Deutsche
Forschungsgemeinschaft

Scientific Library Services and Information Systems (LIS):

DFG Practical Guidelines on Digitisation

Status: April 2009

DFG

DFG Practical Guidelines on Digitisation
for programmes funding Scientific Library Services and Information Systems.
Developed by the Subcommittee on Cultural Heritage on 12 and 13 February 2009,
and passed by the Committee on Scientific Library Services and Information Systems on 20 March 2009.

Deutsche Forschungsgemeinschaft (DFG)
Head Office: Kennedyallee 40,
D-53175 Bonn, Germany
Postal address: D-53170 Bonn, Germany
Telephone: ++49/(0)228/885-2358
Fax: ++49/(0)228/885-2272
E-Mail: lis@dfg.de
Internet: http://www.dfg.de/lis

DFG

Contents

Introduction to the *Practical Guidelines on Digitisation*

The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), in its funding area Scientific Library Services and Information Systems (LIS), supports projects in Germany that help build powerful, networked and supraregional information systems for all research areas. The results of these projects must be accessible to scientists and academics at no charge and for the long term.[1]

The *Practical Guidelines on Digitisation*, for the funding area Scientific Library Services and Information Systems, aim 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 structural metadata, producing full text, or preserving digital contents for the long term.

The following sections 1 through 4 provide a general and a more comprehensive introduction to the issues and methods relevant to projects that aim to digitise printed works and rare documents, which covers the majority of projects currently underway. These sections are especially geared toward those who are planning such projects and may not have any detailed previous knowledge. Section 5 specifies the presentation standards and formats required by the DFG. Section 6 briefly summarises the most important requirements. Deviations from these rules may be permitted if the project proposal is able to justify them. Section 6 also mentions other types of digital resources (esp. AV media) on which very little practical experience is available and which are therefore not included in the preceding sections. Finally, Section 7 lays out important procedural rules for conducting DFG projects.

Even though it will be repeated several times below, one principle is so important that it should be mentioned upfront: The scientifically motivated digitisation of cultural heritage materials is considered standard, not a technical novelty. When it comes to envisioning projects, this means that it continues to be important to create digitised copies whose quality for research purposes is beyond reproach, but also that it is crucial to use effective and cost-conscious methods which can be applied systematically to large amounts of material.

1 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). The DFG also supports projects that improve scientific information infrastructures in Germany. The results of these projects must be accessible to scientists and academics at no charge and for the long term (http://www.dfg.de/lis). It should be noted that an applying entity's defined institutional tasks and financing should not be substituted by the DFG funds granted under this programme. Projects must therefore exceed an institution's ordinary mission, be of limited temporal and topical scope, and focus on outstanding materials with supraregional 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.

1. Objectives and Selection

1.1 Objectives

Digitisation has become instrumental for providing access to printed and written scientific information. Materials that were once difficult to get hold of or vulnerable to damage can now be viewed conveniently at home or on library and archive computers. This has made direct research with sources much easier, even while conserving valuable and sometimes fragile originals. Not only does the digitisation of historic library and archive holdings make copies easily accessible online; it also helps build an infrastructure that turns the Internet into an integral research space for scholars in the humanities and cultural studies. Only by linking these digital documents with other online resources — such as catalogues, encyclopaedias, bibliographies, editions, secondary literature etc. — can the potential of the Internet be fully leveraged. Thus the objective is not only to make these materials available, but also and especially to integrate them into a network.

While there is a broad base of tested 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. However, as a general rule, there is little need for true pilot projects that experiment with novel techniques or workflows in the area of printed works, since a wealth of experience is already available. These recommendations are based on the assumption that in the 21st century, digitisation is a standard service for scientific information centres to provide, rather than a distinctive feature. In the near future, digital access will be the rule rather than the exception. Digitisation of unique specimens or important collections is not at odds with large-scale digitisation efforts.

Because the majority of previously executed and currently planned projects has been focused on manuscripts and old prints, the techniques and parameters pertaining to this area will be discussed below in especially great detail. However, it would be wrong to conclude that the digitisation of modern materials is not possible or not expected by the research community. Requirements that apply to older materials can certainly be transferred to documents from later eras, which are much greater in number. In the medium term, these *Practical Guidelines* will also include additional recommendations on how to handle photographs, films, broadcasts, 3D images etc. The summary in section 6 already provides initial pointers regarding these types of media.

1.2 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. Therefore each digitisation project must decide in advance whether including 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 foundational 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 that can plausibly formulate its own needs. Ideal are cooperative arrangements in which an academic undertaking, e.g. a research or editorial project in philology or legal history, wants to establish an online presence and link back to library or archive holdings, thus enabling two-way linkage. The DFG offers special funding for this type of

project.[2] Alternatively, relevant subject bibliographies that formulate a canon may be used, or a blend of both approaches may be taken. Also of great interest are the concepts of digitisation on demand and digitisation on use, which assure specific demand in each case. Further discussion of selection issues is available in best practice manuals.[3]

Larger digitisation projects should be part of and dovetail with an overall programme. DFG funding for mass digitisation is contingent on the prior existence of high-quality metadata for the printed works (e.g. when digitising on the basis of the VD 16 / VD 17 bibliographies or when digitising materials from Special Subject Collections).

When it comes to choosing selection criteria, organisational and research priorities may be at odds with each other. For instance, while it may make sense from an organisational point of view to proceed according to time segments or regional categories, researchers generally are only interested in their subject area (regardless of whether a book was printed in the 16th or 17th century, in the Netherlands or in Germany). A large-scale campaign must strive to reconcile both aspects. However, the more comprehensive a digitisation project, the more heavily can organisational considerations be expected to weigh in. Conversely, once the centralised bibliographic tools and interfaces, which are currently under development, have been implemented, projects for the digitisation of large amounts of subject-specific literature or topic-specific sources will certainly be viable.

1.3 Duplicate checking and data matching for image digitisation projects

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

1.3.1 Proposals and reports are expected to mention finished 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.

1.3.2 For large-scope projects (over 1,000 printed works), proposals and reports should explain how they relate to commercial digitisation projects that are accessible free of charge (Google, Microsoft, Yahoo etc.). A reference to the Bavarian State Library holdings digitised by Google is also expected.[4] A pragmatic effort should be made to keep the number of duplicate digitisations as small as possible.

1.3.3 For digitisation projects targeting titles published before the year 1601 or materials not in the public domain,[5] items 1.3.1 and 1.3.2 do not apply. For materials published between 1501 and 1700, the VD 16 and VD 17 bibliographies should be consulted to check for existing digitisations. URNs and URLs of digital copies must be reported to these bibliographies. For digitisations of incunabula, the Census of Incunabula for Germany (ISTC) should be consulted.

2 See http://www.dfg.de/lis

3 For further selection criteria, see e.g. the Minerva Good Practice Handbook: http://www.minervaeurope.org/structure/workinggroups/goodpract/document/bestpracticehandbook1_2.pdf.

4 See Bavarian State Library website (http://www.bsb-muenchen.de/Informationen_fuer_Antragsstel.1844.0.html).

5 See guidelines 12.154e.

2. Digitisation of Printed Works and Rare Documents

Digitisation includes creating digital images and/or capturing full text, as well as generating structural data and metadata. In the following recommendations the term digitisation refers to the entire work process (preparation, digitisation proper, cataloguing / indexing or metadata generation, as well as long-term safeguarding / digital preservation).

2.1 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 materials actually available? Are there any conservation-related objections to digitising the originals? Would it therefore be preferable to digitise an existing microfilm? Are there sufficient personnel to draw out the books? 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. If reproduction could expose an original to risk or undue stress,[6] it should be done on the basis of existing microfilm, if possible, or not at all. It is also conceivable for multiple institutions to cooperate and decide amongst them whether a title that cannot be processed at one institution, e.g. due to a narrow fold, may be digitised at another. At any rate, valuable historic prints must be handled with due conservational care, even if this reduces the scan throughput and takes more time.

2.2 Image Digitisation

Retrospective digitisation should at least consist of image digitisation. Even if machine-readable full text is available, image digitisation or the presentation of a digital facsimile should not be omitted, because a wealth of information can be conveyed only in a visual copy of the print or manuscript. Unless film-based digitisation is indicated for conservational or other reasons, older printed materials up to about 1750 should usually and manuscripts should always be reproduced in colour on the basis of the original. Colour imaging is standard in today's digital cameras and scanners, and the cost of storage materials per MB has decreased exponentially. Moreover, colour management turned out to be not nearly as problematic as was once feared. Problems with capturing can be compensated by including targets or using standardised control mechanisms when generating images. Storage of 48-bit images (not to be confused with 48-bit digitisation) currently makes sense only in rare cases. This image quality offers advantages only for the production of extremely high-quality scans, which is not relevant for most material groups.

2.2.1 Digitisation parameters

Digitisation aims to reproduce a print or manuscript as faithfully as possible, according to applicable scientific requirements. Digitisation parameters should be selected with regard to image quality, long-term availability, and interoperability. In addition to the following rules, the recommendations of the Digital Library Federation (DLF)[7] and other relevant institutions may be consulted for comparative orientation.

6 Petersen, Dag-Ernst (1999). Die Mikroform: Chance und Gefahr für das Buch. IADA Preprints 1999, IXth IADA Congress, 16. – 21 August 1999 in Copenhagen, pp. 181 – 183.

7 See recommendations of the Digital Library Federation: Benchmark for Faithful Digital Reproductions of Monographs and Serials. URL: http://www.diglib.org/standards/bmarkfin.htm, PURL: http://purl.oclc.org/DLF/benchrepro0212

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

The digital master forms the basis for all further processes. Its production and storage should therefore be given special attention. But masters too should not automatically be generated in the best technical quality that is available at the given time. Older literature often takes the position, explicitly or implicitly, that due to the demands of the digitisation process the best technically viable quality should always be strived for, so that an item will never have to be digitised again. This argument is not tenable. A sober, if generous, assessment of the types of use that can be expected should therefore be as self-evident as the care for quality that will stand the test of time.

2.2.1.1 Resolution and image quality

For greyscale or colour images, a minimum resolution of 300 dpi, relative to the format of the original, is recommended as a standard. For manuscripts or maps with very fine lines or writing, 400 dpi may be necessary. Bitonal scans require 600 dpi.

Higher resolutions are rarely helpful since the above standard values generally ensure the visibility of all the important information. The situation may be different with special investigations such as the examination of paper structures, which require significant magnification; however, such cases are beyond the scope of these recommendations.

Yet resolution is only one of several aspects that determine image quality; another aspect is technique. The generated image should therefore be carefully checked for colour fidelity and faithfulness to the original. For this purpose, it is recommended to calibrate the monitors[8] and establish a controlled lighting environment, in order to be able to objectively evaluate the images on the screen.

When using scanners, the distance to the target object is fixed. Scanner resolution remains constant up to a set maximum object size (e.g. 300 dpi up to DIN A3). For digital cameras, resolution depends on the distance to the target object. Ensuring a resolution of 300 dpi requires one-time calculation of the maximum object size for a given camera. For example, if the camera matrix has 4,000 × 3,000 pixels (12 million pixels), only objects up to a size of 33.9 × 25.4 cm may be photographed. This is calculated using the following formula:

$$\frac{\text{number of camera pixels (dots)}}{\text{resolution in dots per inch (dpi)}} = \text{max. object size in inches (1 in = 2.54 cm)}$$

For 300 dpi:
long side: 4,000 dots ÷ 300 dpi = 13.3 in = 33.9 cm
short side: 3,000 dots ÷ 300 dpi = 10 in = 25.4 cm

8 This is usually done with special calibration tools: a measuring instrument is attached to the screen with a suction cup and checks the monitor's colour fidelity using a predefined colour scale. Included software then generates a profile that allows correction of divergent colour levels.

When digitising film, the reduction factor of the microfilm or microfiche relative to the original has to be considered to ensure the target resolution of 300 dpi. For example, if an original with a size of 24 × 36 cm has been recorded on a traditional slide sized 24 × 36 mm, then the target resolution has to be multiplied by the factor 10 when scanning from the film; in other words, the slide has to be scanned at 3,000 dpi. To exactly determine the correct resolution for the scan, it is necessary to know the size of the original object or at least to make an educated estimate of it (e.g. folio volume no larger than 40 cm), in order to avoid falling short of the target resolution of at least 300 dpi relative to the original size.

2.2.1.2 Colour depth

Bitonal scans (black/white) use a colour depth of 1 level (1 bit) per pixel. Thus each pixel takes the form of either 1 (= black) or 0 (= white). Greyscale images are digitised at 256 levels per pixel. Colour images use the three colour channels red, green and blue (RGB colours) and are hence a combination of 3 × 256 levels. Thus a colour image is technically a triple greyscale image, in which the colour values for each channel are additively combined and applied to a pixel (e.g. 35 red + 233 green + 186 blue). This results in a total of 16.7 million colours (256 × 256 × 256). Differentiation with 256 levels requires 8 bits, or 1 byte, in the computer (each bit takes the form of either 0 or 1). A colour image thus has a colour depth of usually 24 bits (3 × 8 bits = 3 bytes). Some camera and scanner manufacturers offer even greater colour depth of up to 48 bit, which is not really necessary since 24-bit colour depth is entirely sufficient for computer screen display; however, it may be helpful when editing a scanned image, since colour values can get lost due to level correction. *Capturing* images at 48 bits is often advantageous and may provide a more balanced picture, since cameras see the colour spectrum differently than humans.

Archiving 48-bit images, however, is rarely justified (see 2.2). It should also be considered that a 48-bit image takes up twice the storage space, which can become a real problem for very large archives. Colour depth over 24 bits should therefore be limited to materials that require — for scientific reasons — the most accurate colour rendering possible, specific colour spaces, or potentially comprehensive editing.

2.2.2 File formats

According to current knowledge, image masters of greyscale or colour images should be archived in TIFF (Uncompressed). For bitonal images, TIFF with Group 4 Compression may be used. TIFF has been around since the eighties and has established itself as one of the most important standards. It is expected that all standard programmes will continue to support this format. However, this holds true only for so-called Baseline TIFFs. The factual background behind these terms: The TIFF standard as a whole is extraordinarily rich and allows saving images even with exotic properties — e.g. images split into tiles that allow loading individual parts of the picture independently of each other, which is very useful for areas like high-resolution cartographic imaging. However, a very rich standard is difficult to implement in its entirety. For this reason, the TIFF standard distinguishes between a relatively small core of image properties that *must* be supported by any application claiming to support TIFF, and numerous extensions that *may* be used by applications but can be ignored by other TIFF-supporting software without waiving the “TIFF support” claim.

If storage space is an issue, images could theoretically be archived using a lossless compression method (e.g. LZW-compressed TIFF). However, a compressed format always comes with a risk, because even minimal damage to the file or isolated bit flips may compromise the entire picture. Such damages may occur due to defective storage media or when files are copied. This is especially common with the JPEG standard. Broken JPEGs that show only part of the picture are a familiar sight. Images with this type of damage usually cannot be repaired. The tricky part is that an image may be fine on one medium but become damaged

during the copying process when migrating to a new storage medium. Therefore JPEG is not a suitable format for archiving. If compression is unavoidable for cost reasons, PNG or TIFF (LZW) should be used as file format.

A new development is JPEG2000[9], which offers not only a new and more efficient compression algorithm but is also less vulnerable compared to JPEG. Unlike the latter, JPEG2000 allows lossless storage. Additional advantages are progressive image transmission (the more of an image is loaded, the more details are visible) and the possibility to include metadata. What may prove to be the key benefit of JPEG2000 is its ability to generate from a large image a variety of resolutions and even details, which would make image archives simpler to manage because storing different resolutions would no longer be necessary. Whether or to what extent JPEG2000 is suitable to replace TIFF as a master format remains to be seen. Given that popularity and software support are key criteria when choosing a master format, JPEG2000 cannot currently be recommended for archiving purposes.

For Internet publishing, JPEG and PNG are recommended due to their great popularity. GIF, due to its limited colour palette, is only viable for bitonal and greyscale scans. But ultimately any desired format may be used because all formats that are supported by standard browsers can simply be generated from the master. Thus a poor decision can easily be revised, and changed conditions can quickly be responded to.

2.3 Full text generation

Full text can be searched, analysed quantitatively, and processed. It can be integrated into larger collections of texts, further indexed according to specific criteria, or prepared for new reading devices. Full text includes the characters of the master copy, markup data to identify structural features, and metadata, which are usually part of the same file.

Full text can be generated in two ways: through 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. It is important to establish at the beginning of the project why full text generation is being undertaken. This objective should be kept in mind throughout the digitisation process and referred to again and again, especially when deciding whether and how certain text features should be recorded.

Prior to the actual capture, be it by transcription or OCR, it is often useful to make a copy of the digitised image (in addition to the unmodified master scan) and process it further, either with scanning software or stand-alone applications. This makes it possible to improve text recognition by correcting distortion, sharpening, adjusting contrast, straightening lines etc.

2.3.1 Character encoding

All common operating system support Unicode, and 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, preferably UTF-8 or UTF-16.

However, documents frequently contain characters that are not part of the Unicode standard. There are several options to encode such characters (private use plane in Unicode, entities, graphics).

When encoding, one should also consider whether or not it is essential to the respective subject area that certain typographic details — such as the differences between the long and short s or ligatures (ch, tz etc.) in Fraktur typeface — be preserved.

9 http://www.jpeg.org/jpeg2000/

2.3.2 Markup of structural data of printed works (see also *2.6.2*)

When marking up structural data, one must first decide how and to what extent text-type-specific divisions like chapters, subchapters, bound volumes, articles etc. should be identified. Other possible structural data include: table of contents, register, line break,[10] column break, page break, header / footer / column title, page number, image and image-like element, image caption, marginal note, change of font (e.g. from Fraktur to Antiqua for foreign-language quotations), change of font size, change of font style (regular, italic, bold etc.), formula (e.g. mathematical [MathML] or chemical [CML]), continuation mark (catchword) at the bottom of a page (to connect sheets) etc.

These examples are not requirements but rather aspects that should be considered when undertaking full text digitisation. In any case, the chosen options should be documented.

2.3.3 Layout

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 XML markup language (e.g. XSLT, XSL:FO) that largely ensures independence from special software. If valid reasons prohibit archiving the format with XML techniques, text documents may also archived according to ISO standard 19005-1 (PDF/A), which specifies a safe subset of PDF. But PDF files can never substitute the provision of structural data in XML format, because the PDF format does not allow this.

2.3.4 Text capture

OCR

As of this writing, optical character recognition (OCR) produces acceptable results only beginning with younger Antiqua fonts in the 19th century and Fraktur fonts in the second half of the 19th century. However, with a dynamic provider market and new products becoming available all the time, these *Practical Guidelines* cannot currently give further recommendations on OCR applications and their usability.

Especially for large holdings of uniformly printed text, OCR software may deliver usable results if trained appropriately. Currently no OCR application is able to switch between Fraktur and Antiqua typeface on its own. If the different fonts are segmented by blocks, the switch can be made manually. If fonts alternate within continuous text, only the preset typeface will be recognised reliably. If the master copy contains formulas etc. in addition to text, today's OCR applications are not always able to recognise such elements.

The situation is different if a "dirty" version is to be used (a.k.a. dirty OCR). In those cases, the full text is only utilised for positive searches but not as a reliable text basis. A positive search shows only positive hits; thus a negative result does not guarantee that there are no hits, nor does a positive result necessarily include all theoretically possible hits. This method can be very useful as a transitional solution until better texts are available. A potential drawback is that users may come to erroneous conclusions if documentation is poor. Search conditions should therefore be clearly and prominently spelled out. It is also helpful to display the full text, even if its quality is questionable, in order to enable users to form their own impression of the quality of their research basis.

Manual input / 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

10 Note the following problem: If a separating hyphen is preserved, most search engines will not find the word. Conversely, if separating hyphens are omitted, then the digital text is not completely true to the original text.

discrepancies are filtered out. This allows transcription accuracies of up to 99.997%, i.e. virtually error-free texts. When choosing this type of transcription, one should not be misled by service providers claiming ostensibly high percentages; because results with less than 99.5% accuracy are essentially worthless for manual text entry (99% accuracy means that 1 out of 100 letters is wrong, which comes out to about one error per line).

If transcription is to be outsourced to a service provider, the contract must specify the appropriate target accuracy. Adherence to this standard should be verified by spot checks of the digitised text. If the target accuracy is not met, contractual repercussions should kick in, e.g. price reduction or non-payment.

While manual entry is also error-prone, double keying followed by automatic discrepancy checking can deliver the best text quality. However, this procedure is currently the most costly one. The actual transcription is usually done outside of Germany; however, a contracted digitisation provider should have a representative in Germany because close cooperation and consultations on the details of the text input 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 for this purpose.

2.4 Long-term preservation

There is currently no blanket solution for the long-term preservation of digital contents that works for all types of objects and materials. Key criteria for successful long-term archiving of digital documents are

(1) creating the right organisational and economic framework, and
(2) establishing the right technical environment, along with choosing suitable techniques and strategies.

On the one hand, the long-term safety of the results of digitisation projects depends on the choice of data and metadata formats (see above). On the other hand, it must be ensured that the digital data remain physically available. It should be noted that the DFG will underwrite the costs of long-term archiving only if the data copies created are secure for the long term. Proposals are expected to explain which institutional measures will be taken to safeguard, for the long term, the data generated under the project. Applicants must state that, as part of the overall concept, a budget is provided for the long-term continuation of the software platform required for the digital service, and that an explicit plan exists for the long-term preservation of data.

The issue of archiving is often not given enough attention, and the costs and efforts it entails are often underestimated. The higher the resolution a camera or scanner provides, the more storage space is required for the images created. The digital master of a colour image may have 20 to 80 MB or more. Image archives comprising several terabytes may quickly accumulate and need to be stored.

Currently, four types of carrier media are used for most digital archiving:

(1) Removable optical media like CDs and DVDs

(2) Tape drives (streamers)
(3) Hard disk drives
(4) Microfilm

(1) Storage on removable optical media like CD-Rs or DVD-Rs is not efficient for large-scale digitisation projects and cannot be funded. Besides, CDs and DVDs are not suitable media for long-term preservation, given a rapidly evolving technology and the consistency of the material. CD or DVD storage may be viable for those who are beginning to build a digital archive or want to digitise only a limited number of items.[11] But with larger data volumes, capacity limits are quickly reached. Standard CDs hold 700 MB, DVDs 4.7 GB or, with dual-layer technology, about 8 GB. Storing just 1 terabyte (the equivalent of about 180 books[12]), would require 1429 CDs, 213 DVDs at 4.7 GB, or 125 DVDs at 8 GB. Considering these numbers, archiving on CDs quickly gets to be cumbersome. But even DVDs do not offer a truly satisfying solution for large amounts of data. Discs need to be burned, labelled, and archived, and these tasks can be very time-consuming. Direct access is complicated and location-based, unless robots or CD servers are used. If CDs / DVDs are used for archiving, spot checks should be conducted routinely, and redundant (at least double) storage is highly recommended.

(2) Tape storage offers a somewhat more convenient method for mass archiving; its drawback, however, is that tapes are relatively slow. If digital masters need to be accessed frequently, then tapes are not a good solution. Moreover, they have to be operated regularly to prevent sticking. University libraries — or libraries with organisational ties to universities — which want to implement a tape-based archiving system are strongly recommended to consult with their university's computer centre when deciding on a long-term archiving strategy. Modern tape archiving systems (robots), designed for several hundred terabytes, ensure that several copies are made of each tape, and that the tape cartridges are operated with the necessary frequency.

(3) Especially recommended with a view toward migration is redundant data storage on hard disks (e.g. RAID 5) in the form of Network Attached Storage (NAS) systems or Storage Attached Networks (SAN) in data centres. For safety reasons, a tape backup or an additional hard disk copy has to be made (tape backup or disk copy must not be kept at the same location). This model allows fast and uncomplicated access to data and facilitates any migrations that may become necessary. The hard-drive storage model assumes that data are kept current on an ongoing basis, and that none are ever set aside without hardware or software processes in place that safeguard integrity. In that regard, this model takes a different and as yet untested approach to long-term archiving of digital media — all in all a matter on which the jury is still out. Notable national and international activities in this field include the Nestor[13] and Portico[14] projects.

(4) Another way to archive digital reproductions for the long term is to print them to microfilm as part of a conversion strategy. Compared to other analogue photo materials, microforms are the most durable when stored under optimum conditions. If desired, these films can be converted back to a digital format at a later point with the help of film scanners.[15] There are also solutions available now that allow the printing of high-quality digital colour reproductions to colour microfilm.[16] These newly developed technologies cannot yet be evaluated conclusively.

11 Iraci, Joe (2005). Die relative Haltbarkeit verschiedener optischer Speicherplatten: CD, DVD. Restaurator 26.2, 134 – 150.
12 Realistic estimate for 1 TB: on average 25 MB per image, 220 pages per book; hence 5.5 GB per book, or 1 TB for 182 books.
13 http://www.langzeitarchivierung.de/
14 http://www.portico.org/
15 Such film scanners have been around for some time for b/w microfilms.
16 See e.g. the ARCHE project (http://www.landesarchiv-bw.de/arche). However, there is currently no satisfying solution for efficient film digitisation, although developments along these lines are underway.

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. Smaller institutions may take advantage of the expertise and services of large institutions.

Library projects funded under the scientific information infrastructure programme are encouraged — or may in some cases be legally required — to submit complete data sets to the German National Library.[17]

2.5 In-house or outsourced digitisation?

Digitisation may be done in-house or by contracting a service provider. In the former case, expenses for equipment and project-specific personnel are eligible for DFG funding. Especially for larger quantities it is sometimes less costly and more practical to do the work in-house on an institution's own equipment and by its own staff. Other times it may be better and more economical to hire an external service provider. Lastly, in individual cases it may make sense to request funding for staff or direct project costs in order to use them at another location, e.g. a large library with relevant digitisation expertise.

Thus 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. 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 historic holdings should have an appropriate track record. Since the criteria and guidelines in this area are the same as for film, the issue requires no further discussion.

Checking the quality delivered by a contractor is usually not an easy task. In film digitisation it is often not clear whether poor digitisation is due to a bad film or the service provider's insufficient scanning technique. To facilitate adequate evaluation of service quality, one should make sure that a colour chart and a ruler are included when digitising from originals, to allow reliable assessment of colour fidelity and resolution.

It should be emphasised that while contractors are without doubt inexpensive and effective when it comes to carrying out certain tasks, this does not relieve the contracting library from its responsibility to be knowledgeable about digitisation (to ensure both long-term continuation of digital services as well as effective collaboration with suppliers).

Previous experience has shown that even though outsourcing subtasks of a digitisation project to external service providers *can* be very effective, this is by no means always the case. As a general rule, even the comprehensive use of contractors does not relieve 21st-century libraries and archives from their obligation to maintain in-house expertise in the area of digitisation — if only to be able to negotiate competently with suppliers. It bears repeating that the DFG assumes that the digital reproductions will be maintained for the long term within an institution's in-house infrastructure. An institution submitting a proposal should therefore have sufficient funds available for the staff, equipment and consumables it needs to perform project management and controlling, as well as to select and prepare materials for digitisation. Furthermore, the work of indexing and preparing the digitised materials for users is virtually always done by the applicant libraries, archives and academic institutions themselves.

Given these conditions, hiring external service providers will often be advantageous. 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. Even downstream production stages, like putting data online, selling and marketing offline products, offering specialised printing services, and migrating to long-term storage, can some-

17 http://www.d-nb.de/netzpub

times be taken care of more economically by outside contractors than by libraries and academic institutions.

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

(1) Job parameters must be exactly specified, in particular the requirements and format standards for deliverable raw data. 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.

(2) 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 promise in writing that it will, without delay and free of charge, render substitute performance or rectify defects should this become necessary due to the its non-compliance with the *Practical Guidelines on Digitisation* or other justified quality complaints.

(3) When granting funding, the DFG assumes that digitisation and subsequent use will occur in compliance with copyright regulations, and that the permission of rights holders will be obtained if necessary. It must be ensured that the owner of the original digital reproductions does not cede any rights to contractors.

2.6 Metadata

If data are gathered separately, outside of existing library networks or central portals, metadata must be provided in a software-independent format. 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-independent format — which usually comes down to XML encoding. A project plan is highly problmatic if it delays the creation of manufacturer-independent metadata until late into the project or, using other funds, until after the end of the project. When utilising proprietary software systems, it must therefore be assured from the beginning that data can be output in a manufacturer-independent format.

If the materials generated by a DFG-funded digitisation project are suitable to be integrated with a DFG-funded portal and/or virtual subject library, the proposal is expected to either explain which measures will be taken within the project to ensure this integration for the duration of the project, or to make plausible why such an integration is not necessary or sensible, for topical reasons or due to the effort it would require.

Generally speaking, a distinction is made between descriptive (usually bibliographic or archivistic information), structural (text or document structure), administrative (e.g. rights management), and technical (e.g. file types) metadata. The following discussion relates to descriptive and structural metadata only.

2.6.1 Description of collections and holdings, cataloguing and indexing, descriptive metadata

2.6.1.1 Description of collections and holdings, project information

Even for traditional library and archive services, descriptions of collections and holdings have played an important role in providing users with an overview of the nature and make-up of a historic library or archive. Digitisation projects are expected to present their nature and scope also on an English-language web page. The fact that the project is funded by the DFG should be mentioned. In addition, a standardised description in XML is desirable to facilitate the future merging of this information in national or international portals that enable targeted research. Important in this regard is the Dublin Core Collection Description Application Profile (DC CD),

which adopts elements of the Collection Description Format (CLD) and enriches them with its own elements.[18] Archives should look to the international cataloguing standards ISAD (G)[19] and EAD (Encoded Archival Description)[20].

2.6.1.2 Cataloguing and indexing

Digital reproductions of old prints or archive materials must be assigned at least descriptive (bibliographic) metadata. Digitisation projects that do not provide at least bibliographic metadata or descriptive archive records according to current library and archive standards do not make any sense. Library stock may be recorded either by cataloguing the electronic version or by giving the URL of the image files in the local 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. Digital reproductions must be listed through relevant supraregional portals. It is strongly recommended to provide an OAI interface that delivers METS / MODS in addition to Dublin Core (see Appendix A) to ensure that relevant portals can harvest the data.

Digital reproductions of medieval manuscripts should be recorded in the manuscript database (Manuscripta Mediaevalia). If metadata of medieval manuscripts are recorded or archived separately, then the manuscript format according to TEI P5 should be used.[21]

2.6.2 Structural metadata for image digitisation (see also *2.3.2*)

Worth considering is the use of structural metadata, i.e. encoding a document's structural elements such as dedication, introduction, chapter, or illustration. Inclusion of these aspects takes a cue from analytical bibliographies, which break down the contents of a work along the lines of its chapter and text structure. In many cases, creating such an artificial table of contents is essential to enable users to navigate the digital reproduction. Nobody should have to go through the trouble of trying to find the right place in the alphabet in a 600-page digital dictionary. Therefore the decision whether to generate structural metadata is always object-specific.

If structural metadata are used, it is recommended to consult the list of designations available on the DFG Viewer website.[22] In case additional designations are needed, standardised terms for a given digitisation project should be agreed on, and this typically specialised vocabulary should be published on the project's website to allow others to reuse it.

When assigning structural metadata, the question arises 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, the recommended encoding standard is TEI[23]. For page description with some qualifying features (e.g. illustrations or annotations), the Metadata Encoding and Transmission Standard (METS)[24], as followed by the Library of Congress, could be used. 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 subsequent queries and document displays. However, this advantage comes with the price of greater technical

18 http://www.ukoln.ac.uk/metadata/rslp/schema/
19 Examples of online overviews of holdings can be found on the website of the State Archive of Baden-Württemberg (http://www.landesarchiv-bw.de).
20 For usage of EAD in German archiving, see the German Federal Archive's "daofind" project (www.daofind.de).
21 An modified version based on the DFG guidelines for manuscript cataloguing is available at the Duke August Library Wolfenbüttel (see http://www.manuscripta-mediaevalia.de/hs/kataloge/HSKRICH.htm). For newer manuscripts or literary remains, EAD (http://www.loc.gov/ead/) is recommended.
22 http://www.dfg-viewer.de
23 http://www.tei-c.org
24 http://www.loc.gov/standards/mets

requirements for processing and displaying the documents. 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,[25] 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 them to METS. Since both standards are XML-based and the described features are similar, in can be assumed that conversion will not present a major problem.

2.7 Exchange and dissemination of metadata, publicity

For the development of a decentralised digital library, it is crucial to create a global standard for data and metadata exchange (see chapter 5). However, standards below the level of classic descriptive cataloguing can be developed and established only within each respective community. Identical resources may well be relevant to entirely different inquiries and accordingly require diverging sets of metadata. A generalised procedure for exchanging metadata must therefore be able to handle flexibly not only library metadata but also different metadata formats and community-based specifications. This can be achieved using the protocol of the Open Archive Initiative (OAI)[26]. In terms of old prints and manuscripts,[27] OAI is especially useful as a technical exchange protocol. OAI requires that Dublin Core data be provided as a minimum. While this is insufficient for descriptions of old prints and manuscripts, it is useful as additional information. The OAI standard explicitly provides for the parallel support of additional metadata formats, so that OAI can be combined with any XML-based metadata format (MARCXML, MABxml, EAD, METS / MODS, TEI P5 etc.).

The DFG strongly recommends that metadata be provided via OAI. Metadata according to METS standards (METS / MODS for prints) should be delivered in addition to the mandatory Dublin Core metadata (see also chapter 5). If possible, TEI, EAD or documents with other domain-specific standards should also be delivered within METS, which in this case functions as a wrapper. Through the described OAI functionality, the DFG Viewer can be served. Newly implemented OAI interfaces should be reported to relevant portals and subject libraries. In addition, suitable measures should be taken to ensure that metadata will be found by search engines (e.g. using the Sitemap protocol[28]).

The DFG expects funded projects to undertake targeted efforts for ensuring that the generated resources will be highly visible and frequently used. At a minimum, this should entail the integration of these resources with existing or developing material-specific portals and catalogues as well as with virtual subject libraries.

25 http://www.tei-c.org/Sample_Manuals/bestpractice.htm
26 http://www.openarchives.org/
27 Cf. Diane L. Hillmann, in: Kenny, Anne R. & Rieger, Oya (Eds.). (2000). Moving Theory into Practice: Digital Imaging for Libraries and Archives (pp. 89f.). Mountain View: Research Library Group.
28 http://www.sitemaps.org/

3. Citing Digitised Prints and Manuscripts, Persistent Addressing

When digitisation was in its infancy, the issue of citability of digital resources was frequently underestimated. But it is exactly citability that makes Internet-based digitised sources viable for academic writing. Different from 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 space. Unlike traditional photocopies, digital copies require special citation rules if they are on the Internet. Digital versions on CDs or other non-networked storage media can be handled and quoted just like film or paper copies; but when a copy is online, it needs a unique address so that other documents or databases can link to it. 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.

A positive effect of the net-based citation format, which will usually follow the physical image sequence, 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 corpus proper (cover, endpaper, additional digitised watermarks, partial reproductions of illustrations etc.). The only prerequisite is that a specific image can be unequivocally located in an ascending alphanumeric sequence (e.g. 00001, 00002, 00002a, 00003 etc.; in this example, image 00002a was inserted). Here the image is the reference target. Different mechanisms apply to full text, for which no specific recommendations can be made as yet (XPath and similar techniques are examples of options that allow unequivocal referencing).

As a rule, the highest granularity possible should be strived for. Two functionalities in particular are important for online presentation: the addressability of a work as a whole, and the addressability of individual pages or double-pages within a work. The structure of a reference might look like this (fictitious example):

http://digitalebibliothek.ubique.de?titelid=234&image=0002

At least the accessibility and citability of the work as a whole must be guaranteed. In the future, the work's individual physical pages will also have to be reliably accessible and citable. 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.

It is recommended to generate URNs at the work level via the German National Library.[29]

29 See http://www.persistent-identifier.de/

4. Provision of Digital Prints and Manuscripts to the Public

4.1 Open access

The DFG funds the digitisation of scientific materials in order to make them accessible to researchers in Germany and worldwide. Therefore all projects should be designed such that their results will be available to researchers quickly and for the long term. In virtually all cases, this will entail the provision of digital copies on the Internet.

The DFG is a cosignatory to the Berlin Declaration on Open Access. In the spirit of this declaration, the results of DFG-funded digitisation projects should be accessible free of charge to re-searchers around the world.[30] 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. The digital resource should be provided in a form that allows scientific use in other research contexts (e.g. by offering an image without navigation context). This does not preclude fees for higher-grade copies, derivatives, or other types of media (CD, print etc.). The origin of digital copies should always be clearly identified, also in downstream usage environments.

The DFG expects that the projects it funds include clear credits and make mention of the DFG as funding source on the files provided online. For image digitisations, this is usually done by adding a credits bar to the published user copy (e.g. in JPEG); for full texts, appropriate credits should be included in the header of the text file. Appendix A gives a detailed description of the technical specifications and formats, which apply to all DFG-funded digitisation projects and all types of materials.

For projects that digitise not only public-domain materials and hence cooperate with commercial partners or publishers, delayed publication (“moving wall”) may be agreed on. In these cases, publication may be delayed up to one year after the completion of the project.

4.2 Minimum requirements for provisioning systems of digital libraries

The principles laid out above apply to any kind of project that provides digital content. The following minimal requirements apply specifically to the provisioning of digital files that have the character of digital books or documents. They cover certain basic standards and a catalogue of minimally necessary functionality.

4.2.1 Basic requirements and architecture

The provisioning 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, in order let users navigate documents and collections as they would an open-stack library arranged by subject. 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 provisioning systems of the individual libraries and archives should allow access across institutions, both in navigating digital collections or holdings and in searching indexes. In addition, the transparent linkage of provisioning systems with local catalogue systems and network databases is desirable.

30 http://oa.mpg.de/openaccess-berlin/berlindeclaration.html

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

(1) Metadata are stored centrally in a catalogue system (e.g. the local OPAC or a library network catalogue), while digital document files (incl. electronic table of contents and index) 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.

(2) A document management system (DMS) or Content Management System (CMS) is used.

4.2.2 Functionality requirements

Regardless of the architecture chosen, the following functionalities must be provided as a minimum:

Collections / holdings may be accessible in a variety of ways:

- via the providing institution's website;
- via an OAI interface;
- via a locally implemented or externally operated DFG Viewer (see chapter 5);
- via a search inquiry to the local and regional library catalogue / the local online finding-aids system;
- via the virtual subject libraries' shared portal or one of the DFG-funded material-specific portals that enable integrated access to all digital collections funded under the DFG programme,
- via Internet search engines.

In addition to being able to access specific documents in a targeted way by means of a metadata search, users should also have the option for structured browsing in predefined collections, collection sections 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-fielded search masks that require a solid understanding of the data structure of a given collection or inventory.

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 (for books from 1850 onward)[31]
- Metadata info: View current document information in description fields stored in DMS
- 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 signal to the user the current location within the digital

31 With today's technology, OCR should always be considered for machine-press era prints from 1850 onward.

document. If a server contains materials that users will normally regard as conceptual units (multivolume works), these units must be visible as such.

In addition, the following functions must be implemented:

- Download[32]
- Print as PDF[33]
- Centralised DFG-funded information systems (VD 16, VD 17, subject portals etc.) should first link to a view in the style of the DFG Viewer.

4.2.3 Minimum technical requirements

As far as applicable, servers must be set up to:[34]

(1) 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.

(2) Provide all materials, conversely, in a quality that allows processing via DSL without cumbersome delays.

(3) Enable the free download, for research purposes, of any complete unit as one single file (e.g. of individual printed works).

(4) Support all currently popular browsers, to the extent viable.[35]

32 Download by sections or individual pages should be implemented if the size of the entire file would be unmanageable.
33 Printing by sections or individual pages should be implemented if the size of the entire file would be unmanageable.
34 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 1,600 × 1,200, 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.
35 If a browser does not support a format required by an advanced 3D application, there is no need to bother with developing a suitable plug-in.

5. Presentation Standards (DFG Viewer) and Formats (METS / MODS)

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

(1) Defining a standardised design profile for visualising digital copies that were generated with DFG funding (DFG Viewer).

(2) Creating a defined technical interface on the basis of 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. The goal is to create consistent display and scroll functions that enable homogenous access to decentralised resources even from central search portals, via an XML interface in the METS / MODS format (or, in the future, METS / TEI-P5 for manuscripts and METS / EAD for archive materials) which describes scrolling, metadata display and other basic functions. This interface is compatible with the METS / MODS AP to deliver bibliographic and structural data. If desired, it can be enriched and expanded into a full interface for data delivery, e.g. by OAI. The DFG Viewer can also read OAI data.

For their first supraregional presentations, DFG-funded digitisation projects should serve the aforementioned interfaces and/or implement a DFG-style viewer at their own institution. Appendix A defines the interface's METS / MODS format and the design of the viewer.

6. Checklist for Applicants and Reviewers

As the previous sections demonstrate, there are numerous choices to make when planning digitisation projects, even in the area of digitising conventional library stock. This holds true all the more for projects that aim to provide materials with which there is little experience, such as 3D simulations of buildings, globes etc. Moreover, technology is constantly in flux.

Rigidly prescribing specific standards would therefore unduly restrict the projects to be funded and hamper their continued dynamic development. Then again, the potential trouble spots are well known. The checklist below — which also includes recommendations from areas other than book digitisation — should therefore be understood with the following in mind:

(1) All proposals with digitisation components will be reviewed for their technical concepts, regardless of topical considerations.

(2) Proposals generally have to demonstrate plausibly that the project will be implemented according to the standards listed below.

(3) Any deviation from these standards must be justified in detail.

(4) If a digitisation project plans to exceed the specified standards, the need for doing so should be explained in detail to the extent that this entails higher costs.

(5) The technical preparation of a proposal must be comprehensive enough to allow an overall evaluation of technical requirements and procedures based on 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 conclude prior to the submission of a proposal.

(6) Digitisation projects for cultural heritage materials have been using a generally well-understood technical procedure. It is therefore safe to assume that its costs will continually decline. Applicants and reviewers should keep in mind that later projects should at least not exceed the costs reported by previously concluded projects.

(7) For better comparability, proposals (or work reports) should provide information about estimated (or actual) costs for scanning (raw digitisation) per image.[36]

36 In principle, the DFG wants to know the real costs. However, there are currently no commonly accepted standards for calculating them. The *Practical Guidelines* therefore recommend that the costs of scanning (raw digitisation) be calculated as follows:

(1) For outsourced digitisation:
   (a) Cost 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; 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 digitisation hardware operators.
   (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 are *not* considered costs for scanning (raw digitisation):
   (a) Project management (e.g. selecting and fetching materials to be digitised).
   (b) Metadata entry of any kind (exception: naming and saving files).
   (c) Long-term archiving.
   (d) Indirect costs typically assessed in terms of internal cost accounting.

CHECKLIST for Applicants and Reviewers

6.1 General technical procedures and resources

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

(1) Are the staffing requirements specified both sufficient and necessary? To help answer this question, average available resources (working hours, storage capacity of computers involved in the workflow) per basic volume unit must be stated.

(2) 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, or by the results of self-conducted pretests.

6.2 Data quality and formats

(1) For all materials to be digitised under a project, the quality proposed must permit batch processing without human intervention to generate reproductions meant for immediate publication.

(2) For long-term preservation (archive copies), the following guidelines should be followed (and any deviations[37] justified):

Images:
Masters should be stored as uncompressed Baseline TIFF or PNG files. For monochrome files, the use of TIFF with Group 4 Compression is recommended *(→ 2.2.2).*

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. For most materials, it is assumed that a resolution, relative to the size of the original, of 300 dpi for colour and greyscale images and 600 dpi for monochrome images will ensure this.

Digitisation in the form of monochrome images should be chosen when it is clear that a document contains no pictures or shades of grey. Colour images must be stored as 24-bit images, greyscale images as 8-bit images (→ *2.2.1.1* and *2.2.1.2*).

Proposals that go above or below these guide values must include an explanation and a series of test scans that demonstrate the possibility or necessity of deviating from the norm.

Audio:

Waveform Audio File Format (WAVE) with Linear PCM bitstream (essential: uncompressed) or Audio Interchange File Format (AIFF) with Linear PCM bitstream.

Due to limited experience in this area, no quality recommendations are possible at this time.

37 When deviating from the file format recommendations, which are very restrictive by design, it is strongly recommended to follow the preferences according to http://www.digitalpreservation.gov/formats. The following list reflects, as far as applicable, a subset of these recommendations. Also taken into consideration are the recommendations by the Florida Center for Library Automation, which are based on practical studies: www.fcla.edu/digitalArchive/pdfs/recFormats.pdf.

Moving images (video):
MPEG-1 or MPEG-2 with one of the following profiles: Simple, Main, or 4:2:2. If precise editability of video sequences should be retained for the long term, consider Motion JPEG.

Due to limited experience in this area, no quality recommendations are possible at this time.

3D data:
According to the current state of international standardisation efforts and available tools, X3D and COLLADA are the recommended choices for 3D models that are secure for the long term. The use of VRML is discouraged due to its numerous variations and differently designed viewers and processing tools.

Database content:
If a digitisation project involves databases as tools to access metadata, the requirements specified above must be met. If it involves databases that comprise more than metadata, the following should be noted:

The workflow must be designed such that, in addition to storing data in a given database system, a software-independent version of the database content is generated and will remain available even if the project is terminated abruptly (→ *2.6*). Always suitable for this purpose are XML database extracts on the basis of a documented DTD, and for databases with SQL capability also SQL DDL statements that can be used to create the database (SQL dialects should be avoided if possible).

Text:
Full text can be generated in two ways: by OCR or transcription. It is recommended to save texts in Unicode, preferably UTF-8 or UTF-16.

When presenting a full text, it is important in some cases to secure the document's layout for the long term. These *Practical Guidelines* recommend the use of a suitable XML markup language (e.g. XSLT, XSL:FO) to largely ensure independence from special software.

The decision which important types of structural data to record must be explained and justified.

The choice of text capture method should be justified with regard to the required text quality (→ *2.3*).

6.3 Long-term preservation

The proposal must include a plausible strategy for institutional long-term preservation. Digital reproductions must be archived redundantly. Submission of a complete dataset to the German National Library is encouraged or may in some cases be legally required (→ *2.4*).

6.4 Working with contractors

When it comes to working with contractors, the applicant institution must demonstrate its ability to competently supervise a project. Contracts must exactly specify all services to be provided. The DFG expects that an appropriate percentage of the invoice amount be withheld for security purposes and not paid out until a quality check has been performed. Funding recipients must ensure that no copyrights or other property rights will be infringed (→ *2.5*).

6.5 Metadata

Each digital reproduction, at least at the title level, must be catalogued according to applicable library and archive standards and listed in a central reference system (library network, central portal, virtual subject library etc.). Analogous rules apply to archives. If data are recorded outside of existing library networks or central portals, it is expected that interim results be archived in XML (→ *2.6*).

For the description of text and pages as well as of structural elements of book-like documents, the use of a uniform structural-data format is recommended.[38] EAD should be used for archive materials, and TEI-P5 for medieval manuscripts. In justified cases, it is acceptable to deviate from these recommendations (→ *2.6.1*).

6.6 Exchange and dissemination

Digitised copies must be cataloguable via central portals. It is strongly recommended to provide an OAI interface that delivers METS / MODS in addition to Dublin Core (see Appendix A) to ensure that relevant portals can harvest the data (→ *2.7*).

Digitisation projects must support the METS / MODS format specified in Appendix A and in supraregional contexts link primarily to a web presentation in the style of the DFG Viewer (→ *5*).

6.7 Citation, persistent addressing

At least the accessibility and citability of the work as a whole must be guaranteed. In the future, the work's individual physical pages will also have to be reliably accessible and citable. 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 (→ 3).

6.8 Provision of digital copies, publicly accessible interfaces

Digital reproductions and project results must be provided free over the Internet. A "moving wall" of up to one year may be agreed on *(→ 4.1)*.

As a rule, digital materials should be accessible by a variety of paths:

(1) via the website of the providing library or archive;
(2) via an OAI interface;
(3) via search inquiry to the local and regional library catalogue or archival online finding-aids system;
(4) via the shared portal of the virtual subject libraries, or one of the DFG-funded material-specific portals that allow integrated access to all digital collections funded under the DFG programme (→ *4.1.2*);
(5) via Internet search engines;
(6) from supraregional systems, via a presentation in the style of the DFG Viewer (→ 5).

All materials 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 (→ *4.1.2* und *4.1.3*).

38 See also http://dfg-viewer.de/profil-der-strukturdaten

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:

(1) An independent server that provides the digitalised material along with the tools needed to use it.

(2) 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 (→ *3*).

(3) An interface in the technical sense, to allow DFG-funded material-specific portals to access all metadata generated by a project; particularly an OAI interface that delivers metadata as well as DC METS / MODS (METS / TEI-P5, METS / EAD) to suitable harvesters (→ *2.7*).

(4) Appropriate measures that enable search engines to find the metadata (→ *2.7*).

7. Guidelines for the Implementation of Digitisation Projects

(1) Between the DFG's funding approval and the actual start of work there is usually a certain period of time during which hardware is purchased, service providers are contracted, and the project staff is recruited. However, the work schedules included in project proposals tend to assume that the project team is complete and fully operational. To make it easier for the DFG to monitor projects and counteract undesirable developments, funded projects are therefore expected to carry out such preparatory activities in a start-up phase during which they receive little or no personnel funding. Once these preparations have been completed, the DFG must be notified in writing of the actual project start. All subsequent deadlines are calculated from this date. No more than one year may pass between the funding approval and the actual project start.

(2) All projects are required to present a functional model of their online services at a point in time when it is still possible to counteract undesirable developments (generally after the first year of funding). This model must demonstrate that it meets the minimum requirements specified above. All components of the technical services must be essentially functional at this point. If questions or doubts arise, an on-site demonstration of online services may be required.

(3) Only after the server and its basic functionalities have been approved and any criticised shortcomings have been remedied is it possible to submit a continuation proposal.

(4) All project reports must state:

    (a) what portion of the total volume has been digitised;

    (b) what portion of the total volume is available online;

    (c) which substantive access figures are shown in the server log files.

If the work progresses at a substantially slower pace than described in the project proposal, it is frequently a sign of poor project organisation rather than a justification for a successful continuation proposal. If the discrepancy is large, even funding for previously approved project stages must be reconsidered.

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 supraregional catalogue systems (e.g. VD 16 / VD 17, ZVDD, virtual subject libraries), 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 contents. The DVG Viewer may then link to the special local offerings of any given institution.

The DFG Viewer[39] was built under the “Digitisation of VD 16 / VD 17” line of action by the libraries funded in the first round of proposals. These libraries, in collaboration with additional partners, continue 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 security for proposal planning and ensure that metadata meet the DFG Viewer requirements, metadata generated by such projects should be valid against the XML schema to be read by the Viewer’s website.[40]

METS[41] 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 bibliographic metadata (prints only), the Viewer requires MODS[42]-encoded metadata (see 2. below). To link administrative metadata (e.g. local use, homepage, 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 homepage of the DFG Viewer’s reference application.

These guidelines apply currently only to printed works (METS / MODS). Measures are underway to enable the DFG Viewer to display other materials as well (e.g. manuscripts), using TEI-P5 (METS / TEI-P5) and EAD (METS / EAD) as guidelines.

2. MODS DFG standard set (print holdings)

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

39 http://www.dfg-viewer.de/
40 http://dfg-viewer.de/profil-der-metadaten
41 http://www.loc.gov/standards/mets/
42 http://www.loc.gov/standards/mods/

Element / subelement	*Repeatable*	*Comments*	*Status*
1 Title information			
<titleInfo>	Yes	Title information; if work has no title, a title must be created	Mandatory
<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 “pnd”.	Mandatory if applicable
<name> / <namePart type=“…“>	Yes	Contains name elements of the type specified in @type; possible values are “date”, “family”, “given”, “termsOfAddress”	
<name> / <displayForm>	No	Name in desired display form	Recommended
<name> / <role>	No	Wrapper element for role of person	Mandatory if available
<name> / <role> / <roleTerm type=”code” authority=”marcrelator”>	No	Role of person; <roleTerm> field value is encoded (MARC relator code) [43]	
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>	Yes	Publication information/Imprint; the first <originInfo> block is for information on the source; the second <originInfo> block is for information on the digital edition	Mandatory if available
<originInfo> / <place>	Yes	Contains elements on place of publication	Mandatory if available
<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 available
<originInfo> / <dateIssued keyDate="yes" encoding="w3cdtf">	Yes	Contains year of publication; if year is unknown, write “[o.J.]”	Mandatory

43 A code from the *MARC Value List for Relators and Roles*: http://www.loc.gov/marc/sourcecode/relator/relatorlist.html.

Element / subelement	*Repeatable*	*Comments*	*Status*
5 Edition information			
<originInfo> / <edition>	Yes	Contains name of edition	Mandatory if applicable
6 Physical description			
<physicalDescription>	No	Physical description area / collation statement	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
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 type=”host” 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 attributes 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 on how to populate the respective fields.

3. Example of a METS / MODS dataset according to DFG standard

```
<mets:mets xmlns:mets="http://www.loc.gov/METS/"
           xmlns:xlink="http://www.w3.org/1999/xlink"
           xmlns:mods="http://www.loc.gov/mods/v3"
           xmlns:dv="http://dfg-viewer.de/">
           <mets:dmdSec ID="dmd_586140484">
             <mets:mdWrap MDTYPE="MODS">
               <mets:xmlData>
                 <mods:mods>
                   <mods:recordInfo>
                     <mods:recordIdentifier source="WDB_OPAC"
                       >oai:diglib.hab.de:ppn_586140484</mods:recordIdentifier>
                   </mods:recordInfo>
                   <mods:location>
                    <mods:physicalLocation>Herzog August Bibliothek Wolfenbüttel</mods:physicalLocation>
                    <mods:shelfLocator>M: Gn 4° 1572:1</mods:shelfLocator>
                   </mods:location>
                   <mods:identifier type="purl"
                     >http://diglib.hab.de/drucke/gn-4f-1572-1b/start.htm</mods:identifier>
                   <mods:identifier type="urn"
                     >urn:nbn:de:gbv:23-drucke/gn-4f-1572-1b3</mods:identifier>
                   <mods:language>
                     <mods:languageTerm type="code" authority="iso639-2b"
                     >lat</mods:languageTerm>
                   </mods:language>
                   <mods:name type="personal">
                     <mods:namePart type="family">Leibniz</mods:namePart>
                     <mods:namePart type="given">Gottfried Wilhelm</mods:namePart>
                     <mods:displayForm>Leibniz, Gottfried Wilhelm</mods:displayForm>
                     <mods:namePart type="date">1646-1716</mods:namePart>
                     <mods:role>
                       <mods:roleTerm type="code" authority="marcrelator">ctb</mods:roleTerm>
                     </mods:role>
                   </mods:name>
                   <mods:originInfo>
                     <mods:place>
                       <mods:placeTerm type="text">Hanoverae</mods:placeTerm>
                     </mods:place>
                     <mods:publisher>Foerster</mods:publisher>
                     <mods:dateIssued keyDate="yes" encoding="w3cdtf"
                     >1707</mods:dateIssued>
                   </mods:originInfo>
                   <mods:originInfo>
                     <mods:place>
                       <mods:placeTerm type="text"
                       >Wolfenbüttel</mods:placeTerm>
                     </mods:place>
                     <mods:publisher>Herzog August Bibliothek</mods:publisher>
                     <mods:dateIssued encoding="w3cdtf">2008</mods:dateIssued>
                     <mods:edition>[Electronic ed.]</mods:edition>
                   </mods:originInfo>
                   <mods:physicalDescription>
                     <mods:extent>2°</mods:extent>
                     <mods:extent>[22] Bl., 1004 S., [1] Bl</mods:extent>
                     <mods:digitalOrigin>reformatted digital</mods:digitalOrigin>
                   </mods:physicalDescription>
                   <mods:titleInfo>
                     <mods:title>Scriptores Rervm Brvnsvicensivm Illustrationi
                       Inservientis, Antiqvi Omnes Et Religionis Reformatione
                       Priores ...</mods:title>
                   </mods:titleInfo>
                   <mods:relatedItem type="host">
                     <mods:recordInfo>
```

```
                    <mods:recordIdentifier source="WDB_OPAC"
                      >oai:diglib.hab.de:ppn_58614045X</mods:recordIdentifier>
                  </mods:recordInfo>
                </mods:relatedItem>
                <mods:part type="host" order="1">
                  <mods:detail>
                    <mods:number>[1]</mods:number>
                  </mods:detail>
                </mods:part>
                </mods:mods>
            </mets:xmlData>
          </mets:mdWrap>
        </mets:dmdSec>
        <mets:amdSec ID="amd_586140484">
          <mets:rightsMD ID="amd_dvrights_586140484">
            <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>auskunft@hab.de/</dv:ownerContact>
                </dv:rights>
              </mets:xmlData>
            </mets:mdWrap>
          </mets:rightsMD>
          <mets:digiprovMD ID="amd_dvlinks_586140484">
            <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="DVLINKS">
              <mets:xmlData>
                <dv:links>
                  <dv:reference>
                    http://sunny.biblio.etc.tu-
                    bs.de:8080/DB=2/SET=2/TTL=2/CMD?ACT=SRCHA&amp;IKT=1016&amp;SRT=YO
                    P&amp;TRM=url+diglib.hab.de%5C%2Fdrucke%5C%2Fgn-4f-1572-
                    1b%5C%2Fstart.htm
                    </dv:reference>
                  <dv:presentation>http://diglib.hab.de/drucke/gn-4f-1572-1b/start.htm</dv:presentation>
                </dv:links>
              </mets:xmlData>
            </mets:mdWrap>
          </mets:digiprovMD>
        </mets:amdSec>
        <mets:fileSec>
          <mets:fileGrp USE="DEFAULT">
            <mets:file ID="drucke_gn-4f-1572-1b_00001" MIMETYPE="image/jpeg">
              <mets:FLocat
                xlink:href="http://diglib.hab.de/drucke/gn-4f-1572-1b/00001.jpg"
                LOCTYPE="URL"/>
            </mets:file>
             <!-- Kürzung um  00001.jpg bis  01057.jpg  DEFAULT Auflösung -->
            <mets:file ID="drucke_gn-4f-1572-1b_01058" MIMETYPE="image/jpeg">
              <mets:FLocat
                xlink:href="http://diglib.hab.de/drucke/gn-4f-1572-1b/01058.jpg"
                LOCTYPE="URL"/>
            </mets:file>
          </mets:fileGrp>
          <mets:fileGrp USE="MIN">
            <mets:file ID="min_drucke_gn-4f-1572-1b_00001" MIMETYPE="image/jpeg">
              <mets:FLocat
                xlink:href="http://diglib.hab.de/drucke/gn-4f-1572-1b/min/00001.jpg"
                LOCTYPE="URL"/>
            </mets:file>
            <!-- Kürzung um  00002.jpg bis  01057.jpg  Minimale Auflösung -->
```

```
            <mets:file ID="min_drucke_gn-4f-1572-1b_01058" MIMETYPE="image/jpeg">
              <mets:FLocat
                xlink:href="http://diglib.hab.de/drucke/gn-4f-1572-1b/min/01058.jpg"
                LOCTYPE="URL"/>
            </mets:file>
          </mets:fileGrp>
          <mets:fileGrp USE="MAX">
            <mets:file ID="max_drucke_gn-4f-1572-1b_00001" MIMETYPE="image/jpeg">
              <mets:FLocat
                xlink:href="http://diglib.hab.de/drucke/gn-4f-1572-1b/max/00001.jpg"
                LOCTYPE="URL"/>
            </mets:file>
            <!-- Kürzung um  00002.jpg bis  01057.jpg  Maximale Auflösung -->
            <mets:file ID="max_drucke_gn-4f-1572-1b_01058" MIMETYPE="image/jpeg">
              <mets:FLocat
                xlink:href="http://diglib.hab.de/drucke/gn-4f-1572-1b/max/01058.jpg"
                LOCTYPE="URL"/>
            </mets:file>
          </mets:fileGrp>
          <mets:fileGrp USE="THUMBS">
            <mets:file ID="thumbs_drucke_gn-4f-1572-1b_00001" MIMETYPE="image/jpeg">
              <mets:FLocat
                xlink:href="http://diglib.hab.de/drucke/gn-4f-1572-1b/thumbs/00001.jpg"
                LOCTYPE="URL"/>
            </mets:file>
            <!-- Kürzung um  00002.jpg bis  01057.jpg  Thumbnails -->
            <mets:file ID="thumbs_drucke_gn-4f-1572-1b_01058" MIMETYPE="image/jpeg">
              <mets:FLocat
                xlink:href="http://diglib.hab.de/drucke/gn-4f-1572-1b/thumbs/01058.jpg"
                LOCTYPE="URL"/>
            </mets:file>
          </mets:fileGrp>
        </mets:fileSec>
        <mets:structMap TYPE="LOGICAL">
          <mets:div ID="logMD_586140484" TYPE="Monograph" DMDID="dmd_586140484"
            ADMID="amd_586140484"/>
        </mets:structMap>
        <mets:structMap TYPE="PHYSICAL">
          <mets:div ID="physMD_586140484" TYPE="physSequence">
            <mets:div ID="physMD_586140484_1" TYPE="page" ORDER="1">
              <mets:fptr FILEID="drucke_gn-4f-1572-1b_00001"/>
              <mets:fptr FILEID="min_drucke_gn-4f-1572-1b_00001"/>
              <mets:fptr FILEID="max_drucke_gn-4f-1572-1b_00001"/>
              <mets:fptr FILEID="thumbs_drucke_gn-4f-1572-1b_00001"/>
            </mets:div>
            <!-- Kürzung um  Zuordnungen von 00002 bis  01057  -->
            <mets:div ID="physMD_586140484_1058" TYPE="page" ORDER="1058">
              <mets:fptr FILEID="drucke_gn-4f-1572-1b_01058"/>
              <mets:fptr FILEID="min_drucke_gn-4f-1572-1b_01058"/>
              <mets:fptr FILEID="max_drucke_gn-4f-1572-1b_01058"/>
              <mets:fptr FILEID="thumbs_drucke_gn-4f-1572-1b_01058"/>
            </mets:div>
          </mets:div>
        </mets:structMap>
        <mets:structLink>
          <mets:smLink xlink:from="logMD_586140484" xlink:to="physMD_586140484" />
        </mets:structLink>
      </mets:mets>
```

4. 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 library 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 downloadable PDF files, footers may be added to each image as well, in addition to a cover page. Compare the following examples: