible via **OAI** (see section 3.2.1.4).

### 3.3.1 Indexing, descriptive metadata

The indexing of objects to be digitised with descriptive metadata fulfils different functions depending on the material. The descriptive metadata is what allows the content to be located by a search. The object is classified and placed in its historical context and a material description is added if necessary. Ideally, descriptive metadata should provide links for different enquiries and disciplines. Funding is not available for digitisation projects that do not include metadata classification in line with established community standards.

If a digitisation project is intended to include not just existing descriptive metadata but also project-specific indexing and/or information provision (which will usually be the case for rare

---

<sup>44</sup> <http://www.manuscripta-mediaevalia.de/>

<sup>45</sup> <http://kalliope.staatsbibliothek-berlin.de/en/index.html>

objects at least), the most important enquiries relating to the material should be anticipated and listed in a core field catalogue. The systematic preparation of metadata on the basis of a core field catalogue is essential to the optimum provision of digitised holdings as described in section 5.2.

To ensure the optimally distributed and long-term usability of metadata, indexing should be based on **relevant standards (e.g. RDA<sup>46</sup>, ISAD (G)<sup>47</sup>)** and **reference models** (CIDOC-CRM<sup>48</sup>, possibly also IFLA FRBR/FRBRoo<sup>49</sup>), and linked with **published standard data** wherever possible. To record personal biographical information and for geographic indexing we expressly recommend the **Standard Authority File (GND)** offered by the German National Library. Other controlled vocabularies such as **Iconclass<sup>50</sup> for image classification** should enable national and international links.

Metadata must be made available for further use in accordance with **material-specific standards: METS/MODS** for printed text-based works (see Appendix A), **METS/TEI** for manuscripts (see Appendix B), **EAD(DDB)<sup>51</sup> in connection with METS/MODS mapping** for archive material (see Appendix A), and **LIDO** for (usually rare) pictographic and three-dimensional objects (see Appendix C). The metadata must be valid for the relevant XML schema and must also be checked for semantic correctness.

### 3.3.2 Structural metadata for digital facsimiles

It is worth considering the use of structural metadata to index images, i.e. coding the structural elements of a document such as dedications, prefaces, chapters or illustrations. The inclusion of these aspects takes its cue from analytical bibliographies, which break down the contents of a work along the lines of its chapter and text structure. In some cases the production of structural metadata is of somewhat lesser importance, while in others it is essential to create this artificial table of contents to enable the user to navigate within the digital document. For example, nobody should have to go to the trouble of scrolling through a 600-page digital dictionary to find the right place in the alphabet. For some enquiries, structural metadata are also of interest for search purposes. Therefore the decision whether to generate structural metadata is always **material- and object-specific**.

If structural metadata is used, it is recommended to consult the **structural data list** available on the **DFG Viewer** website. If additional designations are needed, standardised terms for a given digitisation project should be agreed on. This typically specialised vocabulary should be published on the project's website and if appropriate the DFG Viewer website to allow others to reuse it.

When assigning structural metadata, the question arises as to whether document indexing should follow the digital facsimile, the physical page sequence, or the work's text and/or chapter structure. If a transcription or edition will accompany the digital facsimile of an old print or manuscript, the recommended **encoding standard** is **TEI**. For page description with some qualifying features (e.g. illustrations or annotations) the standard **METS**, which is administered

<sup>46</sup> <https://wiki.dnb.de/display/RDAINFO/RDA-Info>.

<sup>47</sup> [http://www.ica.org/sites/default/files/CBPS\\_2000\\_Guidelines\\_ISAD%28G%29\\_Second-edition\\_DE.pdf](http://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_DE.pdf).

<sup>48</sup> CIDOC Conceptual Reference Model: <http://cidoc-crm.org/>

<sup>49</sup> IFLA Functional Requirements for Bibliographic Records: <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>

<sup>50</sup> [www.iconclass.nl/](http://www.iconclass.nl/)

<sup>51</sup> <http://www.landesarchiv-bw.de/web/53401>.

by the Library of Congress, is recommended. There are good arguments for both the page-oriented and the document-oriented model. Usually it is even possible to merge both approaches. A structure that follows the logic of the text tends to be more powerful for retrieval and the display of digital representations. However, this advantage comes with the price of greater technical requirements for processing and displaying the digital representations. It should be pointed out that even encoding on the basis of the physical page sequence, which tends to be more common in libraries, does not rule out the use of TEI,<sup>52</sup> so that it may be possible to effectively combine both aspects.

The standards currently recommended for old prints are **METS** or **TEI**. However, the METS-based **DFG Viewer** should be supported in all cases. Therefore, if TEI is used for structural data, the project must convert it to METS as well. Since both standards are XML-based and the described features are similar the conversion is feasible.

### 3.3.3 Description of collections and holdings

Descriptions of collections and holdings contextualise and locate the individual object and thus enable the user to get an overview of the overall holdings of an institution. The basis for a digitisation project can also be virtual holdings and collections (e.g. subject-specific, material-specific).

Digitisation projects are expected to present **at least the nature and scope of the selection of materials and/or objects on a website**, preferably with an **English-language version**. A **standardised description in XML** is desirable to facilitate the future merging of this information in national or international portals that enable the respective retrieval. This description may be based on the **Dublin Core Collections Application Profile** (see Appendix D) or on the same metadata standard in which the object descriptions are made available: **METS, MODS, TEI headers, EAD(DDB)** and **LIDO** all offer the necessary features. A unique identification and description according to the ISO 27730 standard, Information and Documentation – International Standard Collection Identifier (ISCI), which is based on an institution's ISIL, should be considered as well.

### 3.3.4 Exchange and dissemination of metadata

For the development of comprehensive reference systems it is crucial to create a global **standard for exchanging metadata** of digital copies (see section 6). However, standards can be developed and established only within the respective community. One and the same resource may well be relevant to entirely different research questions. Accordingly, digital resources require diverging sets of metadata. A generalised procedure for exchanging metadata must therefore be able to deal with different metadata formats and community-based specifications. The protocol of the **OAI** allows for flexible use of different metadata formats and community-based specifications. With regard to old prints and manuscripts, OAI is especially useful as a technical exchange protocol. OAI requires that **Dublin Core data** be provided as a minimum. While Dublin Core data are insufficient for content-oriented descriptions of old prints, manuscripts<sup>53</sup> and non-text-based objects<sup>54</sup>, they are useful as additional information. The OAI standard explicitly enables the **parallel support of additional metadata formats**, so that

---

<sup>52</sup> [http://www.tei-c.org/Sample\\_Manuals/bestpractice.htm](http://www.tei-c.org/Sample_Manuals/bestpractice.htm)

<sup>53</sup> See Hillmann, Diane I.: Choices: MARC or Dublin Core? In: Anne R. Kenny/Oya Rieger (eds.): Moving Theory into Practice. Digital Imaging for Libraries and Archives. Mountain View: Research Library Group 2000, p. 89f.

<sup>54</sup> The impermissibly simplified description of museum objects in Dublin Core and their representation in online environments was the main reason why the museum community developed the harvesting format LIDO.

OAI generally can be combined with all XML-based metadata formats (METS/MODS with sector-specific mappings, METS/TEI-HSS, MARCXML, MABxml, EAD(DDB), TEI P5, LIDO, etc.).

The DFG requires that users be provided with **OAI-compliant** metadata. This can be realised via a portal (e.g. ZVDD, DDB, Archivportal-D). In addition to the **Dublin core metadata** required by OAI, material-specific metadata must be provided in **METS/MODS** for old prints, **METS/MODS** for archive material with a mapping addressing the DFG Viewer, **METS/TEI** for medieval manuscripts, and **LIDO** for pictographic and three-dimensional materials (see also section 7 and Appendix A). Any deviation from this requirement must be explicitly justified in the proposal.

### 3.4 Full text generation

A versatile scientific reusability of digitized texts which permits automated searches, quantitative evaluations with text or data mining, semantic analysis, pattern recognition in non-text-based materials, enrichments, contextualisations and further processing – in both physical and virtual research environments – depends on the easy and unhindered usability of the data, appropriate legal permissions and the availability of the full digitised text. Whenever feasible and reasonable, DFG funding for text-based works should be used to generate machine-readable full texts. The OCR-D Coordination Project,<sup>55</sup> which tests the technique of OCR-supported full-text capture and develops prototypical workflows, should be considered when planning full text generation via OCR. The results of this project will be incorporated into the next version of the Practical Guidelines. Please note: For **printed works dating from 1850 or later**, the **full text must be generated** in addition to image digitisation.

Full text can be generated in two ways: via **OCR** or **transcription**; which method to choose depends on a number of factors, including the age and condition of the master copy and the acceptable error tolerance considering the intended purpose.

#### 3.4.1 Text capture

##### Text accuracy

Whether a text is being generated via OCR or manual transcription, it is important to decide what quality is needed for what purpose and what costs are appropriate. Text quality requirements will vary from one project to another. An edition project will impose the highest standards, while a mass digitisation project involving thousands of titles will have to do without the comparatively expensive process of manual transcription. What standard is considered adequate and what costs are acceptable for what quality level must be carefully justified on a case-by-case basis depending on the material and scholarly requirements. The quality of OCR-generated texts is usually given as a percentage. There is no consensus as to the measurement criteria and processes which should be applied. Accuracy may refer to the correctness of individual letters or of whole words. In the first case, 99% means that 1 in 100 letters is wrong, and in the second case 1 in 100 words. Whether incorrect layout information (marginal notes correctly recognised but placed in the wrong position) or missing syllabification are considered errors depends on the project and its criteria.

---

<sup>55</sup> OCR-D: Coordination Project for the Development of Methods for Optical Character Recognition (<http://www.ocr-d.de/>)

To create some uniformity on which to base the evaluation of **accuracy**, applicants are requested to state the **letter accuracy**, i.e. incorrect syllabification and layout errors should be ignored. Applicants should therefore check how many characters in the source text were correctly recognised, including punctuation characters. Ideally, measurements should be made on the basis of reliable reference texts, but as these are not always available, **random sampling** may have to be used.

For now, a **statistical method** is recommended to check the accuracy of transcribed or OCR-generated texts. Future findings, e.g. based on the results of the OCR-D project, might lead to a different recommendation. Statistical methods should be used to assess whether the recognition rate claimed by a service provider can be relied upon. The tests should be based on a random sample. The probability of error should be kept as low as possible while keeping the size of the random sample manageable. The recommended statistical method is the so-called **Bernoulli trial**. As the calculation is fairly complicated, we recommend to use the following specification. The recommended basis is a random sample of 500 characters. To select the position of the characters we recommend the use of a random generator (1st character: page 15, line 24, character 7. 2nd character: page 73, line 3, character 32 etc.). To check the accuracy, the following table applies:

Claimed recognition rate	Minimum number of correctly recognised characters (random sample size = 500)
95%	485
96 %	489
97 %	493
98 %	496
99 %	499
Over 99%	500

The left column shows the claimed recognition rate. The right column shows the minimum number of correctly identified characters that must occur in the random sample to verify that a claimed recognition rate is correct. Hence, if a service provider claims a text has an accuracy of 96%, at least 489 characters must be correctly recognised in the random sample of 500 characters with a 2.5% probability of error in order for the service provider's claim to be accepted. An accuracy of less than 95% should preferably not be accepted.

If the recognition rate is claimed to be more than 99%, the size of the random sample would have to be increased. The two tables specify the minimum number of correctly recognised characters in relation to the size of the random sample for texts with claimed recognition rates of 99.5% and 99.7%:

Claimed accuracy: 99.5 %

Random sample size	Minimum number of correctly recognised characters
500	500
1000	999
2000	1996
5000	4985
10000	9960

Claimed accuracy: 99.7 %

Random sample size	Minimum number of correctly recognised characters
500	500
1000	1000
2000	1998
5000	4995
10000	9990

It goes without saying that there are pragmatic limits to sampling. The question as to what level of accuracy is good, sufficient or insufficient depends on the use of the text and the concrete project requirements.

### OCR

OCR technology has made considerable progress in recent years, and significant improvements have been made on Fraktur fonts, both from the hand press and the machine press era. However, considering the dynamic development and the results expected from the OCR-D project, these Practical Guidelines cannot make any conclusive recommendations concerning OCR software and its usability for now.

The OCR process itself consists of the following stages that are based on each other: Preprocessing to prepare the image (cropping, despeckling, deskewing, binarisation), optical layout recognition (segmentation or identification of image and text parts, structural analysis), optical character recognition (actual text recognition), and postprocessing (text correction). The quality of the original image has a decisive impact on the binarisation process, which in turn is the basis for quality of character recognition; for this reason only digital images of sufficient quality should be used for OCR processing. Problems may arise as a result of intrinsic phenomena such as dirty marks, shadows of the reverse side, manual underlining and annotations, which could have an adverse effect on the OCR process. On the basis of binarisation the software identifies the text areas in an image and separates the actual text from illustrations or other pictorial elements (=segmentation). Marginal notes in older printed works and complex newspaper layouts may interfere with the segmentation process. Binarisation and segmentation are the basis for the actual OCR process of text recognition, which may be supported by a subsequent process of text improvement (automatic use of dictionaries or manual correction).

The ALTO (Analyzed Layout and Text Object) standard, which is maintained by the Library of Congress, is recommended to facilitate the use of data from OCR.<sup>56</sup>

### Double keying

There are two methods used to transcribe texts: the single-key and the double-key method. In the latter, a text is transcribed twice, then the two versions are compared automatically and any discrepancies are filtered out. This allows for transcription accuracies of up to 99.97%, i.e. virtually error-free texts. When choosing this type of transcription, one should not be misled by

---

<sup>56</sup> <http://www.loc.gov/standards/alto/>

service providers claiming ostensibly high accuracy rates; results with less than 99.5% accuracy are inadequate for manual transcription.

If transcription is to be outsourced to a service provider, the contract must specify the contracted accuracy rate. Adherence to the contracted rate should be verified by random checks of the digitised text (see above).

Manual transcription has the advantage of high accuracy but is costly. Advantages and disadvantages compared to OCR capture must be weighed. The manual transcription is often done outside of Germany; however, a contracted digitisation provider should have a representative in Germany since close cooperation and discussion on transcription details are usually needed.

As a first step, a digitisation project must determine which properties of the master copy should be captured by a structural markup. Only features that are graphically distinct can be marked up. Simple structures can be recognised automatically by the service provider; further details must be marked in the images before the materials are handed over to the contractor. This requires a certain amount of labour, which must be taken into account when calculating the project budget.

Because most service providers invoice based on the number of characters including markups, it is advisable to use a markup language with few characters<sup>57</sup> for this purpose.

### 3.4.2 Character encoding

All common operating systems support Unicode. Unicode is also the character encoding format for XML, which forms the basis of the most important structural data markup systems. It is therefore recommended to save the texts in **Unicode**. Preference should be given to **UTF-8**, which is more economical with European languages. Characters which are not found in the Unicode standard can be depicted by using the private use areas in Unicode<sup>58</sup> and represented by appropriate graphics or fonts. Options for standardisation should always be investigated.

The question of whether to encode the long and short s in Fraktur fonts, ligatures in Fraktur fonts (ch, tz etc.) and diphthongs (æ etc.) depends on subject-specific requirements or editorial decisions which are beyond the scope of these recommendations, but which should be borne in mind when generating full text and should ideally be documented in the encoding description of a TEI header.

### 3.4.3 Markup of full texts

Unless there are cogent reasons not to, **full texts of prints and manuscripts** must be **encoded and marked up** following the model of the **TEI**. As a transparent XML format, TEI is also prospectively the best choice for long-term archiving as long as it is carefully documented. PDF/A should be avoided, in spite of existing ISO standards (19005-1:2005 and 19005-2:2011), since its use, especially in the digitally working humanities and cultural studies, is limited by the lack of structural markup. However, as a derivative format, PDF – and increasingly ePub for mobile devices – is suitable for dynamically generated reading versions or print-ready text, and

---

<sup>57</sup> For example teitite: [http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei\\_tite.doc.html](http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_tite.doc.html)

<sup>58</sup> <http://de.wikipedia.org/wiki/Unicode>

should be offered as an extra option by digital libraries because of its widespread popularity (see for example the material available at [archive.org](http://archive.org)).

When encoding XML structures in TEI documents, it must first be decided how, and to what extent, text-type specific divisions such as chapter, section, volume or article should be taken into account. The same applies to other possible structural features such as tables of contents, indexes, line breaks, column breaks, page breaks, headers / footers / running titles, page numbers, images or image-like elements, captions, marginal notes, changes of font, e.g. from Fraktur to Antiqua (e.g. for foreign quotations), changes of font size, changes of font style (normal, italic, bold etc.), and others, formulas, e.g. mathematical (MathML) or chemical (CML) formulas, continuation marks (catchwords) at the bottom of a page (referring to the following page) and so on.

The choice of markup is generally dependent on the particular project. To ensure that full texts, marked up in this way, can be exchanged and reused, the XML elements and attributes used should be documented in the TEI header. Please note the efforts of the TEI community to develop a strictly formulated exchange format: TEI-Simple.<sup>59</sup>

#### 3.4.4 Layout

In some cases, when presenting a full text, it is important to preserve the layout of a document for the long term. The Practical Guidelines recommend using a **suitable formatting language** (e.g. XSLT, CSS), which largely ensures independence from special software. If valid reasons prohibit archiving the format with XML techniques, layout information for text documents may also be archived in PDF according to ISO standard 19005-1. However, as explained in [3.4.2.](#), PDF files cannot replace the provision of marked-up full text in XML.

A publication in XML + formatting language allows dynamic views to be generated according to the user's intended purpose. This should be borne in mind for the purposes of presentation and as wide a spectrum as possible should be offered. Typical output formats include HTML/XHTML, PDF, ePub and plain text.

### 3.5 Long-term availability

**Long-term preservation** and **archiving** of digital content at the level of bitstream archiving is well studied and can be viewed as solved. However, problems still persist in the preservation at the level of information representation (according to OAIS). We can distinguish between the archiving of genuinely digital information and the long-term preservation of digital copies of existing analogue objects. Different criteria can be applied regarding these types of archiving.

For long-term preservation, files are stored in stable, migratable formats in a technically and organisationally secure storage system. Digital data is archived within a similar storage system but with more extensive technical and organisational measures that cover not only the physical retention of the data, but also strategies for providing access to the data, also in the context of existing and future information systems. The storage systems must be designed for redundancy.

---

<sup>59</sup> <https://github.com/TEIC/TEI-Simple>

Technical information and information about the change history of an object are both important in long-term archiving.<sup>60</sup> Particularly with regard to the change history of an object, PREMIS (Preservation Metadata: Implementation Strategies)<sup>61</sup> is the data model of choice.

Whether long-term preservation or archiving is chosen for generated digital content will depend on the strategy of the institution, which must be outlined in the proposal.<sup>62</sup> Key criteria for successful long-term preservation and archiving of digital documents are as follows:

- (1) Creating the necessary organisational and financial framework,
- (2) Creating the necessary technical environment and choosing suitable techniques and strategies.

The Open Archival Information System (OAIS) should be used as a reference model for the archiving of digital data.<sup>63</sup> The “Criteria for Trusted Digital Repositories” are essential. They define organisational framework, legal basis, quality management and authenticity for a trusted repository.<sup>64</sup> Long-term preservation is implemented in digital magazines or reproduction administrations based on the OAIS model.

The long-term availability of the results of digitisation projects depends firstly on the choice of data and metadata formats. These have been outlined in previous sections. Secondly, it must be ensured that the digital data remains physically available. It should be noted that in DFG-funded digitisation projects, the **costs of project-specific data preservation** are expected to be borne by the institution for the duration of the project. **DFG funding is not** available for these costs.

Long-term preservation and archiving is an integral part of any digitisation project. The costs and efforts it entails should not be underestimated. The long-term costs of storage, which, depending on the project, may amount to several terabytes, and the long-term effort involved in physical retention must both be taken into account.

It should be noted that the DFG views digitisation projects as **endeavours undertaken by the entire institution**: It is assumed that the department in charge of the project will be supported by the **in-house IT infrastructure**. We also encourage smaller institutions to take advantage of the **expertise and services of larger institutions**.

Proposals must contain **convincing statements as to institutional long-term preservation and archiving**.

---

<sup>60</sup> <http://www.langzeitarchivierung.de/Subsites/nesstor/DE/Standardisierung/Metadaten.html:jsessionid=5239B697C2522475FBEC-D446877A114F.prod-worker5>

<sup>61</sup> [http://www.loc.gov/standards/premis/understanding\\_premis\\_german.pdf](http://www.loc.gov/standards/premis/understanding_premis_german.pdf)

<sup>62</sup> For example, an institution may specify in its strategy that born-digital objects should be archived and digital images of analogue objects should be stored in long-term storage systems. It may also be advisable to archive the digital copies of the original if the originals are fragile, for example.

<sup>63</sup> The OAIS reference model has been approved as an ISO 14721 standard:  
<http://public.ccsds.org/publications/archive/650x0m2.pdf>

<sup>64</sup> See Schoger, Astrid/Susanne Dobratz/Reinhard Althenhöner: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Frankfurt am Main, 2008. See: <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>. See also: DIN 31644:2012-04.

## 4. Organisational Issues – In-House vs. Outsourced Digitisation

Digitisation may be done in-house or outsourced. In the first case, in addition to personnel costs, equipment costs can be funded if they are project-specific and not part of the core infrastructure to be provided by the institution.

The **decision** whether to undertake digitisation as an **in-house project** or to **outsource** it is always **specific to the project**, taking into account all practical and financial considerations, and exclusively the applicant's responsibility. Using contractors for direct digitisation is, above all, a matter of trust. Unlike with film digitisation, where the originals are not at risk, the service providers hired to handle rare material or historic library holdings should have an appropriate track record.

Outsourcing is possible even if the materials to be digitised may not be taken off the premises; on larger projects, service providers will often work on site, bringing their own people and equipment.

The following issues should be considered when drafting agreements with suppliers:

- **Job parameters must be exactly specified**, in particular the requirements and format standards for deliverable raw data. These must be formulated in a project-specific specification in accordance with the Practical Guidelines on Digitisation. Contracted suppliers should be able to demonstrate certified quality assurance procedures. The outsourcer is required to perform careful quality control on deliverables before settling invoices in full.
- The DFG expects that **an appropriate percentage of the invoice amount be withheld for security purposes and not paid out to the business providing the service until a quality check has been performed**. In addition, the business should be required to provide a written undertaking that it will, without delay and free of charge, render substitute performance or rectify defects should this become necessary due to its non-compliance with the specification or other justified quality complaints.

## 5. Citing Digital Resources, Persistent Addressing

When digitisation was in its infancy, the issue of citability of digital resources was frequently underestimated. But it is citability that makes Internet-based digitised sources viable for academic use. In contrast to previous secondary formats, like microfilm or paper printouts, an Internet resource is not just a copy of the original, which can be treated and hence quoted like the original, but rather an independent object in a dynamic integral research environment. When a digital copy is online it needs a **unique address**, so that other documents or databases can link to it. In addition to the customary analogue citation format, which can and should still be given via the navigation software, this requires the specification and online documentation of **addressing techniques** according to the granularities required in a research project (book, page, collection, file, object, object part, etc.)

A positive effect of the net-based citation format is that **referencing becomes unequivocal** – something that usually cannot be said about old prints, because of the many mistakes they contain, or documents like incunabula, which lack pagination or foliation. Therefore the content-based citation format (e.g. p.8, a4, 213r etc.) should be joined by a formal citation based on the image sequence. This also allows for the unequivocal citation of images that are not part of the

work proper (cover, endpaper, additional digitised watermarks, partial reproductions of illustrations etc.).

Different mechanisms apply to full text, for which no specific recommendations can be made as yet (XPointer, ID issuing and similar techniques are examples of options that allow unequivocal referencing).

For printed works and manuscripts, the accessibility and citability of the work as a whole as well as of its individual physical pages must be guaranteed by means of URI or IRI. Institutions should implement suitable mechanisms (PURL, URN, DOI, Handle etc.) to ensure the **persistence and linkability of a resource**, thus reliably providing sources for scientific research. Important to point out is also the benefit of unambiguous citability by URI of resources in terms of the further development of the Semantic Web and Linked Open Data (LOD).

It is strongly recommended to generate **URNs** at least at the work level via the German National Library.<sup>65</sup>

## 6. Provision of Project Results to the Public

### 6.1 Rights, licensing and open access

The DFG is a cosignatory to the Berlin Declaration on Open Access.<sup>66</sup> In the spirit of this declaration, the results of DFG-funded digitisation projects should generally be accessible free of charge to researchers around the world. Thus it is expected that digital copies will be available online at no cost, in a quality sufficient for the bulk of typical research purposes. If a project will deviate from this requirement, specific reasons must be stated in the proposal. Restrictions to open access to the generated data may be justified for reasons such as privacy, copyright or archival law, but may not affect more than 5% of the total material to be digitised.<sup>67</sup>

The DFG funds the digitisation of scientific material in order to make this material available to **researchers in Germany and the rest of the world**. All projects should be designed to make **research results available promptly and for the long term**. In virtually all cases, this will entail the provision of digital copies on the Internet. If, for legal reasons, metadata and/or digital copies cannot be freely made available via open access soon after their creation, it must be ensured that this will occur promptly after any legal restrictions cease to apply.

Whereas, **prior to digitisation**, the rights to the material to be digitised must be determined and, if necessary, obtained for digitisation and provision (→ [1.8](#); [2.2](#)), it must be ensured **during digitisation** that any ensuing rights fall to the preserving institution. Rights may arise from the creation of a digital image of the original. This is generally not the case with mass digitisation of flatware.

Generally no rights arise from the **creation of metadata** with regard to the individual data sets.<sup>68</sup> If existing catalogue information is digitised that contains not only formal data but also

---

<sup>65</sup> <http://www.persistent-identifier.de/?lang=en>

<sup>66</sup> <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>

<sup>67</sup> The DFG cannot fund the processing of inaccessible material. A plan for providing access, in terms of the technical solution as well as the temporal dimension, is a prerequisite for funding.

<sup>68</sup> See Klimpel, Das Eigentum an Metadaten, in Euler / Hagedorn-Saupe / Maier; Handbuch Kulturportale. Online-Angebote aus Kultur und Wissenschaft, Berlin/Boston 2015, pp. 57 – 64.

descriptive text, and which therefore may be legally protected, any rights necessary for the provision and use of the digital copies must be obtained.

In keeping with the principles of open access and open source, all results – metadata, digital images and full texts, or the XML underlying the full texts, as well as XSLT scripts and DTDs or XML schema files – must be made available for reuse as freely as legally possible, either by marking them as public domain or using free Creative Commons licences.<sup>69</sup> When digitising public-domain material, the digital copy must be marked as public domain. For all results where this is not possible because rights exist or have been generated in the process of digitisation, it is likewise necessary to ensure that the data are made available and licensed via open access as freely as legally possible. This can be done by completely waiving any ensuing rights with a CC0 mark (e.g. in the case of metadata) or by using of the free licence types CC BY or CC BY SA (→ [1.8](#)). If protected material is digitised and the digital copy cannot be freely licensed because of rights to the source material that are still in effect, the legal status should be indicated with standardised rights statements.<sup>70</sup> For projects that digitise not only public-domain materials and/or that involve collaboration with commercial partners or publishers, **delayed open-access publication (moving wall) up to one year after the end of the project** can be agreed.

The primary aim of making results available as freely as possible and of standardised rights statements is to allow digital images, metadata and full texts to be analysed in usage contexts other than the immediate project environment (e.g. for data mining or data aggregation). This requires downloading, re-indexing and provision in separate research and presentation contexts. Images should be provided in a form that generally allows **full scholarly use** in other research contexts (e.g. through the use of individual images in frames, as with the DFG Viewer or Google iFrames, or by issuing a general permission to display preview images on third-party servers). For this purpose, either highly resolved derivatives in TIFF format or fully resolved JPEG images with a compression of 90, maximum 80, in combination with the full scan resolution should be made available. This does not affect the levying of charges for copies in other qualities, derivatives or the production of other formats (CD, printouts and so forth).

The DFG expects that data made available online as part of DFG-funded projects should include **a clear reference to its origin and, if possible, a reference to DFG funding**. In the case of digitised images, this is usually done by **adding an acknowledgement to the published user copy** (for example in JPEG). For full texts an appropriate note may be added to the header of the text file; acknowledgements may also be incorporated in image headers. But in all cases the acknowledgement of origin and the funding credit must be included in the accompanying metadata and output in the data provision system.

## 6.2 Minimum requirements for provision systems for digitised material

### 6.2.1 Functionality requirements

Collections / holdings must generally be made accessible in a variety of ways:

- via the providing institution's **website**

---

<sup>69</sup> Always use the latest version of Creative Commons licences: <https://creativecommons.org/>.

<sup>70</sup> [www.rightsstatements.org](http://www.rightsstatements.org)

- via a locally implemented or externally operated **DFG Viewer**, if applicable to the material (→ [7.3](#))
- via a search inquiry to the **local and regional library catalogue** / the **local and regional archive portal** / the **relevant material-specific online application** and the **nationwide reference and presentation systems** (DDB, Archivportal-D)
- if available, via one of the **DFG-funded material-specific portals** that enable integrated access to all digital collections funded under the DFG programme, or a shared subject-specific portal, e.g. of the specialised information services
- via **Internet search engines**
- via an **OAI interface**

The **DFG Viewer** should be addressed via the **OAI functionality** (if the respective format is supported). The data should no longer be provided solely as XML files, for the benefit of OAI functionalities. Newly created OAI interfaces should be registered with potentially relevant portals and specialised information services. In addition, suitable measures should be taken to ensure that metadata can be found by search engines (e.g. with the sitemap protocol<sup>71</sup>).

As well as being able to access specific documents in a targeted way by means of a metadata search, users should also have the option of structured browsing in predefined collections, part-collections or holdings. Regarding the search engine it should be noted that simple, Google-style search tools tend to serve a larger user community than multi-field search masks that require a solid understanding of the data structure of a given collection or inventory. Ideally, both models should be combined in a faceted search, which enables users to whittle down long hit lists using defined criteria (facets). The highly specific multi-field search is still, however, desirable as an additional tool for a highly specialised community.

The following functions should also be implemented:

- Download function<sup>72</sup>
- Print function for output document view<sup>73</sup>
- Centralised DFG-funded information systems (VD16, VD17, subject-specific portals, material-specific portals, etc.) should first link to a view in the style of the DFG Viewer.
- All DFG-funded provision systems are expected to provide an **automatic option** for users to give **feedback** on the digital resource. This function should be set up on the project page or centrally in the digital provision system.

### 6.2.2 Minimum technical requirements

As far as applicable, servers must be set up to:<sup>74</sup>

- Provide all materials in a quality that allows their convenient use for research purposes on typical university equipment. This entails, for instance, providing a type size that is easy to read.

---

<sup>71</sup> <http://www.sitemaps.org/>

<sup>72</sup> If the size of the complete file would be unmanageable, the file may be split into sections or individual pages for download purposes. Provision to researchers in the original size of the image master (at least JPEG 80% compression) should generally be made possible free of charge.

<sup>73</sup> If the size of the complete file would be unmanageable, the file may be split into sections or individual pages for print purposes.

<sup>74</sup> The key criterion is practicality, not the implementation of abstract desirables. If objects in a project by their nature cannot be meaningfully displayed with a resolution under 1600 × 1200, there is no need to bother with pseudo solutions; if an object cannot be processed meaningfully under 3 MB, it does not violate the criterion of DSL compatibility not to provide smaller versions.

- Provide all materials, conversely, in a quality that allows processing via DSL without cumbersome delays.
- Enable the free download, for research purposes, of any complete unit as one single file (e.g. of individual printed works).
- Support all currently popular browsers, to the extent viable.<sup>75</sup>

## 7. Presentation Standards for Text-Based Works, DFG-Viewer

The principles laid out above apply to any kind of project that provides digital content. Additionally, the following minimal requirements apply to the provision of digital files that have the character of digital books or multi-page documents. They cover certain basic standards and the minimum with regard to functionality.

### 7.1 Basic requirements and architecture

The provision system combines digitised image or full-text files into a document structure to enable users to navigate a document. Furthermore, it establishes connections between digital documents, or parts thereof (e.g. chapters, pages), and metadata, to allow users to access the individual document or certain document parts based on a metadata search. Finally, it organises digital documents into digital collections or holdings according to subject matter or origin, to let users navigate documents and collections in a scientifically appropriate fashion. It provides user interfaces for searching, navigating, accessing and retrieving metadata, documents, collections and holdings, and it supports largely automated export and import of standards-compliant raw data. The provision systems of the individual information infrastructure facilities should allow access across institutions, both in navigating digital collections or holdings and in searching indexes. In addition, the transparent linkage of provision systems with local catalogue / information systems as well as comprehensive information systems is desirable.

Various system architectures can be used to accomplish these tasks. The following basic alternatives are viable:

- Metadata are stored centrally in an online information system (e.g. the local OPAC or online search system, or a comprehensive information system such as a library network catalogue or Archivportal-D), while digital document files including XML-encoded structure data are provided in a hierarchically organised file system on a separate document server for online access. The structure of the digitised collection, or the internal structure of the digitised documents, can be mirrored by the hierarchy of the file system.
- A document management system (DMS) or Content Management System (CMS) is used, in which both the metadata and the digital data are stored in the database system.

Typically, the first variant is used, which allows for a distributed and transparent information infrastructure.

---

<sup>75</sup> If a browser does not support a format required by an advanced 3D application, there is no need to go to the trouble of developing a suitable plug-in.

## 7.2 Functionality requirements

A key benchmark of functional quality is the comfort with which users can navigate within a found document. The following **navigation functions** are considered the basic standard:

- Go to any desired image
- Home: Jump to beginning of document
- End: Jump to end of document
- Forward: Go forward one page
- Back: Go back one page
- Full text search within digital copies (mandatory for books from 1850 onward)<sup>76</sup>
- Metadata info: View current document information in description fields stored in information system.
- Help: Help menu should provide detailed descriptions with examples for navigation and for searching the digital library.

Whenever possible and appropriate, tables of contents, structure trees or functional equivalents should be included and designed to be searchable. Navigation aids are desirable, e.g. graphic representations in a header that indicate to the user the current location within the digital document. If an information system contains materials that users will normally regard as conceptual units (multivolume works), these units must be visible as such.

## 7.3 DFG Viewer

In addition to the differently designed and locally managed web information systems of individual institutions, scientific users should have standardised access to the data (contents) of all DFG-funded digitised works. To this end, the DFG currently pursues two complementary strategies:

- Defining a standardised design profile for visualising digital copies that were generated with DFG funding (DFG Viewer)
- Creating an OAI interface based on the METS standard. The primary purpose of this interface is to display images and their metadata in a uniform manner for all DFG-funded projects and to deliver METS/MODS for printed works and archive materials, and METS/TEI for manuscripts. The goal is to describe scrolling, metadata display and other basic functions and thus create consistent display and scroll functions that enable homogeneous access to decentralised resources even within central search portals (harvesting).

For their primary nationwide presentations, DFG-funded digitisation projects should serve the aforementioned interfaces and/or implement a DFG-style viewer at their own institution. Appendixes A and B define the interface's METS/MODS and METS/TEI format and the design of the viewer.

---

<sup>76</sup> With today's technology, OCR should always be considered for machine-press era printed works from 1850 onward.

## Appendix A: METS/MODS Profile for DFG Viewer Display and Transmission by OAI

### 1. DFG Viewer

In order to achieve a uniform presentation when local digital offerings are accessed through nationwide catalogue systems (e.g. VD16 / VD17, ZVDD, Deutsche Digitale Bibliothek, Archivportal-D), DFG-funded projects should use the browser display known as DFG Viewer and serve the interfaces on which it is based. The purpose is to make it easier for researchers to use digital content. The DFG Viewer may then link to the special local information systems of any given institution.

The DFG Viewer<sup>77</sup> was developed in the context of the funding scheme “Digitisation of VD16 / VD17” by the libraries funded in the first round of proposals. The SLUB Dresden, in collaboration with additional partners, continues to develop the Viewer on an ongoing basis. The Viewer’s reference application is currently hosted by the SLUB Dresden.

To give DFG-funded projects maximum certainty for proposal planning and ensure that metadata meets the DFG Viewer requirements, metadata generated by such projects should be valid against one of the application profiles on the Viewer’s website.<sup>78</sup>

METS<sup>79</sup> is used to display metadata in the DFG Viewer. It serves as a frame format (wrapper) within which descriptive, administrative and structural metadata as well as resources (e.g. images, full texts) are recorded. To display descriptive metadata (prints, journals, newspapers, archival materials), the Viewer requires MODS<sup>80</sup>-encoded metadata (see 2. below). To link administrative metadata (e.g. local application, website, institution logo), a special format (namespace dv), developed specifically for the Viewer, is used.

Detailed documentation on how to implement the METS format can be found on the website of the DFG Viewer’s reference application.

These guidelines apply in principle to all digitised media types. In individual cases, however, alternative formats may be used instead of METS/MODS (METS/TEI for medieval and early-modern manuscripts, EAD for coding archival finding aids, LIDO for museum objects – although without the option of DFG Viewer display).

---

<sup>77</sup> <http://dfg-viewer.de/en/regarding-the-project/>

<sup>78</sup> <http://dfg-viewer.de/en/profile-of-the-metadata/>

<sup>79</sup> <http://www.loc.gov/standards/mets/>

<sup>80</sup> <http://www.loc.gov/standards/mods/>

## 2. MODS DFG standard set (library holdings)

The MODS standard offers a simplified subset of MARC21, which should facilitate automatic conversions from common catalogues. For DFG Viewer display, only a few mandatory fields are required (see table below).

<b>Element / subelement</b>	<b>Repeatable</b>	<b>Comments</b>	<b>Status</b>
-----------------------------	-------------------	-----------------	---------------

### 1 Title information

<titleInfo>	Yes	Title information; if work has no title, a title must be created. In the case of multivolume works without individual titles for each volume, it is sufficient to state the title of the whole work in the element <code>mods:relatedItem</code> . However, it is then mandatory to state the volume numbers in the element <code>mods:part</code> and the subelements <code>mods:detail</code> and <code>mods:number</code> .	Mandatory (except multivolume works without individual titles)
<titleInfo> / <title>	No	Contains main title of work.	Mandatory
<titleInfo> / <subTitle>	No	Contains subtitle / addition to main title of work.	Mandatory if applicable

### 2 Person

<name type="personal" authority="...">	Yes	Person related to work (e.g. author). The @authority attribute contains the code for the set of rules according to which the person has been identified; usually "GND". Optionally, a URI can be used for the person's name in @valueURI.	Mandatory if applicable
<name> / <namePart type="...">	Yes	Contains name elements of the type specified in @type; possible values are "date", "family", "given", "termsOfAddress".	Mandatory if applicable
<name> / <displayForm>	No	Name in desired display form	Recommended
<name> / <role>	No	Wrapper element for role of person	Mandatory if applicable
<name> / <role> / <roleTerm type="code" authority="marcrelator">	No	Role of person; <roleTerm> field value is encoded (MARC relator code) <sup>81</sup> .	Mandatory if applicable

<sup>81</sup> A code from the MARC Value List for Relators and Roles: <http://www.loc.gov/marc/sourcecode/relator/relatorlist.html>.

<b>Element / subelement</b>	<b>Repeat able</b>	<b>Comments</b>	<b>Status</b>
-----------------------------	--------------------	-----------------	---------------

### 3 Corporate body

<name type="corporate" authority="">	Yes	Corporate body related to work.	Mandatory if applicable
<name> / <namePart>	Yes	See above	
<name> / <role>	No	See above	
<name> / <role> / <roleTerm type="code" authority="marcrelator">	No	See above	

### 4 Publication information / Imprint

<originInfo eventType="publication">	Yes	Publication information/Imprint; the first <originInfo> block is for information on the source; the second <originInfo> block is for including information on the digital edition by indicating eventType="digitisation".	Mandatory if applicable
<originInfo> / <place>	Yes	Contains elements on place of publication.	Mandatory if applicable
<originInfo> / <place> / <placeTerm type="text">	No	Contains place of publication; if place of publication is unknown, write "[o.O]".	Mandatory
<originInfo> / <publisher>	Yes	Contains publisher / print shop.	Mandatory if applicable
<originInfo> / <dateIssued keyDate="yes" encoding="w3cdtf">	Yes	Contains year of publication; if year is unknown, write "[o.J]".	Mandatory

### 5 Edition information

<originInfo> / <edition>	Yes	Contains name of edition (see also above, information about electronic edition).	Mandatory if applicable
--------------------------	-----	--	-------------------------

### 6 Physical description

<physicalDescription>	No	Physical description area / collation	Mandatory
<physicalDescription> / <extent>	Yes	Contains information on pagination, size and illustrations.	Mandatory if applicable
<physicalDescription> / <digitalOrigin>	No	For digitised printed works, the <digitalOrigin> field usually states "reformatted digital".	Mandatory

<b>Element / subelement</b>	<b>Repeat able</b>	<b>Comments</b>	<b>Status</b>
-----------------------------	--------------------	-----------------	---------------

**7 Superior work level**

<relatedItem type="host"> / <recordInfo> / <recordIdentifier>	No	<recordIdentifier> is an identifier that permits linkage to hierarchically superior / superordinate datasets.	Mandatory if hierarchy exists
---	----	---	-------------------------------

**8 Volume information**

<part order=""> / <detail> / <number>	No	Part information; value of @order attribute is any numeric value that ensures correct order of parts; <number> states the volume.	Mandatory if hierarchy exists
--	----	---	-------------------------------

**9 Language**

<language>	Yes	Language information	Mandatory if applicable
<language> / <languageTerm type="code" authority="iso639-2b">	No	Contains language of work in ISO 639-2/B code.	

**10 Citable identifier**

<identifier type="...">	Yes	Worldwide unique identifier of resource (@type attribute e.g. URN, PURL, DOI, Handle, URI etc.). If available, GW and VD numbers of print must also be given (@type attributes VD16, VD17, GW).	Mandatory
-------------------------	-----	---	-----------

**11 Database ID**

<recordInfo> / <recordIdentifier>	No	Dataset identifier for unique identification within a database system, e.g. PICA production number	Mandatory
--------------------------------------	----	--	-----------

**12 Shelfmark**

<location> <physicalLocation> <shelfLocator>	Yes	Location and call number of original	Recommended
--	-----	--------------------------------------	-------------

If these data should be suitable for OAI harvesting, an expansion of this basic set should be considered, depending on the type of material and the design of the project. The DFG Viewer website offers more differentiated format definitions for this purpose, including additional explanations of how to use the various fields.

### 3. Notes for archives on converting EAD(DDB) to MODS

Mapping EAD(DDB) to corresponding MODS elements is generally easy. For simplified processing of archival data sets, tools are available in the archiving community for use by interested institutions.<sup>82</sup>

The following elements are recommended for archival data, although institutions are free to specify additional indexing information according to the MODS application profile:

<i>Field description</i>	<i>EAD(DDB)</i>	<i>MODS</i>	<i>Status</i>
Title	c/did/unittitle	<p>mods:titleInfo, if appropriate with subelements</p> <p>MODS definition:</p> <p>A word, phrase, character, or group of characters, normally appearing in a resource, that names it or the work contained in it.</p> <p>Comment:</p> <p>For the root structure element in a METS/MODS set, at least one title must be specified using the mods:titleInfo element with the mods:title subelement.</p>	Mandatory
Institution	ancestor::archdesc/did/repository/corpname	<p>mods:physicalLocation</p> <p>MODS definition:</p> <p>The institution or repository that holds the resource or where it is available.</p>	Mandatory

<sup>82</sup> For further information see <https://pro.deutsche-digitale-bibliothek.de/dokumente-links-fuer-archive>.

Signature	c/did/unitid[not(@type)]	<p>mods:shelfLocator</p> <p>MODS definition:</p> <p>Shelfmark or other shelving designation that indicates the location identifier for a copy.</p>	Mandatory
Duration	c/did/unitdate	<p>mods:dateCreated</p> <p>MODS definition:</p> <p>The date of creation of the resource</p> <p>Comment: Date of creation or duration of the analogue source material.</p>	Mandatory if available
Context <sup>83</sup>	ancestor::c/did/unittitle	<p>mods:relatedItem, if appropriate with subelements</p> <p>MODS-Definition:</p> <p>Information that identifies other resources related to the one being described.</p> <p>mods:relatedItem can be used to show not only hierarchical relationships, but a variety of relationships using the type attribute.</p> <p>To create a relationship between documents or document parts with MODS, unique identifiers must be used.</p>	Recommended

---

<sup>83</sup> This refers to the inventory context of the resource, i.e. the higher levels of description in terms of tectonics and classification.

ID	c/@id	mods:recordIdentifier  MODS-Definition: Contains the system control number assigned by the organization creating, using, or distributing the record.	Mandatory
----	-------	--	-----------

## 4. Examples of METS/MODS datasets according to the DFG standard

### 4.1 For library material:

```

<?xml version="1.0" encoding="UTF-8"?>
<mets:mets xmlns:mets="http://www.loc.gov/METS/" xmlns:tei="http://www.tei-c.org/ns/1.0"
  xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mods="http://www.loc.gov/mods/v3" xmlns:dv="http://dfg-viewer.de/"
  xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/mets/mets.xsd">
  <mets:metsHdr>
    <mets:agent ROLE="IOWNER">
      <mets:name>Herzog August Bibliothek Wolfenbüttel</mets:name>
    </mets:agent>
    <mets:agent ROLE="CREATOR">
      <mets:name>Herzog August Bibliothek Wolfenbüttel</mets:name>
    </mets:agent>
  </mets:metsHdr>
  <mets:dmdSec ID="dmdsec">
    <mets:mdWrap MDTYPE="MODS">
      <mets:xmlData>
        <mods:mods
          xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-
5.xsd">
          <mods:titleInfo>
            <mods:title>Occulta Philosophia</mods:title>
            <mods:subTitle>Von den verborgenen Philosophischen Geheimnissen [...]</mods:subTitle>
          </mods:titleInfo>
          <mods:identifier type="fingerprint">inl- t?h- esin Woab 3 1613R</mods:identifier>
          <mods:identifier type="vd17">39:122887Y</mods:identifier>
          <mods:originInfo eventType="publication">
            <mods:dateIssued keyDate="yes" encoding="iso8601">1613</mods:dateIssued>
            <mods:place>
              <mods:placeTerm type="text">Franckfurt am Mayn</mods:placeTerm>
            </mods:place>
            <mods:publisher>Bringer</mods:publisher>
          </mods:originInfo>
          <mods:name type="personal" valueURI="http://d-nb.info/gnd/119068982">
            <mods:namePart type="termsOfAddress">Trismegistus</mods:namePart>
            <mods:namePart type="given">Hermes</mods:namePart>
            <mods:displayForm>Hermes &lt;Trismgivenegistus&gt;</mods:displayForm>
            <mods:role>
              <mods:roleTerm authority="marcrelator" type="code"
valueURI="http://id.loc.gov/vocabulary/relators/oth">oth</mods:roleTerm>
            </mods:role>
          </mods:name>
          <mods:name type="personal" valueURI="http://d-nb.info/gnd/118507036">
            <mods:namePart type="termsOfAddress">Valentinus</mods:namePart>
            <mods:namePart type="given">Basilius</mods:namePart>
            <mods:displayForm>Basilius &lt;Valentinus&gt;</mods:displayForm>
            <mods:role>
              <mods:roleTerm authority="marcrelator" type="code"
valueURI="http://id.loc.gov/vocabulary/relators/oth">oth</mods:roleTerm>
            </mods:role>
          </mods:name>
        </mods:mods>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:dmdSec>
</mets:mets>

```

```

    <mods:name type="personal">
      <mods:namePart type="given">Ludwig Georg</mods:namePart>
      <mods:namePart type="family">Stolberg</mods:namePart>
      <mods:displayForm>Stolberg, Ludwig Georg</mods:displayForm>
      <mods:role>
        <mods:roleTerm authority="marcrelator" type="code"
valueURI="http://id.loc.gov/vocabulary/relators/dte">dte</mods:roleTerm>
      </mods:role>
    </mods:name>
    <mods:name type="personal">
      <mods:namePart type="given">Johann</mods:namePart>
      <mods:namePart type="family">Bringer</mods:namePart>
      <mods:displayForm>Bringer, Johann</mods:displayForm>
      <mods:role>
        <mods:roleTerm authority="marcrelator" type="code"
valueURI="http://id.loc.gov/vocabulary/relators/clb">clb</mods:roleTerm>
      </mods:role>
    </mods:name>
    <mods:name type="personal" valueURI="http://d-nb.info/gnd/1037506073">
      <mods:namePart type="given">Johann</mods:namePart>
      <mods:namePart type="family">Bringer</mods:namePart>
      <mods:displayForm>Bringer, Johann</mods:displayForm>
      <mods:role>
        <mods:roleTerm authority="marcrelator" type="code"
valueURI="http://id.loc.gov/vocabulary/relators/prt">prt</mods:roleTerm>
      </mods:role>
    </mods:name>
    <mods:language>
      <mods:languageTerm type="code" authority="iso639-2b">ger</mods:languageTerm>
    </mods:language>
    <mods:genre authority="aad">Alchemie</mods:genre>
    <mods:location>
      <mods:physicalLocation authority="marcorg">DE-23</mods:physicalLocation>
      <mods:shelfLocator>M: Nd 14 (1)</mods:shelfLocator>
    </mods:location>
    <mods:physicalDescription>
      <mods:extent>[4] Bl., 78 S.</mods:extent>
      <mods:extent>4°</mods:extent>
      <mods:extent>Tbl. r&amp;s, Ill. (Holzschn.)</mods:extent>
    </mods:physicalDescription>
    <mods:recordInfo>
      <mods:recordIdentifier source="DE-23">092275664</mods:recordIdentifier>
      <mods:recordOrigin xml:lang="en">Converted from PICA using a local XSL transformation
script</mods:recordOrigin>
      <mods:recordContentSource authority="marcorg">DE-23</mods:recordContentSource>
    </mods:recordInfo>
  </mods:mods>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>
<mets:amdSec ID="amdsec">
  <mets:rightsMD ID="rights">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="DVRIGHTS">
      <mets:xmlData>
        <dv:rights>

```

```

    <dv:owner>Herzog August Bibliothek Wolfenbüttel</dv:owner>
    <dv:ownerLogo>http://www.hab.de/images/logo_dfg_viewer.gif</dv:ownerLogo>
    <dv:ownerSiteURL>http://www.hab.de/</dv:ownerSiteURL>
    <dv:ownerContact>mailto:auskunft@hab.de</dv:ownerContact>
  </dv:rights>
</mets:xmlData>
</mets:mdWrap>
</mets:rightsMD>
<mets:digiprovMD ID="digiprov">
  <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="DVLINKS">
    <mets:xmlData>
      <dv:links>
        <dv:reference>http://opac.lbs-braunschweig.gbv.de/DB=2/PPN?PPN=092275664</dv:reference>
        <dv:presentation>http://diglib.hab.de/drucke/nd-14-1s/start.htm</dv:presentation>
      </dv:links>
    </mets:xmlData>
  </mets:mdWrap>
</mets:digiprovMD>
</mets:amdSec>
<mets:fileSec>
  <mets:fileGrp USE="DEFAULT">
    <mets:file ID="file--idp65681520" MIMETYPE="image/jpeg">
      <mets:FLocat xlink:href="http://diglib.hab.de/drucke/nd-14-1s/00001.jpg"
        LOCTYPE="URL"/>
    </mets:file>
    <!-- weitere Verweise auf Images aus Umfangsgründen ausgelassen -->
    <mets:file ID="file--idp65467888" MIMETYPE="image/jpeg">
      <mets:FLocat xlink:href="http://diglib.hab.de/drucke/nd-14-1s/eb04.jpg"
        LOCTYPE="URL"/>
    </mets:file>
  </mets:fileGrp>
  <mets:fileGrp USE="MAX">
    <mets:file ID="file-max-idp65681520" MIMETYPE="image/jpeg">
      <mets:FLocat xlink:href="http://diglib.hab.de/drucke/nd-14-1s/max/00001.jpg"
        LOCTYPE="URL"/>
    </mets:file>
    <!-- weitere Verweise auf Images aus Umfangsgründen ausgelassen -->
    <mets:file ID="file-max-idp65467888" MIMETYPE="image/jpeg">
      <mets:FLocat xlink:href="http://diglib.hab.de/drucke/nd-14-1s/max/eb04.jpg"
        LOCTYPE="URL"/>
    </mets:file>
  </mets:fileGrp>
  <mets:fileGrp USE="MIN">
    <mets:file ID="file-min-idp65681520" MIMETYPE="image/jpeg">
      <mets:FLocat xlink:href="http://diglib.hab.de/drucke/nd-14-1s/min/00001.jpg"
        LOCTYPE="URL"/>
    </mets:file>
    <!-- weitere Verweise auf Images aus Umfangsgründen ausgelassen -->
    <mets:file ID="file-min-idp65467888" MIMETYPE="image/jpeg">
      <mets:FLocat xlink:href="http://diglib.hab.de/drucke/nd-14-1s/min/eb04.jpg"
        LOCTYPE="URL"/>
    </mets:file>
  </mets:fileGrp>
  <mets:fileGrp USE="THUMBS">
    <mets:file ID="file-thumbs-idp65681520" MIMETYPE="image/jpeg">

```

```

    <mets:FLocat xlink:href="http://diglib.hab.de/drucke/nd-14-1s/thumbs/00001.jpg"
      LOCTYPE="URL"/>
  </mets:file>
  <!-- weitere Verweise auf Images aus Umfangsgründen ausgelassen -->
  <mets:file ID="file-thumbs-idp65467888" MIMETYPE="image/jpeg">
    <mets:FLocat xlink:href="http://diglib.hab.de/drucke/nd-14-1s/thumbs/eb04.jpg"
      LOCTYPE="URL"/>
  </mets:file>
</mets:fileGrp>
</mets:fileSec>
<mets:structMap TYPE="PHYSICAL">
  <mets:div ID="struct-physical" TYPE="physSequence" DMDID="dmdsec" ADMID="amdsec">
    <mets:div ID="struct-physical-idp65681520"
      CONTENTIDS="http://diglib.hab.de/drucke/nd-14-1s/start.htm?image=00001" TYPE="page"
      ORDER="1">
      <mets:fptr FILEID="file--idp65681520"/>
      <mets:fptr FILEID="file-min-idp65681520"/>
      <mets:fptr FILEID="file-max-idp65681520"/>
      <mets:fptr FILEID="file-thumbs-idp65681520"/>
    </mets:div>
    <!-- weitere Verweise auf Strukturen aus Umfangsgründen ausgelassen -->
    <mets:div ID="struct-physical-idp65467888"
      CONTENTIDS="http://diglib.hab.de/drucke/nd-14-1s/start.htm?image=eb04" TYPE="page"
      ORDER="94">
      <mets:fptr FILEID="file--idp65467888"/>
      <mets:fptr FILEID="file-min-idp65467888"/>
      <mets:fptr FILEID="file-max-idp65467888"/>
      <mets:fptr FILEID="file-thumbs-idp65467888"/>
    </mets:div>
  </mets:div>
</mets:structMap>
<mets:structMap TYPE="LOGICAL">
  <mets:div ID="struct-logical" TYPE="Monograph" DMDID="dmdsec" ADMID="amdsec"
    CONTENTIDS="http://diglib.hab.de/drucke/nd-14-1s/start.htm">
    <mets:div ID="struct-logical-idp65475312" TYPE="binding">
      <mets:div ID="struct-logical-idp65474288" TYPE="endsheet"/>
    </mets:div>
    <mets:div ID="struct-logical-idp65472496" TYPE="section">
      <mets:div ID="struct-logical-idp65477616" TYPE="title_page"/>
      <mets:div ID="struct-logical-idp65477104" TYPE="dedication"
        LABEL="Dem Wolgebornen Graffen und Herrn/ Herrn Lufwig Georgen/ Graffen zu Stolberg /
        Königstein / Roschefort/ Weringeroda und Hohenstein / ...">
        <mets:div ID="struct-logical-idp65478128" TYPE="ornament"/>
      </mets:div>
      <mets:div ID="struct-logical-idp65474032" TYPE="section"
        LABEL="Das Erste Theil., Ein Colloquium oder Gespräch zwischen einem alten und jungen
        Studenten.">
        <mets:div ID="struct-logical-idp65483504" TYPE="section"
          LABEL="Adolphus. Mein Freundlichen Gruß und alle Wolfahrt wünsche ich Euch lieber alter
          Senior: ...">
          <mets:div ID="struct-logical-idp65482352" TYPE="ornament"/>
        </mets:div>
        <mets:div ID="struct-logical-idp65494384" TYPE="section"
          LABEL="und Abgöttisch ist / ein ander Werck neben Gottes zusetzen/ ..."/>
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>

```

```

    <mets:div ID="struct-logical-idp65498224" TYPE="section"
      LABEL="greiff das ewig Leben/ dazu du auch beruffen bist/ ..."/>
  </mets:div>
  <mets:div ID="struct-logical-idp65499248" TYPE="section"
    LABEL="Das ander Theil Aureliae Occultae Philosophorum.">
    <mets:div ID="struct-logical-idp65498352" TYPE="illustration"/>
    <mets:div ID="struct-logical-idp65505904" TYPE="section"
      LABEL="Schmaragt Tafel Hermetis.">
      <mets:div ID="struct-logical-idp65511152" TYPE="illustration"/>
      <mets:div ID="struct-logical-idp65510640" TYPE="illustration"/>
      <mets:div ID="struct-logical-idp65507952" TYPE="illustration"/>
      <mets:div ID="struct-logical-idp65516400" TYPE="illustration"/>
    </mets:div>
    <mets:div ID="struct-logical-idp65516656" TYPE="section"
      LABEL="Würckung deß Philosophischen Wercks/ erst Figur.">
      <mets:div ID="struct-logical-idp65521520" TYPE="illustration"/>
    </mets:div>
    <mets:div ID="struct-logical-idp65513584" TYPE="section"
      LABEL="Die ander Figur deß Wercks.">
      <mets:div ID="struct-logical-idp65520240" TYPE="illustration"/>
    </mets:div>
    <!-- weitere Strukturmerkmale aus Umfangsgründen ausgelassen-->
  </mets:div>
  <mets:div ID="struct-logical-idp65539824" TYPE="binding">
    <mets:div ID="struct-logical-idp65543536" TYPE="endsheet"/>
  </mets:div>
  <mets:div ID="struct-logical-idp65538672" TYPE="binding">
    <mets:div ID="struct-logical-idp65538928" TYPE="cover"/>
    <mets:div ID="struct-logical-idp65541232" TYPE="paste_down"/>
    <mets:div ID="struct-logical-idp65539312" TYPE="annotation"/>
    <mets:div ID="struct-logical-idp65549296" TYPE="paste_down"/>
    <mets:div ID="struct-logical-idp65549424" TYPE="cover"/>
  </mets:div>
</mets:div>
</mets:structMap>
<mets:structLink>
  <mets:smLink xlink:from="struct-logical" xlink:to="struct-physical"/>
  <mets:smLink xlink:from="struct-logical-idp65475312" xlink:to="struct-physical-idp65681520"/>
  <!-- weitere Verknüpfungslinks aus Umfangsgründen ausgelassen-->
  <mets:smLink xlink:from="struct-logical-idp65549424" xlink:to="struct-physical-idp65467888"
  />
</mets:structLink>
</mets:mets>

```

## 4.2 For archive material:

```

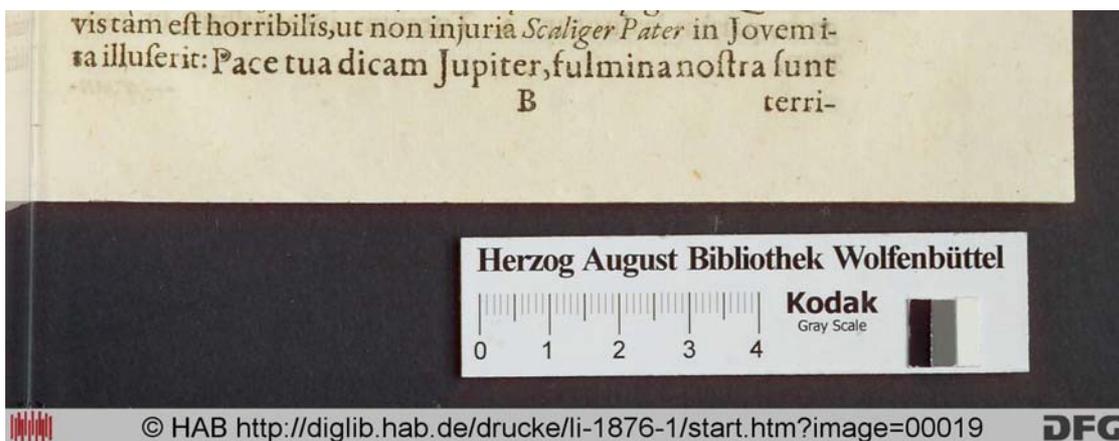
<mets:dmdSec ID="DMDLOG_1">
  <mets:mdWrap MDTYPE="MODS">
    <mets:xmlData>
      <mods:mods>
        <mods:recordInfo>
          <mods:recordIdentifier>labw-1-1263080</mods:recordIdentifier>
          <mods:descriptionStandard>DFG-Viewer/Archiv v2.3</mods:descriptionStandard>
        </mods:recordInfo>
        <mods:location>
          <mods:physicalLocation>Landesarchiv Baden-Württemberg, Abt. Hauptstaatsarchiv
Stuttgart</mods:physicalLocation>
          <mods:shelfLocator>H 51 U 589</mods:shelfLocator>
        </mods:location>
        <mods:titleInfo>
          <mods:title>Kaiser Karl publiziert die dreiundzwanzig ersten Kapitel der Goldenen Bulle</mods:title>
        </mods:titleInfo>
        <mods:originInfo eventType="production">
          <mods:dateCreated point="end" qualifier="approximate" encoding="iso8601">1356 Januar
10</mods:dateCreated>
        </mods:originInfo>
        <mods:relatedItem type="host" displayLabel="Gliederung">
          <mods:titleInfo>
            <mods:title>Urkunden</mods:title>
          </mods:titleInfo>
          <mods:relatedItem type="host" displayLabel="Bestand">
            <mods:titleInfo>
              <mods:title>Kaiserselekt</mods:title>
            </mods:titleInfo>
          </mods:relatedItem>
        </mods:relatedItem>
      </mods:mods>
    </mets:xmlData>
  </mets:mdWrap>
</mets:dmdSec>
<mets:amdSec ID="amd1">
  <mets:rightsMD ID="rights1">
    <mets:mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="DVRIGHTS">
      <mets:xmlData>
        <dv:rights>
          <dv:owner>Landesarchiv Baden-Württemberg, Abt. Hauptstaatsarchiv Stuttgart</dv:owner>
          <dv:ownerSiteURL>http://www.landesarchiv-bw.de/hstas</dv:ownerSiteURL>
          <dv:ownerContact>mailto:landesarchiv@la-bw.de</dv:ownerContact>
          <dv:ownerLogo>https://www.landesarchiv-bw.de/dfgviewer/logo.jpg</dv:ownerLogo>
        </dv:rights>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:rightsMD>
  <mets:digiprovMD ID="digiprov1">
    <mets:mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" OTHERMDTYPE="DVLINKS">
      <mets:xmlData>
        <dv:links>
          <dv:reference>http://www.landesarchiv-bw.de/plink/?f=1-1263080</dv:reference>
          <dv:presentation>http://www.landesarchiv-bw.de/plink/?f=1-1263080-1</dv:presentation>
        </dv:links>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:digiprovMD>

```

```
</dv:links>
</mets:xmlData>
</mets:mdWrap>
</mets:digiprovMD>
</mets:amdSec>
<mets:fileSec>
  <mets:fileGrp USE="DEFAULT">
    <mets:file ID="file-full-image_1-1263080-1" MIMETYPE="image/jpeg">
      <mets:FLocat LOCTYPE="URL"
        xlink:href="http://www.landesarchiv-bw.de/plink/?f=1-1263080-1&ext=456788&format=jpg"/>
    </mets:file>
    <mets:file ID="file-full-image_1-1263080-2" MIMETYPE="image/jpeg">
      <mets:FLocat LOCTYPE="URL"
        xlink:href="http://www.landesarchiv-bw.de/plink/?f=1-1263080-2&ext=456788&format=jpg"/>
    </mets:file>
    <mets:file ID="file-full-image_1-1263080-3" MIMETYPE="image/jpeg">
      <mets:FLocat LOCTYPE="URL"
        xlink:href="http://www.landesarchiv-bw.de/plink/?f=1-1263080-6&ext=456788&format=jpg"/>
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>
<mets:structMap TYPE="PHYSICAL">
  <mets:div ID="phys1" TYPE="physSequence">
    <mets:div ID="image_1-1263080-1" ORDER="1" TYPE="page">
      <mets:fptr FILEID="file-full-image_1-1263080-1"/>
    </mets:div>
    <mets:div ID="image_1-1263080-2" ORDER="2" TYPE="page">
      <mets:fptr FILEID="file-full-image_1-1263080-2"/>
    </mets:div>
    <mets:div ID="image_1-1263080-3" ORDER="3" TYPE="page">
      <mets:fptr FILEID="file-full-image_1-1263080-3"/>
    </mets:div>
  </mets:div>
</mets:structMap>
<mets:structMap TYPE="LOGICAL">
  <mets:div ID="log1" TYPE="act" DMDID="DMDLOG_1" ADMID="amd1"
    LABEL="Kaiser Karl publiziert die dreiundzwanzig ersten Kapitel der Goldenen Bulle" ORDER="1">
    <mets:div ID="Bild1" TYPE="section" LABEL="Schmuckkassette" ORDER="2"/>
    <mets:div ID="Bild2" TYPE="section" LABEL="Einband Vorderseite" ORDER="3"/>
    <mets:div ID="Bild3" TYPE="section" LABEL="fol 01r" ORDER="4"/>
  </mets:div>
</mets:structMap>
<mets:structLink>
  <mets:smLink xlink:from="Bild1" xlink:to="image_1-1263080-1"/>
  <mets:smLink xlink:from="Bild2" xlink:to="image_1-1263080-2"/>
  <mets:smLink xlink:from="Bild3" xlink:to="image_1-1263080-3"/>
</mets:structLink>
</mets:mets>
```

## 5. Adding a footer

For formats provided to users or on the Internet, the footer should be appended to the lower edge of the image. The institution logo should be shown on the left side of the footer; for DFG-funded projects, the DFG logo should be added on the right side, if possible. It is recommended to provide a citable URL in the middle area (in addition to including it in the Viewer XML file). Text and logos must be scaled according to the resolution. In PDF files used for the download footers may be added to each image as well, in addition to a cover page. See the following examples:



## Appendix B: METS/TEI Specification for the Display of Digitised Manuscripts

### 1. DFG Viewer

The objective of this METS/TEI specification for digitised medieval and early-modern manuscripts is the display of the digitised data within the unified interface of the DFG Viewer. The DFG Viewer, which was originally created to display digitised printed works, is thus extended to another media type without abandoning the familiar user interface. It is merely supplemented by a few extra functions to meet the specific requirements of the new media type.

Since this involves not a fundamentally new implementation of the DFG viewer, but merely the support of an additional application profile for METS/TEI, the information given in Annex A also applies to digitised manuscripts – in particular the general information on the DFG Viewer in section 1, as well as the instructions for adding a footer in section 5.

More information on the data formats and functions supported by the DFG Viewer as well as on how to integrate the web service in your own information systems can be found on the DFG Viewer website.<sup>84</sup>

### 2. TEI-DFG standard set (manuscript holdings)

The following information on how to code a digitised manuscript in TEI is limited to the necessary mandatory fields and does not cover the complete spectrum of possibilities. Detailed documentation of all options can be found on the DFG Viewer website.

The basis of the application profile is the TEI specification of the Text Encoding Initiative<sup>85</sup> as well as the preparatory work of the Europeana Regia<sup>86</sup> project, with which the DFG standard set is compatible.

Element / subelement	repeat-able	Comments / explanations	Status
----------------------	-------------	-------------------------	--------

#### 1 Identification data

All elements described in this section are subelements of

`teiHeader/fileDesc/sourceDesc/msDesc`

<code>&lt;head&gt; / &lt;title&gt;</code>	no	Contains the title of the manuscript according to the DFG Guidelines for Cataloguing Manuscripts.	Mandatory if available
<code>&lt;msIdentifier&gt; / &lt;settlement&gt;</code>	no	Contains the geographic location where the manuscript is held.	Mandatory
<code>&lt;msIdentifier&gt; / &lt;repository&gt;</code>	no	Contains the name of the institution where the manuscript is held.	Mandatory

<sup>84</sup> <http://dfg-viewer.de/>

<sup>85</sup> <http://www.tei-c.org/>

<sup>86</sup> <http://www.europeanaregia.eu/>

<msIdentifier> / <idno>	no	Contains a unique identifier within the institution where the manuscript is held, usually the signature, possibly supplemented by an indication of the folios across which the logical structure extends.	Mandatory
<msIdentifier> / <msName>	yes	Contains a non-canonical title of the manuscript.	Recommended

## 2 Information on the origin

All elements described in this section are subelements of

`teiHeader/fileDesc/sourceDesc/msDesc/history`.

<origin> / <origPlace>	no	Contains the place of origin of the manuscript as it should appear in the DFG Viewer.	Recommended
<origin> / <origDate>	no	Contains information about the time of origin of the manuscript as it should appear in the DFG Viewer. Exact numeric entries are made in the attributes <i>when</i> , <i>notBefore</i> , <i>notAfter</i> , <i>from</i> , and <i>to</i> , if possible.	Recommended

## 3 Information on the physical condition

All elements described in this section are subelements of

`teiHeader/fileDesc/sourceDesc/msDesc/physDesc/objectDesc`.

<supportDesc material="...">	no	The material is specified in the <code>material</code> attribute. It is recommended to use the attribute values <i>paper</i> , <i>parchment</i> , <i>papyrus</i> , and <i>mixed</i> .	Mandatory if available
<supportDesc> / <support>	no	Contains information on the writing material as it should appear in the DFG Viewer.	Mandatory if available
<supportDesc> / <extent> / <measure>	no	Contains quantitative information such as the number of folios, which is also displayed in the DFG viewer.	Mandatory if available
<supportDesc> / <extent> /<dimensions>	no	Contains information on physical dimensions, such as the sheet format, which is also displayed in the DFG Viewer.  The specific dimensions are given in the subelements <code>depth</code> , <code>height</code> and <code>width</code> .	Mandatory if available

#### 4 Information on content

All elements described in this section are subelements of

`teiHeader/fileDesc/sourceDesc/msDesc/msContents`.

<code>&lt;summary&gt;</code>	no	Contains a summary description of the content.	Recommended
<code>&lt;textLang mainLang="..." otherLangs="..."&gt;</code>	no	Contains information on the languages in which the texts of the manuscript are written.  The <code>mainLang</code> attribute should specify the most commonly used language, while all other languages used should be coded in the <code>otherLangs</code> attribute, separated by spaces. The attribute values must be language codes according to the standards ISO 639-2 or ISO 639-3.	Recommended

#### 5 Administrative information on the digital copy

All elements described in this section are subelements of `teiHeader/fileDesc`.

<code>&lt;titleStmt&gt; / &lt;title&gt;</code>	no	Contains a title for the digital copy. As a rule, this is the headline according to DFG guidelines, consisting of the signature of the digitised manuscript and its title. This field should not be confused with the title and does not appear in the DFG Viewer. The requirement is based on the general TEI specification and is not specific to the DFG Viewer.	Mandatory
<code>&lt;titleStmt&gt; / &lt;funder&gt;</code>	yes	Contains a funder of the digitisation, e.g. "Deutsche Forschungsgemeinschaft". It could also be the digitising institution.	Recommended
<code>&lt;publicationStmt&gt; / &lt;publisher&gt;</code>	yes	Contains the publisher of the digital copy, which is typically the digitising institution. At least one subelement of <code>publicationStmt</code> is mandatory.	Mandatory if available
<code>&lt;publicationStmt&gt; / &lt;pubPlace&gt; / &lt;ptr target="..." cRef="..." type="..." /&gt;</code>	yes	Contains a persistent identifier of the digital copy. If the identifier is a URL, it is specified in the <code>target</code> attribute, otherwise in the <code>cRef</code> attribute. In addition, the type of the identifier must	Mandatory if available

		be specified in the <code>type</code> attribute with the values <i>urn</i> or <i>purl</i> . At least one subelement of <code>publicationStmnt</code> is mandatory.	
<code>&lt;publicationStmnt&gt; / &lt;idno&gt;</code>	yes	Contains an identifier of the dataset. Optionally, its type can be specified in the <code>type</code> attribute.	Mandatory if available

### 3. Example of a METS/TEI dataset according to the DFG standard

The following example shows the METS/TEI dataset, due to its length only in excerpts. The excerpt represents a complete descriptive metadata set that contains all the mandatory and optional fields as described above. This excerpt is located within a `dmdSec` section of a METS file.

```

<tei:teiHeader>
  <tei:fileDesc>
    <tei:titleStmt>
      <!-- Fingierter Titel für das Digitalisat (obligatorisch)
      Es handelt sich hierbei nicht um den bibliografischen Titel der
      Handschrift, sondern um die nach DFG-Richtlinien gebildete Schlagzeile
      Bestehend aus Signatur und wissenschaftlichem Titel. Der Titel kommt im
      DFG-Viewer nicht zur Anzeige, ist aber aufgrund der TEI-
      Spezifikation dennoch verpflichtend. -->
      <tei:title>
        Zeugbuch Kaiser Maximilians I. - BSB Cod.icon. 222
      </tei:title>
      <!-- Förderer der Digitalisierung (optional) -->
      <tei:funder>Deutsche Forschungsgemeinschaft</tei:funder>
    </tei:titleStmt>
    <tei:publicationStmt>
      <!-- Herausgeber des Digitalats (obligatorisch, sofern bekannt) -->
      <tei:publisher>Bayerische Staatsbibliothek</tei:publisher>
      <!-- Persistenter Identifikator für das Digitalisat (obligatorisch) -->
      <tei:pubPlace>
        <tei:ptr
          target="http://daten.digitale-sammlungen.de/~db/0002/bsb00020956/"
          type="purl"/>
        </tei:pubPlace>
      <!--Identifikator für den Datensatz (obligatorisch) -->
      <tei:idno type="zend">bsb00020956</tei:idno>
    </tei:publicationStmt>
    <tei:sourceDesc>
      <tei:msDesc>
        <tei:msIdentifier>
          <!-- Geographischer Aufbewahrungsorts (obligatorisch) -->
          <tei:settlement>München</tei:settlement>
          <!-- Name der bewahrenden Institution (obligatorisch) -->
          <tei:repository>Bayerische Staatsbibliothek</tei:repository>
          <!-- Signatur o. eindeutiger Identifikator (obligatorisch) -->
          <tei:idno>Cod.icon. 222</tei:idno>
          <!-- Nicht-kanonischer Titel (optional, wiederholbar) -->
          <tei:msName>Zeugbuch Kaiser Maximilians I.</tei:msName>
        </tei:msIdentifier>
        <tei:head>
          <!-- Handschriftentitel (obligatorisch, sofern vorhanden) -->
          <tei:title/>
        </tei:head>
        <tei:history>
          <tei:origin>
            <!-- Entstehungsort (empfohlen, sofern bekannt) -->
            <tei:origPlace>Innsbruck</tei:origPlace>
            <!-- Entstehungszeit (empfohlen, sofern bekannt) -->
            <tei:origDate notBefore="1500" notAfter="1505">
              um 1502
            </tei:origDate>
          </tei:origin>
        </tei:history>
      </tei:msDesc>
    </tei:sourceDesc>
  </tei:fileDesc>
</tei:teiHeader>

```

```

    <tei:physDesc>
      <tei:objectDesc>
        <!-- Angabe des Materials (obligatorisch, sofern ermittelbar) -->
        <tei:supportDesc material="mixed">
          <!-- Angaben zum Beschreibstoff (obligatorisch, sofern
            ermittelbar) -->
          <tei:support>
            Pergament (2-8, 11-30, 34, 48, 295, 296) und Papier
          </tei:support>
        </tei:supportDesc>
        <tei:extent>
          <!-- Angaben zum Umfang wie z.B. die Blattzahl
            (obligatorisch, sofern ermittelbar) -->
          <tei:measure type="leavesCount">2 + 276 Blätter</tei:measure>
          <!-- Blattformat (obligatorisch, sofern ermittelbar) -->
          <tei:dimensions unit="mm" type="leaves">
            <tei:height>420</tei:height>
            <tei:width>285</tei:width>
          </tei:dimensions>
        </tei:extent>
      </tei:objectDesc>
    </tei:physDesc>
    <tei:msContents>
      <!-- Summarische Beschreibung des Inhalts (empfohlen) -->
      <tei:summary/>
      <!-- Textsprache (empfohlen, sofern bekannt) -->
      <tei:textLang mainLang="ger">Deutsch</tei:textLang>
    </tei:msContents>
  </tei:msDesc>
</tei:sourceDesc>
</tei:fileDesc>
</tei:teiHeader>

```

The rest of the structure of the METS file is analogous to the documentation for printed works. Within the `structMap` element with the `TYPE=LOGICAL` attribute, the logical structure of the manuscript is coded in the METS file. Each logical structural unit (e.g. a text fragment) may, in turn, refer to its own section with descriptive metadata, which must also follow the structure described above.

## Appendix C: LIDO core elements for publication

### 1. LIDO core elements for pictorial and three-dimensional material

The LIDO standard is a harvesting format for collating descriptive and administrative metadata for searching and presenting rare, pictorial and three-dimensional objects in online environments.

Only a few mandatory fields are required. The general description elements can be differentiated for different object types by using a type attribute, always retaining the standard defined semantics of the element. The basic set may be extended depending on the material type and design of the project.

For further explanations about the use of individual fields, please refer to the LIDO specification and other information on the LIDO website at <http://www.lido-schema.org>.

*Preliminary remarks:* Many LIDO elements (e.g. object type, material, technique) are assigned data values that are ideally defined as concepts or “units of thought” in controlled vocabularies (thesauri, classification systems or word lists)<sup>87</sup>. These elements, which are indicated below by the word *Concept*, always consist of the subelements *conceptID* to specify the identifier in the controlled vocabulary and *term* to specify a name for the concept.

#### Descriptive metadata

<i>Element / subelement</i>	<i>Repeat able</i>	<i>Comments</i>	<i>Status</i>
-----------------------------	--------------------	-----------------	---------------

#### **1 LIDO identification number**

<lidoRecordID>	Yes	Unique identification number for the LIDO data set, typically comprising the data provider’s ISIL number and the local data set number. (For ISIL numbers for museums, see <a href="http://www.museen-in-deutschland.de/">http://www.museen-in-deutschland.de/</a> )	Mandatory
----------------	-----	---	-----------

<i>Element / subelement</i>	<i>Repeat able</i>	<i>Comments</i>	<i>Status</i>
-----------------------------	--------------------	-----------------	---------------

<sup>87</sup> The vocabularies may be published and centrally maintained or unpublished and locally maintained. The first option is generally to be preferred for networking purposes.

**2 Object type**

<objectWorkType>	Yes	Object or work type Concept	Mandatory
------------------	-----	--------------------------------	-----------

**3 Classification**

<classification type="...">	Yes	A concept that categorises an object within a larger context. The (optional) type attribute qualifies the type of classification, e.g. class or biological system. Concept	Mandatory if applicable
-----------------------------	-----	---	-------------------------

**4 Title / object name**

<titleSet>	Yes	Envelope for title information; if the object has no title then one must be generated, e.g. from the object type.	Mandatory
< titleSet> / <appellationValue>	For language variants only	Title or name given to an object.	

**5 Repository / Location**

<repositorySet type="current">	Yes	Envelope for information about current storage location	Mandatory if applicable
<repositorySet> / <repositoryName> / <legalBodyName> / <appellationValue>	For language variants only	Name of institution holding the object	
<repositorySet> / <workID>	Yes	A unique numerical or alphanumerical ID number given to the object by the institution that holds it; typically the inventory number	
<repositorySet> / <repositoryLocation> / <namePlaceSet> / <appellationValue>	For language variants only	Name of location (particularly important for architectural objects)	

<b>Element / subelement</b>	<b>Repeatable</b>	<b>Comments</b>	<b>Status</b>
-----------------------------	-------------------	-----------------	---------------

**6 Object description**

<objectDescriptionSet type="...">	Yes	Envelope for descriptive texts and their sources. The (optional) type attribute qualifies the type of description, e.g. object history or catalogue text.	Recommended
< objectDescriptionSet> / <descriptiveNoteValue>	For language variants	Brief essay-like text describing the object.	

**7 Measurements**

<objectMeasurementsSet >	Yes	Envelope for measurements	Mandatory if applicable
<objectMeasurementsSet > / <displayObjectMeasurements>	For language variants only	Measurements given in text form	
<objectMeasurementsSet > / <objectMeasurements> / <measurementsSet>	No	Measurements given in structured form with the subelements measurementType, measurementUnit, measurementValue	Recommended

**8 Event**

<eventSet>	Yes	Envelope for information about an event in the life-cycle of the object	Mandatory if applicable
<eventSet> / <event> / <eventType>	No	Nature of relationship between object and event. Depending on the object type, typical events might be: production, discovery, use. For recommended event types, see <a href="http://terminology.lido-schema.org/eventType">http://terminology.lido-schema.org/eventType</a> Concept	

<b>Element / subelement</b>	<b>Repeatable</b>	<b>Comments</b>	<b>Status</b>
<b>Identifiable actors involved in the event</b>			
<eventSet> / <event> / <eventActor>	Yes	Envelope for information about the involvement of a person, institution, group or family in the event	Mandatory if applicable
<eventSet> / <event> / <eventActor> / <actorInRole> / <actor> / <actorID>	Yes	Identifier for the actor involved. Ideally references an authority file. GND should be used for DFG projects.	
<eventSet> / <event> / <eventActor> / <actorInRole> / <actor> / <nameActorSet> / <appellationValue>	For name variants only	Name of actor involved	
<eventSet> / <event> / <eventActor> / <actorInRole> / <attributionQualifierActor>	Yes	Information for attributing an object to an actor; particularly relevant for artists	Recommended
<b>Cultural context of event</b>			
<culture>	Yes	Cultural context of event; mainly relevant for production and use, if no specific individual can be named as an actor Concept	Mandatory if applicable
<b>Date of event</b>			
<eventDate>	No	Date of event	Mandatory if applicable
<eventDate > / <displayDate>	For language variants only	Date in text form, which allows any uncertainties to be noted.	
<eventDate > / <earliestDate>	No	Year or exact date representing earliest date on which the event happened or started.	
<eventDate > / <latestDate>	No	Year or exact date representing latest date on which the event happened or ended.	

<b>Element / subelement</b>	<b>Repeatable</b>	<b>Comments</b>	<b>Status</b>
<b>Time period of event</b>			
<periodName>	Only for stating the earliest / latest period	Time period of event; mostly relevant in archaeological and natural history contexts Concept	Mandatory if applicable
<b>Place where event happened</b>			
<eventPlace>	Yes	Envelope for information about place of event	Mandatory if applicable
<eventPlace > / <place> / <namePlaceSet> <appellationValue>	For name variants only	Name of place	
<eventPlace > / <place> / <gml>	Yes	Georeferences of location	Recommended
<b>Information about material and technique relating to the event</b>			
<eventMaterialsTech>	Yes	Envelope for information about material and technique relating to the event; may vary for different parts of the object.	Mandatory if applicable
< eventMaterialsTech > / <materialsTech> / <termMaterialsTech type="...">	Yes	A material or technique The type attribute qualifies the material or technique Concept	

**9 Theme / content**

<subjectSet>	Yes	Envelope for information about the theme or content of an object Note that different entities (concept, actor, place, event, object) have their own subelements.	Mandatory if applicable
<subjectSet> / <subject> / <subjectConcept>	Yes	Keywords for topic or content of an object Controls ICONCLASS etc. Concept	

**Administrative metadata**

**10 Rights relating to object**

<rightsWorkSet>	Yes	Envelope for information about rights relating to the object / work.	Mandatory if applicable
<rightsWorkSet > / <rightsHolder> / <legalBodyName>/ <appellationValue>	For language variants only	Name of rights-holder.	

**11 Data set**

<recordID>	Yes	Unique identification number in data provider's (local) system.	Mandatory
<recordType>	No	States whether the data set describes a single object, collection, series, group of objects etc. Concept.	Mandatory
< recordSource>	Yes	Source of data set, usually the institution that provided the data	Mandatory
<recordSource> / <legalBodyName>/ <appellationValue>	For language variants only	Name of data set source.	
<recordRights>	Yes	Envelope for information about rights relating to the data set.	Mandatory if applicable
<recordRights> / <rightsHolder> / <legalBodyName>/ <appellationValue>	For language variants only	Name of rights-holder.	
<recordInfoSet> / <recordInfoLink>	Yes	Link to object description.	Mandatory if applicable

**12 Reproductions**

<resourceSet>	Yes	Envelope for information about (digital) reproductions of the object.	Mandatory if applicable
<resourceSet> / <resourceRepresentation >	Yes	Link(s) to digital reproduction(s) of the object, including resolution(s) and technical information on format (audio, video)	
<resourceSet> / <rightsResource> / <rightsHolder>	Yes	Envelope for information about rights relating to digital reproduction.	
<resourceSet> / <rightsResource> / <rightsHolder> / <legalBodyName>/ <appellationValue>	For language variants only	Name of rights-holder.	

## 2. A sample LIDO data set

```

<?xml version="1.0" encoding="UTF-8"?>
<lido:lido xmlns:lido="http://www.lido-schema.org" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.lido-schema.org http://www.lido-
schema.org/schema/v1.0/lido-v1.0.xsd">
  <lido:lidoRecID lido:type="local" lido:source="Deutsches Dokumentationszentrum für Kunstgeschichte -
Bildarchiv Foto Marburg">DE-Mb112/lido/obj/00154983</lido:lidoRecID>
  <lido:descriptiveMetadata xml:lang="de">
    <lido:objectClassificationWrap>
      <lido:objectWorkTypeWrap>
        <lido:objectWorkType lido:type="Sachbegriff">
          <lido:term>Gemälde</lido:term>
        </lido:objectWorkType>
      </lido:objectWorkTypeWrap>
    </lido:objectClassificationWrap>
    <lido:classificationWrap>
      <lido:classification lido:type="Gattung">
        <lido:term>Tafelmalerei</lido:term>
      </lido:classification>
    </lido:classificationWrap>
    <lido:objectIdentificationWrap>
      <lido:titleWrap>
        <lido:titleSet>
          <lido:appellationValue lido:pref="preferred">La Primavera / Der Frühling</lido:appellationValue>
        </lido:titleSet>
      </lido:titleWrap>
      <lido:repositoryWrap>
        <lido:repositorySet lido:type="current">
          <lido:repositoryName>
            <lido:legalBodyName>
              <lido:appellationValue>Galleria degli Uffizi — Pinacoteca (Florenz)</lido:appellationValue>
            </lido:legalBodyName>
          </lido:repositoryName>
          <lido:workID lido:type="Inventarnummer">8360 (Inv. 1890)</lido:workID>
          <lido:repositoryLocation>
            <lido:namePlaceSet>
              <lido:appellationValue>Florenz</lido:appellationValue>
            </lido:namePlaceSet>
          </lido:repositoryLocation>
        </lido:repositorySet>
      </lido:repositoryWrap>
    </lido:objectIdentificationWrap>
    <lido:objectMeasurementsWrap>
      <lido:objectMeasurementsSet>
        <lido:displayObjectMeasurements>Höhe x Breite: 203 x 314 cm</lido:displayObjectMeasurements>
        <lido:objectMeasurements>
          <lido:measurementsSet>
            <lido:measurementType>Höhe x Breite</lido:measurementType>
            <lido:measurementUnit>cm</lido:measurementUnit>
            <lido:measurementValue>203 x 314</lido:measurementValue>
          </lido:measurementsSet>
        </lido:objectMeasurements>
      </lido:objectMeasurementsSet>
    </lido:objectMeasurementsWrap>
    <lido:eventWrap>
      <lido:eventSet>
        <lido:event>
          <lido:eventType>
            <lido:conceptID lido:type="URI" lido:source="http://terminology.lido-
schema.org/eventType">http://terminology.lido-schema.org/lido00007</lido:conceptID>
            <lido:term>Herstellung</lido:term>
          </lido:eventType>
        </lido:event>
      </lido:eventSet>
    </lido:eventWrap>
  </lido:descriptiveMetadata>
</lido:lido>

```

```

</lido:eventType>
<lido:eventActor>
  <lido:actorInRole>
    <lido:actor lido:type="person">
      <lido:actorID lido:type="local" lido:source="Bildindex">kue 02553338</lido:actorID>
      <lido:actorID lido:type="URL"
lido:source="http://www.bildindex.de">http://www.bildindex.de/dokumente/html/kue02553338</lido:actor
ID>
      <lido:actorID lido:type="URI" lido:source="http://d-nb.info/gnd">http://d-
nb.info/gnd/118514008</lido:actorID>
      <lido:nameActorSet>
        <lido:appellationValue lido:pref="preferred">Botticelli, Sandro</lido:appellationValue>
      </lido:nameActorSet>
      <lido:nameActorSet>
        <lido:appellationValue lido:pref="alternative">Filipepi, Alessandro</lido:appellationValue>
      </lido:nameActorSet>
      <lido:nameActorSet>
        <lido:appellationValue lido:pref="alternative">Filipepi, Sandro</lido:appellationValue>
      </lido:nameActorSet>
      <lido:nationalityActor>
        <lido:term>Italien</lido:term>
      </lido:nationalityActor>
      <lido:vitalDatesActor>
        <lido:earliestDate lido:type="evidenceDate">1445</lido:earliestDate>
        <lido:latestDate lido:type="evidenceDate">1510-05-17</lido:latestDate>
      </lido:vitalDatesActor>
      <lido:genderActor>männlich</lido:genderActor>
    </lido:actor>
    <lido:roleActor>
      <lido:term>Maler</lido:term>
    </lido:roleActor>
  </lido:actorInRole>
</lido:eventActor>
<lido:eventDate>
  <lido:displayDate>um 1482</lido:displayDate>
  <lido:date>
    <lido:earliestDate>1472</lido:earliestDate>
    <lido:latestDate>1492</lido:latestDate>
  </lido:date>
</lido:eventDate>
<lido:eventMaterialsTech>
  <lido:materialsTech>
    <lido:termMaterialsTech lido:type="Material">
      <lido:term>Tempera</lido:term>
    </lido:termMaterialsTech>
    <lido:termMaterialsTech lido:type="Material">
      <lido:term>Pappelholz</lido:term>
    </lido:termMaterialsTech>
  </lido:materialsTech>
</lido:eventMaterialsTech>
</lido:event>
</lido:eventSet>
<lido:eventSet>
  <lido:event>
    <lido:eventType>
      <lido:conceptID lido:type="URI" lido:source="http://terminology.lido-
schema.org/eventType">http://terminology.lido-schema.org/lido00034</lido:conceptID>
      <lido:term>Restaurierung</lido:term>
    </lido:eventType>
    <lido:eventDate>
      <lido:displayDate>1982</lido:displayDate>
      <lido:date/>
    </lido:eventDate>

```

```

</lido:event>
</lido:eventSet>
</lido:eventWrap>
<lido:objectRelationWrap>
<lido:subjectWrap>
<lido:subjectSet>
<lido:subject lido:type="Ikonographie">
<lido:subjectConcept>
<lido:conceptID lido:type="URI"
lido:source="http://iconclass.org">http://iconclass.org/23D42</lido:conceptID>
<lido:term lido:pref="preferred">23 D 42</lido:term>
<lido:term lido:pref="alternative">Frühling; Ripa: Ver; Primavera</lido:term>
</lido:subjectConcept>
</lido:subject>
</lido:subjectSet>
<lido:subjectSet>
<lido:subject lido:type="Ikonographie">
<lido:subjectConcept>
<lido:conceptID lido:type="URI"
lido:source="http://iconclass.org">http://iconclass.org/57A9:</lido:conceptID>
<lido:term lido:pref="preferred">57 A 9</lido:term>
<lido:term lido:pref="alternative">Menschlichkeit, Höflichkeit; Ripa: Cortesia, Humanità</lido:term>
</lido:subjectConcept>
</lido:subject>
</lido:subjectSet>
<lido:subjectSet>
<lido:subject lido:type="Ikonographie">
<lido:subjectConcept>
<lido:conceptID lido:type="URI"
lido:source="http://iconclass.org">http://iconclass.org/92C47</lido:conceptID>
<lido:term lido:pref="preferred">92 C 47</lido:term>
<lido:term lido:pref="alternative">spezifische Darstellungsformen, allegorische Darstellungsformen;
Venus als Schutzgottheit</lido:term>
</lido:subjectConcept>
</lido:subject>
</lido:subjectSet>
<lido:subjectSet>
<lido:subject lido:type="Ikonographie">
<lido:subjectConcept>
<lido:conceptID lido:type="URI"
lido:source="http://iconclass.org">http://iconclass.org/92D3</lido:conceptID>
<lido:term lido:pref="preferred">92 D 3</lido:term>
<lido:term lido:pref="alternative">Grazien (Chariten), meist als Dreiergespann; Ripa:
Gratie</lido:term>
</lido:subjectConcept>
</lido:subject>
</lido:subjectSet>
<lido:subjectSet>
<lido:subject lido:type="Ikonographie">
<lido:subjectConcept>
<lido:conceptID lido:type="URI"
lido:source="http://iconclass.org">http://iconclass.org/92E5421</lido:conceptID>
<lido:term lido:pref="preferred">92 E 54 21</lido:term>
<lido:term lido:pref="alternative">Zephyrus entführt Flora (oder Chloris), die in der Regel Blumen
streut</lido:term>
</lido:subjectConcept>
</lido:subject>
</lido:subjectSet>
<lido:subjectSet>
<lido:subject lido:type="Ikonographie">
<lido:subjectConcept>
<lido:conceptID lido:type="URI"
lido:source="http://iconclass.org">http://iconclass.org/96A2351</lido:conceptID>

```

```

    <lido:term lido:pref="preferred">96 A 23 51</lido:term>
    <lido:term lido:pref="alternative">Flora in ihrem Königreich: ein Garten voller Blumen, der ihr von
Zephyrus geschenkt worden ist</lido:term>
  </lido:subjectConcept>
</lido:subject>
</lido:subjectSet>
</lido:subjectSet>
<lido:subject lido:type="Ikonographie">
  <lido:subjectConcept>
    <lido:conceptID lido:type="URI"
lido:source="http://iconclass.org">http://iconclass.org/92D1521</lido:conceptID>
    <lido:term lido:pref="preferred">92 D 15 21</lido:term>
    <lido:term lido:pref="alternative">Cupido schießt einen Pfeil</lido:term>
  </lido:subjectConcept>
</lido:subject>
</lido:subjectSet>
</lido:subjectWrap>
</lido:objectRelationWrap>
</lido:descriptiveMetadata>
<lido:administrativeMetadata xml:lang="de">
  <lido:rightsWorkWrap>
  <lido:rightsWorkSet>
  <lido:rightsHolder>
  <lido:legalBodyName>
    <lido:appellationValue>Galleria degli Uffizi — Pinacoteca (Florenz)</lido:appellationValue>
  </lido:legalBodyName>
  </lido:rightsHolder>
</lido:rightsWorkSet>
</lido:rightsWorkWrap>
  <lido:recordWrap>
  <lido:recordID lido:type="local">obj/00154983</lido:recordID>
  <lido:recordType>
  <lido:term>Einzelobjekt</lido:term>
</lido:recordType>
  <lido:recordSource>
  <lido:legalBodyID lido:type="URI" lido:source="ISIL (ISO 15511)">info:isil/DE-
Mb112</lido:legalBodyID>
  <lido:legalBodyName>
  <lido:appellationValue>Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto
Marburg</lido:appellationValue>
  </lido:legalBodyName>
  <lido:legalBodyWeblink>http://www.fotomarburg.de</lido:legalBodyWeblink>
</lido:recordSource>
  <lido:recordRights>
  <lido:rightsType>
  <lido:conceptID
lido:type="URI">http://creativecommons.org/publicdomain/zero/1.0/</lido:conceptID>
  <lido:term>CC0 1.0 Public Domain Dedication</lido:term>
  </lido:rightsType>
  <lido:rightsHolder>
  <lido:legalBodyID lido:type="URI" lido:source="ISIL (ISO 15511)">info:isil/DE-
Mb112</lido:legalBodyID>
  <lido:legalBodyName>
  <lido:appellationValue>Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto
Marburg</lido:appellationValue>
  </lido:legalBodyName>
  </lido:rightsHolder>
</lido:recordRights>
  <lido:recordInfoSet>
  <lido:recordInfoLink
lido:formatResource="html">http://www.bildindex.de/dokumente/html/obj00154983</lido:recordInfoLink>
</lido:recordInfoSet>
</lido:recordInfoSet>

```

```

    <lido:recordInfoLink lido:formatResource="xml">http://oai.bildindex.de/db/apsisa.dll/oai-
    pmh?verb=GetRecord&identifier=DE-
    Mb112/lido/obj/00154983&metadataPrefix=lido</lido:recordInfoLink>
  </lido:recordInfoSet>
</lido:recordWrap>
<lido:resourceWrap>
  <lido:resourceSet>
    <lido:resourceID lido:type="local">C 654.591</lido:resourceID>
    <lido:resourceRepresentation lido:type="image_thumb">
      <lido:linkResource
lido:formatResource="jpg">http://www.bildindex.de/bilder/fmc654591b.jpg</lido:linkResource>
      </lido:resourceRepresentation>
      <lido:resourceType>
        <lido:term>Negativ</lido:term>
      </lido:resourceType>
      <lido:resourceDateTaken>
        <lido:displayDate>1949</lido:displayDate>
      </lido:resourceDateTaken>
      <lido:rightsResource>
        <lido:rightsType>
          <lido:conceptID lido:type="URI">http://www.europeana.eu/rights/rr-f/</lido:conceptID>
          <lido:term>Rights Reserved - Free Access</lido:term>
        </lido:rightsType>
        <lido:rightsHolder>
          <lido:legalBodyID lido:type="URI" lido:source="ISIL (ISO 15511)">info:isil/DE-
          Mb112</lido:legalBodyID>
          <lido:legalBodyName>
            <lido:appellationValue>Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto
            Marburg</lido:appellationValue>
          </lido:legalBodyName>
        </lido:rightsHolder>
      </lido:rightsResource>
    </lido:resourceSet>
  </lido:resourceWrap>
</lido:administrativeMetadata>
</lido:lido>

```

## Appendix D: Describing Collections and Holdings

### 1. Core elements for describing collections and holdings

Descriptions of collections and holdings may be prepared in accordance with the Dublin Core Collections Application Profile<sup>88</sup> or in the same metadata standard in which the object descriptions are made available: METS, MODS, TEI headers, EAD/DDB and LIDO all offer the necessary features.

On the basis of the description fields specified in ISO 27730, Information and Documentation – International Standard Collection Identifier (ISCI) for a collection defined in the Dublin Core Collections Application Profile, the following core elements are recommended for describing collections and holdings:

<b>Element</b>	<b>Repeatable</b>	<b>Comments</b>	<b>Status</b>
<dc:identifier>	Yes	Collection identifier	Mandatory
<dc:type>	Yes	Type of resource described: Collection, holding	Mandatory
<dc:title>	Yes	Name of collection	Mandatory
<dc:language>	Yes	Language of collection objects	Mandatory if applicable
<dcterms:abstract >	Yes	Summary description of collection	Mandatory
<cld:isLocatedAt >	Yes	Institution which holds the collection	Mandatory
<cld:isAccessedVia >	Yes	Means of access to collection objects	Mandatory
<dcterms:provenance >	Yes	Provenance of collection	Mandatory if applicable
<dc:subject>	Yes	Subject of collection	Recommended
<dc:creator>	Yes	Creator of collection	Recommended
<cld:itemType>	Yes	Type of individual objects in collection	Recommended
<cld:itemFormat>	Yes	Format of individual objects in collection	Recommended
<dcterms:hasPart>	Yes	Part-collection	Recommended
<dcterms:isPartOf>	Yes	Larger collection of which the collection is part	Recommended
<cld:associatedCollection >	Yes	Other collection associated with the collection described	Recommended

<sup>88</sup> <http://dublincore.org/groups/collections/collection-application-profile/>

## 2.1 Example 1 of a collection description in Dublin Core Collections Application Profile

```

<?xml version="1.0" encoding="UTF-8"?>
<cld:cld xmlns:cld="http://purl.org/cld/terms/" xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/" xmlns:marcrel="http://www.loc.gov/loc/terms/relators/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <dc:identifier>DE-Mb112/Bickell1977</dc:identifier>
  <dc:type>Sammlung</dc:type>
  <dc:title>Bickell, Ludwig</dc:title>
  <dcterms:abstract>Ludwig Bickell (1838-1901) war seit 1892 Bezirkskonservator von
Kurhessen und außerdem Begründer des Marburger Universitätsmuseums. Der Bestand umfaßt seinen
fotografischen Nachlaß, etwa 3.000 Glasplatten (teilweise 30 x 40), überwiegend aus dem 19.
Jahrhundert. Im Archiv der Archive befinden sich zusätzlich 5 Abzüge der Photographischen Anstalt
L.BICKELL, weitere konfektionierte Vintageabzüge (Großformat) von 1893 im
Plattenarchiv.</dcterms:abstract>
  <cld:isLocatedAt>Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto
Marburg</cld:isLocatedAt>
  <cld:isAccessedVia>http://www.bildindex.de/?+qfotograf:"bickell ludwig"</cld:isAccessedVia>
  <dcterms:provenance>Marburg, Verein für hessische Geschichte und Landeskunde,
Zweigverein Marburg e. V.</dcterms:provenance>
  <dcterms:provenance>Der Bestand wurde dem Bildarchiv vom &quot;Verein für hessische
Geschichte und Landeskunde, Zweigverein Marburg e. V.&quot; 1977 als Dauerleihgabe anvertraut.
Hinweis: Der sonstige Nachlaß wird im Hessischen Staatsarchiv Marburg verwaltet. (Staatsarchiv
Marburg, StaM 340, Bickell)</dcterms:provenance>
  <dc:subject>Kunst</dc:subject>
  <dc:subject>Architektur</dc:subject>
  <dc:subject>Hessen</dc:subject>
  <dc:creator>Bickell, Ludwig</dc:creator>
  <cld:itemType>Glasnegativ</cld:itemType>
  <cld:itemType>Abzug</cld:itemType>
  <cld:dateTimeItemsCreated>1869-1900</cld:dateTimeItemsCreated>
  <cld:itemFormat>18 x 24</cld:itemFormat>
  <cld:itemFormat>30 x 40</cld:itemFormat>
</cld:cld>

```

## 2.2 Example 2 of a collection description in Dublin Core Collections Application Profile (project description)

```

<?xml version="1.0" encoding="UTF-8"?>
<cld:cld xmlns:cld="http://purl.org/cld/terms/" xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/" xmlns:marcrel="http://www.loc.gov/loc/terms/relators/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <dc:identifi er>http://diglib.hab.de/?link=023</dc:identifi er>
  <dc:type>Virtuelle Sammlung</dc:type>
  <dc:title>Helmstedter Drucke Online</dc:title>
  <dcterms:abstract>Ziel des Projektes ist die Image- und Volltextdigitalisierung der in Helmstedt
gedruckten Werke, die sich in der Herzog August Bibliothek befinden. Die Herzog August Bibliothek
besitzt etwa 10.000 Drucke, die bis zur Schließung der Universität Helmstedt 1810 in Helmstedt
gedruckt worden sind. In der ersten zweijährigen Phase des Projektes sollen die auf das 16. und 17. Jh.
entfallenden Drucke als Image präsentiert, mit Strukturdaten zur Navigation versehen und im
Verbundkatalog sowie im VD 16 und VD 17 nachgewiesen werden. Zugleich wird erstmals
experimentell eine größere Menge von 120.000 Seiten per OCR bearbeitet und im Volltext angeboten
werden. In einer zweiten Phase soll auch der aus dem 18 Jh. stammende Anteil digitalisiert, auf
nationalbibliographischem Niveau katalogisiert und zur Nachnutzung einem möglichen VD 18 zur
Verfügung gestellt werden.
Mit der nahezu kompletten Digitalisierung der Produktion eines der wichtigsten norddeutschen
Universitätsdruckorte - Helmstedt findet sich unter den 10 am häufigsten nachgewiesenen Druckorten
im VD 17 - wird nicht nur ein substantieller Beitrag zur Komplettdigitalisierung des deutsche gedruckten
Kulturerbes geleistet, sondern erstmals ein Überblick über die Druckproduktion einer bedeutenden
frühneuzeitlichen Universität geschaffen. Das Projekt flankiert den programmatischen und auf mehrere
Jahre angelegten Forschungsschwerpunkt der HAB zur Erforschung der Universitätsgeschichte
Helmstedts und wird unmittelbar mehreren Forschungsprojekten in diesem Feld zugute
kommen.</dcterms:abstract>
  <cld:isLocatedAt>Herzog August Bibliothek Wolfenbüttel</cld:isLocatedAt>
  <cld:isAccessedVia>http://diglib.hab.de/?link=023</cld:isAccessedVia>
  <dcterms:provenance>ehem. Universitätsbibliothek Helmstedt</dcterms:provenance>
  <cld:itemType>Images</cld:itemType>
    <cld:itemType>Fulltext</cld:itemType>
  <cld:datelItemsCreated>2010-2014</cld:datelItemsCreated>
</cld:cld>

```